# ENCYCLOPEDIA OF
# Physical Science
# AND Technology

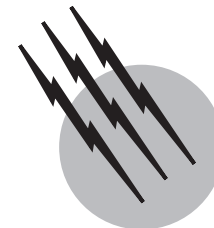## THIRD EDITION

## Energy

# Coal Preparation

## Robert A. Meyers

*Ramtech Limited*

## Anthony D. Walters

*Kilborn Engineering Limited*

## Janusz S. Laskowski

*University of British Columbia*

## GLOSSARY

**Dense media (or heavy liquids)** Fluids used in dense-medium separation of coal and gangue particles by their relative densities. The medium can be any suitable fluid, but in commercial coal preparation operations, it is usually a suspension of fine magnetite in water.

**Density (gravity) separation** Separation methods based on differences in density of separated minerals, such as dense-medium separation and jigging.

**Float-and-sink tests** Tests carried out to determine coal washability, by which the coal is separated into various density fractions.

**Floatability** Description of the behavior of coal particles in froth flotation.

**Hardgrove grindability index** Measure of the ease by which the size of a coal can be reduced. Values decrease with increasing resistance to grinding.

**Metallurgical coal** Coal used in the manufacture of coke for the steel industry.

**Near-density material** Percentage of material in the feed within $\pm 0.1$ density range from the separation density.

**Organic efficiency (recovery efficiency)** Measure of separating efficiency, calculated as

$$\frac{\text{actual clean coal yield}}{\text{theoretical clean coal yield}} \times 100$$

with both yields at the same ash content.

**Probable error $E_p$ (Ecart probable moyen)** Measure of separating efficiency of a separator (jig, cyclone, etc.). Calculated from a Tromp curve.

**RRB particle size distribution** Particle size distribution developed by Rossin and Rammler and found by Bennett to be useful in describing the particle size distribution of run-of-mine coal.

**Separation cut point $\delta_{50}$** Density of particles reporting equally to floating and sinking fractions (heavy media overflow and underflow in the dense media cyclone). The $\delta_{50}$ is also referred to as partition density.

**Separation density $\delta_s$** Actual density of the dense medium.

**Thermal coals** Coals used as a fuel, mainly for power generation. These are lower-rank coals (high volatile bituminous, subbituminous, and liginites).

**Tromp curve (partition curve, distribution curve)** Most widely used method of determining graphically the value of $E_p$ as a measure of separation efficiency. Synonyms: partition curve, distribution curve.

**Washability curves** Graphical presentation of the float-sink test results. Two types of plots are in use: Henry–Reinhard washability curves and M-curves.

**Washing of coal (cleaning, beneficiation)** Term denoting the most important coal preparation unit operation in which coal particles are separated from inorganic gangue in the processes based on differences in density (gravity methods) or surface properties (flotation). Cleaning then increases the heating value of raw coal.

**COAL PREPARATION** is the stage in coal production—preceding its end use as a fuel, reductant, or conversion plant feed—at which the run-of-mine (ROM) coal, consisting of particles, different in size and mineralogical composition, is made into a clean, graded, and consistent product suitable for the market; coal preparation includes physical processes that upgrade the quality of coal by regulating its size and reducing the content of mineral matter (expressed as ash, sulfur, etc.). The major unit operations are screening, cleaning (washing, beneficiation), crushing, and mechanical and thermal dewatering.

## I. COAL CHARACTERISTICS RELATED TO COAL PREPARATION

Coal, an organic sedimentary rock, contains combustible organic matter in the form of macerals and inorganic matter mostly in the form of minerals.

Coal preparation upgrades raw coal by reducing its content of impurities (the inorganic matter). The most common criterion of processing quality is that of ash, which is not removed as such from coal during beneficiation processes, but particles with a lower inorganic matter content are separated from those with a higher inorganic matter content. The constituents of ash do not occur as such in coal but are formed as a result of chemical changes that take place in mineral matter during the combustion process. The ash is sometimes defined as all elements in coal except carbon, hydrogen, nitrogen, oxygen, and sulphur.

Coal is heterogeneous at a number of levels. At the simplest level it is a mixture of organic and inorganic phases, but because the mineral matter of coal originated in the inorganic constituents of the precursor plant, from other organic materials, and from the inorganic components transported to the coal bed, its textures and liberation characteristics differ. The levels of heterogeneity can then be set out as follows (Table I):

1. At the seam level, a large portion of mineral matter in coal arises from the inclusion during mining of roof or floor rock.
2. At the ply and lithotype level, the mineral matter may occur as deposits in cracks and cleats or as veins.
3. At the macerals level, the mineral matter may be present in the form of very finely disseminated discrete mineral matter particles.
4. At the submicroscopic level, the mineral matter may be present as strongly, chemically bonded elements.

Even in the ROM coal, a large portion of both coal and shale is already liberated to permit immediate concentration. This is so with heterogeneity level 1 and to some extent with level 2; at heterogeneity level 3 only crushing and very fine grinding can liberate mineral matter, while at level 4, which includes chemically bonded elements and probably syngenetic mineral matter, separation is possible only by chemical methods.

Recent findings indicate that most of the mineral matter in coal down to the micron particle-size range is indeed a distinct separable phase that can be liberated by fine crushing and grinding.

The terms *extraneous mineral matter* and *inherent mineral matter* were usually used to describe an ash-forming material, separable and nonseparable from coal

**TABLE I  Coal Inorganic Impurities***a*

| Type | Origin | Examples | Physical separation |
|---|---|---|---|
| Strongly chemically bonded elements | From coal-forming organic tissue material | Organic sulphur, nitrogen | No |
| Adsorbed and weakly bonded groups | Ash-forming components in pure water, adsorbed on the coal surface | Various salts | Very limited |
| Mineral matter a. Epiclastic | Minerals washed or blown into the peat during its formation | Clays, quartz | Partly separable by physical methods |
| b. Syngenetic | Incorporated into coal from the very earliest peat-accumulation stage | Pyrite, siderite, some clay minerals | Intimately intergrown with coal macerals |
| c. Epigenetic | Stage subsequent to syngenetic; migration of the minerals-forming solutions through coal fractures | Carbonates, pyrite, kaolinite | Vein type mineralization; epigenetic minerals concentrated along cleats, preferentially exposed during breakage; separable by physical methods |

*a* Adapted from Cook, A. C. (1981). *Sep. Sci. Technol.* **16**(10), 1545.

by physical methods. Traditionally in coal preparation processes, only the mineral matter at the first and, to some extent, the second levels of heterogeneity was liberated, the rest remained unliberated and, left with the cleaned coal, contributed to the inherent mineral matter. Recent very fine grinding, which also liberates the mineral matter at the third level of heterogeneity, has changed the old meanings of the terms *inherent* and *extraneous*. The content of the "true" inherent part of ash-forming material (i.e., the part left in coal after liberating and removing the mineral matter at the first, second, and third levels of heterogeneity) is usually less than 1%.

In recent years, the emphasis in coal preparation has been placed on reducing sulfur content of coal and on recovering the combustible material. Sulfur in coal is present in both organic and inorganic forms. The dominant form of inorganic sulfur is pyrite, but marcasite has also been reported in many coals. Pyrite occurs as discrete particles, often of microscopic size. It comprises 30–70% of total sulfur in most coals. Other forms of inorganic sulfur that may be present are gypsum and iron sulfates. The sulfate level in fresh unoxidized coals is generally less than 0.2%.

Organic sulfur in coal is believed to be contained in groups such as



THIOPHENE

R—S—R′          R—SH          R—SS—R′
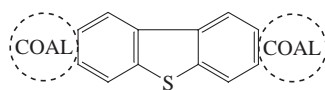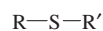
ORGANIC       MERCAPTAN        ORGANIC
SULFIDE                       DISULFIDE

The organic sulfur content in coals range from 0.5 to 2.5% w/w.

Physical cleaning methods can remove inorganic sulfates (gypsum) and most of the coarse pyrite; the finely disseminated microcrystalline pyrite and organic sulfur are usually not separable by such processes. This means that in the case of coal containing 70% of sulfur in pyritic form and 30% as organic sulfur, the physical cleaning can reduce the sulfur content by about 50%.

## II. BREAKING, CRUSHING, AND GRINDING

The primary objectives of crushing coal are

1. Reduction of the top size of ROM coal to make it suitable for the treatment process
2. liberation of coal from middlings, and
3. size reduction of clean coal to meet market specification.

Size reduction of coal plays a major role in enabling ROM coal to be used to the fullest possible extent for power generation, production of coke, and production of synthetic fuels. ROM coal is the "as-received" coal from the mining process. Because the types of mining processes are varied, and size reduction actually begins at the face in the mining operation, it is quite understandable that the characteristics of the products from these various processes differ widely. The type of mining process directly affects the top size and particle size distribution of the mine product.

During the beneficiation of coal, the problem of treating middlings sometimes arises. This material is not of sufficient quality to be included with the high-quality clean coal product, yet it contains potentially recoverable coal. If this material is simply recirculated through the cleaning

circuit, little or no quality upgrading can be achieved. Liberation can be accomplished if the nominal top size of the material is reduced, which permits recovery of the coal in the cleaning unit. Hammer mills, ring mills, impactors, and roll crushers in open or closed circuits can be used to reduce the top size of the middlings. Normally, reducing the nominal top size to the range 20–6 mm is practiced.

*Breaking* is the term applied to size operations on large material (say +75 mm) and *crushing* to particle size reduction below 75 mm; the term *grinding* covers the size reduction of material to below about 6 mm. However, these terms are loosely employed. A general term for all equipment is *size reduction equipment*, and because the term *comminution* means size reduction, another general term for the equipment is *comminution equipment.*

## A. Breaking and Crushing

Primary breakers that treat ROM coal are usually rotary breakers, jaw crushers, or roll crushers; some of these are listed below.

*Rotary Breakers* (Fig. 1). The rotary breaker serves two functions—namely, reduction in top size of ROM and rejection of oversize rock. It is an autogenous size-reduction device in which the feed material acts as crushing media.

*Jaw Crusher.* This type of primary crusher is usually used for crushing shale to reduce it to a size suitable for handling.

*Roll Crusher.* For a given reduction ratio, single-roll crushers are capable of reducing ROM material to a product with a top size in the range of 200–18 mm in a single pass, depending upon the top size of the feed coal. Double-roll crushers consist of two rolls that rotate in opposite directions. Normally, one roll is fixed while the other roll is movable against spring pressure. This permits the passage of tramp material without damage to the unit. The drive units are normally equipped with shear pins for overload protection.

*Hammer Mills.* The swinging hammer mill (instead of having teeth as on a single-roll crusher) has hammers that are mounted on a rotating shaft so that they have a swinging movement.

## B. Grinding

The most common grindability index used in conjunction with coal size reduction, the Hardgrove Grindability Index, is determined by grinding 50 g of $16 \times 30$ mesh dried coal in a standardized ball and race mill for 60 revolutions at an upper grinding ring speed of 20 rpm. Then the sample is removed and sieved at 200 mesh to determine $W$, the amount of material passing through the 200 mesh sieve. The index is calculated from the following formula:

$$HGI = 13.6 + 6.93W. \tag{1}$$

From the above formula, it can be deduced that as the resistance of the coal to grinding increases, the HGI decreases. The HGI can be used to predict the particle size distribution, that is, the Rossin–Rammler–Bennett curve [see Eq. (4)].

The distribution modulus ($m$) and the size modulus $d_{63.2}$ must be known to determine the size distribution of a particular coal.

It has been shown that for Australian coals, the distribution modulus can be calculated from the HGI by the following equation:
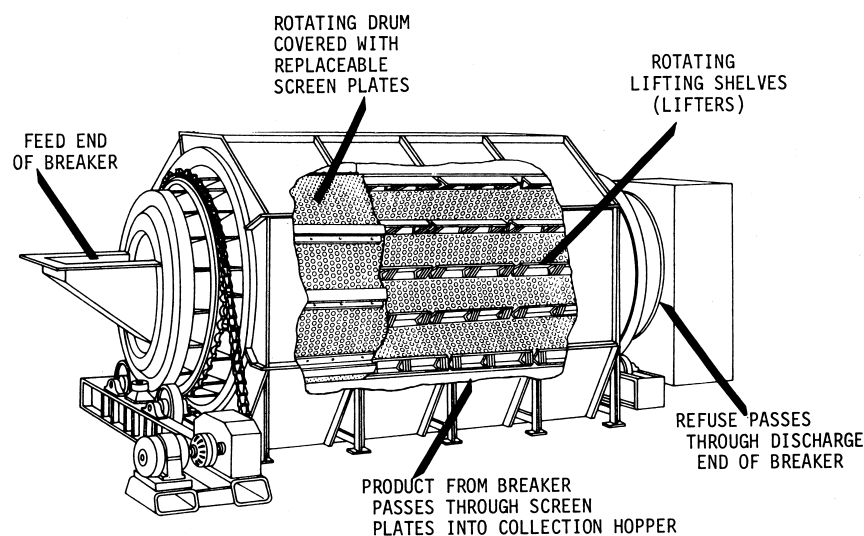


**FIGURE 1** Cutaway view of rotary breaker.

$$\text{HGI} = 35.5m^{-1.54}. \qquad (2)$$

The value of $d_{63.2}$ is a function of the degree of breakage that the coal has undergone during and after mining. Having selected $d_{63.2}$ and the value of $m$ from the HGI, one can predict the size distribution.

For the use of coal at power stations and for treatment by some of the newer beneficiation techniques (e.g., for coal/water slurries), grinding is employed to further reduce the top size and produce material with a given particle size distribution. The principal equipment used for coal grinding is the following:

1. air-swept ball mills,
2. roll or ball-and-race type mills,
3. air-swept hammer mills, and
4. wet overflow ball mills.

## III. COAL SCREENING

Sizing is one of the most important unit operations in coal preparation and is defined as the separation of a heterogeneous mixture of particle sizes into fractions in which all particles range between a certain maximum and minimum size. Screening operations are performed for the following purposes:

1. Scalping off large run-of-mine coal for initial size reduction
2. Sizing of raw coal for cleaning in different processes
3. Removal of magnetic (dense medium) from clean coal and refuse
4. Dewatering
5. Separation of product coal into commercial sizes

The size of an irregular particle is defined as the smallest aperture through which that particle will pass.

In a practical screening operation the feed material forms a bed on the screen deck (Fig. 2) and is subjected to mechanical agitation so that the particles repeatedly approach the deck and are given the opportunity of passing
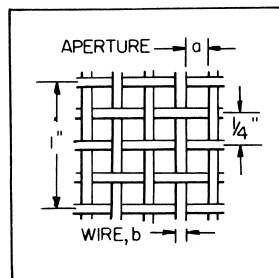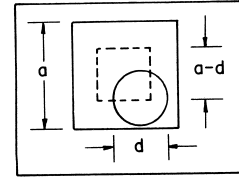


**FIGURE 2** Four-mesh wire screen.



**FIGURE 3** Passage of a spherical particle through a square aperature.

through. In the simplest case (Fig. 3), a spherical particle of diameter $d$ will only pass through if it does not touch the sides of the aperture. The condition for passing is that the center of the sphere falls within the inner square, side $a - d$. The probability $P$ of passing is thus given by the ratio between the areas of the inner and outer squares, that is,

$$P = (a - d)^2/a^2 = (1 - d/a)^2. \qquad (3)$$

The assumption that passage will be achieved only if there is no contact with the aperture sides is too restrictive. They can and do collide with the screen deck while passing through. In particular, the following factors contribute to the probability of passing:

1. Percentage open area
2. Particle shape
3. Angle of approach
4. Screen deck area
5. Bed motion
6. Size distribution of feed.

To select the correct screen for an application in a coal preparation plant, a detailed knowledge of the size distribution of the feed is necessary. Size distributions are usually presented graphically, and it is useful to use a straight-line plot, because curve fitting and subsequent interpolation and extrapolation can be carried out with greater confidence. In addition, if a function can be found that gives an acceptable straight-line graph, the function itself or the parameter derived from it can be used to describe the size distribution. This facilitates data comparison, transfer, or storage, and also offers major advantages in computer modeling or control. The particle size distribution used most commonly in coal preparation is the Rossin–Rammler–Bennett,

$$F(d) = 100(1 - \exp[(-d/d_{63.2})^m], \qquad (4)$$

where $F(d)$ is the cumulative percent passing on size $d$, $d_{63.2}$ is the size modulus ($d_{63.2}$ is that aperture through which 63.2% of the sample would pass), and $m$ is the distribution modulus (the slope of the curve on the RRB graph paper). A sample of the RRB plot is shown in Fig. 4.

In Fig. 4 one can read off the slope $m$; the scale for $S \cdot d_{63.2}$ is also provided. If $d_{63.2}$ is the particle size (in
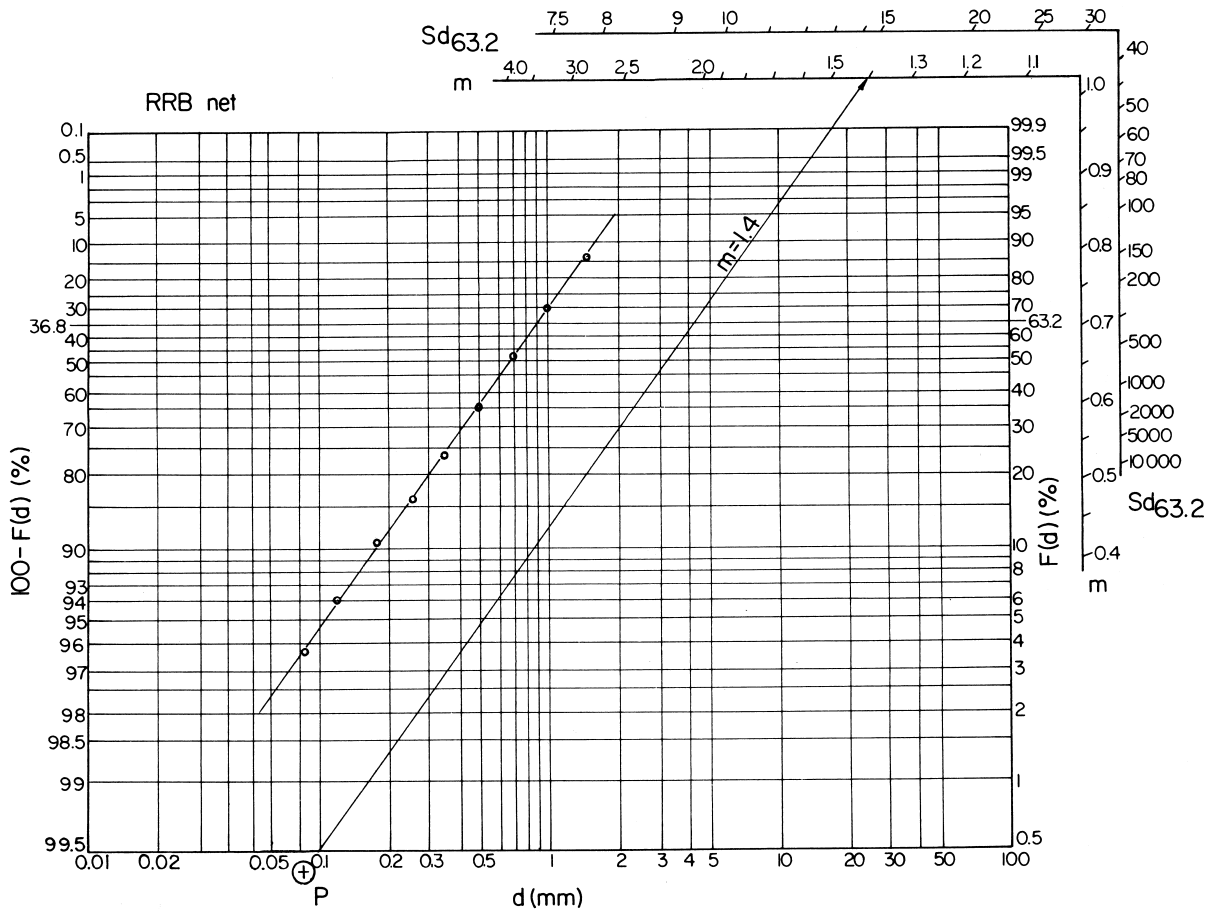
**FIGURE 4** Rossin–Rammler–Bennett net with additional scales. Specific surface $S$ is in square meters per cubic decimeter, $d_{63.2}$ in millimeters.

millimeters) belonging to $F(d) = 63.2\%$, then the scales furnish the specific surface $S$ in m²/dm³.

The types of screens used in coal preparation plants generally fall into the following categories:

1. Fixed screens and grizzlies (for coarse coal).
2. Fixed sieves (for fine coal). These are used for dewatering and/or size separation of slurried coal or rejects. The sieve bend is the most commonly used in coal preparation plants for this application (Fig. 5).
3. Shaking screens. These are normally operated by camshafts with eccentric bearings. They can be mounted horizontally or inclined, and operate at low speeds with fairly long strokes (speeds up to 300 rpm with strokes of 1–3 in).
4. Vibrating screens. These are the most commonly used screens in coal preparation and are to be found in virtually all aspects of operations. A summary of their application is shown in Table II. A recent development in vibrating screens has been the

introduction of very large screens (5.5 m wide, 6.4 m long). Combined with large tonnage [1000 metric tons per hour (tph) per screen], screens have been developed (banana screens) with varying slope: first section 34°, second section 24°, and third section 12°.
5. Resonance screens. These have been designed to save energy consumption. The screen deck is mounted on flexible hanger strips and attached to a balance frame, which is three or four times heavier than the screen itself.
6. Electromechanical screens. This type of screen operates with a high-frequency motion of very small throw. The motion is usually caused by a moving magnet which strikes a stop.

## A. Recent Developments

A new process for the screening of raw coal of high moisture content at fine sizes has been developed by the National Coal Board. Their Rotating Probability Screen
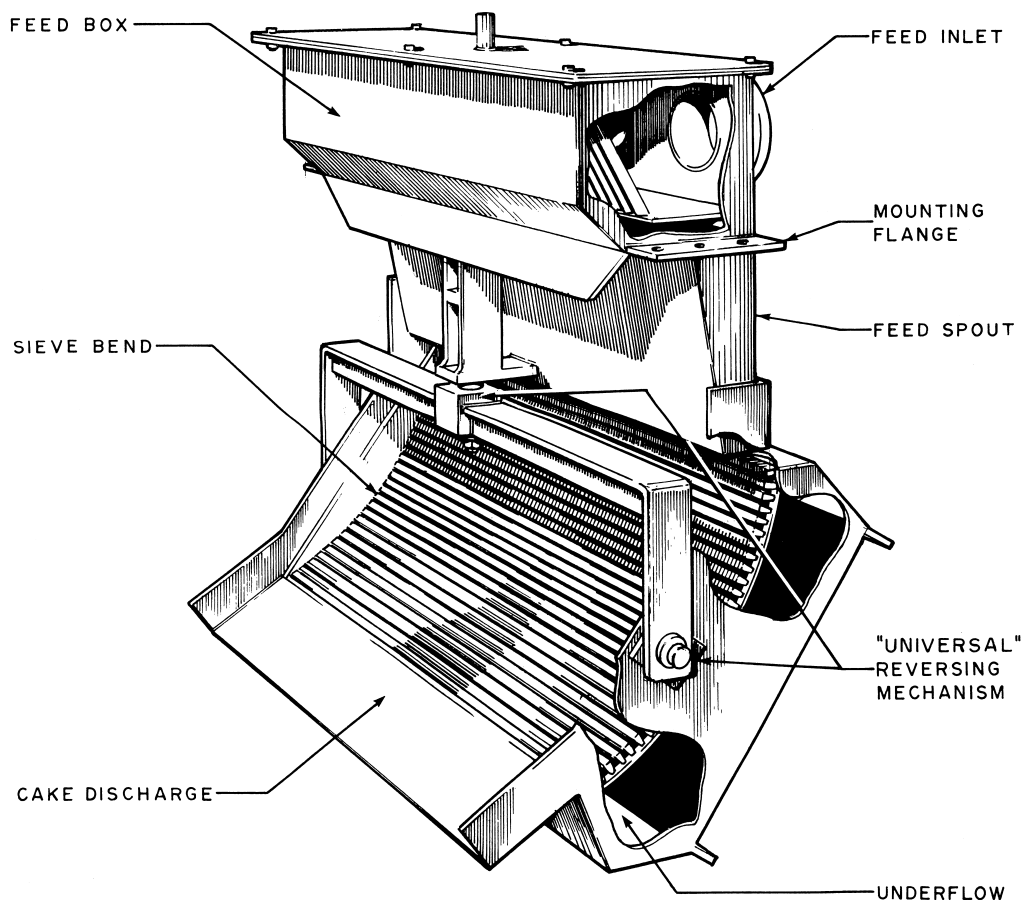
FIGURE 5   Sieve bend.

## IV. FLOAT-AND-SINK ANALYSIS

(Fig. 6) has the advantage that the effective screening aperture can be changed while the screen is running. The screen "deck" is made up of small diameter stainless steel rods radiating from a central hub. The hub is rotated, and the coal falls onto the rotating spokes. The undersize coal passes through the spokes; the oversize coal is deflected over them. The speed of rotation dictates the screening aperture. The ability to screen coal with high levels of moisture has taken precedence over the accuracy of size separation. A screening operation by which the proportion of underflow product can be controlled while the machine is in operation represents an important advance in the preparation of blended coals in treatment plants where the fines are not cleaned.

## IV. FLOAT-AND-SINK ANALYSIS

Most coal cleaning processes that are used to remove inorganic impurities from coal are based on the gravity separation of the coal from its associated gangue. ROM coal consists of some proportion of both coal and shale particles, already sufficiently liberated, together with coal particles with inclusions of gangue (i.e., bands of shale). Commercially cleaned coal contains only very disseminated impurities and has a density ranging from 1.2 to 1.6. Carbonaceous shale density ranges from 2.0 to 2.6, and pure shale, clay, and sandstone have a density of about 2.6. The density of pyrite is about 5.0. The difference in density between pure coal and these impurities, in a liberated state, is sufficient to enable an almost complete separation to be achieved fairly easily. However, it has been shown that the inorganic impurity content, and hence, the ash content, ranges from pure coal containing only microscopic impurities to shale, which is entirely free from carbonaceous matter. Generally speaking, the mineral matter content of any coal particle is proportional to its density and inversely proportional to its calorific value.

For the prediction of concentration results (design of flowsheet) and for the control of preparation plant operations, a measure of concentrating operations is needed.

**TABLE II  Vibrating-Screen Applications in Coal Preparation Plants**

| Type | Number of decks | Installation angle | Aperture[a] | Screen deck type | Accessories |
|---|---|---|---|---|---|
| Run-of-mine scalper | Single | 17°–25° | 6 in. | Manganese skid bars, AR perforated plate with skid bars | Feed box with liners, extra high side plates, drive guard enclosures |
| Raw-coal sizing screen | Double | 17°–25° | 1 in. | AR steel perforated polyurethane, rubber | Dust enclosures, drive plate, guard enclosures |
| | | | $\frac{5}{16}$ in. | Polyurethane, wire, 304 stainless steel profile deck, rubber | Feed box with liners |
| Pre-wet screen | Double | Horizontal | 1 in. | Wire, polyurethane, rubber | Water spray bar, side plate drip angles, drive guard enclosures, feed box liners |
| | | | 1 mm | Stainless steel profile deck, polyurethane | |
| Dense-medium drain and rinse screen (coarse coal) | Double | Horizontal | 1 in. | Wire, polyurethane, rubber | Side plate drip angles, spray bars, shower box cross flow screen or sieve bend, drip lip angles, drive guard enclosures |
| | | | 1 mm | 304 stainless steel profile deck, polyurethane | |
| Dewatering screen (coarse coal) | Single | Horizontal | 1 mm | 304 stainless steel profile deck, polyurethane | Sieve bend or cross flow screen, dam, discharge drip lip angles, drive guard enclosures |
| Desliming screen | Single | Horizontal | 0.5 mm | 304 stainless steel profile deck, polyurethane | Sieve bend or cross flow screen, spray bars, shower box, drive guard enclosures |
| Classifying screen (fine coal) | Single | 28° | 100 mesh | Stainless steel woven wire sandwich screens | Three-way slurry distributor and feel system |
| Dense-medium drain and rinse screen (fine coal) | Single | Horizontal | 0.5 mm | 304 stainless steel profile deck, polyurethane | Sieve bend or cross flow screen, spray bars, shower box, drip lip angles, drive guard enclosures |
| Dewatering screen (fine coal) | Single | Horizontal or 27°–29° | 0.5 mm | 304 stainless steel profile deck or woven wire, rubber, polyurethane | Sieve bend or cross flow screen, dam, drip lip angles, drive guard enclosures |

[a] Typical application.

The best known means of investigating and predicting theoretical beneficiation results are the so-called washability curves, which represent graphically the experimental separation data obtained under ideal conditions in so-called float–sink tests. Float-and-sink analysis is also used to determine the Tromp curve, which measures the practical results of a density separation. The practical results of separation can then be compared with the ideal and a measure of efficiency calculated.

The principle of float–sink testing procedure is as follows: A weighed amount of a given size fraction is gradually introduced into the heavy liquid of the lowest density. The floating fraction is separated from the fraction that sinks. The procedure is repeated successively with liquids extending over the desired range of densities. The fraction that sinks in the liquid of highest density is also obtained. The weight and ash contents of each density fraction are determined.

In the example shown in Fig. 7, five heavy liquids with densities from 1.3 to 1.8 are used. The weight yields of six density fractions are calculated ($\gamma_1, \gamma_2, \ldots, \gamma_6$), and their ash contents are determined ($\lambda_1, \lambda_2, \ldots, \lambda_6$). The results are set out graphically in a series of curves referred to as washability curves (Henry–Reinhard washability curves or mean-value curve, M-curve, introduced by Mayer).

The construction of the primary washability curve (Henry–Reinhard plot) is shown in Fig. 8. It is noteworthy that the area below the primary curve (shaded) represents ash in the sample. The shaded area, changed into the rectangle of the same surface area, gives the mean-ash content in the sample ($\alpha = 16.44\%$ in our example).

The shape of the primary curve reveals the proportions of raw coal within various limits of ash content and so shows the proportions of free impurities and middlings present. Because the relative ease or difficulty of cleaning a raw coal to give the theoretical yield and ash content
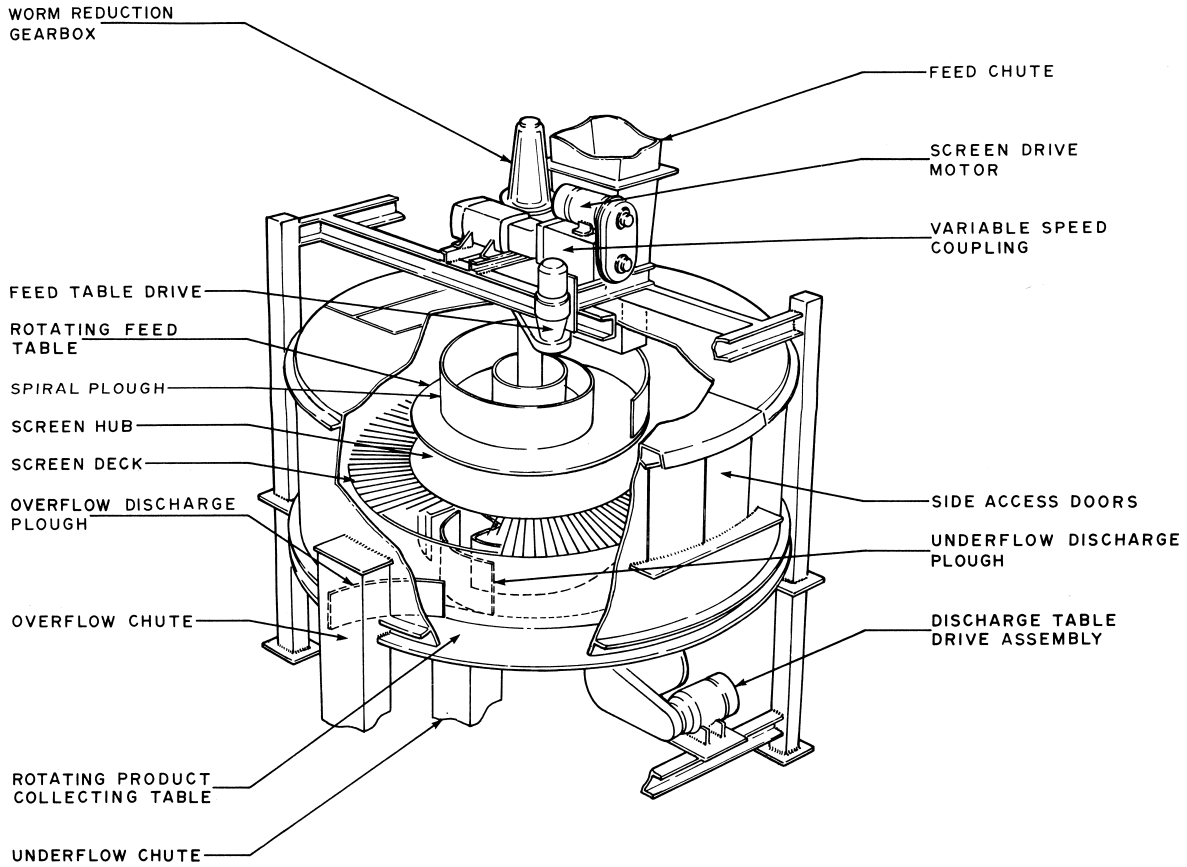
**FIGURE 6**  Rotating probability screen.

depends on the proportion of middlings, the shape of the primary curve also indicates whether the coal is easy or difficult to clean.

The construction of the mean-value curve (M-curve), also referred to as Mayer's curve, is shown in Fig. 9. The point where the curve intersects the abscissa gives the average ash content of the raw coal ($\alpha$).

The shape of the primary washability curve and the M-curve is an indication of the ease or difficulty of cleaning the coal. The more the shape approximates the letter $L$ the easier the cleaning process will be. (This is further illustrated in Fig. 10.)

Figure 10 shows four different cases: (a) ideal separation, (b) easy cleanability, (c) difficult cleanability, and (d) separation impossible.

The primary washability curve for the difficult-to-clean coal (c) exhibits only a gradual change in slope revealing a large proportion of middlings.

It can be seen from the example that the further the M-curve is from the line connecting the zero yield point
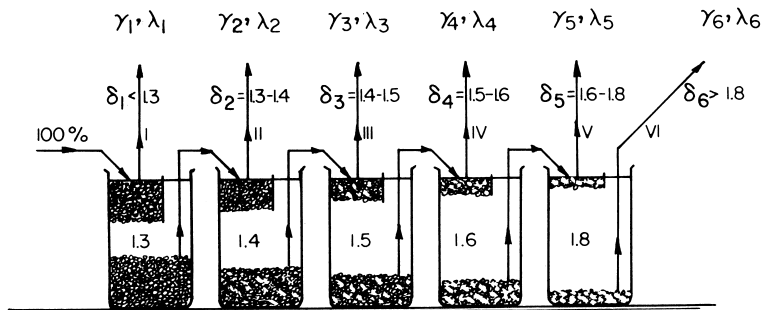


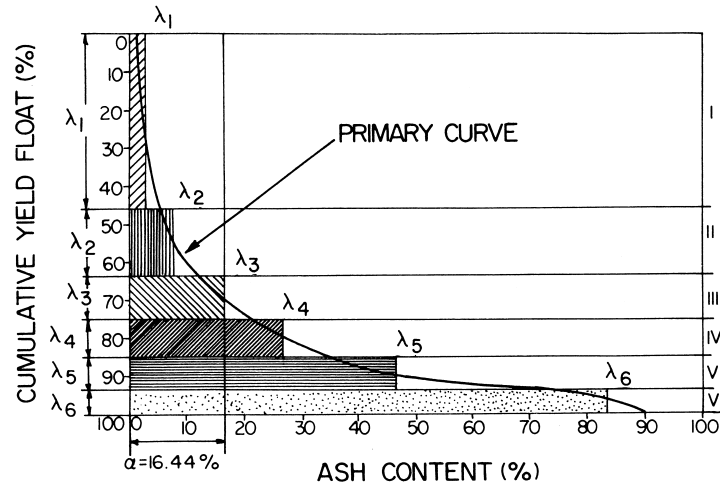**FIGURE 7**  Float–sink analysis procedure.

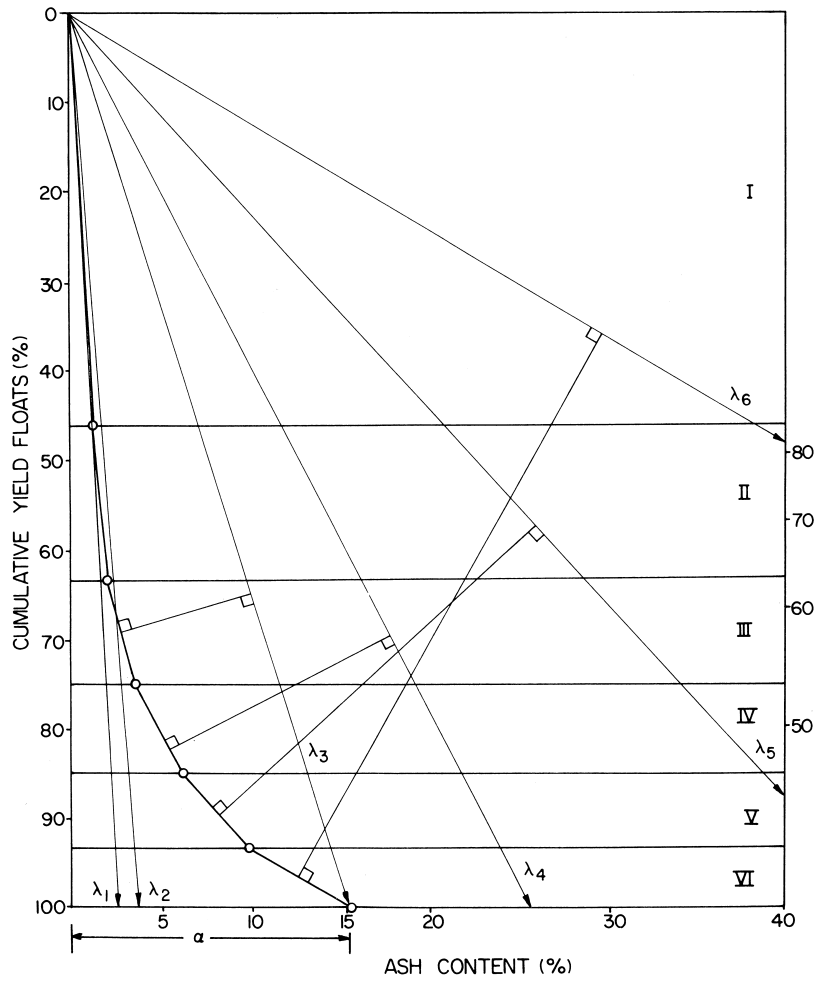**FIGURE 8** Construction of the primary washability curve (Henry–Reinhard washability diagram).
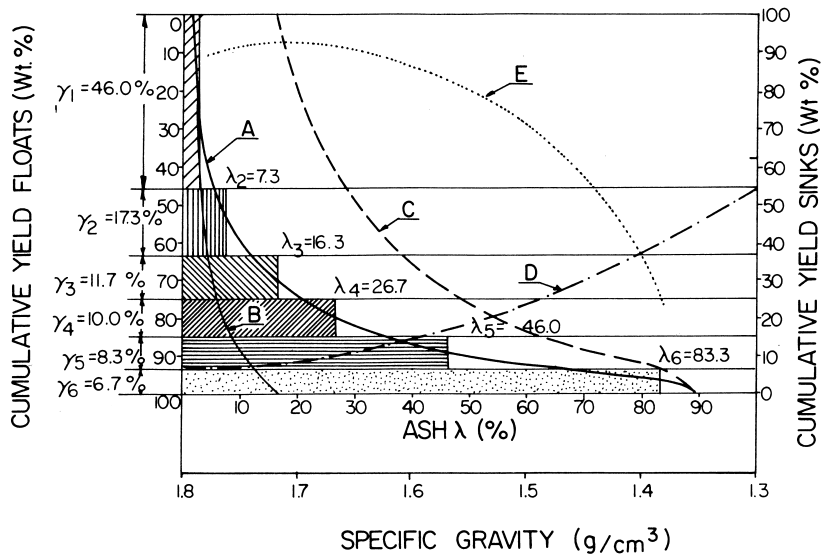
**FIGURE 9** Construction of the M-value curve.

**FIGURE 10** Primary washability and M-value curves for (a) ideal separation, (b) easy-to-clean coal, (c) difficult-to-clean coal, and (d) impossible separation.

with $\alpha$ on the $\lambda$ axis, the greater the range of variation of ash content in the coal. The more gradual change in slope of an M-curve indicates more difficult washability characteristics.

The washability curve can be used in conjunction with the specific-gravity-yield curve (curve D) and the ±0.1 near-density curve (curve E) as demonstrated in the example given in Table III and plotted in Fig. 11.

Curve A is the primary washability curve. Curve B is the clean-coal curve and shows the theoretical percent ash of the clean-coal product at any given yield. Curve C (the cumulative sink ash) shows the theoretical ash content of the refuse at any yield. Curve D, plotted directly from the cumulative percent yield of floats versus density, gives the yield of floats at any separation density. Curve E, the curve of near-density material, gives the amount of ma-

terial within ±0.1 specific-gravity unit, and is used as an indication of the difficulty of separation. The greater the yield of the ±0.1 fraction, the more difficult will be the separation at this separation density. For the example given, the separation process, as indicated by the ±0.1 density curve, will be easier at higher densities than at lower ones.

## V. DENSE-MEDIUM SEPARATION

The use of liquids of varying density in a simple float–sink test leads to separation of the raw coal into different density fractions. In a heavy liquid (dense medium), particles having a density lower than the liquid will float, and those having a higher density will sink. The commercial dense-media separation process is based on the same principle. Since the specific gravity of coal particles varies

**TABLE III  Set of Float–Sink Test Results**

| Specific gravity (g/cm³) (1) | Direct Yield $\gamma$ (wt. %) (2) | Direct Ash $\lambda$ (%) (3) | Weight of ash of total (%) (4) | Cumulative weight of ash (%) (5) | Cumulative floats Yield (%) (6) | Cumulative floats Ash (%) (7) | Sink weight of ash (%) (8) | Cumulative sinks Yield (%) (9) | Cumulative sinks Ash (%) (10) | ±0.1 Specific gravity distribution Specific gravity (g/cm³) (11) | ±0.1 Specific gravity distribution Yield (%) (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| <1.3 | 46.0 | 2.7 | 1.24 | 1.24 | 46.0 | 2.7 | 15.2 | 54.0 | 16.44 | — | — |
| 1.3–1.4 | 17.3 | 7.3 | 1.26 | 2.50 | 63.3 | 3.9 | 13.94 | 36.7 | 28.22 | 1.4 | 63.3 |
| 1.4–1.5 | 11.7 | 16.3 | 1.87 | 4.37 | 75.0 | 5.8 | 12.07 | 25.0 | 38.00 | 1.5 | 29.0 |
| 1.5–1.6 | 10.0 | 26.7 | 2.67 | 7.04 | 85.0 | 8.3 | 9.4 | 15.0 | 48.28 | 1.6 | 14.15 |
| 1.6–1.8 | 8.3 | 46.0 | 3.82 | 10.86 | 93.3 | 11.6 | 5.58 | 6.7 | 62.66 | 1.7 | 8.3 |
| >1.8 | 6.7 | 83.3 | 5.58 | 16.44 | 100.0 | 16.44 | — | — | 83.29 | 1.8 | 10.85 |
|  | 100.0 |  | 16.44 |  |  |  |  |  |  |  |  |

**FIGURE 11**  Complete set of washability curves (for data given in Table III).

from about 1.2 g/cm$^3$ for low-ash particles to about 2.0 g/cm$^3$ for the high-ash particles, the liquids within this range of density provide conditions sufficient for the heavy-medium separation of coals.

Of the four types of dense media that can be considered in such a process—namely, organic liquids (carbon tetrachloride, bromoform), aqueous solutions ($ZnCl_2$, $CaCl_2$), aqueous suspensions of high-density solids (ferrosilicon, magnetite, barite, quartz sand), and air fluidized bed suspension (sand)—the first two are used only in laboratory washability studies, and only the third has found wide industrial applications.

The main difference between the former two and the latter two is stability: The first two are homogeneous liquids, while the latter are composed of fine solid particles suspended in water (or air), and as such are highly un-

stable. Magnetite has now become the standard industrial dense medium.

To achieve densities in the range 1.5–1.9, the concentration of magnetite in water must be relatively high. At this level of concentration, the suspension exhibits the characteristics of a non-Newtonian liquid, resulting in a formidable task in characterizing its rheological properties.

The principles of separation in dense media are depicted in Fig. 12, which shows a rotating drum dense-medium separator. The classification of static dense-medium separators offered 40 years ago by T. Laskowski is reproduced in Table IV.

Theoretically, particles of any size can be treated in the static dense-medium separators; but practically, the treatable sizes range from a few millimeters to about 150 mm.

**TABLE IV**  **Classification of Dense-Medium Separators$^a$**

| | Static | | | | | |
|---|---|---|---|---|---|---|
| | Shallow | | | Deep | | |
| Rotating-drum separates | Separation products removal with paddles | Separation products removal by belt conveyor | Separation products removal with scraper chains | Hydraulic removal of Separation products | Separation products removal by airlift | Mechanical removal of Separation products |
| | Examples | | | | | |
| Wemco drum | Link belt; SKB-Teska; Neldco (Nelson–Davis); Disa; Norwalt; Drewboy | Ridley–Sholes bath; Daniels bath | Tromp shallow; Dutch State Mines bath; McNally Tromp | McNally static bath; Potasse d'Alsace | Wemco cone; Humboldt | Chance; Barvoys; Tromp |

$^a$ After Laskowski, T. (1958). "Dense Medium Separation of Minerals," Wyd. Gorn. Hutnicze, Katowice. Polish text.
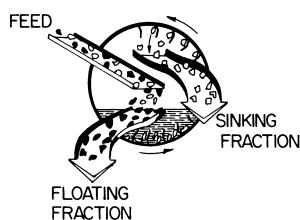
**FIGURE 12**  Principles of separation in dense media.

Since this process is unable to treat the full size range of raw coal, it is most important to remove the fines from the feed, because their presence in the circuit can increase the medium viscosity and increase the loss of media per ton of treated coal. The loss of magnetite in the coarse-coal cleaning is usually much below 1 kg/metric ton of washed product.

Dense-media static separators have a capability of handling high throughputs of up to 1000 metric ton/h.

Dynamic separators, which are the most efficient devices for cleaning intermediate size coal (50 down to 0.5 mm), include the dense-medium cyclone, two- and three-product Dynawhirlpool separators, the Vorsyl dense-medium separator, and the Swirl cyclone.

The most common are dense-medium cyclones (DMS) (Fig. 13a). Because the raw coal and medium are introduced tangentially into the cyclone, and the forces acting on the particles are proportional to $V^2/r$, where $V$ is the tangential velocity and $r$ the radius of the cylindrical section, the centrifugal acceleration is about 20 times greater than the gravity acceleration acting upon particles in the static dense-medium separator (this acceleration approaches 200 times greater than the gravity acceleration at the cyclone apex). These large forces account for the high throughputs of the cyclone and its ability to treat fine coal.

A DMS cyclone cut point $\delta_{50}$ is greater than the density of the medium and is closely related to the density of the cyclone underflow. The size distribution of the magnetite particles is very important for dense-medium cyclones, and magnetite that is 90% below 325 mesh (45 $\mu$m) was
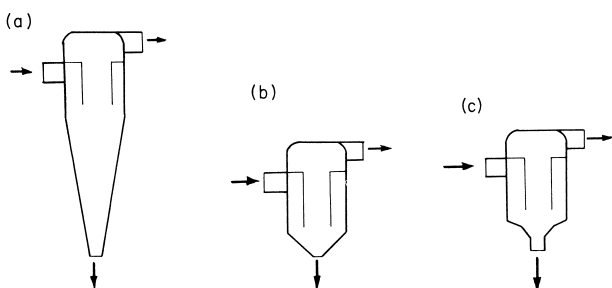


**FIGURE 13**  Cyclones in coal preparation (a) dense-medium cyclone; (b) water-only cyclone; (c) water-compound cyclone (Visman tricone).

found to provide the best results. Even finer magnetite (50% below 10 $\mu$m) was found to be essential for treating $-0.5 + 0.075$ mm coal. In treating such fine coal in a DMS, which technically is quite possible, the main problem that still awaits solving is medium recovery at reasonable cost.

Innovative dynamic separators are now being introduced that are able to treat the full size range (100–0.5 mm) in a single separating vessel. The Larcodems (*LA*rge, *CO*al, *DE*nse, *ME*dium, *SE*parator) is basically a cylindrical chamber that is mounted at 30° to the horizontal. Feed medium is introduced under pressure by an involute inlet at the bottom, and the raw coal enters separately through an axial inlet at the top. Clean coal is discharged through an axial inlet at the bottom end and reject expelled from the top via an involute outlet connected to a vortextractor.

The Tri-flo separator is a large-diameter three-product separator similar in operation to a Dyna Whirlpool.

## A. Dense Media

Any suitable fluid can be used as a dense medium, but only fine solid particles in water suspensions have found wide industrial applications. A good medium must be chemically inert, and must resist degradation by abrasion and corrosion, have high inherent density, be easily recovered from the separation products for reuse, and be cheap.

There is a close relationship between the medium density and viscosity: the lower the medium density (i.e., the solids content in suspension), the lower its viscosity. This, in turn, is related to medium stability (usually defined as the reciprocal of the settling rate) through the size and shape of solid particles: the finer the medium particles, the lower the settling rate (hence, the greater stability), but the higher the viscosity. Medium recovery is easier for coarser medium particles. Thus the medium density cannot be changed without the viscosity and stability being affected. For constant medium density, a higher specific gravity solid used as the medium reduces solid concentration in the dense medium and then reduces viscosity and stability; both are also affected by the selection of a spherical medium type (i.e., granulated ferrosilicon).

The term *viscosity*, as used above, is not accurate. Dense media at higher solid concentrations exhibit characteristics of non-Newtonian liquids, namely, Bingham plastic or pseudo-plastic behavior, as shown in Fig. 14.[1]

---

[1]The use of the modified rheoviscometer, which enables handling of unstable mineral suspensions, has recently revealed that the Casson Equation fits the flow curve for the magnetite suspension better than the typically used Bingham plastic model [see the special issue of *Coal Preparation* entirely devoted to magnetite dense media: [*Coal Preparation* (1990). **8**(3–4).]
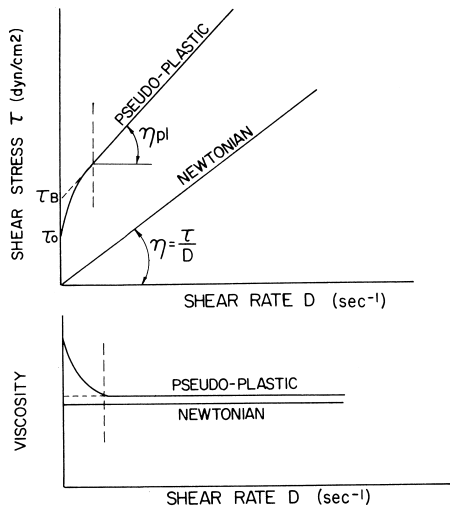
**FIGURE 14** Rheological curves for Newtonian and pseudo-plastic liquids.

As the rheological curves for magnetite dense-medium show, for such systems the shear stress is not proportional to the shear rate; therefore, such systems are not characterized by one simple value of viscosity, as in the case of Newtonian liquids.

The plastic and pseudo-plastic systems are described by the Bingham equation,

$$\tau = \tau_B + \eta_{pl}D, \tag{5}$$

and so both values $\tau_0$ ($\tau_B$) and $\eta_{pl}$, which can be obtained only from the full rheological curves, are needed to characterize dense-medium viscosity.

Figure 15 shows the effect of magnetite particle size and concentration on the medium rheology. As seen, the finer
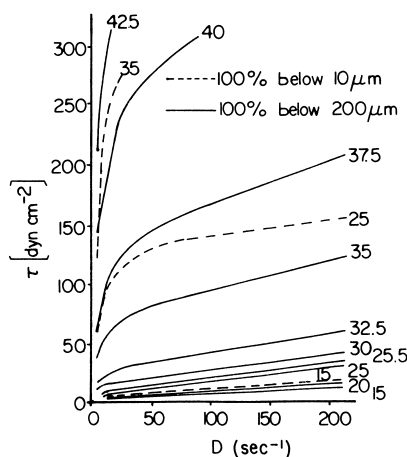


**FIGURE 15** Rheological curves for magnetite dense media. Solid line, magnetite particle size 100% below 200 $\mu$m; broken line, magnetite size 100% below 10 $\mu$m. [Adapted from Berghöfer, W. (1959). *Bergbauwissenschaften*, **6**(20), 493.]

magnetite suspension exhibits much higher viscosity. It is also seen that while magnetite dense-medium behaves as a Newtonian liquid at magnetite concentration lower than 15% by volume, it is clearly pseudoplastic at higher concentrations.

In a heavy-medium separation process, coal particles whose density is higher than the medium sink while the lower-density coal particles float; the separation efficiency and the through put of the separation device depend on the velocity of coal particles in a dense medium. The viscosity of a medium has little effect on the low-density coal particles or the high-density gangue particles, but becomes critical in the separation of material of a density equal to or near that of the medium; hence, a low viscosity must be maintained to separate near-density material at a high rate of feed.

The efficiency of separation ($E_p$) in dense-medium baths was found to depend on the plastic viscosity of the media, and a high yield stress ($\tau_0$) of the medium is claimed to cause elutriation of the finer particles into the float as they (and near-density particles) are unable to overcome the threshold shear stress required before the movement takes place. When a particle is held in a suspension, the yield stress is responsible for it; but when it is moving, its velocity is a function of plastic viscosity. It is understandable then that dispersing agents used to decrease medium viscosity may improve separation efficiency quite significantly. They may also decrease magnitite loses.

As the research on coal/water slurries shows, the suspension viscosity can also be reduced by close control of the particle size distribution; bimode particle distributions seem to be the most beneficial.

## VI. SEPARATION IN A WATER MEDIUM

### A. Jigs

Jigging is a process of particle stratification in which the particle rearrangement results from an alternate expansion and contraction of a bed of particles by pulsating fluid flow. The vertical direction of fluid flow is reversed periodically. Jigging results in layers of particles arranged by increasing density from the top to the bottom of the bed.

Pulsating water currents (i.e., both upward and downward currents) lead to stratification of the particles. The factors that affect the stratification are the following:

1. *Differential Acceleration.* When the water is at its highest position, the discard particles will initially fall faster than the coal.

2. *Hindered Settling.* Because the particles are crowded together, they mutually interfere with one another's

settling rate. Some of the lighter particles will thus be slowed by hindrance from other particles. This hindered settling effect is an essential feature, in that it helps to make separation faster; once the separation has been achieved, it helps to ensure that the material remains stratified.

3. *Consolidation Trickling.* Toward the end of the downward stroke, the particles are so crowded together that their free movement ceases, and the bed begins to compact. The further settling of the particles that leads to this compaction is referred to as *consolidation.* The larger particles lock together first, then the smaller particles consolidate in the interstices. Therefore, the larger particles cease to settle first, while the smaller particles can trickle through the interstices of the bed as consolidation proceeds.

### 1. Discard Extraction Mechanism

The method of discard extraction varies in different types of jig. Discard is discharged into the boot of the bucket elevator, and the perforated buckets carry it farther out. Some fine discard passes through the screen deck and is transported to the boot by screw conveyor, but the major portion is extracted over the screen plate. There are many methods of automatic control of discard.

### 2. Baum Jig

The most common jig used in coal preparation is the Baum jig, which consists of U-shaped cells (Fig. 16a). One limb of the cell is open and contains the perforated deck; the other terminates in a closed chamber, to and from which air is repeatedly admitted and extracted by rotary valves coupled to an air compressor. Particle separation is effected on the perforated deck.

### 3. Batac Jigs

For many years, jig technology was largely unchanged (i.e., Baum type), except for advances in the design of air valves and discard extraction mechanisms. A new concept, the Batac jig, in which the air chamber is placed beneath the screen plate, was developed initially in Japan and later in Germany. The principal feature of the design (Fig. 16b) is that instead of the box being U-shaped, the air chambers are placed beneath the washbox screen plates. In this way, the width of the jig can be greatly increased and its performance improved by more equal jigging action over the width of the bed.

Jigs traditionally treat coals in the size range 100–0.6 mm; however, for optimum results, coarse and fine coals should be treated in separate jigs. Fine-coal jigs have their own "artificial" permanent bed, usually feldspar.

## B. Concentrating Tables

A concentrating table (Fig. 17) consists of a riffled rubber deck carried on a supporting mechanism, connected to a head mechanism that imparts a rapid reciprocating motion
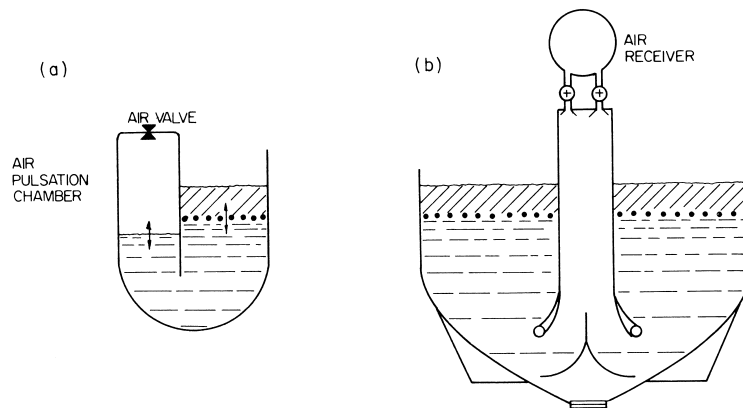


**FIGURE 16** Comparison of (a) Baum and (b) Batac jigs. [After Williams, D. G. (1981). *In* "Coal Handbook" (Meyers, R. A. ed.), p. 265, Marcel Dekker, New York.]

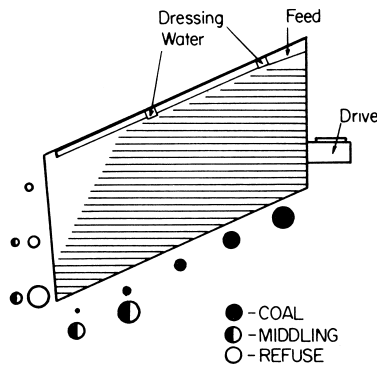| Type | Baum jig | Batac jig |
|---|---|---|
| Year introduced | 1892 | 1965 |
| Present width | 2.1 m | 7.0 m |
| Present area | 8.5 m$^2$ | 42.0 m$^2$ |
| Present capacity | 450 metric tons/hr | 730 metric ton/hr. |

**FIGURE 17** Distribution of concentrating table products by particle size and specific gravity. ●, coal; ◑, middling; ○, refuse.

in a direction parallel to the riffler. The side slope of the table can be adjusted. A cross flow of water (dressing water) is provided by means of a launder mounted along the upper side of the deck. The feed enters just ahead of the water supply and is fanned out over the table deck by differential motion and gravitational flow. The particles are stratified in layers by three forces:

1. Friction or adhesion between deck and particle
2. Pressure of the moving water
3. Differential acceleration due to the table action.

The clean coal overflows the lower side of the table, and the discard is removed at the far end. The device cleans efficiently over the size range 5–0.5 mm.

## C. Water-Only Cyclones

The water-only cyclone (WOC) (first developed by the Dutch State Mines) is a cylindro-conical unit with an included apex angle up to 120° (Fig. 13b). Unlike the dense-medium cyclone, it uses no medium other than water; for this reason it is sometimes referred to as the autogenous cyclone. Geometrically, the principal differences from the dense-medium cyclone are the greater apex angle and the much longer vortex finder of the water-only cyclone.

The compound water cyclone (Visman Tricon) (Fig. 13c) is a variation of the water-only cyclone. Its distinguishing features are the three sections of the cone part with different apex angles, the first being 135°, the second 75°, and the third 20°.

Water-only cyclones do not make a sharp separation, but are commonly used as a preconcentrator or "rougher." Their efficiency can be increased by the addition of a second stage. There is also a sizing effect with water-only cyclones, and particles finer than 100 $\mu$m tend to report to the overflow regardless of ash content or relative density.

## D. Spirals

Spiral separators have been used for many years, since the introduction of the Humphrey unit. The early spirals had simple profiles and were used in easy separations. Later development of Reichert spirals, suitable for a wide variety of applications, increased the use of spirals.

The effective size range for coal spiral operation ($-3$ mm $+0.075$ $\mu$m) coincides largely with the sizes that lie between those most effectively treated by heavy-media cyclones and those best treated by froth flotation. Thus, the principal areas of application might be substitution for water-only cyclones, for fine heavy-media cyclone separation, and for coarse froth flotation.

The capacity of Reichert coal spirals is about 2 tph. $E_p$ values for Reichert Mark 9 and Mark 10 spirals have recently been reported in the range of 0.14–0.18, but they were found to be much lower for the combined product plus middlings than for the product alone. Reichert spirals offer simple construction requiring little maintenance plus low capital cost and low operating costs. Reichert-type LD Mark 10 and Vickers spirals have recently been shown to be able to achieve lower $E_p$ values for the finer fractions (0.1 mm) but higher $E_p$ for the coarse fractions (1 mm) than water-only cyclone. In general, the spiral is less affected by a change in particle size than WOC, but most importantly, the spirals tolerate increasing amounts of clay in the feed with little or no change in separation efficiency, in contrast to the WOC performance, which deteriorates with increasing amounts of clays. This indicates that the spiral is less influenced by increase in viscosity and is better suited for raw coals with significant proportions of clays.

## VII. FLOTATION

Typically, coal fines below 28 mesh (0.6 mm) are cleaned by flotation. In some plants the coarser part of such a feed [+100 mesh (150 $\mu$m)] is treated in water-only cyclones, with only the very fine material going to flotation.

The coal flotation process is based on differences in surface properties between hydrophobic low-ash coal particles and hydrophilic high-ash gangue. Coals of different rank have various chemical composition and physical structure, and therefore their surface properties and floatability change with coalification. While metallurgical coals float easily and may require only a frother, flotation of lower-rank subbituminous coal and lignites (as well as high-rank anthracites) may be very difficult. In such a process one may need not only large amounts of oily collector but also a third reagent, the so-called promoter (Table V).

**TABLE V   Coal Flotation Reagents**[a]

| Type | Examples | Remarks |
|---|---|---|
| Collectors | Insoluble in water, oily hydrocarbons, kerosene, fuel oil | Used in so-called emulsion flotation of coal, in which collector droplets must attach to coal particles |
| Frothers | Water-soluble surfactants; aliphatic alcohols; MIBC | To stabilize froth; adsorb at oil/water interface and onto coal; have some collecting abilities |
| Promoters | Emulsifiers | Facilitate emulsification of oily collector and the attachment of the collector droplets to oxidized and/or low-rank coal particles |
| Depressants/dispersants | Organic colloids: dextrin, carboxymethyl cellulose, etc. | Adsorb onto coal and make it hydrophilic |
| Inorganic salts | NaCl, CaCl$_2$, Na$_2$SO$_4$, etc. | Improve floatability; in so-called salt flotation process may be used to float metallurgical coals, even without any organic reagents; are coagulants in dewatering. |

[a] Adapted from Klassen, V. I. (1963). "Coal Flotation," Gogortiekhizdat, Moscow. (In Russian.)

The surface of coal is a hydrophobic matrix that contains some polar groups as well as hydrophobic inorganic impurities. Coal surface properties are determined by

1. The coal hydrocarbon skeleton (related to the rank)
2. The active oxygen content (carboxylic and phenolic groups)
3. Inorganic matter impurities

Only the third term, the content of inorganic matter, is related to coal density. Therefore, in general there is no relationship between coal washability and coal floatability; such a relationship can be observed only in some particular cases.

Results obtained by various researchers agree that the most hydrophobic are metallurgical, bituminous coals. The contact angle versus volatile matter content curve, which shows wettability as a function of the rank (Fig. 18), is reproduced here after Klassen. As seen, lower-rank coals are more hydrophilic, which correlates quite well with oxygen content in coal, shown here after Ihnatowicz (Fig. 19). Comparison of Fig. 18 with Fig. 19 also indicates that the more hydrophilic character of anthracites cannot be explained on the same basis.
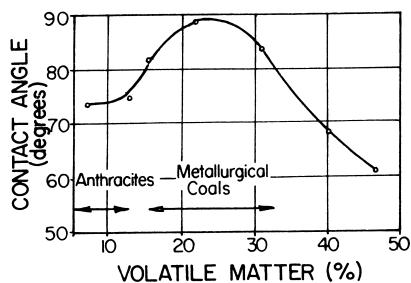
Two types of reagents are traditionally used in the flotation of coal:

Water-insoluble, oily hydrocarbons, as collectors
Water-soluble surfactants, as frothers

Insolubility of the collector in coal flotation requires prior emulsification or long conditioning time. On the other hand, the time of contact with the frother should be as short as possible to avoid unnecessary adsorption of frother by highly porous solids, such as coal.

In the flotation of coal, the *frother* (a surface-active agent soluble in water) adsorbs not only at the water/gas interface but also at the coal/water and the oil/water interfaces; it may facilitate, to some extent, the attachment of
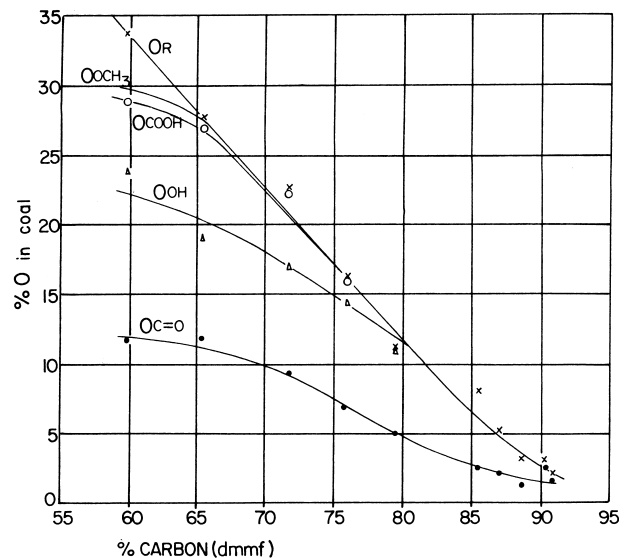


**FIGURE 18**   Effect of the rank on coal wettability. [From Klassen, V. I. (1963). "Coal Flotation," Gosgortiekhizdat, Moscow. (In Russian).]



**FIGURE 19**   Oxygen functional groups in coals. (Large O stand for oxygen.) [From Ihnatowicz, A. (1952). *Bull. Central Research Mining Institute*, Katowice, No. 125. (In Polish).]
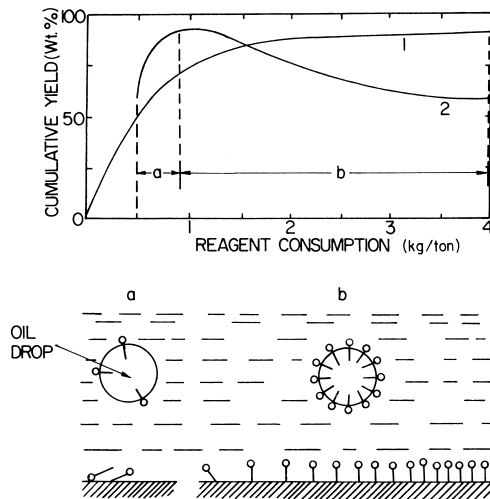
**FIGURE 20** Effect of surface-active agent (frother) on the flotation of coal with water-insoluble oily collector. Curve 1 is for flotation with kerosene; curve 2 for kerosene and *n*-octyl alcohol. [From Melik-Gaykazian, V. I., Plaksin, I. N., and Voronchikhina, V. V. (1967). *Dokl. Akad. Nauk SSSR*, **173,** 883.]

oily droplets to coal particles. This can lead to a substantial improvement in flotation, as shown in Fig. 20.

In order to float oxidized and/or low-rank coals, new reagents called *promotors* have recently been tested and introduced into the industry. These are emulsifying agents that facilitate emulsification of the oily collector and the attachment of oily droplets to coal particles.

The reduction of sulfur is one of the principal benefits of the froth flotation process. Of the two types of sulfur in coal (i.e., inorganic and organic), only the inorganic sulfur (mainly represented by pyrite) can be separated from coal by physical methods. The floatability of coal-pyrite is, however, somewhat different from that of ore-pyrite, and its separation from coal presents a difficult problem. The research based on the behavior of ore-pyrite in the flotation process—namely, its poor flotability under alkaline conditions—did not result in a development of a coal-pyrite selective flotation process. The two-stage reverse flotation process proved to be much more successful. In this process, the first stage is conventional flotation with the froth product comprising both coal and pyrite. This product is reconditioned with dextrin, which depresses coal in the second stage, and is followed by flotation of pyrite with xanthates.

The fine-coal cleaning circuits in the newest plants frequently comprise water-only cyclones and flotation (Fig. 21). Such an arrangement is especially desirable in cleaning high-sulfur coals, because the difference in specific gravity between coal particles (1.25–1.6) and pyrite (5) is extremely large, and flotation does not discriminate well between coal and pyrite. Advantages of such an arrangement are clearly seen in Table VI, quoted after Miller, Podgursky, and Aikman.

## VIII. SEPARATION EFFICIENCY

The yield and quality of the clean-coal product from an industrial coal preparation plant and the theoretical yield and quality determined from washability curves are known to be different. In the ideal cleaning process, all coal particles lower in density than the density of separation would be recovered in the clean product, while all material of greater density would be rejected as refuse. Under these conditions the product yield and quality from the actual concentration process and the yield and quality expected from the washability curves would be identical.

The performance of separators is, however, never ideal. As a result, some coal particles of lower than the separation density report to rejects, and some high-ash particles of higher than the separation density report to clean coal. These are referred to as *misplaced material*.

Coal particles of density well below the density of separation and mineral particles of density well above the density of separation report to their proper products: clean coal and refuse. But as the density of separation is approached, the proportion of the misplaced material reporting to an improper product increases rapidly.

Tromp, in a study of jig washing, observed that the displacement of migrating particles was a normal or
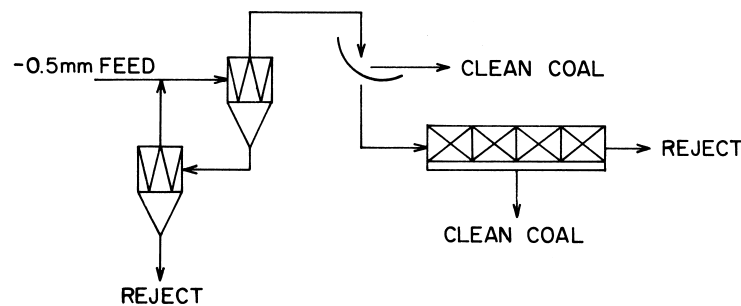


**FIGURE 21** Fine-coal cleaning circuit.

**TABLE VI  Effectiveness of Hydrocycloning and Flotation in Ash and Sulfur Removal**[a]

| Particle size, mesh | Ash removal (%) | | Sulfur removal (%) | |
|---|---|---|---|---|
| | Hydrocyclone | Flotation (single stage) | Hydrocyclone | Flotation (single stage) |
| $28 \times 100$ | 60–65 | 50–60 | 70–80 | 50–60 |
| $100 \times 200$ | 40–45 | 40–45 | 60–70 | 30–40 |
| $200 \times 325$ | 15–18 | 40–45 | 35–40 | 25–30 |
| $325 \times 0$ | 0–5 | 50–55 | 8–10 | 20–25 |

[a] After Miller, F. G., Podgursky, J. M., and Aikman, R. P. (1967). *Trans. AIME* **238,** 276.

near-normal frequency (gaussian curve), and from this observation the partition curve (distribution, Tromp curve) in the form of an ogive was evolved.

The partition curve, the solid line in Fig. 22a, illustrates the ideal separation case ($E_p = 0$), and the broken-line curve represents the performance of a true separating device. The shaded areas represent the misplaced material.
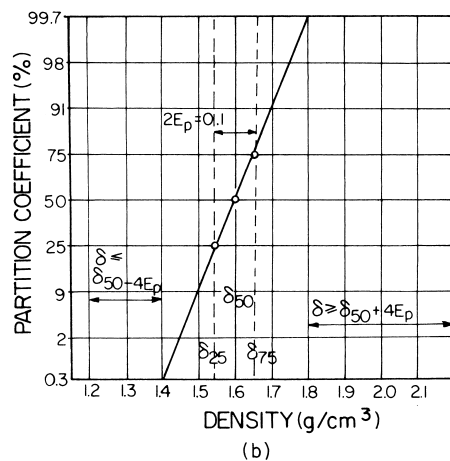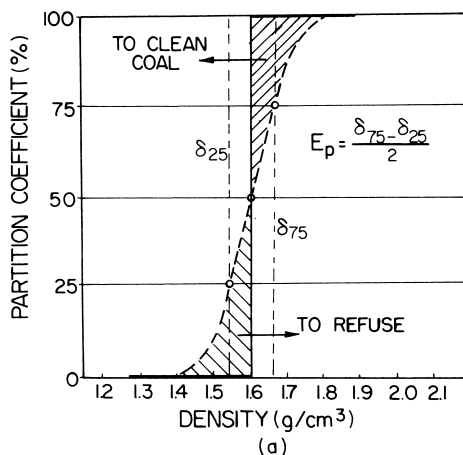


FIGURE 22  (a) Partition curve plotted on linear graph paper and (b) its anamorphosis plotted on a probability net.

The curves are plotted according to European convention and represent the percent of feed reporting to reject. In American practice, Tromp curves usually give the percent of feed reporting to washed coal.

The Tromp curve from the mathematical point of view is a cumulative distribution curve and as such can be linearized on probability graph paper. Such anamorphosis is produced by plotting the partition coefficients on a probability scale versus specific gravity on a linear scale (Fig. 22b) for dense-media separation, and versus log ($\delta - 1$) for jigs.

To determine the partition curve for a cleaning operation, one needs the yield of clean coal from this operation and the results of float–sink tests for both products—that is, for the clean coal and the refuse. Such data allow the reconstituted feed to be calculated, and from this can be found the partition coefficients, which give the percentage of each density fraction reporting to reject. As seen in Fig. 22b, the particles with densities below $\delta_{50} - 4E_p$ and the particles of density above $\delta_{50} + 4E_p$ report entirely to their proper products. The density fractions within $\delta_{50} \pm 4E_p$ are misplaced. Material of density very close to $\delta_{50}$ (near-density material) is misplaced the most. As postulated by Tromp, 37.5% of fractions within $\delta_{50}–\delta_{25}$ and $\delta_{75}–\delta_{50}$ are misplaced (this corresponds to $\delta_{50} \pm E_p$), and this percentage falls off drastically with the distance of the actual density fraction from the $\delta_{50}$ density.

Figure 23 shows partition curves for the major U.S. coal-cleaning devices. As seen, the sharpness of separation in dense-media separators is much better than in jigs or water-only hydrocyclones. Figure 24 shows $E_p$ values plotted versus the size of treated particles for various cleaning devices. $E_p$ values for the dense-medium bath and dense-medium cyclone are in the range of 0.02–0.04 for particles larger than 5 mm; $E_p$ values of jigs and concentrating tables are in the range of 0.08–0.15; and for water-only cyclones, $E_p$ values exceed 0.2. As seen, in all cases the efficiency of separation as given by $E_p$ values decreases sharply for finer particles.
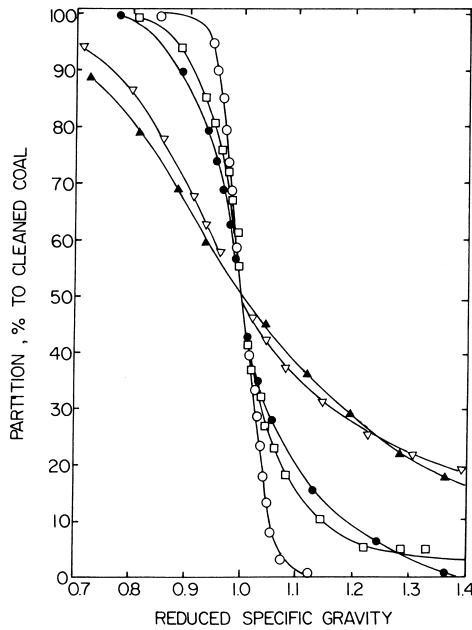
**FIGURE 23** Performance of gravity separators. ▲, hydrocyclones, $\frac{1}{4}$ in. $\times$ 200 mesh; ▽, air tables, 2 in. $\times$ 200 mesh; ●, jigs: Baum, 6 in. $\times \frac{1}{4}$ in.; Batac $\frac{3}{4}$ in. $\times$ 28 mesh; □, concentrating tables, $\frac{3}{4}$ in. $\times$ 200 mesh; ○, dense-medium separators: cyclone, $\frac{3}{4}$ in. $\times$ 28 mesh; vessel, 6 in. $\times \frac{1}{4}$ in. [From Killmeyer, R. P. "Performance characteristics of coal-washing equipment: Baum and Batac jigs," U.S. Department of Energy, RI-PMTC-9(80).]

Another index developed to characterize the sharpness of separation is the imperfection $I$.

$$I = E_{\mathrm{p}}/\delta_{50} \tag{6}$$

for dense-medium separation, and

$$I = E_{\mathrm{p}}/\delta_{50} - 1 \tag{7}$$

for jigs.

According to some authors, $I$ values vary from 0.07 for dense-medium cyclones and 0.175 for jigs to above 0.2 for flotation machines.

Conventional float and sink methods for the derivation of the partition curve are expensive and time-consuming. In the diamond and iron ore industries, the density tracer technique has been developed to evaluate the separation efficiency. This technique has also been adopted for studies of coal separation.

The tracers, usually plastic cubes prepared to known specific gravities rendered identifiable through color coding, are introduced into a separating device, and on recovery from the product and reject streams, are sorted into the appropriate specific gravity fractions and counted. This allows the points for the partition curve to be calculated. The technique, as described above, was adopted at the Julius Kruttschnitt Mineral Research Centre, Brisbane, while tracers made from plastic–metal composites that can

be detected with metal detectors mounted over conveyer belts, known as the Sentrex system, were developed in the United States.

## IX. ANCILLARY OPERATIONS

There are many ancillary operations in coal preparation that are similar to operations in the mineral processing industry—namely

1. Mechanical dewatering
   a. Vibrating basket centrifuges
   b. Screen bowl centrifuges
   c. Solid bowl centrifuges
   d. Disk filters
   e. Drum filters
   f. Belt filters
   g. Plate and frame filter presses
   h. Thickeners, conventional and high capacity
2. Thermal dewatering
   a. Fluidized bed drying
   b. Rotary dryers
3. Stocking and blending systems
4. Automatic product loadout systems
5. Computer startup and process monitoring

## X. COAL PREPARATION FLOWSHEETS

In ROM coal, a large portion of both coal and shale is already liberated sufficiently to permit immediate concentration. Because dewatering of coarse coal is much more efficient and cheaper than that of fines, and because many users prefer larger size fractions for their particular processes, the preparation of feed before cleaning has traditionally consisted of as little crushing as possible. Therefore, coal preparation consists mostly of gravity concentration and to some extent flotation, complemented by screening and solid/liquid separation auxiliary processes.

The precrushing needed to reduce the top size of the material is widely applied by means of rotary breakers. This process, as already pointed out, is based on the selective breakage of fragile coal and hard inorganic rock, and combines two operations: size reduction and preconcentration (because it rejects larger pieces of hard rock). A further development in coal preparation has been the application of the three-product separators to treat coarse and intermediate size fractions. The recrushed middlings produced in such a cleaning are reprocessed, together with the fine part of the feed, increasing the total coal recovery.

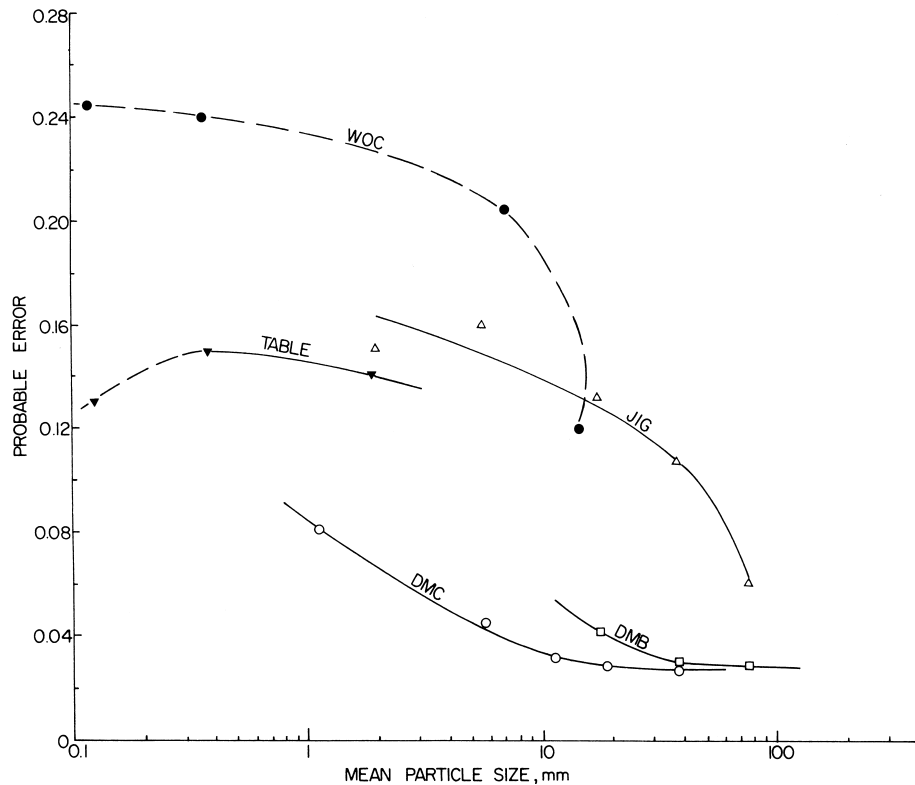This complete concentration sequence is typical for processing a metallurgical coal.

**FIGURE 24**  Probable error $E_p$ vs mean size of the treated coal for dense-media baths (DMB), dense-media cyclones (DMC), jigs, concentrating tables, and water-only cyclones (WOC). [After Mikhail, M. W., Picard, J. L., and Humeniuk, O. E. (1982). "Performance evaluation of gravity separators in Canadian washeries." Paper presented at *2nd Technical Conference on Western Canadian Coals*, Edmonton, June 3–5.].

Coal preparation plants are classified according to the extent to which coal is cleaned, as follows (see Fig. 25).

*Type I.* Only the large size coal is reduced and screened into various size fractions.

*Type II.* The plant cleans coarse coal by gravity methods. The gangue content in coal is reduced in the form of coarse reject from the Bradford breaker and high-gravity rejects. The yield of saleable product is 75–85%.

*Type III.* The plant cleans coarse and intermediate size coal, leaving the fines untreated.

*Type IV.* These modern coal preparation plants clean the full size-range of raw coal.

*Type V.* The plant produces different saleable products that vary in quality, including deep-cleaned coal that is low in sulfur and ash.

*Type VI.* The plant incorporates fine grinding to achieve a high degree of liberation, using froth flotation (oil agglomeration) and water-only cyclones (or spirals).

Type VI represents a very important stage in the development of coal preparation strategy. Types I–V are all predominantly based on coarse coal cleaning, with the amount of fines reduced to a minimum; but in the type VI plant all coarse coal cleaning operations are aimed at eliminating coarse hard gangue and constitute a pretreatment stage, with the final cleaning following after fine crushing and grinding. This future development in coal preparation technology is aimed at cleaning and utilization of thermal coal, and is mainly a result of the new environmental-protection restrictions.

In the future, type VI coal preparation plants will be combined with coal/water slurry technology, and such slurries will to some extent replace liquid fuels.

## XI. ON-LINE ANALYSIS

On-line measurements of parameters such as ash, moisture, and sulfur use an inferential technique in which some property can be measured that in turn relates to the parameter in the material.

Most on-line techniques use some form of electromagnetic radiation. When radiation passes through matter, there is a loss of intensity by absorption and scattering.
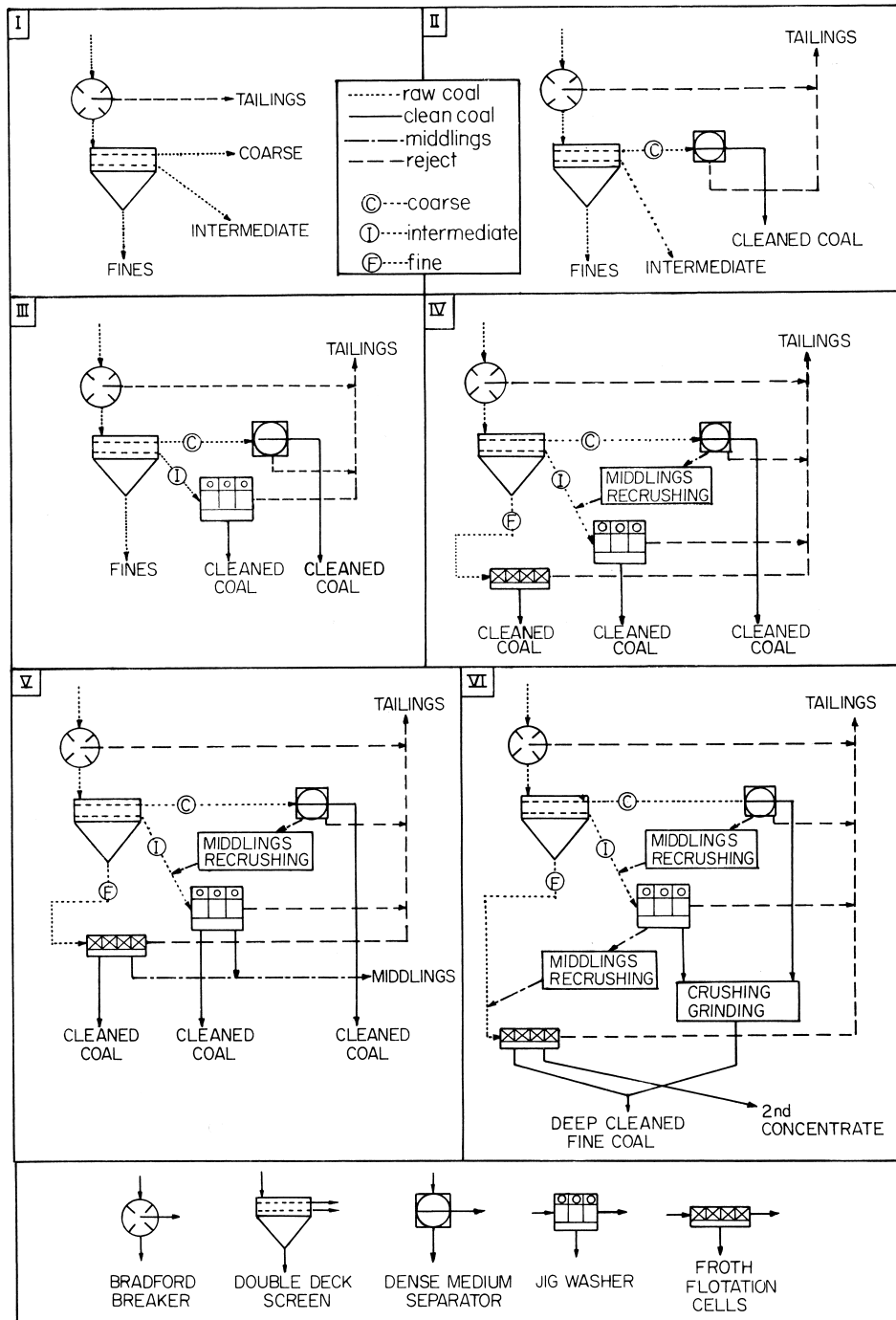
**FIGURE 25** Classification of coal preparation circuits.

## A. Ash Analysis

Several devices irradiate the coal with $\gamma$- or X-rays, and the backscatter is measured. At relatively low energies (<100 keV) the change in scattering coefficient with atomic number more than offsets the change in mass absorption coefficient; thus the backscatter can be correlated to mean atomic number and to the ash for a given coal. The backscattered radiation intensity is a function of both bulk density and mean atomic number. Thus, instruments using back scatter must be designed to ensure that a consistently packed sample of coal is presented for analysis. Some of the gauges using this principle are the Gunston Sortex, the Berthold–Humbolt–Wedag, and the Wultex.

When $\gamma$ radiation above energy levels of 1022 keV interacts with matter, positron–electron pairs are generated. The positron is rapidly annihilated by collision with an electron and releases gamma rays, the intensity of which is strongly dependent upon atomic number. The gauge that uses the pair production principle is the COALSCAN Model 4500, which uses Radium (226) emitting $\gamma$ radiation at both the 1760 and 2200 keV levels. The devices that use absorption use a low energy level source that is sensitive to the mean atomic number (2) of the material. There is an eightfold difference in the mass absorption coefficient between the organic portion of coal ($Z = 6$) and the mineral matter ($Z = 12$). By comparing the differences in absorption between low and high energy beams, the mean atomic number can be determined, and from this the ash composition can be inferred. This is the underlying principle behind such ash gauges as the COALSCAN Model 3500, the SAI 400, and the ASHSCAN slurry gauge.

Another form of radiation used for coal analysis is neutron activation. This method, known as Prompt Gamma Neutron Activation Analysis (PGNAA), has the distinct advantage of being able to conduct a complete elemental analysis rather than just total ash. This method permits the determination of the sulfur, carbon, and hydrogen composition, which also permits the calculation of specific energy. Several commercial PGNAA gauges are available—SAI Models 100 and 200, MDH Motherwell, and the Gamma Metrics.

The determination of ash in coal slurries uses the basic principle of measurement of those for solid coal. The extra problem is that the entrained air could seriously affect the accuracy of these devices. Various methods are used to compensate for this, such as pressurizing the system to collapse the air bubbles (ASHSCAN) or using a neutron attenuation measurement to correct for possible air entrainment (AMDEL).

## B. Moisture Analysis

There are several methods of on-line moisture analysis:

1. Microwave attenuation/phase shift: Involves transmitting a microwave beam through a given layer of coal (conveyor belt). The maximum absorption of the energy occurs at 20 GHz and the energy absorbed by the material is calibrated against moisture content. However, simple attenuation is found to be sensitive to other properties such as particle size, bulk density, and salt content. To overcome these problems, the phase shift of the microwave energy is measured after passage through the sample or the attenuation is measured at more than one frequency.

2. Capacitance measurements: Based on the principle that electrical conductivity of coal increases with the quantity of moisture. However, it is only applicable for the measurement of surface moisture.

3. Nuclear magnetic resonance: This method measures the hydrogen in the coal, and is able to distinguish between hydrogen in water and in hydrocarbons.

4. Neutron moisture gauges: Neutron transmission and scatter methods can be used to measure total hydrogen content, and assumes a constant hydrocarbon hydrogen.

## C. Sulfur Analysis

The PGNAA method is generally used to measure elemental sulfur. Commercial gauges available are the MDH-Motherwell ELAN, Science Application International-Nucoalyzer, and the Gamma-Metrics.

## XII. RESEARCH INTO NEW BENEFICIATION PROCESSES

### A. Gravity

#### 1. Otisca Process

The Otisca process is a waterless separation process employing $CCl_3F$ (with certain additives) as a heavy liquid. In this process, raw coal with surface moisture up to 10% is subjected to separation in a static bath. The medium is recovered by evaporation from the separation products and is recycled. The Otisca process plant is completely enclosed, and the chemicals used are nonflammable, odor free, and noncorrosive.

The $E_p$ calculated from the 8-hr plant tests for different size feeds gave the results (shown in Table VII). As seen, $E_p$ values increase for very fine material ($28 \times 100$ mesh and $100 \times 325$ mesh) but are still better than for many other methods. This indicates very favorable low-viscosity conditions in the process, which were confirmed by rheological measurements. The favorable rheology of the true organic heavy-liquid slurries has further been confirmed in experiments on the use of Freon-113 (1,1,2-trichloro-1,2,2-trifluoro ethane) in dense-medium

**TABLE VII** $E_p$ **Values for Otisca Process Calculated for Different Size Feeds**[a]

| Composite feed | Organic efficiency % | Probable error, $E_p$ | Specific gravity |
|---|---|---|---|
| $\frac{3}{8} \times \frac{1}{4}$ in. | 100 | 0.008 | 1.48 |
| $\frac{1}{4}$ in. $\times$ 28 mesh | 99 | 0.015 | 1.49 |
| $28 \times 100$ mesh | 98 | 0.175 | 1.57 |
| $100 \times 325$ mesh | 96 | 0.260 | 1.80 |
| $\frac{3}{8}$ in. $\times$ 325 mesh | 98 | 0.023 | 1.49 |

[a] After Keller, D. V. (1982). *In* "Physical Cleaning of Coal" (Liu, Y. A., ed.), p. 75, Marcel Dekker, New York.

cyclones. The separation in cyclones, with use of Freon-113 as a heavy medium, gave $E_p = 0.009$ for 28 mesh $\times$ 0 coal, $E_p = 0.045$ for 200 mesh $\times$ 0, and $E_p = 0.100$ for 400 mesh $\times$ 0 coal.

## B. Magnetic Separation

This process is being investigated in the separation of coal and pyrite. The difference in magnetic susceptibility of coal and pyrite is very small; therefore, the most promising application is for high-gradient magnetic separators (HGMS) or the pretreatment of coal, which could increase the susceptibility of coal mineral matter.

Because pyrite liberation usually requires very fine grinding (e.g., below 400 mesh), high-gradient magnetic separation, the process developed to upgrade kaolin clays by removing iron and titanium oxide impurities, seems to be particularly suitable for coal desulfurization. Wet magnetic separation processes offer better selectivity for the fine dry coal that tends to form agglomerates. Experiments on high-gradient wet magnetic separation have so far indicated better sulfur rejection at fine grinding.

Perhaps the most interesting studies are on the effect of coal pretreatment, which enhances the magnetic properties of the mineral matter, on the subsequent magnetic separation. It has, for example, been shown that microwave treatment heats selectively pyrite particles in coal without loss of coal volatiles, and converts pyrite to pyrrhotite, and as a result, facilitates the subsequent magnetic separation.

In the Magnex process, crushed coal is treated with vapors of iron carbonyl [$Fe(CO)_5$], followed by the removal of pyrite and other high-ash impurities by dry magnetic separation. Iron carbonyl in this process decomposes on the surfaces of ash-forming minerals and forms strongly magnetic iron coatings; the reaction with pyrite leads to the formation of pyrrhotite-like material.

In 1970, Krukiewicz and Laskowski described a magnetizing alkali leaching process, which was applied to convert siderite particles into magnetic $\gamma$-$Fe_2O_3$ and $Fe_3O_4$. A similar process has recently been tested for high-sulfur coal, and it was reported that, after pretreatment of the coal in 0.5-M NaOH at 85°C and 930 kPa (135 psi) air pressure for 25 min, more than 50% of the coal sulfur was rejected under the same conditions in which the same high-gradient magnetic separation had removed only 5% sulfur without the pretreatment.

## C. Processes Based on Surface Properties

### 1. Oil Agglomeration

As in coal flotation, oil agglomeration takes advantage of the difference between the surface properties of low-ash coal and high-ash gangue particles, and can cope with even finer particles than flotation. In this process, coal particles are agglomerated under conditions of intense agitation. The following separation of the agglomerates from the suspension of the hydrophilic gangue is carried out by screening.

The most important operating parameters in this process are the amount of oil, intensity of agitation (agitation time and agitation speed), and oil characteristics.

The amount of oil that is required is in the range of 5–10% by weight of solids. Published data indicate that the importance of agitation time increases as oil density and viscosity increase, and that the conditioning time required to form satisfactory coal agglomerates decreases as the agitation is intensified. Because the agitation initially serves to disperse the bridging oil to contact the oil droplets and coal particles, higher shear mixing with a lower viscosity bridging liquid is desirable in the first stage (microagglomeration), and less intense agitation with the addition of higher viscosity oil (macroagglomeration) is desirable in the second stage. Viscous oil may produce larger agglomerates that retain less moisture. With larger oil additions (20% by weight of solids), the moisture content of the agglomerated product can be well below 20% and may be reduced even further if tumbling is used in the second stage instead of agitation.

The National Research Council of Canada developed the spherical agglomeration process in the 1960s. This process takes place in two stages: First, the coal slurry is agitated with light oil in high shear blenders where microagglomerates are formed; then the microagglomerates are subjected to dewatering on screen and additional pelletizing with heavy oil.

Shell developed a novel mixing device to condition oil with suspension. Application of the Shell Pelletizing Separator to coal cleaning yielded very hard, uniform in size, and simple to dewater pellets at high coal recoveries. The German Oilfloc Process was developed to treat the high-clay, $-400$-mesh fraction of coal, which is the product of flotation feed desliming. In the process developed by the Central Fuel Research Institute of India, coal slurry is treated with diesel oil (2% additions) in mills and then agglomerated with 8–12% additions of heavy oil.

It is known that low rank and/or oxidized coals are not a suitable feedstock for beneficiation by the oil agglomeration method. The research carried out at the Alberta Research Council has shown, however, that bridging liquids, comprising mainly bitumen and heavy refinery residues are very efficient in agglomeration of thermal bituminous coals. Similar results had earlier been reported in the flotation of low rank coals; the process was much improved when 20% of no. 6 heavy oil was added to 2 fuel oil.

### 2. Otisca T-Process

This is another oil agglomeration process that can cope with extremely fine particles. In this process, fine raw coal, crushed below 10 cm, is comminuted in hammer crushers to below 250 $\mu$m and mixed with water to make a 50% by weight suspension; this is further ground below 15 $\mu$m and then diluted with water to 15% solids by weight. Such a feed is agglomerated with the use of Freon-113, and the coal agglomerates and dispersed mineral matter are separated over screen. The separated coal-agglomerated product retains 10–40% water and is subjected to thermal drying; Freon-113, with its boiling point at 47°C, evaporates, and after condensing is returned liquified to the circuit. The product coal may retain 50 ppm of Freon and 30–40% water.

Various coals cleaned in the Otisca T-Process contained in most cases below 1% ash, with the carbonaceous material recovery claimed to be almost 100%. Such a low ash content in the product indicates that very fine grinding liberates even micromineral matter (the third level of heterogeneity); it also shows Freon-113 to be an exceptionally selective agglomerant.

### 3. Selective Agglomeration During Pipelining of Slurries

In some countries, for example in Western Canada, the major obstacles to the development of a coal mining industry are transportation and the beneficiation/utilization of fines. Selective agglomeration during pipelining offers an interesting solution in such cases. Since, according to some assessments, pipelining is the least expensive means for coal transportation over long distances, this ingenious invention combines cheap transportation with very efficient beneficiation and dewatering. The Alberta Research Council experiments showed that selective agglomeration of coal can be accomplished in a pipeline operated under certain conditions. Compared with conventional oil agglomeration in stirred tanks, the long-distance pipeline agglomeration yields a superior product in terms of water and oil content as well as the mechanical properties of the agglomerates. The agglomerated coal can be separated over a 0.7-mm screen from the slurry. The water content in agglomerates was found to be 2–8% for metallurgical coals, 6–15% for thermal coals (high-volatile bituminous Alberta), and 7–23% for subbituminous coal. The ash content of the raw metallurgical coal was 18.9–39.8%, and the ash content of agglomerates was 8–15.4%. For thermal coals the agglomeration reduced the ash content from 19.8–48.0 to 5–12.8%, which, of course, is accompanied by a drastic increase in coal calorific value. Besides transportation and beneficiation, the agglomeration also facilitates material handling; the experiments showed

that the agglomerates can be pipelined over distances of 1000–2000 km.

### 4. Flotation

Flotation has progressed and developed over the years; recent trends to achieve better liberation by fine grinding have intensified the search for more advanced means of improving selectivity. This involves not only more selective flotation agents but also better flotation equipment. Since the froth product in conventional flotation machines contains entrained fine gangue, which is carried into the froth with feed water, the use of froth spraying was suggested in the late 1950s to eliminate this type of froth "contamination." The flotation column patented in Canada in the early 1960s and marketed by the Column Flotation Company of Canada, Ltd., combines these ideas in the form of wash water supplied to the froth. The countercurrent wash water introduced at the top of a long column prevents the feed water and the slimes that it carries from entering an upper layer of the froth, thus enhancing selectivity.

Of many variations of the second generation columns, perhaps the best known is the Flotaire column marketed by the Deister Concentrator Company.

The microbubble flotation column (Microcel) developed at Virginia Tech is based on the basic premise that the rate ($k$) at which fine particles collide with bubbles increases as the inverse cube of the bubble size ($D_b$), i.e., $k \sim 1/D_b^3$. In the Microcel, small bubbles in the range of 100–500 $\mu$m are generated by pumping a slurry through an in-line mixer while introducing air into the slurry at the front end of the mixer. The microbubbles generated as such are injected into the bottom of the column slightly above the section from which the slurry is with drawn for bubble generation. The microbubbles rise along the height of the column, pick up the coal particles along the way, and form a layer of froth at the top section of the column. Like most other columns, it utilizes wash water added to the froth phase to remove the entrained ash-forming minerals. Advantages of the Microcel$^{TM}$ are that the bubble generators are external to the column, allowing for easy maintenance, and that the bubble generators are nonplugging. An 8-ft diameter column uses four 4-in. in-line mixers to produce 5–6 tons of clean coal from a cyclone overflow containing 50% finer than 500 mesh.

Another interesting and quite different column was developed at Michigan Tech. It is referred to as a static tube flotation machine, and it incorporates a packed-bed column filled with a stack of corrugated plates. The packing elements arranged in blocks positioned at right angles to each other break bubbles into small sizes and obviate the need for a sparger. Wash water descends through the same flow passages as air (but countercurrently) and removes

entrained particles from the froth product. It was shown in both the laboratory and the process demonstration unit that this device handles extremely well fine below 500-mesh material.

The Jameson Column developed at the University of Newcastle is considerably shorter, and its height may be only slightly greater than its diameter.

Another novel concept is the Air-Sparged Hydrocyclone developed at the University of Utah. In this device, the slurry fed tangentially through the cyclone header into the porous cylinder to develop a swirl flow pattern intersects with air sparged through the jacketed porous cylinder. The froth product is discharged through the overflow stream.

## D. Chemical Desulfurization

Chemical desulfurization has been investigated as a method to selectively remove inorganic sulfur (which consists of pyritic sulfur with trace amounts of sulfate) and organic sulfur without need to discard part of the coal rich in sulfur and ash as with conventional coal preparation. If this could be accomplished without loss of coal product heat content, a clean burning coal alternative to flue gas desulfurization would be at hand. Several methods have been tested for the chemical removal of pyritic or organic sulfur from coal. Treatment with aqueous sodium hydroxide or aqueous sodium hydroxide with added lime at temperatures of 200–325°C and pressures up to 12500 psi removes only 45–95% of the inorganic sulfur and none of the organic sulfur, while significantly reducing the heat content of the product coal due to hydrothermal oxidation. Sulfur dioxide, potassium nitrate with complexing agents, nitrogen dioxide, nitric acid, air oxidation with and without the bacterium *Thiobacillus ferrooxidans*, hydrogen peroxide, chlorine, and organic solvents have all been tested. In each case, there was only minimal sulfur removal (based on sulfur content per unit heat content of
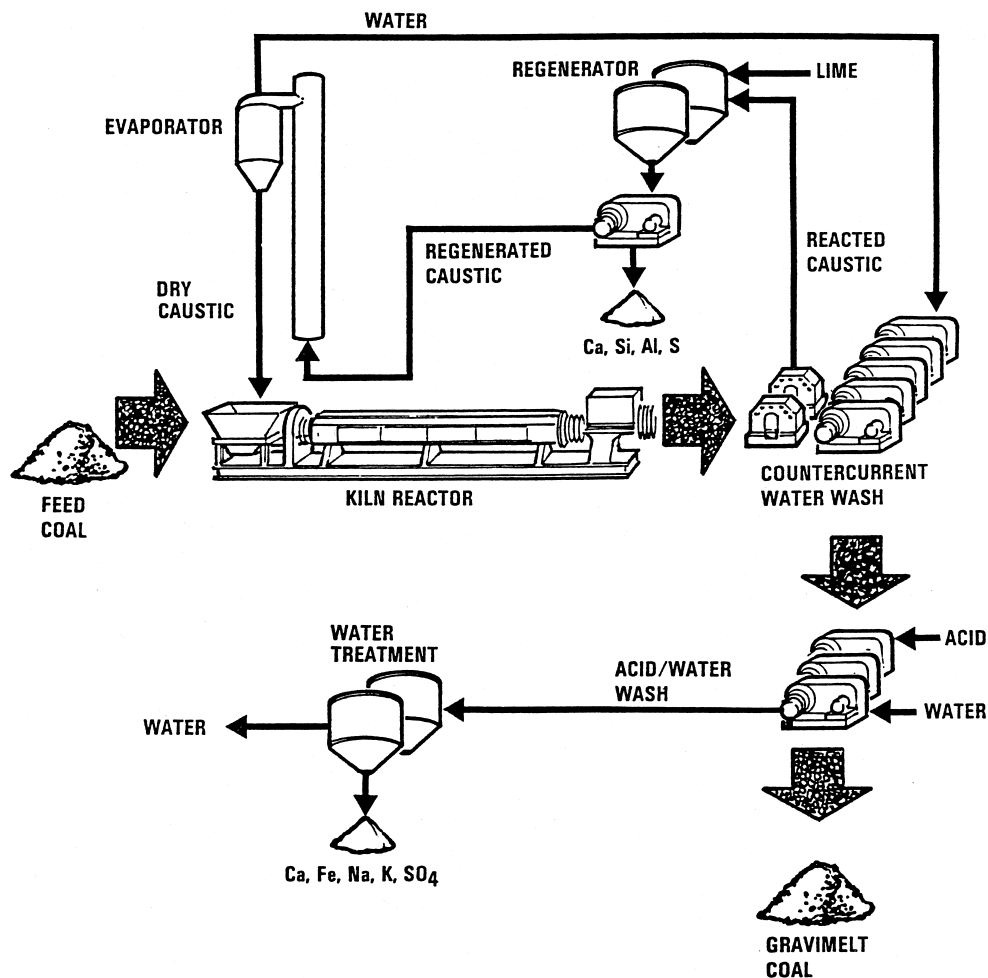


**FIGURE 26** Molten–Caustic–Leaching (Gravimelt Process) flow schematic.

product coal) and considerable loss of coal heat content and/or dilution of the organic coal matrix by bonding with reactant chlorine, nitrate, etc.

However, the use of aqueous ferric sulfate has proved to be effective in selectively removing virtually all of the inorganic sulfur while the heat content increases in direct proportion to the amount of iron pyrite removed. The ferric sulfate sulfur removal technology, known as the Meyers process, was developed through pilot plant demonstration under the auspicies of the Environmental Protection Agency. Typically, 90–95% of the pyritic sulfur was removed and the ferric salt was simultaneously regenerated with air oxidation. Elemental sulfur is produced as a byproduct, and must be extracted from the coal by use of an organic solvent or steam distillation. The process has not yet been utilized commercially as removal of pyritic sulfur alone (usually only half of the total sulfur, the remainder being mainly organic

sulfur) does not meet present air pollution control standards.

Subsequently, simultaneous removal of organic sulfur and pyritic sulfur was demonstrated in the laboratory and then in a test plant utilizing molten sodium and potassium hydroxide followed by a countercurrent water wash to recover sodium, potassium, and about half of the original coal mineral matter. A dilute sulfuric acid wash removes the remainder of the coal mineral matter (ash). This process, termed Molten–Caustic–Leaching, or the Gravimelt Process, results in removal of 90–99% of the coal inorganic and organic sulfur as well as 95–99% of the coal mineral matter. The product coal meets all current air pollution control standards for sulfur and ash emissions, and has an increased heat content, usually in the range of 14000 btu/lb. This coal is also very efficient as a powdered activated carbon for water purification. This is due to the high internal pore structure



**FIGURE 27** Rotary kiln reactor in operation at Gravimelt Test Plant Facility. The reactor consists of a 12-ft long furnace section, a 3-ft cooling zone, and a discharge area.

introduced by removal of sulfur and ash. The coal also functions as a cation exchange medium for water cleanup due attachment of carboxy groups during the caustic treatment.

A schematic of the Gravimelt Process is shown in Fig. 26. Feed coal is premixed with anhydrous sodium hydroxide or mixed sodium and potassium hydroxides, and then fed to a rotary kiln reactor where the mixture is heated to reaction temperatures of 325–415°C, causing the caustic to (1) melt and become sorbed in the coal matrix, (2) react with the coal sulfur and mineral matter, and (3) dissolve the reaction products containing sulfur and inorganic components. The kiln reactor can be seen in operation in Figure 27.

The reaction mixture is cooled in the cooling section of the kiln such that the mixture exits the rotary kiln in the form of pellets that are then washed with water in a seven-stage countercurrent separation system consisting of two filters and five centrifuges. As an example, four weights of wash water per weight of coal may be used, resulting in a 50% aqueous reacted caustic solution exiting the first stage of filtration. A large part of the coal-derived iron and a small amount of sodium and/or potassium remain with the coal cake exiting this countercurrent washing system.

The coal cake is next washed to remove the residual iron and caustic in a countercurrent three-stage centrifuge system. Sulfuric acid is added to the first stage to produce a pH of about 2 in order to dissolve the residual alkali and iron hydroxide. As an example, three weights of wash water per weight of coal may be used, resulting in an acid extract leaving the first stage containing 5900 ppm of mixed sodium and potassium sulfates and 12,500 ppm of iron sulfate. The product coal exiting this acidwashing system contains about one weight of sorbed water per weight of coal.

All of the iron sulfate and a major portion of the alkali sulfates can be removed from the acid extract water by a sequence of heating and lime treatment to form insoluble minerals such as gypsum. This water can then be recycled to the water wash train. Under some (very carefully controlled) conditions, the precipitate formed can be a jarosite-like double salt, which can be easily disposed of or land-filled.

A newer process configuration, with potentially large cost savings to the process, was tested in which the effluent from the acid wash train (without acid addition) was used as the makeup wash water for the water wash train. This configuration resulted in a product coal with low sulfur and moderate ash content suitable for those commercial applications that do not require low ash coal.

The aqueous caustic recovered from the first countercurrent washing circuit is limed to remove the coal-derived mineral matter, sulfur compounds, and carbonates. The mixture is provided with sufficient residence time to permit precipitation of the impurities before being centrifuged. The purified liquid is preheated and sent to a caustic evaporator where the water is recovered for recycle to the first wash train, while producing anhydrous caustic (as either a molten liquid or as flakes) for reuse in the initial leaching of the coal.

## SEE ALSO THE FOLLOWING ARTICLES

Coal Geology • Coal Structure and Reactivity • Fossil Fuel Power Stations, Coal Utilization • Froth Flotation • Fuels • Mining Engineering • Petroleum Refining

## BIBLIOGRAPHY

Deurbrouck, A. W., and Hucko, R. E. (1981). *In* "Chemistry of Coal Utilization" (M. A. Elliot, ed.), 2nd Suppl. vol., Chap. 10, Wiley, New York.

Gochin, R. J., and Smith, M. R. (1983). *Min. Mag.*, Dec., p. 453–467.

Horsfall, D. W., ed. (1980). "Coal Preparation for Plant Operators," South African Coal Processing Society, Cape Town.

Klassen, V. I. (1963). "Coal Flotation," Gosgorttiekhizdat, Moscow (Russian text).

Laskowski, T., Blaszczyński, S., and Slusarek, M. (1979). "Dense Medium Separation of Minerals," 2nd ed. Wyd. Slask, Katowice (Polish text).

Leonard, J. W., ed. (1979). "Coal Preparation," 4th ed., American Institute of Mechanical Engineering, New York.

Liu, Y. A., ed. (1982). "Physical Cleaning of Coal," Dekker, New York.

Meyers, R. A. (1977). "Coal Desulfurization," Marcel Dekker, New York.

Misra, S. K., and Klimpel, R. R., eds. (1987). "Fine Coal Processing," Noyes Publications, Park Ridge.

Osborne, D. G. (1988). "Coal Preparation Technology," Volumes 1 and 2, Graham and Trotman, London.

Speight, J. G. (1994). "The Chemistry and Technology of Coal," 2nd., revised and Expanded, Marcel Dekker, New York.

TRW Space and Electronics Group, Redondo Beach, CA Applied Technology Division. (1993). "Molten–Caustic–Leaching (Gravimelt) system integration project, final report," Department of Energy, Washington, DC, NTIS Order Number: DE930412591NZ, Performing Organization's Report Number: DOE/PC/91257-T18.

U.S. Department of Energy (1985). "Coal Slurry Fuels Preparation and Utilization," U.S. Department of Energy, Pittsburgh.

Wheelock, T. D., ed. (1977). "Coal Desulfurization," ACS Symp. Ser. No. 64, American Chemical Society, Washington, DC.

# Coal Structure and Reactivity

**John W. Larsen**

*Lehigh University*

**Martin L. Gorbaty**

*Exxon Research and Engineering Company*

## GLOSSARY

**Anthracite** Coals with less than ~10% volatile matter; the highest coal rank.

**Bituminous** Coals that have between about 10 and 30% volatile matter and coals that agglomerate on heating.

**Diagenesis** Initial biological decomposition of coal-forming plant material.

**Grade** Suitability of a coal for a specific use.

**Lignite** Stage of coalification following peat; high-oxygen, low-heating-value, young coals.

**Lithotypes** Macroscopic rock types occurring in coals, usually in banded structures.

**Macerals** Homogeneous microscopic constituents of coals, analogous to minerals in inorganic rocks.

**Metamorphism** Chemical changes responsible for the conversion of fossil plant materials to coal.

**Rank** Degree of maturation of a coal.

**Type** Origin and composition of the components of a coal.

**IN THE BROADEST TERMS**, coal may be defined as a naturally occurring, heterogeneous, sedimentary organic rock formed by geological processes (temperature, pressure, time) from partially decomposed plant matter under first aerobic and then anaerobic conditions in the presence of water and consisting largely of carbonaceous material and minor amounts of inorganic substances. The main use for coal worlwide is combustion as a source of heat for generating steam to make electricity. Coals used for this purpose are called steam coals. Certain other coals are used to produce the metallurgical coke required for making iron from iron ore. Coal can also serve as the raw material for making clean, usable liquid fuels, such as gasoline, and chemicals.

**TABLE I   Worldwide Coal Deposits**[a]

| Country | Estimated resources (billions of tons) | Proven resources (billions of tons) |
|---|---|---|
| Australia | 600 | 32.8 |
| Canada | 323 | 4.2 |
| Federal Republic of Germany | 247 | 34.4 |
| India | 81 | 12.4 |
| People's Republic of China | 1,438 | 98.9 |
| Poland | 139 | 59.6 |
| South Africa | 72 | 43 |
| United Kingdom | 190 | 45 |
| United States of America | 2,570 | 167 |
| Russia | 4,860 | 110 |
| Other | 229 | 55.7 |
| | 10,749 | 663 |

[a] Data from Wilson, C. L. (1980). "Coal—Bridge to the Future, Report of the World Coal Study," p. 161, copyright Ballinger Publishing Co., Cambridge, MA.

## I. INTRODUCTION

Coal is perhaps the most abundant fossil fuel, and large deposits are distributed worldwide. As with other natural resources, the amounts are estimated on two bases: *estimated resources* and *proven reserves*. Estimated resources are based on geological studies (e.g., seismic surveys and exploratory drilling) and are the best estimate of the amount of coal believed to be present in a given area. Proven reserves are those amounts of coal known to be economically recoverable by current technology in a given area. Proven reserves represent a small fraction of estimated resources. Table I contains resource and reserve data for some of the major worldwide coal deposits expressed in billions of tons. It must be recognized that these estimates vary from year to year, as do the methods for obtaining the data. Therefore, the data in the table should not be viewed as being absolute. According to these data, ∼663 billion tons, or 6% of worldwide resources of over 10 trillion tons, are recoverable by current technology. The worldwide reserves of coal have an energy content equivalent to ∼2 trillion barrels of petroleum.

## II. GEOLOGY: ORIGIN AND DEPOSITION

Coals may be viewed as fossilized plant matter. The processes by which plant matter was converted to coal are complex and involve at least three steps: accumulation, diagenesis, and metamorphism. The starting material for coal was a diversity of plant types and parts. For the plant matter to accumulate, a specific set of conditions had to prevail. The deposition rate of plant matter had to be approximately equal to the subsidence of the land on which it grew. Furthermore, there had to be water present to cover and partially preserve the plant matter. Initially, aerobic organisms began to decompose the cellulosic parts of the plant, as well as proteinaceous and porphyrin materials. This biological decomposition is called *diagenesis*. As the accumulated, partly decayed biomass was buried deeper, aerobic activity ceased, and under the increasing temperature and pressure conditions of burial it was chemically transformed to coal. This chemical transformation is called *metamorphism*. From the least serve to the most severe conditions, the coals that resulted are arranged as follows: peat → lignite → subbituminous → bituminous → anthracite. Chemically, the transformations can be viewed as combinations of deoxygenation, dehydrogenation, aromatization, and oligomerization. These transformations are inferred from the elemental analyses shown in Table II. The data in this table are typical; each class includes a range of values.

All coals contain minor amounts of inorganic material, called *mineral matter*, ranging from about 2 to 30% and averaging ∼10% by weight. Inorganics were incorporated into coal as the inorganic matter present in the original vegetation, mineral detritus deposited by water or wind flowing through the decaying biomass (detrital), and by precipitation or ion exchange of soluble ions from water percolating through the coal bed during metamorphism (authigenic). The former two are called syngenetic, the latter epigenetic to differentiate the processes occurring during diagenesis from those occurring during and after metamorphism.

The major components of mineral matter include aluminosilicate clays, silica (quartz), carbonates (usually of calcium, magnesium, or iron), and sulfides (usually as pyrite and/or marcasite). Many other inorganics are present, but only in trace quantities of the order of parts per million.

**TABLE II   Elemental Analysis Characterizing Metamorphism as a Deoxygenation/Aromatization Process**

| Coal or precursor | Carbon (%) | Hydrogen (%) | Oxygen (%) |
|---|---|---|---|
| Wood | 49.3 | 6.7 | 44.4 |
| Peat | 60.5 | 5.6 | 33.8 |
| Lignite | 69.8 | 4.7 | 25.5 |
| Subbituminous | 73.3 | 5.1 | 18.4 |
| Bituminous | 82.9 | 5.7 | 9.9 |
| Anthracite | 93.7 | 2.0 | 2.2 |

It is important to differentiate between mineral matter and ash. The former consists of inorganics present in the coal as mined; the latter is an oxide residue from combustion.

## III. COAL CLASSIFICATIONS

The diverse nature of deposited plant and mineral matter and the burial conditions make it clear that coal is not one substance but a wide range of heterogeneous materials, each with considerably different chemical and physical properties from the others. The heterogeneity of coals can be seen on the macroscopic and microscopic levels, and the science of classifying coals on the basis of these difference is called *coal petrography*.

On the macroscopic level, coals, particularly bituminous coals, appear to be layered or banded. These layers, called *lithotypes*, fall into four main classes: vitrain (coherent, black, and high gloss), clarain (layered, glossy), durain (granular, dull sheen), and fusain (soft, friable, charcoal-like).

On the microscopic level, the lithotypes just described are seen to be composites made up of discrete entities called *macerals*. Optically differentiated by microscopic examination of coals under reflected light or by the study of thin sections of coal using transmitted light (Fig. 1), macerals are classified into three major groups: vitrinite, liptinite (often termed exinite), and inertinite. The *vitrinite* class appears orange-red in transmitted light and is derived from fossilized lignin. *Liptinites* appear yellow and are the remains of spores, waxy exines of leaves, resin bodies, and waxes. *Inertinites* appear black and are believed to be the remnants of carbonized wood and unspecified detrital matter. Chemically, the liptinites are richest in hydrogen, followed by the vitrinites. Both the liptinites and vitrinites are reactive and give off a significant amount of volatiles when pyrolyzed, while the inertinites are relatively unreactive. It is not possible to generalize with accuracy the proportional maceral composition of all coals; however, it is fair to say that most bituminous coals contain about 70% vitrinite, 20% liptinite, and 10% inertinite.

It is generally believed that maceral composition ultimately determines a coal's reactivity in a particular process. For example, coals are blended together on a commercial scale to make optimum feedstocks for metallurgical coke based largely on their maceral contents. In the past, macerals were separated tediously by hand under a microscope. Today, new techniques are available to separate a coal into its component macerals much more easily, and this has renewed interest in maceral characterization and reactivity. It may even be possible one day to tailor coal feedstocks for specific processes such as pyrolysis, gasification, or liquefaction by separation and blending of macerals.
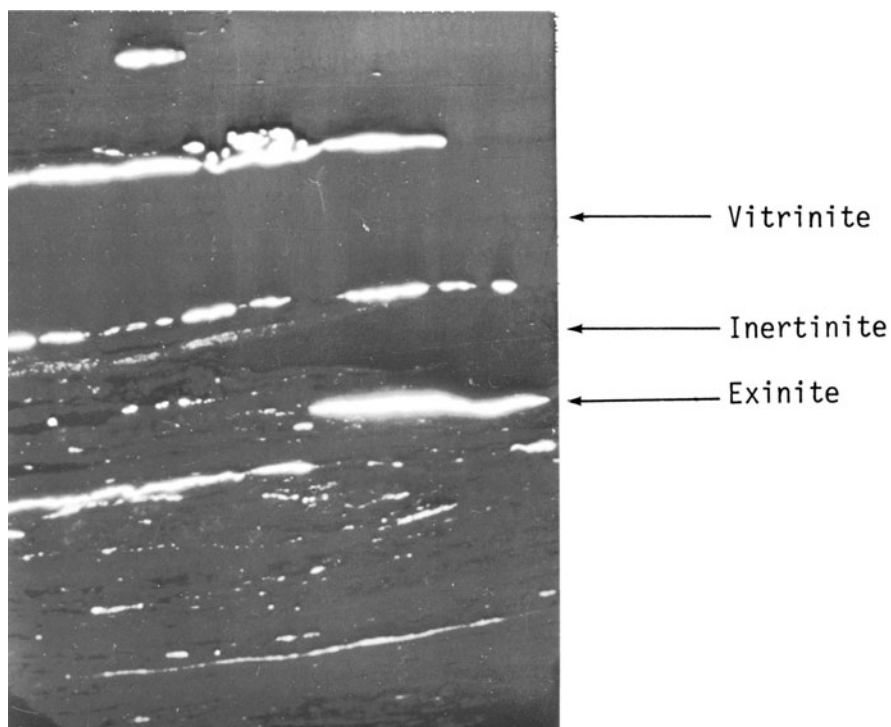


**FIGURE 1**  Transmission photomicrograph of bituminous coal. (Courtesy of D. Brenner, Exxon Research and Engineering Company.)

Another method of classifying coals is by the degree of metamorphism they have undergone. This classification is called *rank*—the higher the rank, the more mature the coal. Since metamorphism is a deoxygenation–aromatization process, rank correlates generally with carbon content and heat content of the coal, although there are many other, more specific methods of determining rank. Lignites and subbituminous coals are generally considered to be low rank, while bituminous coals and anthracite are considered to be high rank. The term "*rank*" must be distinguished from type and grade. Type refers to a coal's composition, the origin of the organic portion, and its organic components. Grade refers to the quality or suitability of a coal for a particular purpose. These distinctions are important because coal properties are not solely attributable to metamorphic alteration; the nature of the source material and its diagenetic alterations are also important in determining coal properties.

In general, carbon content, aromaticity (i.e., the percentage of carbon that is aromatic), number and size of condensed rings, and calorific content increase with increasing rank, while volatile matter, oxygen content, oxidizability, and solubility in aqueous caustic decrease with increasing rank. Hardness and plastic properties increase to a maximum, then decrease with increasing rank, while porosity (i.e., moisture-holding capacity) and density decrease to a minimum, then increase with increasing rank. Typical values are shown in Table III. Within each rank classification, a range of values is found.

## IV. PHYSICAL PROPERTIES

Physical properties are an important consideration in all uses of coals. Most vary with rank and coal type. A brief discussion of the most important physical properties follows.

### A. Mechanical Properties

#### 1. Elasticity

Because of the heterogeneous nature and the natural occurrence of cracks and pores, determining elastic moduli is experimentally difficult. Under normal conditions, coals are brittle solids. For bituminous coals, the elastic properties are independent of rank and similar to those of phenol–formaldehyde resins. At ∼92% carbon, the moduli increase rapidly and become anisotropic.

#### 2. Hardness (Grindability)

Grindability is normally measured by determining the number of revolutions in a pulverizer that are required to achieve a given size reduction. This is an indirect measure of the amount of work required for that size reduction. Grindability increases slowly to a maximum at ∼90% carbon and decreases rapidly beyond this point. Hardness is ascertained by measuring the size of the indentation left by a penetrator of specified shape pressed onto the coal with a specified force for a specified time. It changes only slightly between 70 and 90% carbon and increases rapidly beyond 90% carbon.

#### 3. Plastic Deformation

Coals heated to above 400°C plastically deform, as do those that have been swollen with pyridine and other solvents that are good hydrogen bond acceptors. Very small coal particles flow and deform at room temperature under very high stresses.

### B. Thermal Properties

#### 1. Heat of Combustion

Heat of combustion is normally measured calorimetrically, but it can also be calculated with good accuracy from

**TABLE III  Properties of Coals of Different Rank**

| Property[a] | Lignite | Subbituminous | Bituminous | Anthracite |
|---|---|---|---|---|
| Moisture capacity (wt%) | 40 | 25 | 10 | <5 |
| Carbon (wt%), DMMF | 69 | 75 | 83 | 94 |
| Hydrogen (wt%), DMMF | 5.0 | 5.1 | 5.5 | 3.0 |
| Oxygen (wt%), DMMF | 24 | 19 | 10 | 2.5 |
| Volatile matter (wt%), DMMF | 53 | 48 | 38 | 6 |
| Aromatic carbon (mol fraction) | 0.7 | 0.78 | 0.84 | 0.98 |
| Density (helium) (g/cm$^3$) | 1.43 | 1.39 | 1.30 | 1.5 |
| Grindability (Hardgrove) | 48 | 51 | 61 | 40 |
| Heat content (btu/lb), DMMF | 11.600 | 12.700 | 14.700 | 15.200 |

[a] DMMF: dry, mineral-matter-free.

the elemental composition of the coal. It is the heat given off when a given amount of coal is completely burned.

### 2. Specific Heat

The specific heat records the heat necessary to cause a given temperature rise in a coal. It is rank dependent, decreasing from ~0.35 cal/g at 50% carbon to ~0.23 cal/g at 90% carbon. Above 90% carbon, it decreases quite rapidly.

### 3. Thermal Conductivity

The capacity of a coal to conduct heat is strongly dependent on the pore structure and moisture content. It has been widely studied. When coals soften thermally, between 400 and 600°C, the thermal conductivity increases as the pores collapse.

## C. Electrical Properties

### 1. Dielectric Constant

The dielectric constant is a sensitive function of the amount of water present in a coal. Bituminous coals are insulators with static dielectric constants between 3 and 5. Above 88% carbon, the static dielectric constant increases very rapidly.

### 2. Electric Conductivity

The most interesting feature here is that anthracites are natural semiconductors. Conductivity increases sharply above 88% carbon, presumably due to increasing graphitic character.

### 3. Magnetic Susceptibility

Coals contain numerous unshared electrons, of the order of $10^{18}$ spins per gram. This gives rise to the bulk property called paramagnetic susceptibility, the magnitude of the force exerted on a sample in a strong magnetic field.

### 4. Reflectance

Coal samples are often geologically characterized by very accurate measurements of the reflectance of light from their components. There is a vast literature dealing with this empirical technique.

## D. Density

The measurement of coal densities is complicated by the presence of pores in coals. A number of techniques for dealing with this problem have been developed, and accurate data are available for whole coals and macerals. Density is rank dependent in a regular way. Vitrinite density decreases slowly from ~1.5 g/cm$^3$ at 50% carbon to a minimum of 1.27 at 88% carbon, then increases rapidly, reaching values as high as 1.6–1.8. Fusinites and micrinites are more dense than vitrinites, while liptinites are less dense. The density differences provide a good way of separating macerals.

## V. PHYSICAL STRUCTURE

Coals are very complex mixtures of several macromolecular components and a number of inorganic materials. This fact makes the discussion of physical properties difficult. Each individual component can be isolated and its properties determined, but will the property of the whole coal be the sum or some weighted average of the individual component properties? Perhaps the easiest approach is the phenomenological one: simply to report the properties measured for the whole coal while realizing that these are the result of some complex addition of the properties of many different components. This approach is often perfectly satisfactory. If the heat of combustion of a coal is desired, it makes little difference what the individual component contributions to that heat are. The approach used here will be phenomenological, but the reader should know that this obscures the many differences among individual components.

## A. Pore Structure

Coals are highly porous materials, many having void volumes as high as 20% of thier total volume. To describe the pore structure we need to know the size distribution of the pores, the size ranges of the pores, and the population of each range. The shape of the pores is important, as is the total surface area of the coal.

A reagent coming in contact with a coal first encounters its surface. Often, it reacts there. To understand reactivity patterns, we need to know the surface area. Even if the reagent must diffuse into the coal, the rate at which it does this is strongly influenced by the amount of surface it contacts. The size and shape of the pores control whether a molecule can enter and contact the surface and so are an important part of any description of the pore network.

Most information about pores comes from adsorption measurements. A sample of coal that has been evacuated to clean its surface is exposed to vapors of a gas (e.g., $CO_2$), which adsorb and cover the surface. By means of a very sensitive balance, the amount of adsorption is measured at several pressures. Thus, the amount of gas required to
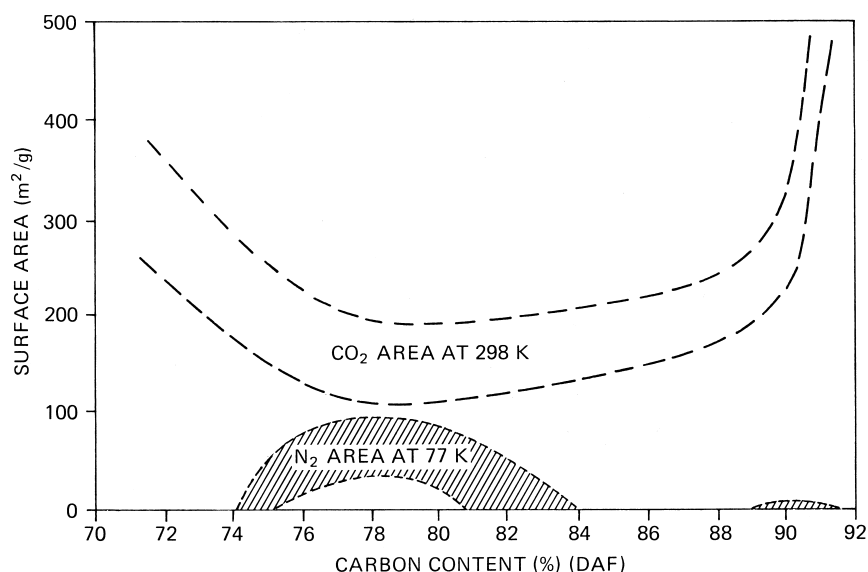
**FIGURE 2**  Dependence of coal surface area for dry, ash-free (DAF) coal on coal rank and technique. [Reproduced with permission from Gan, H., Nandi, S. P., and Walker, P. L. (1972). *Fuel* **51,** 272.]

cover the surface is determined. The cross-sectional area of the gas is known. This is the amount of surface that each gas molecule will cover. From these two quantities, the surface area can be calculated.

Figure 2 shows how the surface areas of coals vary with carbon content. Carbon dioxide measures the entire surface area, pore surfaces as well as external surfaces. Nitrogen cannot penetrate the pores (it does not have enough energy at these low temperatures), and so only the external surface area is measured. The surface areas are enormous, demonstrating that the coals must be very porous.

The pore volume can be estimated from some elaborate gas adsorption studies or measured directly. If the density of coal particles is measured in a vacuum and in a gas that fills the pores, different values are obtained. Density is mass over volume, and for both measurements the volume is the same. In one case (vaccum), the mass of the coal is recorded. In the other case, the mass of the coal plus the mass of the gas in the pores is recorded. If the density of the gas is known, the pore volumes are easily obtained. Some typical values are given in Table IV.

An analysis of the weight increase due to gas adsorption as the pressure of the gas increases can give the size distribution of the pores. At low pressures, pores of very small size fill by capillary condensation, while larger pores are still only surface coated. The results of such an analysis for a series of coals are given in Table IV. For bituminous coals and anthracites, most of the surface area is due to micropores, which also make the largest contribution to the pore volume. Lignites have a much more open pore structure, with the largest pores predominating. Penetration of

the pores by liquid mercury at high pressures can also provide information about the size distribution of macropores and some of the transitional pores. We shall return briefly to pore structure later to discuss micropores as packing imperfections between macromolecules.

There remains only an analysis of pore shape. This issue has been probed mainly by looking at the dependence of molecular absorption on the shape of the absorbed

**TABLE IV   Illustrative Gross Pore Distributions in Coal[a]**

| Rank[b] | Pore volume (total), 4–30,000 Å ($cm^3/g$) | Pores | | |
|---|---|---|---|---|
| | | Macro, 300–30,000 Å ($cm^3/g$) | Transitional, 12–300 Å ($cm^3/g$) | Micro, 4–12 Å ($cm^3/g$) |
| Anthracite | 0.076 | 0.009 | 0.010 | 0.057 |
| LV bituminous | 0.052 | 0.014 | 0.000 | 0.038 |
| MV bituminous | 0.042 | 0.016 | 0.000 | 0.026 |
| HVA bituminous | 0.033 | 0.017 | 0.000 | 0.016 |
| HVB bituminous | 0.144 | 0.036 | 0.065 | 0.043 |
| HVC bituminous | 0.083 | 0.017 | 0.027 | 0.039 |
| HVC bituminous | 0.158 | 0.031 | 0.061 | 0.066 |
| HVB bituminous | 0.105 | 0.022 | 0.013 | 0.070 |
| HVC bituminous | 0.232 | 0.043 | 0.132 | 0.070 |
| Lignite | 0.114 | 0.088 | 0.004 | 0.022 |
| | 0.105 | 0.062 | 0.000 | 0.043 |
| | 0.073 | 0.064 | 0.000 | 0.009 |

[a] Reproduced with permission from Gan, H., Nandi, S. P., and Walker, P. L. (1972). *Fuel* **51,** 272.

[b] LV, Low volatile; MV, medium volatile; HVA, high volatile A; HVB, high volatile B; HVC, high volatile C.

molecules. There are fewer data here than there are on pore volume or surface area. Nevertheless, there is general agreement that the pores are often slits, oblate elipsoids, that often have constricted necks.

Coals are penetrated by an extensive network of very tiny pores and, because of this, have enormous surface areas. The smaller pores are about the same size as small molecules, so coals are molecular sieves, capable of trapping small molecules in their pores while denying access to larger molecules.

## B. Macroscopic Structure

The best analogy for the macroscopic structure of coal is a fruitcake. The dough corresponds to the principal maceral, vitrinite, while the various goodies correspond to the different macerals spread throughout. The ground nuts play the role of the mineral matter. The composition of the fruitcake is not rank dependent, but rather is due to the original coal deposition and depositional environment. The macerals become more like one another as rank increases, but their proportions are not rank dependent. In fact, their relative proportions vary with both vertical and horizontal positions in a coal seam, with the vertical variation occurring over a much shorter distance scale.

To make our fruitcake a more accurate model, we must alter it somewhat. Coals tend to have banded structures, with macerals grouped in horizontal bands. A fruitcake made by a lazy cook who did not stir the batter throughly is an improved model. The nuts must also be ground to cover an enormous size range, from large chunks to micrometer-size particles. Further more, they must stick to the cake. Even extremely fine grinding in a fluid energy mill will not entirely separate the mineral matter from the organics. To separate the macerals on the basis of their different densities, the mineral matter must first be dissolved in acid. We know almost nothing about the binding forces or binding mechanism between the mineral matter and the coal.

## C. Extractable Material

As much as 25% of many coals consists of small molecules that will dissolve in a favorable solvent and can thereby be removed from the insoluble portion. The amount of coal that dissolves is a function of the nature of the coal, the solvent used, and the extraction conditions. First, let us consider the solubility of coals in various solvents at their boiling points at atmospheric pressure. Typical data are contained in Table V.

Certain solvents stand out as being unusually effective, pyridine and ethylenediamine being the most noteworthy. They are not unique, however, and any solvent having a Hildebrand solubility parameter of ∼11 and the capacity to serve as an effective hydrogen bond acceptor will be a

**TABLE V   Yields of Coal Extracts in Solvents at Their Boiling Points[a]**

| Solvent | Yield of coal extract (wt% of dry, ash-free coal) |
|---|---|
| *n*-Hexane | 0.0 |
| Water | 0.0 |
| Formamide | 0.0 |
| Acetonitrile | 0.0 |
| Nitromethane | 0.0 |
| Isopropanol | 0.0 |
| Acetic acid | 0.9 |
| Methanol | 0.1 |
| Benzene | 0.1 |
| Ethanol | 0.2 |
| Chloroform | 0.35 |
| Dioxane | 1.3 |
| Acetone | 1.7 |
| Tetrahydrofuran | 8.0 |
| Pyridine | 12.5 |
| Dimethyl sulfoxide | 12.8 |
| Dimethylformamide | 15.2 |
| Ethylenediamine | 22.4 |
| 1-Methyl-2-pyrrolidone | 35.0 |

[a] Data from Marzec, A., Juzwa, M., Betlej, K., and Sobkowiah, M. (1979). *Fuel Proc. Technol.* **2**, 35, and refer to a specific coal; they are illustrative but not representative.

good solvent. The dependence on coal rank of the amount that can be extracted by pyridine is shown in Fig. 3. The pyridine extract seems to resemble the parent coal quite closely, and the structural characteristics of the insoluble coal can be taken as present in the soluble portion, which is much easier to analyze.

Other solvents generally dissolve less of the coal, with solvents like toluene dissolving only a small percentage and hexane being ineffective. Less material can be extracted from coals by chloroform than the portion of the pyridine extract which is soluble in chloroform. If a coal and the pyridine extract of a coal are extracted with chloroform, different amounts of material go into solution. The greater percentage of the starting coal is dissolved by first extracting with pyridine. This will be explained in the next section.

The solvent power of dense, supercritical fluids is often greater than that of the corresponding liquid. Supercritical organic fluids at temperatures above 350°C are often quite good solvents for coals, and, indeed, a conversion process based on supercritical fluid extraction has received extensive development in Britain. It seems probable that some chemical decomposition is occurring that leads to the high
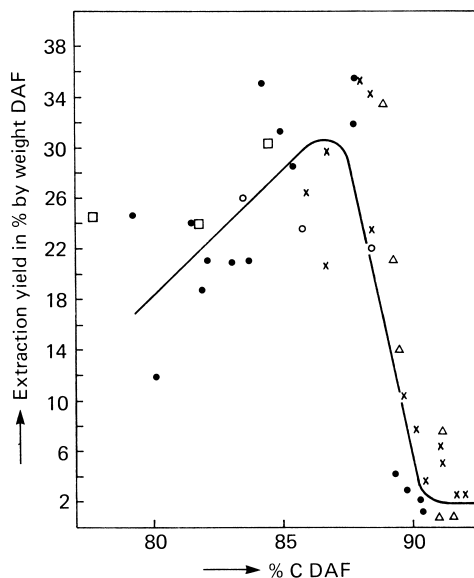
**FIGURE 3** Extraction yield in pyridine at 115°C vs rank. [Reprinted with permission from Van Krevelen, D. W. (1981). "Coal," Elsevier, New York.]

extractabilities observed, often approaching 40% of the coal when toluene is the solvent.

### D. Macromolecular Nature of Coals

Coals are believed to be three-dimensionally cross-linked macromolecular networks containing dissolved organic material that can be removed by extraction. This model offers the most detailed and complete explanation of the chemical and mechanical behavior of coals. It is a relatively recent model and is somewhat controversial at this writing. The insoluble portion of the coal comprises the cross-linked network, one extraordinarily large molecule linked in a three-dimensional array. This network is held together by covalent bonds and hydrogen bonds, the weak interactions that play such a large role in the association of biological molecules. The extractable portion of the coal is simply dissolved in this solid, insoluble framework. A solvent like pyridine, which forms strong hydrogen bonds, breaks the weaker hydrogen bonds within the coal structure. In the presence of pyridine, the network is less tightly held together. Molecules not tied into the network by covalent bonds can be removed. A solvent like chloroform, which does not break hydrogen bonds in coals, cannot extract all of the material that is soluble in it, because some of the soluble molecules are too large to move through the network and be removed. They do come out of the coal in pyridine, which expands the network. Once removed from the network, they readily dissolve in chloroform.

The macromolecular structure controls many important coal properties but has been relatively little studied. The native coal is a hard, brittle, viscoelastic material. This is because it is a highly cross-linked solid. The network is held together by both covalent bonds and a large number of hydrogen bonds, which tie otherwise independent parts of the network together. The network is frozen in place and is probably glassy. Diffusion rates are low because the network cannot move to allow passage of the diffusing molecule.

If the internal hydrogen bonds are destroyed, either by chemical derivatization or by complexation with a solvent that is a strong hydrogen bond acceptor, the properties of coals change enormously. The network is now quite flexible, and the coal is a rubbery solid, which is no longer brittle. Diffusion rates are enormously enhanced. All mechanical properties have been altered. All of this is due to the freeing of those portions of the network that were held to one another by the hydrogen bonds. This greatly increased freedom of motion changes the fundamental character of the solid.

## VI. CHEMICAL STRUCTURE

It is impossible to represent accurately and usefully a material as complex as coal using classical chemical structures. A model large enough to show the distribution of the very many structures present in a coal would be too large and unwiedly to be useful. We shall discuss the occurrence of the various coal functional groups while ignoring their immediate environment. A phenolic hydroxyl will be listed as a phenolic hydroxyl irrespective of whether it is the only functional group on a benzene ring or whether it shares an anthracene nucleus with several other hydroxyl groups. Since the distribution of hydroxyl environments is unknown, there is no alternative.

The elements of interest are, in increasing order of structural importance, nitrogen, sulfur, oxygen, and carbon. The chemistry of hydrogen, the remaining major component, is governed by its position on one of these elements and need not be discussed separately. Trace elements will be ignored, and general mineral mater composition is covered in Section II.

### A. Nitrogen

Nitrogen is a minor constituent of most coals, being less than 2% by weight. Its principal importance is that it may lead to the formation of nitrogen oxides on combustion. Its origin is believed to be amino acids and porphyrins in the plant materials from which coals were derived. Recently, X-ray methods have been developed which allowed
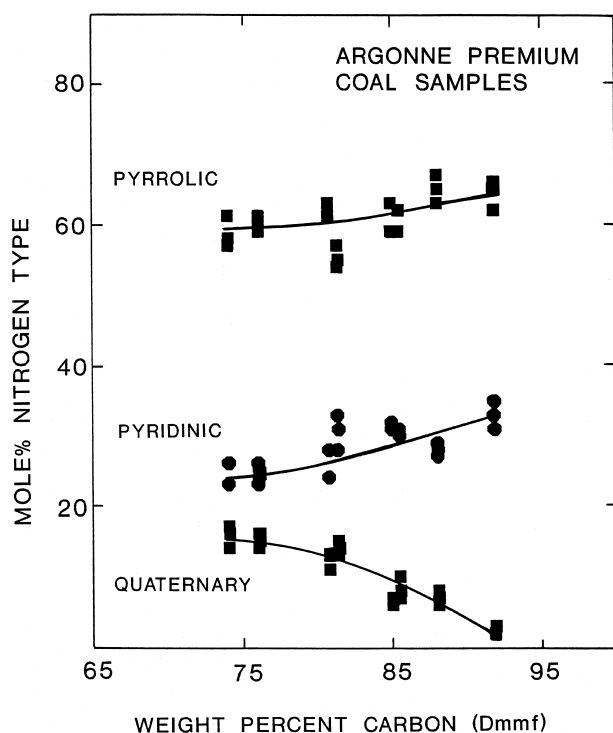
**FIGURE 4** Organic nitrogen forms as a function of rank for Argonne Premium Coals. [Reprinted with permission from Kelemen, S. R., Gorbaty, M. L., and Kwiatek, P. J. (1994). *Energy Fuels* **8,** 896. Copyright American Chemical Society.]

the accurate speciation and quantification of the forms of organically bound nitrogen, including the basic pyridines and quinolines and the weakly acidic or neutral indoles and carbazoles. X-ray photoelectron spectroscopy (XPS) data showing the quantitative distributions of nitrogen types in a suite of well-preserved coals representative of many ranks (Argonne Premium Coals) are shown in Fig. 4 as a function of weight percent carbon. Pyrrolic nitrogen was found to be the most abundant form of organically bound nitrogen, followed by pyridinic, and then quaternary types. It is clear from Fig. 4 that the distributions are rank dependent, with the highest abundance of quaternary nitrogen in the lowest rank coals. The quaternary species are attributed to protonated pyridinic or basic nitrogen species associated with hydroxyl groups from carboxylic acids or phenols. The concentrations of quaternary nitrogen decrease, while pyridinic or basic nitrogen forms appear to increase correspondingly as a function of increasing rank.

## B. Sulfur

The presence of sulfur in coals has great economic consequences. When coal is burned, the resulting $SO_2$ and $SO_3$ must be removed from the stack gasses at great expense or discharged to the atmosphere at great cost to the environment. The presence of sulfur in coals has been the driving force for a great deal of research directed toward its removal. Numerous processes for removing mineral matter (including inorganic sulfur compounds) from coal are in use, as are processes for removing sulfur oxides from stack gasses. Much work has been done on chemical methods for removing sulfur from coals before burning, but none are in large-scale use.

The origin of sulfur in most coals is believed to be sulfate ion, derived from seawater. During the earliest stages of coal formation, bacterial decomposition of the coal-forming plant deposits occurs. Some of these bacteria reduce sulfate to sulfide. This immediately reacts with iron to form pyrite, the principal inorganic form of sulfur in coals. It is also incorporated into the organic portion of the coal. The amount and form of sulfur in coals depend much more on the coal's depositional environment than on its age or rank. In this sense, it is largely a coal-type parameter, not a rank parameter.

Coals contain a mixture of organic and inorganic sulfur. The inorganic sulfur is chiefly pyrite (fool's gold). Exposure of coal seams to air and water results in the oxidation of the pyrite to sulfate and sulfuric acid, causing acid mine drainage from open, wet coal seams. Much of the pyrite can often be removed by grinding the coal and carrying out a physical separation, usually based on the difference in density between the organic portion of the coal and the more dense mineral matter. Coal cleaning is quite common, is often economically beneficial, and is increasing. The organic sulfur in coals, that is, the sulfur bonded to carbon, is very difficult to remove. Although many processes have been developed for this purpose, none has been commercially successful.

In the last decade, major advances have been made in speciating and quantifying forms of organically bound sulfur in coals into aliphatic (i.e., dialkylsulfides) and aromatic (i.e., thiophenes and diarylsulfides) forms. Representative results on a well-preserved suite of coals obtained from XPS and two different sulfur K-edge X-ray absorption near edge structure spectroscopy (S-XANES) analysis methods are plotted together in Fig. 5. The accuracy of the latter methods is reported to be $\pm 10$ mol%. In Fig. 5, aromatic sulfur forms are plotted against weight percent carbon. It is seen that in low-rank coals there are relatively low levels of aromatic sulfur (i.e., significant amounts of aliphatic sulfur) and that levels of aromatic sulfur increase directly as a function of increasing rank. The X-ray methods have been used to follow the chemistry of organically bound sulfur under mild oxidation in air, where it was shown that the aliphatic sulfur oxidizes in air much more rapidly than the aromatic sulfur, *ex situ*
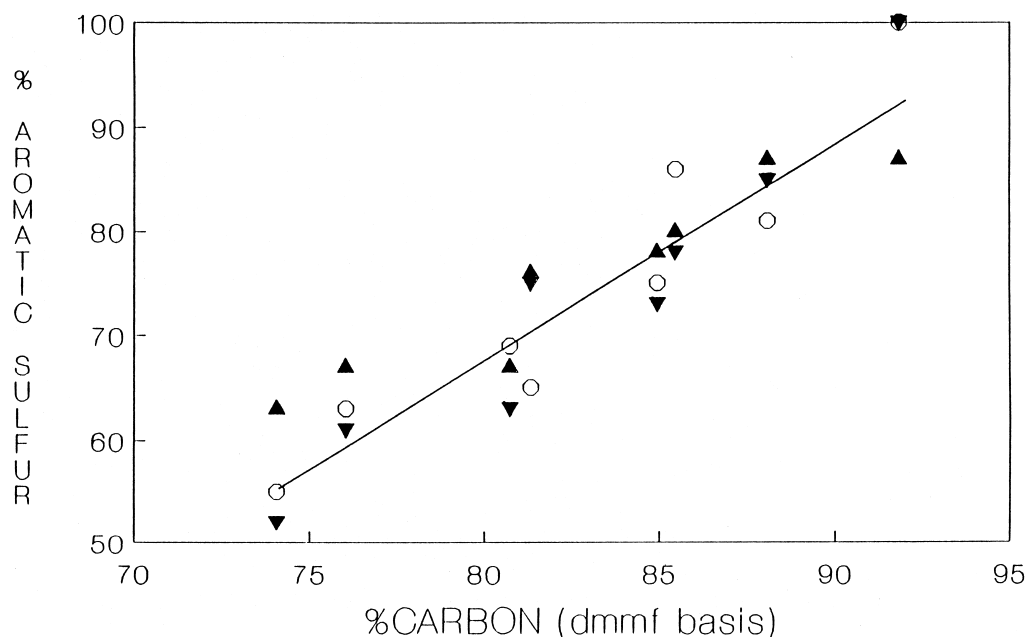
**FIGURE 5** Aromatic sulfur forms as a function of rank for Argonne Premium Coals: ▲, ▼ XANES data; ○ XPS data. [Reprinted with permission from Gorbaty, M. L. (1994). *Fuel* **73,** 1819.]

and *in situ* pyrolysis, single electron transfer, and strong base conditions.

## C. Oxygen

Oxygen is a major component of the plant materials that form coals. The principal precursor of coals is probably lignin, the structural material of plants. Most of the cellulose and starch is probably destroyed in the earliest stages of deposition and diagenesis, although a controversy over this continues. The oxygen content of coals decreases steadily as the rank of the coals increases. Good analytical techniques exist for the important oxygen functional groups, so their populations as a function of rank are known (Fig. 6). The detailed chemistry of the loss and transformations of these functional groups remains largely unknown.

The carboxylate (—COOH) group is the major oxygen functionality in lignites but is absent in many bituminous coals. The mechanism for its rapid loss is unknown. Its presence contributes much to the chemistry of lignites and low-rank coals. It is a weak acid and in many coals is present as the anion (—COO$^-$) complexed with a metal ion. Low-rank coals are naturally occurring ion-exchange resins. Certain lignites are respectable uranium ores, having removed the uranium from ground water flowing through the seam by ion exchange.

Lignites contain water and can be considered to be hydrous gels. Removing the water causes irreversible changes in their structure. The water is an integral part of the native structure of the lignites, not just something complexed to the surface or present in pores. The nature of the water in the structure is unknown. Its presence is undoubtedly due to the presence of a large number of polar oxygen functional groups in the coal.

Hydroxyl (—OH) is present in phenols or carboxylic acids. It is present in all coals except anthracites. It is slowly lost during the coalification process. The phenols are much less acidic than the carboxylic acids and remain un-ionized in native coals.
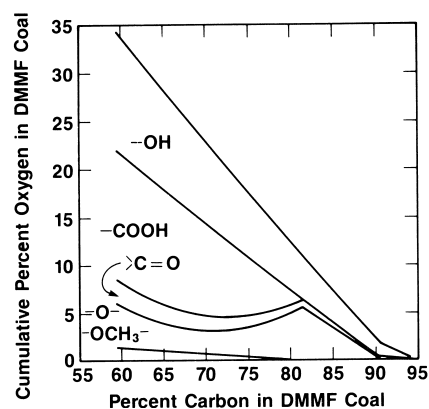


**FIGURE 6** Distribution of oxygen functionality in coals: DMMF, dry, mineral-matter-free. [Reprinted with permission from Whitehurst, D. D., Mitchell, T. O., and Farcasiu, M. (1980). "Coal Liquefaction: The Chemistry and Technology of thermal Processes," Academic Press, New York.]

The methoxyl group (—OCH$_3$) is a minor component whose loss parallels that of carboxylate. The variety of groups containing the carbonyl function ($\supset$C=0) comprise a small portion of the oxygen.

In some ways, the ether functionality (—O—) is the most interesting. Its loss during coalification is not continuous. The pathway by which it is created is unknown. It is also uncertain how much of the ether oxygen is bound to aliphatic carbon and how much to aromatic carbon, but techniques capable of providing an answer to this question have just arrived on the scene.

## D. Carbon

There are two principal types of carbon in coals: aliphatic, the same type of carbon that is found in kerosene and that predominates in petroleum, and aromatic, the kind of carbon found in graphite. Many techniques have been used to estimate the relative amounts of the two types of carbon in coals, but only recently has the direct measurement of the two been possible. Their variation with rank is illustrated in Fig. 7.

The carbon in coals is predominantly aromatic and is found in aromatic ring structures, examples of which are



**FIGURE 7** Plot of fraction of aromatic carbon ( $f_a$ ) vs percent carbon for 63 coals and coal macerals. [Reprinted with permission from Wilson, M. A., Pugmire. R. J., Karas, J., Alemany, L. B., Wolfenden, W. R., Grant, D. M., and Given, P. H. (1984). *Anal. Chem.* **56**, 933.]

given below. It is very desirable to know the frequency of the occurrence of these ring systems in coals, and many techniques have been applied to this problem. Most of the techniques involve rather vigorous degradation chemistry, advanced instrumental techniques, and many assumptions.

The distribution of aromatic ring systems in coals has been estimated for the same Argonne Premium Coal samples discussed previously, based on $^{13}$C NMR experiments, and these data, expressed as number of aromatic carbons per cluster (error $\approx \pm 3$), are shown in Fig. 8. For samples in this suite varying from lignite to low volatile bituminous, the results show that the average number of aromatic carbons per ring cluster ranged from between 9 (lignite) and 20 carbons (lvb), with most subbituminous and bituminous coals of this suite containing 14. This indicates an average size of 2–3 rings per cluster for most of these coals, with the lignite having perhaps an average of 1–2 and the highest rank coal having 3–4.

Our knowledge of the aliphatic structures in coals is currently inadequate. In many low-rank coals, there is a significant amount of long-chain aliphatic material. There is a debate as to whether this material is bound to the coal macromolecular structure or whether it is simply tangled up with it and thus trapped.

Much of the aliphatic carbon is present in ring systems adjacent to aromatic rings. Some is bonded to oxygen and in short chains of —CH$_2$— groups linking together aromatic ring systems. No accurate distribution of aliphatic carbon for any coal is known.

## E. Maceral Chemical Structure

There are four principal maceral groups. The vitrinite macerals are most important in North American coals, comprising 50–90% of the organic material. They are derived primarily from woody plant tissue. Since they are the principal component of most coals, vitrinite maceral chemistry usually dominates the chemistry of the whole coal. Inertinite maceral family members comprise between 5 and 40% of North American coals. These materials are not at all chemically inert, as the name implies. They are usually less reactive than the other macerals, but their chemical reactivity can still be quite high. They are thought to be derived from degraded woody tissue. One member of this family, fusinite, looks like charcoal and may be the result of ancient fires at the time of the original deposition of the plant material. The liptinite maceral group makes up 5–20% of the coals in question. Its origin is plant resins, spores, and pollens, resinous and waxy materials that give rise to macerals rich in hydrogen and aliphatic structures. Terpenes and plant lipid resins give rise to the varied group of resinite macerals. From some unusual western
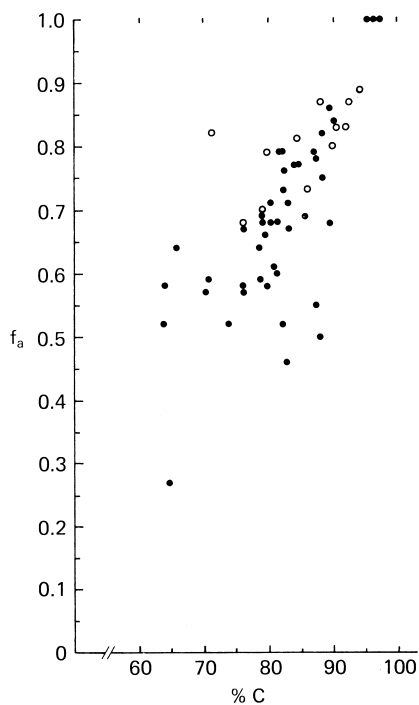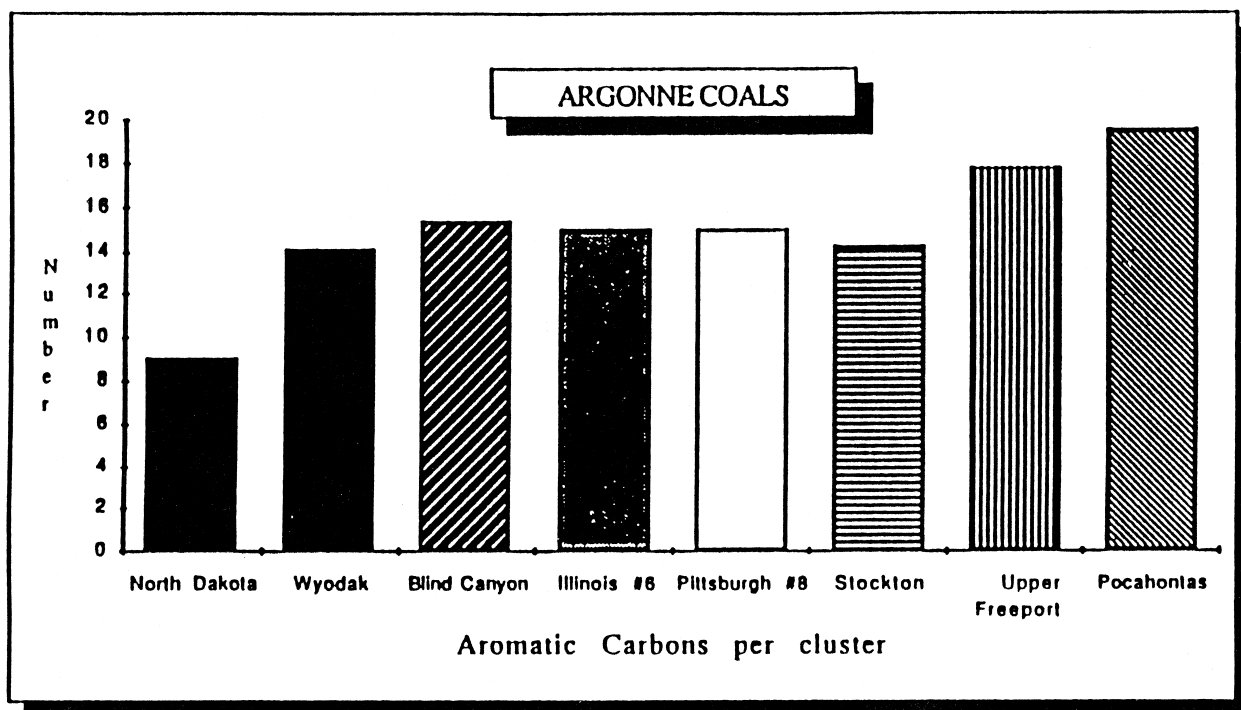
**FIGURE 8** Aromatic cluster sizes in Argonne Premium Coals. [Reprinted with permission from Solum, M. S., Pugmire, R. J., Grant, D. M. (1989). *Energy Fuels* **3,** 187. Copyright American Chemical Society.]

U.S. coals, resinites have been isolated and commercially marketed.

The rank dependence of the elemental composition of the maceral groups is shown in Fig. 9. The inertinites are the most aromatic, followed by the vitrinites and then the liptinites. Their oxygen contents decrease in the order vit-
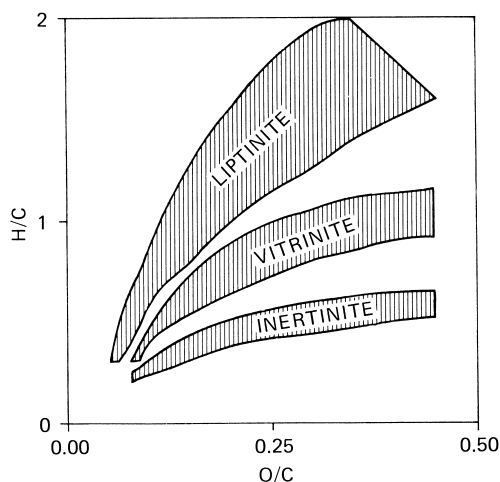


**FIGURE 9** Elemental composition of the three main maceral groups. [Reprinted with permission from Winans, R. E., and Crelling, J. C. (1984). "Chemistry and characterization of coal macerals." *ACS Symp. Ser.* **252,** 1.]

rinite > liptinite > inertinite. It is probably impossible to come up with a universally accurate reactivity scale for these materials. The conditions and structures vary too widely. *Reactivity* refers to the capacity of a substance to undergo a chemical reaction and usually refers to the speed with which the reaction occurs. The faster the reaction, the greater is the reactivity. The reactivity of different compounds changes with the chemical reaction occurring and thus must be defined with respect to a given reaction or a related set of reactions. The reactivity of macerals as substrates for direct liquefaction and their reactivity as hydrogen donors are parallel and in the order liptinite > vitrinite > inertinite > resinite. The solubility of the resinites is generally high, so their low reactivity makes little difference for most coal processing.

## F. Structure Models

A number of individuals have used the structural information available to create representative coal structures. All these representations contain the groups thought to occur most frequently in coals of the rank and composition portrayed. Formulating any such structure necessarily involves making many guesses. Taken in the proper spirit, they are useful for predicting some types of chemistry and are a gauge of current knowledge. One such structure is shown in Fig. 10. It is typical of recent structure proposals
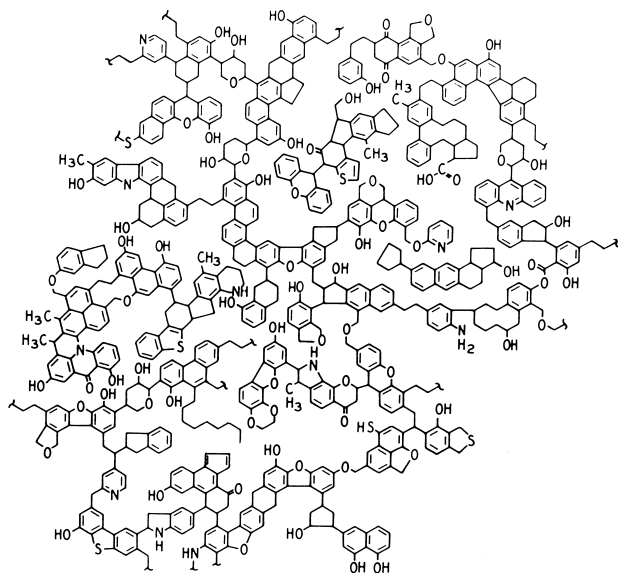
**FIGURE 10** Molecular structure for a bituminous coal. [Reprinted with permission from Shinn, J. (1984). *Fuel* **63,** 1284.]

for bituminous coals. These structures must be used cautiously. They are meant to be representative; they are not complete structures containing all groups actually present.

## VII. COAL CHEMISTRY

The basic chemical problem in coal conversion to gaseous or liquid products can be viewed as management of the hydrogen-to-carbon atomic ratio (H/C). As Fig. 11 shows, conventional liquid and gaseous fuels have atomic H/C ratios higher than that of coal. Thus, in very general terms, to convert coal to liquid or gaseous fuels one must either add hydrogen or remove carbon.

### A. Gasification

Coal can be converted to a combustible gas by reaction with steam,

$$C + H_2O \longrightarrow CO + H_2,$$
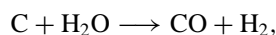
reports of which can be traced back to at least 1780. The reaction is endothermic by 32 kcal/mol and requires temperatures greater than 800°C for reasonable reaction rates.

The major carbon gasification reactions are shown in Table VI, and selected relative rates are shown in Table VII. The products of the carbon–steam reaction, carbon monoxide and hydrogen, called *synthesis gas* or *syngas,* can be burned or converted to methane:

| Name | Reaction | $\Delta H_{298}$ | Reaction temp. (°C) |
|---|---|---|---|
| Carbon–steam | $2C + 2H_2O \xrightarrow{\quad\quad} 2CO + 2H_2$ | 62.76 | 950 |
| Water–gas shift | $CO + H_2O \xrightarrow{\text{catalyst}} CO_2 + H_2$ | −9.83 | 325 |
| Methanation | $CO + 3H_2 \xrightarrow{\text{catalyst}} CH_4 + H_2O$ | −49.27 | 375 |
| | $2C + 2H_2O \xrightarrow{\quad\quad} CH_4 + CO_2$ | 3.66 | |

In this scheme, hydrogen from water is added to carbon, and the oxygen of the water is rejected as carbon dioxide.

To make this chemistry a viable technology, many other steps are involved, a key one being acid gas and ammonia removal from raw syngas. Coal is not made only of carbon, hydrogen, and oxygen, but also contains sulfur and nitrogen. During gasification the sulfur is converted to hydrogen sulfide and the nitrogen to ammonia. Carbon dioxide is also produced. These gases must be removed from the raw syngas because they adversely affect the catalysts used for subsequent reactions.

Another issue in coal gasification is thermal efficiency. The thermodynamics of the reactions written above indicate that the overall $C \rightarrow CH_4$ conversion should be thermoneutral. However, in practice, the exothermic heat of the water–gas shift and methanation reactions is generated at a lower temperature than the carbon–steam reaction and therefore cannot be utilized efficiently. Recent catalytic gasification approaches are aimed at overcoming these thermal inefficiencies.

Coal gasification plants are in operation today and are used mainly for the production of hydrogen for ammonia synthesis and the production of carbon monoxide for chemical processes and for syngas feed for hydrocarbon synthesis.

### B. Liquefaction

As already discussed, coal is a solid, three-dimensional, cross-linked network. Converting it to liquids requires the breaking of covalent bonds and the removal of carbon or the addition of hydrogen. The former method of producing

**TABLE VI  Carbon Gasification Reactions**

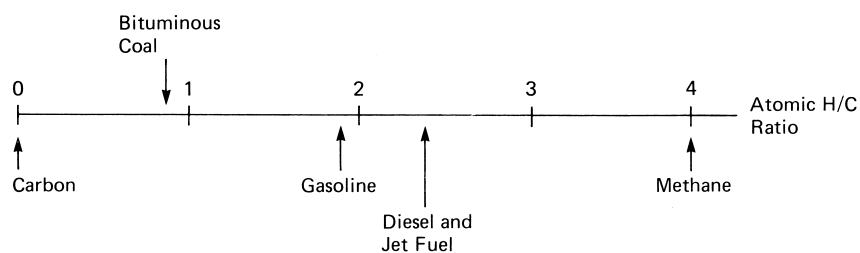| Name | Reaction | $\Delta H_{298}$ (kcal/mol) |
|---|---|---|
| Combustion | $C + O_2 \rightarrow CO_2$ | −94.03 |
| | $C + \frac{1}{2}O_2 \rightarrow CO$ | −26.62 |
| | $CO + \frac{1}{2}O_2 \rightarrow CO_2$ | −67.41 |
| Carbon–steam | $C + H_2O \rightarrow CO + H_2$ | 31.38 |
| Shift | $CO + H_2O \rightarrow CO_2 + H_2$ | −9.83 |
| Boudouard | $C + CO_2 \rightarrow 2CO$ | 41.21 |
| Hydrogenation | $C + 2H_2 \rightarrow CH_4$ | −17.87 |
| Methanation | $CO + 3H_2 \rightarrow CH_4 + H_2O$ | −49.27 |

**FIGURE 11**  Relationship of resource to end products on the basis of the ratio of atomic hydrogen to carbon.

liquid is called *pyrolysis*, the latter *hydroliquefaction*. Coal pyrolysis, or destructive distillation, is an old technology that started on a commercial scale during the industrial revolution. Hydroliquefaction, the reaction of hydrogen with coal to make liquids, was first reported in 1869 and was developed into a commercial process in the period from 1910 to 1920.

In general, both pyrolysis and hydroliquefaction reactions begin with the same step, thermal homolytic bond cleavage to produce free radicals, effecting a molecular weight reduction of the parent macromolecule (Fig. 12). If the radicals are healed, smaller neutral molecules result, leading to liquid and gaseous products. Those radicals that are not healed recombine to form products of the same or higher molecular weight than the parent, leading eventually to a highly cross-linked carbonaceous network called *coke* (if the material passed through a plastic phase) or *char*. In pyrolysis, radicals are healed by whatever hydrogen is present in the starting coal. In hydroliquefaction, excess hydrogen is usually added as molecular hydrogen and/or as molecules (such as 1,2,3,4-tetrahydronaphthalene) that are able to donate hydrogen to the system. Thus, hydroliquefaction produces larger amounts of liquid and gaseous products than pyrolysis at the expense of additional hydrogen consumption. As shown in Table VIII, conventional pyrolysis takes place at temperatures higher than those of hydroliquefaction, but hydroliquefaction requires much higher pressures.

Many techniques have been used in hydroliquefaction. They all share the same thermal initiation step but differ in how hydrogen is provided: from molecular hydrogen, either catalytically or noncatalytically, or from organic donor molecules. Obviously, rank and type determine a particular coal's response to pyrolysis and hydroliquefaction (Table IX), and the severity of the processing conditions determines the extent of conversion and product selectivity and quality.

A large number of pyrolysis and hydroliquefaction processes have been and continue to be developed. Pyrolysis is commercial today in the sense that metallurgical coke production is a pyrolysis process. Hydroliquefaction is not commercial, because the economics today are unfavorable.

## C. Hydrocarbon Synthesis

Liquids can be made from coal indirectly by first gasifying the coal to make carbon monoxide and hydrogen, followed by hydrocarbon synthesis using chemistry discovered by Fischer and Tropsch in the early 1920s. Today, commercialsize plants in South Africa use this approach for making liquids and chemicals. Variations of hydrocarbon synthesis have been developed for the production of alcohols, specifically methanol. In this context, methanation and methanol synthesis are special cases of a more general reaction. Catalyst, temperature, pressure, and reactor design (i.e., fixed or fluid bed) are all critical variables in determinig conversions and product selectivities (Table X).

$$n\,CO + (2n+1)H_2O \xrightarrow{\text{catalyst}} C_nH_{2n+2} + n\,H_2O$$

$$2n\,CO + (n+1)H_2 \xrightarrow{\text{catalyst}} C_nH_{2n+2} + n\,CO_2$$

$$n\,CO + 2n\,H_2 \xrightarrow{\text{catalyst}} C_nH_{2n} + n\,H_2O$$

$$n\,CO + 2n\,H_2 \xrightarrow{\text{catalyst}} C_nH_{2n+1}OH + (n-1)H_2O$$
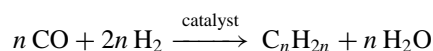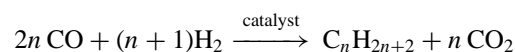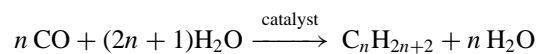
**TABLE VII  Approximate Relative Rates of Carbon Gasification Reactions (800°C, 0.1 atm)**

| Reaction | Relative rate |
|---|---|
| $C + O_2 \longrightarrow CO_2$ | $1 \times 10^5$ |
| $C + H_2O \longrightarrow CO + H_2$ | 3 |
| $C + CO_2 \longrightarrow 2\,CO$ | 1 |
| $C + 2\,H_2 \longrightarrow CH_4$ | $3 \times 10^{-3}$ |

**TABLE VIII  Typical Coal Liquefaction Conductions**

| Reaction | Temperature (°C) | Pressure (psig) |
|---|---|---|
| Pyrolysis | 500–650 | 50 |
| Hydroliquefaction | 400–480 | 1500–2500 |

FIGURE 12 Schematic representation of coal liquefaction.

Clearly, a broad spectrum of products can be obtained, and research today is aimed at defining catalysts and conditions to improve the product selectivity. One recent advance in selectivity improvement involves gasoline synthesis over zeolite catalyst from methanol, which is made selectively from syngas. The good selectivity to gasoline is ascribed to the shape-selective zeolite ZSM-5, which controls the size of the molecules synthesized. Called MTG (methanol to gasoline), the process operates at about 400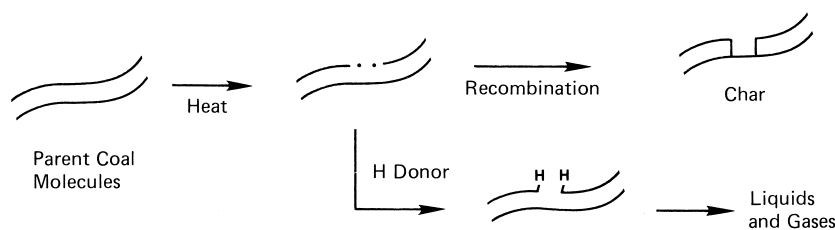°C and 25 psig. Large-scale demonstration plants are currently under construction in New Zealand and the Federal Republic of Germany.

## D. Coke Formation

The most important nonfuel use of coal is the formation of metallurgical coke. Only a few coals form commercially useful cokes, while many more agglomerate when heated. Good coking coals sell at a premium price. The supply of coals suitable for commercial coke making is sufficiently small so that blends of good and less good coking coals are used. Desirable coke properties are also obtained by blending. The coking ability of the blends can be estimated from the maceral composition and rank of the coals used in the blends. The best coking coals are bituminous coals having $\sim$86% carbon. Other requirements for a commercial coking coal or coal blend are low sulfur and mineral matter contents and the absence of certain elements that would have a deleterious effect on the iron or steel produced.

In coke making, a long, tall, narrow (18-in.) oven is filled with the selected coal or coal blend, and it is heated very slowly from the sides. At $\sim$350°C, the coal starts to soften and contract. When the temperature increases to between 50 and 75°C above that, the coal begins to soften and swell. Gases and small organic molecules (coal tar) are emitted as the coal decomposes to form an easily deformable plastic mass. It actually forms a viscous, structured plastic phase containing highly organized liquid crystalline regions called mesophase. With further heating, the coal polymerizes to form a solid, which on further heating forms metallurgical coke. The chemical process is extraordinarily complex.

**TABLE X Hydrocarbon Synthesis Product Distributions (Reduced Iron Catalyst)**

| Variable | Fixed bed | Fluid bed |
|---|---|---|
| Temperature | 220–240°C | 320–340°C |
| $H_2$/CO feed ratio | 1.7 | 3.5 |
| Product selectivity (%) | | |
|   Gases ($C_1$–$C_4$) | 22 | 52.7 |
|   Liquids | 75.7 | 39 |
|   Nonacid chemicals | 2.3 | 7.3 |
|   Acids | — | 1.0 |
| Gases (% total product) | | |
|   $C_1$ | 7.8 | 13.1 |
|   $C_2$ | 3.2 | 10.2 |
|   $C_3$ | 6.1 | 16.2 |
|   $C_4$ | 4.9 | 13.2 |
| Liquids (as % of liquid products) | | |
|   $C_3$–$C_4$ (liquified petroleum gas) | 5.6 | 7.7 |
|   $C_5$–$C_{11}$ (gasoline) | 33.4 | 72.3 |
|   $C_{12}$–$C_{20}$ (diesel) | 16.6 | 3.4 |
|   Waxy oils | 10.3 | 3.0 |
|   Medium wax (mp 60–65°C) | 11.8 | — |
|   Hard wax (mp 95–100°C) | 18.0 | — |
|   Alcohols and ketones | 4.3 | 12.6 |
|   Organic acids | — | 1.0 |
| Paraffin/olefin ratio | $\sim$1:1 | $\sim$0.25:1 |

**TABLE IX Variation of Hydroliquefaction Conversion with Rank**

| | Batch hydroliquefaction | |
|---|---|---|
| Coal rank | Conversion to liquids and gas (% of coal, dry ash-free basis) | Barrels oil/ton coal |
| Lignite | 86 | 3.54 |
| High-volatile bituminous-1 | 87 | 4.68 |
| High-volatile bituminous-2 | 92 | 4.90 |
| Medium-volatile bituminous | 44 | 2.10 |
| Anthracite | 9 | 0.44 |

The initial softening is probably due to a mix of physical and chemical changes in the coal. At the temperature where swelling begins, decomposition is occurring and the coal structure is breaking down. Heat-induced decomposition continues and the three-dimensionally linked structure is destroyed, producing a variety of mobile smaller molecules that are oriented by the large planar aromatic systems they contain. They tend to form parallel stacks. Eventually, these repolymerize, forming coke.

## SEE ALSO THE FOLLOWING ARTICLES

COAL GEOLOGY • COAL PREPARATION • GEOCHEMISTRY, ORGANIC • FOSSIL FUEL POWER STATIONS—COAL UTILIZATION • MINING ENGINEERING • POLLUTION, AIR

## BIBLIOGRAPHY

Berkowitz, N. (1979). "As Introduction to Coal Technology," Academic Press, New York.
Elliott, M. A., ed. (1981). "Chemistry of Coal Utilization," 2nd suppl. vol., Wiley (Interscience), New York.
Francis, W. (1961). "Coal: Its Formation and Composition," Arnold, London.
George, G. N., Gorbaty, M. L., Kelemen, S. R., and Sansone, M. (1991). "Direct determination and quantification of sulfur forms in coals from the Argonne Premium Sample Program," *Energy Fuels* **5,** 93–97.
Schobert, H. H., Bartle, K. D., and Lynch, L. J. (1991). "Coal Science II," ACS Symposium Series 461, American Chemical Society, Washington, DC.
Gorbaty, M. L. (1994). "Prominent frontiers of coal science: Past, present and future," *Fuel* **73,** 1819.
Van Krevelen, D. W. (1961). "Coal," Elsevier, Amsterdam.
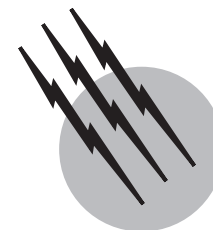
# Energy Efficiency Comparisons Among Countries

**Dian Phylipsen**

*Ecofys*

## GLOSSARY

**Energy efficiency** A measure of the amount of energy consumed per unit of a clearly defined human activity.

**Energy efficiency indicator** An approximation of energy efficiency such as the energy intensity or the specific energy consumption. Most efficiency indicators measure the inverse efficiency; if the efficiency increases, the value of the indicator decreases.

**Energy intensity** An energy efficiency indicator, usually based on economic measures of activity, such as value added. Energy intensity generally measures a combination of energy efficiency and structure. The energy intensity is often used at a higher level of aggregation than the specific energy consumption.

**Specific energy consumption (SEC)** An energy efficiency indicator, usually based on a physical measure for activity, such as the production volume. The SEC is also called energy intensity when it measures a combination of energy efficiency and sector structure.

**Structure** The mix of activities or products within a country or sector.

**ENERGY EFFICIENCY** comparison in the manufacturing industry is a tool that can be used by industry to assess its performance relative to that of its competitors. For this purpose it has been used for many years by the petrochemical industry and refineries, and interest on the part of other industries is growing. The tool can also be used by policy makers when designing policies to reduce greenhouse gas emissions and prioritizing energy savings options. International comparisons of energy efficiency can provide a benchmark against which a country's performance can be measured and policies can be evaluated. This is done both by national governments (such as

the Netherlands and South Korea) and by international organizations such as the IEA (the International Energy Agency) and APERC (the Asian Pacific Energy Research Center).

## I. MEASURING ENERGY CONSUMPTION

### A. System Boundaries

The numerator of the energy efficiency indicator measures energy consumption, which can be defined in many ways:

*End-use energy* measures the energy used for each final application. With surveys of energy consumption and the equipment used (as well as information on frequency of use, home size, etc.), many of the key end uses in homes, buildings, vehicles, and even in complex industrial processes can be determined.

*Final energy* is the energy purchased by final energy users, excluding on-site conversion losses, measured as purchased energy − the consumption by in-plant conversion processes + the production by in-plant conversion processes. The difference between end-use energy and final energy is actually that more conversions are taken into account for the latter.

*Purchased energy* the amount of energy bought from suppliers. In cases where extraction of energy carriers takes place within the industry itself (e.g., wood in the paper industry or hydropower), purchased energy will underestimate the actual energy consumption. Purchased energy overestimates actual energy consumption in cases where part of the purchased energy is exported to other sectors after conversion (e.g., coke oven gas in steel production, excess electricity from combined heat and power generation).

*Net available energy* is the amount of energy that actually becomes available to the user, including both net purchased energy, stock changes, and the user's extraction of energy forms.

*Demand for primary energy carriers* cannot be measured but must be calculated on the basis of, for example, final energy consumption and knowledge of flows in the energy sector itself. The simplest approach used is that of adding sectoral fuel consumption to the electricity consumption using a conversion efficiency for electric power plants. More sophisticated methods also take into account the losses in production, conversion, and transportation of the fuels. Whether to use uniform conversion, factors for all countries and for different points in time or country-dependent and time-dependent factors depends on the goal of the analysis.

### B. Units

Energy is measured in many different units around the world. In addition to the SI (Système Internationale) unit, the joule, energy can measured in calories, British thermal units (Btu), tonnes of oil equivalent (toe), tonnes of coal equivalent (tce), kilowatt-hours (kWh), and in physical units (e.g., cubic meters of gas, barrels of oil). The use of these physical units poses a problem, because the energy density of energy carriers can vary substantially between countries. For the other units, it suffices to list the SI unit, if necessary, in combination with a local unit.

### C. Heating Values

Energy may be measured in terms of the lower heating value (LHV) or the higher heating value (HHV). Higher heating values measure the heat that is freed by the combustion of fuels when the chemically formed water is condensed. Lower heating values measure the heat of combustion when the water formed remains gaseous. Although the energy difference between HHV and LHV may not always be technically recoverable, HHV is a better measure of the energy inefficiency of processes.

## II. MEASURING ECONOMIC ACTIVITY

Economic activity within the industry sector can be measured in either physical or economic terms.

### A. Economic Measures

Economic indicators that can be used to measure the activity in the industry sector are as follows:

*Gross output* is the most comprehensive measure of production, including sales or receipts and other operating income plus inventory changes.

*The value of shipments* is a measure of the gross economic activity of a particular sector, including total receipts for manufactured products, products bought and sold without further manufacturing, and rendered services. As such, it includes the costs of inputs used in that sector.

*The value of production* measures the value of the amount of goods produced, instead of the amount of goods sold and is calculated by correcting the value of shipments for inventory changes.

*Value added* is defined as the incremental value added by sectors or industrial processes, usually measured by the price differential between total cost of the input and price of the output (including services). This is also called *gross value added*. A *net value added* can also be defined by subtracting expenditures for depreciation from gross

value added. Another distinction can be made into (net or gross) value added at market prices or at factor costs. The latter is calculated by subtracting indirect taxes and subsidies from the former. Usually gross value added at market prices is reported.

Value added represents the net economic output for a given sector or process, and as such the value added for sectors and processes can be added without double counting. The total value added of all sectors more or less equals GDP.

A *production index* reports the ratio of production in a given sector in a given year to the production in the same sector in a base year. Often the index is partly physical, partly economic: Physical production is weighed on the basis of the economic value of the different products. This measure is mainly used for studying changes in one country over time.

To be able to compare activity measured in economic terms between countries, national currencies have to be converted to one single currency. This can be done by using market exchange rates or by using purchasing power parities. The first approach simply converts currencies into the reference currency (e.g., U.S. dollars) by applying average annual market exchange rates. Purchasing power parities (PPP) are the rates of currency conversion used to equalize the purchasing power of different currencies. For developing countries, PPP-based, estimates of GDP may be much (2 to 3 times) higher than market exchange-based rates. This means that the same amount of value added would comprise more goods and activities in developing countries than it would in OECD countries. A certain amount of money, converted at PPP rates to national currencies, will buy the same amount of goods and services in all countries.

If comparisons are made for various years, the devaluation of currencies also has to be taken into account. Time series are therefore usually expressed in fixed-year prices; actual expenditures are converted to, for example, 1990 dollars. Because inflation rates vary from country to country and between industries, national currencies should be corrected for domestic inflation rates before they are converted into the reference currency.

## B. Physical Measures

Physical production represents the amount of salable product. The most widely used physical measure of activity is volume based (i.e., metric tonnes). Physical measures of activity are the same for different countries or are easily convertible (e.g., short tonnes to metric tonnes). The measure also varies less over time. Therefore, physical mea-

sures of activity are more suitable for energy-intensive industries, such as the iron and steel industry and the cement industry. For light industry sectors, an economic measure of activity may be more appropriate because the mix of products is very diverse.

## III. ENERGY EFFICIENCY INDICATORS

An energy efficiency indicator gives the amount of energy that is required per unit of human activity or production. Based on the type of demand indicators used, either economic or physical energy efficiency indicators are distinguished.

### A. Economic Efficiency Indicators

The most aggregate economic energy efficiency indicator is generally called the *energy intensity*. A commonly used economic indicator in industry is based on value added. Most national statistics produce figures for value added. Because activities in various sectors are expressed in one measure, it is possible to add them, which is an important advantage of this type of indicator.

The most important shortcoming of economic indicators is the lack of ability to reflect the structural differences among countries, or within a country in time. To some extent this can be corrected for by using decomposition analysis (see Section IV.C). However, usually only intersectoral differences in structure are accounted for in this type of analysis. In addition, economic efficiency indicators can be strongly influenced by changes in product prices, feedstock prices, etc. Furthermore, parts of the economy do not generate value added (as measured in the official statistics), such as trade in informal economies (black markets) or barter economies, intermediate products that are used within the same company, etc. For historical analyses, we also have the a problem that planned economies did not use concepts like GDP and value added.

### B. Physical Efficiency Indicators

The physical energy efficiency indicator is called *specific energy consumption* (SEC). It is defined as the ratio of energy consumption to a measure of human activity in physical terms, i.e., per weight unit of product.

The advantage of physical indicators is that there is a direct relationship with energy efficiency technology: The effect of energy saving measures, in general, is expressed in terms of a decrease of the SEC. Since physical energy efficiency indicators are necessarily developed for a lower aggregation level than economic efficiency indicators, the occurrence of structural effects is by definition less of a

problem than for economic indicators. However, at the level of product mix structural differences may also play a role for physical indicators. In cases where the SEC measures a combination of energy efficiency and structure, it can also be referred to as *energy intensity*.

A problem with physical efficiency indicators is that it is not easy to count together different types of activity. Furthermore, there is no one-to-one relation between physical indicators and the monetary quantities that form the basis of economic models.

### C. Comparing Economic and Physical Energy Efficiency Indicators

Studies have shown that for the heavy industries the correlation between different economic indicators is poor. Differences can be so large that one type of economic indicator may suggest an increase in energy efficiency, while another type of economic indicator suggests a decrease in energy efficiency. Comparisons between physical and economic energy efficiency indicators also show large differences. No one single economic efficiency indicator correlates best with the physical. Rather, correlation varies by country and by sector. Generally, physical energy efficiency indicators are considered to provide a better estimate of actual energy efficiency developments in the heavy industry.

## IV. ACCOUNTING FOR SECTOR STRUCTURE

At the national level, structure is defined by the share the different economic sectors have in gross domestic product, for exampale, the share of manufacturing versus the share of the service sector. At a lower level of aggregation, structure is defined by the mix of products and those aspects that influence product mix and product quality (such as feedstock).

Because of the influence of sector structure on energy intensity, cross-country or cross-time comparisons cannot be made based solely on trends in the absolute value of the energy efficiency indicator (SEC) for each country. To be able to compare developments in energy efficiency between countries and over time, differences and changes in economic structure have to be taken into account. Two approaches are used for this, the reference plant approach and decomposition analysis.

### A. Reference Plant Approach

In the reference plant approach the actual SEC is compared to a reference SEC that is based on a country's sector struc-

ture. A commonly used definition of the reference SEC is the lowest SEC of any commercially operating plant worldwide. Such a best plant is defined for each process and a country's sectoral reference SEC is established by calculating the weighted average of the reference SECs of individual processes and/or products present in that country. This means that both the actual SEC and the reference SEC are similarly affected by changes in sector structure.

The difference between the actual SEC and the reference SEC is used as a measure of energy efficiency, because it shows which level of energy efficiency is achievable in a country with a particular sector structure. Note that this is a somewhat conservative estimate, since the reference SEC is usually not an absolute minimum requirement, such as a thermodynamic minimum. Usually, the SEC of the best plant observed worldwide can, technically speaking, still be reduced, but the necessary measures may not be implemented for a variety of reasons, such as the continuity of production, economic considerations, etc. It must also be noted that technological and economic limitations change over time, resulting in a lower reference SEC.

The relative differences between actual and reference SEC can be compared between countries. Usually this is done by calculating an energy efficiency index (EEI): the ratio between actual SEC and reference SEC. The closer the EEI approaches 1, the higher the efficiency is.

Figure 1 explains the methodology with an example for the pulp and paper industry. The SEC is depicted as a function of the ratio between the amount of pulp produced versus the amount of paper produced. This accounts for differences between countries in the use of waste paper as a feedstock for paper production and in import/export patterns of the energy-intensive intermediate pulp. In the graph, called the structure/efficiency plot, both the actual (national average) SEC and the reference SEC are shown (respectively, the upper end and the lower end of the vertical lines in Fig. 1). The diagonal line represents the reference SEC at various pulp/paper ratios for a given product mix (such as high-quality printing paper, wrapping paper or newsprint). The reference level of individual countries may differ from this line if the manufactured product mix within a country is different (such as for country C).

If the countries in Fig. 1 (A, B, C, and D) were to be compared on the basis of the SEC alone, one would tend to conclude that countries A and B are comparably efficient, country C is less efficient, and country D is, by far, the least efficient of the four countries shown. However, the s/e plot shows that country D's high specific energy consumption is largely caused by a more energy-intensive product mix.

A comparison of the relative differences between actual and reference SEC shows that countries A and D are equally efficient. Country B is the most efficient, while
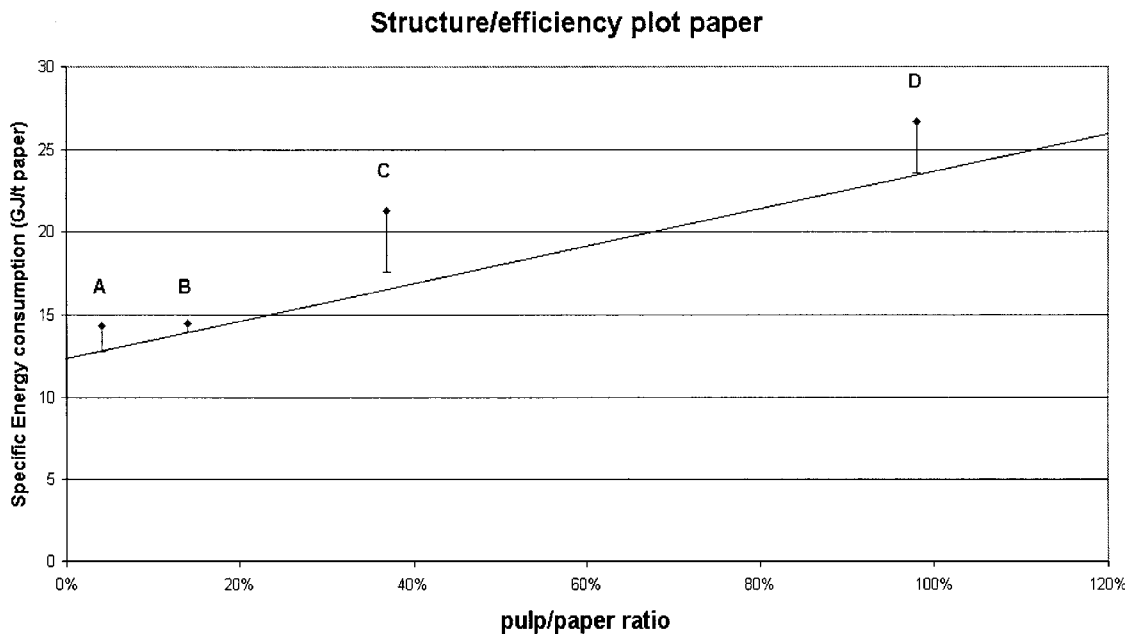
## Structure/efficiency plot paper



**FIGURE 1** A structure/efficiency (s/e) plot for the paper industry.

country C is the least efficient country. The EEIs of A, B, C, and D are 112, 105, 113, and 121, respectively. An EEI of 105 means that the SEC is 5% higher than the reference level, so that 5% of energy could be saved at the given sector structure by implementing the reference level technology.

## B. Decomposition Analysis

Methods have been developed in which changes in overall energy consumption are decomposed into changes in activity level, sector structure and energy efficiency. Such methods are called *decomposition analysis*. These methods are based on the following relation:

$$E_t = A \sum S_{i,t} \bullet I_{i,t},$$

where

$E_t$ = sectoral energy consumption in year $t$
$A$ = sectoral output
$S_{i,t}$ = structure of subsector $i$ in year $t$, measured by $A_i/A$
$A_{i,t}$ = output of subsector $i$ in year $t$
$I_{i,t}$ = energy intensity of activity $i$ in year $t$

In decomposition analysis, indices are constructed that show what sectoral energy consumption would have been in year $t$ if only one of the above parameters ($A$, $S$, or $I$) had changed, while the others remained the same as the level in the base year (year 0). Depending on the ex-

act methodology used, these indices are called Divisia or Laspeyres indices. The main differences between methodologies occur in which base year is chosen (e.g., first year, last year, or the middle year of the analysis) and whether the base year is fixed or not. Fixed year Laspeyres indices can be expressed as follows:

$$LA_t = \frac{A_t \sum S_{i,0} \bullet I_{i,0}}{E_0},$$

$$LS_t = \frac{A_0 \sum S_{i,t} \bullet I_{i,0}}{E_0},$$

$$LI_t = \frac{A_0 \sum S_{i,0} \bullet I_{i,t}}{E_0}.$$

Here, $LA$ measures the change in energy use that would have occurred given actual changes in output, but at fixed structure and intensity. $LS$ measures the change that would have occurred due to structural change if output and intensity had both remained constant; $LI$ shows the development energy consumption that would have existed if output and structure did not change.

As shown by the use of the term *energy intensity* in the equation, decomposition analysis is usually carried out on the basis of economic energy efficiency indicators. This also brings with it the problems associated with those type of indicators as mentioned in Section III. However, recently a method has been developed to carry out a decomposition analysis on the basis of physical indicators.

In this method, activity is measured in physical units. Also the effects of a change in energy efficiency are not

estimated using the energy intensity, but on the basis of a physical production index:

$$\text{PPI} = \sum P_x \bullet w_x$$

where

PPI = physical production index
 $P_x$ = production of product $x$ (kg)
$w_x$ = weighting factor

The weighting factors are chosen in such a way that they represent the amount of energy needed to produce each of the products. An appropriate set of weighting factors may be based on the best plant SECs as used in the reference plant.

When the weighting factors are the same for all the years in the analysis, the development of the PPI represents a frozen efficiency development of the energy consumption. The energy consumption of a sector can then also be written according to:

$$\sum E = \sum P \times \frac{\text{PPI}}{\sum P} \times \frac{\sum E}{\text{PPI}}.$$

Here, $\sum P$ is the parameter for activity (total amount produced in kilograms), PPI/$\sum P$ is related to the composition of the product mix (structure) and $\sum E$/PPI is an indicator of the energy efficiency (on a physical basis). Note that this ratio is actually the same as the energy efficiency index, EEI, as defined in Section IV.A. Figure 2 shows one example of a decomposition analysis on a physical basis.

Note that the countries are both indexed to their own reference situation (in 1973), not to one common reference

situation. This means that an index of 100% for one country represents a different situation than it does for the other country. In the case of the two countries listed in Fig. 2, the Japanese paper industry was more efficient in 1973 than the U.S. paper industry. As a result, the 1993 differences in energy efficiency are larger than Fig. 2 suggests.

## V. TIME SERIES

The reference plant approach described in Section IV.A cannot only be used for a cross-country comparison of energy efficiency, but also for a cross-time analysis of changes in energy efficiency in one or more countries.

To construct a time series, both actual SEC and reference SEC are determined for different points in time. Here too, actual comparison is not based on the absolute level of the actual SEC alone, but on the relative difference between actual and reference SEC, the EEI.

Figure 3 shows a cross-time and cross-country comparison based on this approach. As stated before, a country becomes more efficient as EEI approaches 1. Note that an EEI of 1 can represent a different specific energy consumption for different points in time (and for different countries).

## VI. REAGGREGATION OF ENERGY EFFICIENCY INDICATORS

After energy efficiency indicators have been determined at a low level of aggregation, often further reduction of
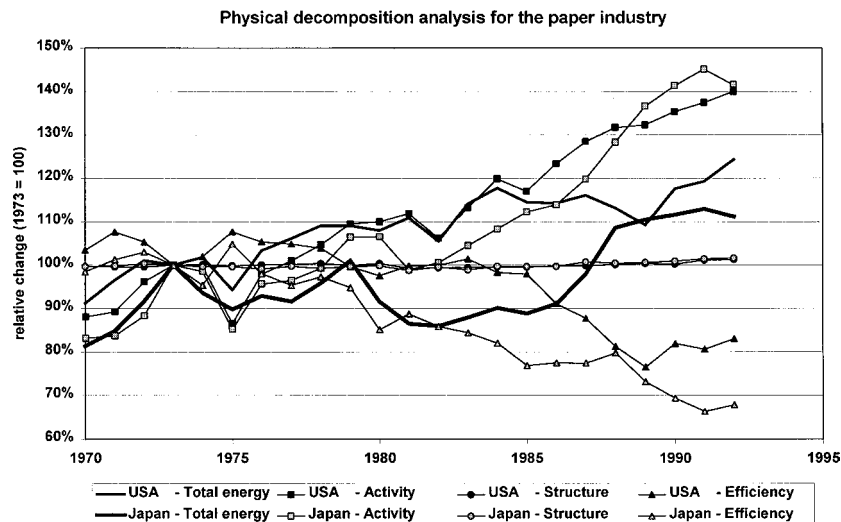


**FIGURE 2**  A decomposition analysis based on physical indicators for the pulp and paper industry. The relative change (compared to 1973) is shown for total energy consumption in the industry, and for the underlying changes in activity level, sector structure (product mix), and energy efficiency.
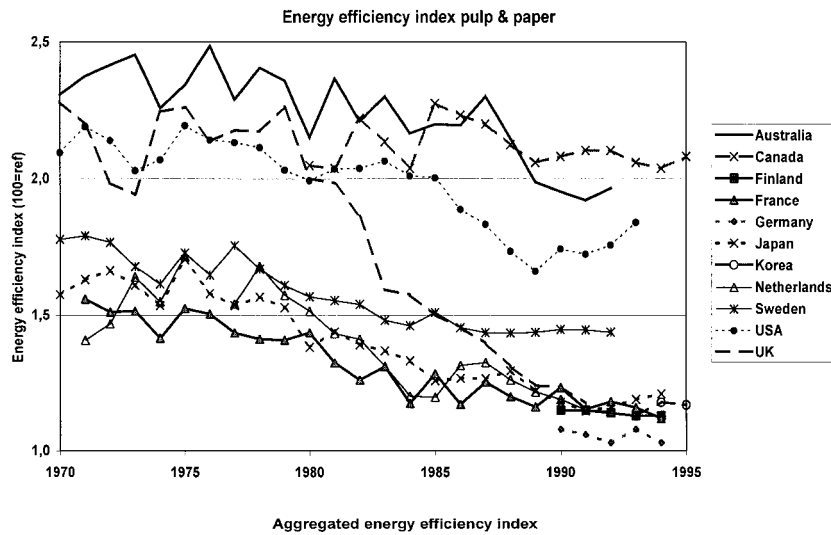
**Energy efficiency index pulp & paper**



**Aggregated energy efficiency index**

**FIGURE 3** The development of the EEI in the pulp and paper industry over time for various countries. Countries' actual primary energy consumption is indexed to the primary energy consumption of the best plant observed worldwide.

the number of indicators into more aggregate figures is desirable (e.g., one per sector).

One approach for this reaggregation could be to calculate an aggregate SEC, based on the actual SEC and the reference SEC of all individual products and processes according to the following equation. In doing this, each of the products is weighed by their relevance for energy consumption (which is, for this aim, more appropriate than weighing by economic value or product quantity):

$$\sec_{ag} = \frac{\sum m_i \bullet \mathrm{SEC}_i}{\sum m_i \bullet \mathrm{SEC}_{\mathrm{ref},i}} = \frac{\sum E_i}{\sum m_i \bullet \mathrm{SEC}_{\mathrm{ref},i}},$$

where

$\sec_{ag}$ = aggregate specific energy consumption (dimensionless)

$E_i$ = energy consumption for product $i$

$m_i$ = production quantity of product $i$

$\mathrm{SEC}_i$ = specific energy consumption of product $i$

$\mathrm{SEC}_{\mathrm{ref},i}$ = a reference specific energy consumption of product $i$

The aggregated SEC equals 1 if all products have a SEC that is equal to the reference SEC. The higher the SECs of each individual products are, the higher the aggregate value (and the higher the room for improvement). Figure 4

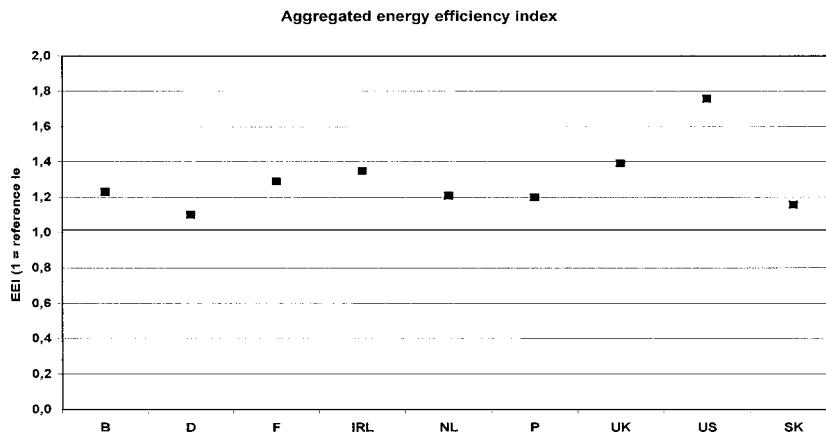**Aggregated energy efficiency index**



**FIGURE 4** The aggregated energy efficiency index for a number of countries. The sectors included are the iron and steel industry, the aluminium industry, the pulp and paper industry, the cement industry, and a major part of the chemical industry (the production of petrochemicals, ammonia, and chlorine).

shows how such a reaggregation might look for a number of countries.

Another possibility is to define a physical indicator according to a "basket of goods" such as 1 tonne of several primary industrial products. Aggregate specific energy consumption would then be defined as:

$$\text{SEC}_{ag} = \sum (W_i \bullet \text{SEC}_i),$$

where $W_i$ is a weight factor equivalent to the energy consumption of industry $i$, divided by the total energy consumption of the basket. One problem with this approach arises when a country does not produce all the products in the basket.

## SEE ALSO THE FOLLOWING ARTICLES

CARBON CYCLE • ENERGY FLOWS IN ECOLOGY AND IN THE ECONOMY • ENERGY RESOURCES AND RESERVES • GREENHOUSE EFFECT AND CLIMATE DATA • GREENHOUSE WARMING RESEARCH

## BIBLIOGRAPHY

Phylipsen, G. J. M., Blok, K., and Worrell, E. (1998). "Handbook on International Comparisons of Energy Efficiency in the Manufacturing Industry," Department of Science, Technology and Society, Utrecht University, The Netherlands.

Phylipsen, G. J. M., Blok, K., and Worrell, E. (1998). "Benchmarking the Energy Efficiency of the Dutch Energy-Intensive Industry; A Preliminary Assessment of the Effects on Energy Consumption and $CO_2$ Emissions," Department of Science, Technology and Society, Utrecht University, The Netherlands.

Schipper, L., Meyers, S., Howard, R. B., and Steiner, R. (1992). "Energy Efficiency and Human Activity: Past Trends, Future Prospects," Cambridge University Press, Cambridge.

Schipper, L. (1997). "Indicators of Energy Use and Efficiency: Understanding the Link Between Energy and Human Activity," International Energy Agency, Paris.

Schipper, L., and Haas, R. (1997). "Energy Policy Special Issue on Cross-Country Comparisons of Indicators of Energy Use, Energy Efficiency and $CO_2$ Emissions," June/July.
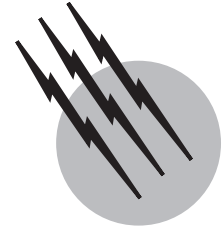
# Energy Flows in Ecology and in the Economy

**Sergio Ulgiati**

*University of Siena*

## GLOSSARY

**Cumulative exergy consumption** The "total consumption of exergy of natural resources connected with the fabrication of the considered product and appearing in all the links of the network of production processes," according to J. Szargut.

**Ecological footprint** A concept introduced to portray the environmental impact of a process or an economy, by measuring the amount of ecologically productive land that is needed to support it.

The concept has been refined by suggesting that the environmental support to a process (i.e., the resources and environmental services) is measured in terms of solar emergy.

**Emergy** The amount of exergy of one form that is directly or indirectly required to provide a given flow or storage of exergy or matter. In so doing, inputs to a process are measured in common units. Emergy is measured in *emjoules* (eJ). When inputs are measured in units of solar exergy, the solar Emergy is calculated, measured in *solar emergy joules* (seJ).

Emergy is a *memory* of the available energy (exergy) previously used up in a process and is used when the global support from biosphere to the process is investigated.

Converting flows of different nature into flows of one kind requires conversion coefficients called *trans-formities* (seJ/J). Emergy accounting acknowledges that different forms of exergy may require a different environmental support and may show different properties.

**Exergy** The amount of work obtainable when some matter is brought to a state of thermodynamic equilibrium with the common components of the natural surroundings by means of *reversible* processes, involving interaction only with the above-mentioned components of nature. Exergy is measured in joules or kcal.

Real processes are *irreversible*. Therefore, measures of actual work compared with theoretical value provide indicators of exergy losses due to the irreversibility of a process. Exergy is therefore a thermodynamic assessment of potential improvements at the scale of the investigated process.

**Life-cycle analysis** A technique for assessing the environmental aspects and potential impacts associated with a product, by (1) compiling an inventory of relevant inputs and outputs of a process, from the extraction of raw resources to the final disposal of the product when it is no longer usable (so called "from cradle to grave" analysis); (2) evaluating the potential environmental impacts associated with those inputs and outputs; and (3) interpreting the results of the inventory and impact phases in relation to the objectives of the study.

**Reversible/irreversible** A reversible change in a process is a change that can be reversed by an infinitesimal modification of any driving force acting on the system. The key word "infinitesimal" sharpens the everyday meaning of the word "reversible" as something that can change direction. In other words, the state of the system should always be infinitesimally close to the equilibrium state, and the process should be described as a succession of quasi-equilibrium states.

A system is in equilibrium with its surroundings if an infinitesimal change in the conditions in opposite directions results in opposite changes in its state.

It can be proved that a system does maximum work when it is working reversibly. Real processes are irreversible, therefore there is always a loss of available energy due to chemical and physical irreversibilities.

**Sustainable/sustainability** Sustainable development is a form of development that integrates the production process with resources conservation and environmental enhancement. It should meet the needs of the present without compromising our ability to meet those of the future.

Some of the aspects involved are the relationship between renewable and nonrenewable resources supporting society, environmental impact in relation to economic actions, intergenerational equity, and indirect impacts of trade.

**PATTERNS** of energy conversion and use, including energy quality evaluation and sustainable use of the energy that is available, are not easy issues to deal with. A general consensus is still far from being reached among energy experts and economists about evaluation procedures and energy policy making. The rationale underlying energy evaluation methodologies and the need for integration of different approaches for a comprehensive assessment are presented in this paper. How the evaluation approaches and space-time scales of the investigations are linked is stressed. Finally, a case study (industrial production of hydrogen and fuel cells) is presented and discussed by using an integrated biophysical approach, to show what each approach is able to contribute. The impossibility of achieving a universal theory of value is emphasized, but synergies achieved by integration of different evaluation procedures are pointed out.

## I. INTRODUCTION

According to Howard Odum; "A tank of gasoline drives a car the same distance regardless of what people are willing to pay for it." It is as a result of Odum's pioneering work in the sixties and subsequent work that the importance of energy as the simultaneous driving force of our economies accompanied by changing ecosystems has been increasingly recognized. Until recently, neoclassical economics has been unsuitable for understanding the role of the biophysical support provided at no cost by the environment to the development of economic processes. Whatever we may be willing to pay for our gallon of gasoline, if we cannot afford the cost, we are unable to fill the tank. Money flows parallel to biophysical flows supporting them, so that they must both be accounted for. The problem is that while money flows are easily dealt with by using market mechanisms, biophysical flows are not. Economists and policy makers must recognize the role of energy and other resource flows in support of human societies. Integration between economic and biophysical tools for a better management of available resources is urgently needed and may lead to a very different way of dealing with production processes and their relationship to the environment.

What do we pay for when we trade dollars for gasoline? Dollars are a cumulative measure of labor and services involved through the whole chain of processes from resource extraction to refining, storage, and transport to the

end user. We pay people for what they do to supply us with gasoline or food. In none of the steps of any of these processes do we pay nature for making the raw material or for concentrating it from dispersed states. "We do not pay nature for each acre of land taken out of biological production, nor do we pay nature for the millions of years of work it did in making coal or oil," M. Gilliland wrote in 1979.

The price of gasoline is highly variable, depending on oil-supply constraints at the sources, the cost of labor and refining processes, taxation in a given country (these taxes are used for services provided by other sectors of the economy), etc. In May 2000, the U.S. price was about $1.50 per gallon, and the price was about U.S. $4 per gallon in Italy, where heavy taxes (including a carbon tax) are levied on oil products. Refining oil may be much more expensive (cost more money) in Italy than in Saudi Arabia, but the process of refining takes about the same amount of energy per liter of gasoline in both countries, since essentially the same technology is used. Energy costs to make things are more constraining than current dollar costs. In principle, we may always find a country where the manufacturing process is cheaper (i.e., where labor is less expensive), but we are unlikely to find a country where the energy cost is greatly lowered, because the best technology is universally preferred.

Why does money appear to be so important and energy not? It is because money appears to be a scarce resource and depends on the amount of economic activity a given country or a person is able to control. Governments and people are not universally convinced that this scarcity also applies to energy and other biophysical resources.

## II. PARADOXES ABOUT ENERGY AND RESOURCE USE

T. R. Malthus recognized in 1798 that when a resource is available, it is used with a boosting effect on system growth. Systems grow using a resource until the resource becomes scarce due to the increasing size of the system. In other words, resource availability increases demand which, in turn, reduces the resource availability because of increasing resource use.

Let us assume now that we are able to save some energy by means of a more efficient domestic heater. In so doing, we save a corresponding amount of money. A part of this money (say the money saved in the first two to three years) will cover the cost of the new heater and the safe disposal of the old one. After that, we have some money available for other uses. We may therefore decide to purchase new books, to go more frequently to the movies, to buy new clothes, or to go to an expensive restaurant.

We might invest money in traveling for leisure or in remodeling our apartment. All of these activities are driven by labor, material, and energy inputs, resulting in making new goods and services available to us. Labor and material are ultimately supported by an additional energy input. The end of the story is that the energy saved in the heating is expended on the energy invested in making the new heater and the energy which supports the new activities and goods.

Depending on the kinds of commodities and services and their market prices (i.e., the costs of labor and services supporting them), we may cause the expenditure of more energy than the amount previously saved by using a more efficient heater.

A different strategy is also possible. The higher efficiency of the new heater makes it possible to supply more heat for the same amount of fuel input. We may therefore decide to heat our apartment more than in the past to feel more comfortable and eventually wear very light clothes at home. We will get used to a higher temperature and will continue this lifestyle even as the heater gets older and has reduced efficiency. This trend requires increased energy input over time. The final result may be the same or even a higher energy expenditure, in addition to the energy cost of making the new heater. The increased use of a resource as a consequence of increased efficiency is the well-known Jevons paradox.

Assume, finally, that increased costs of energy (due to taxes purposefully aimed at decreasing energy use) force us to run the heater at a lower temperature or for a shorter time (this happens in large Italian cities as the result of efforts to lower the environmental impacts of combustion). We are now unable to save money and therefore we are unable to invest this money in different kinds of energy-expensive goods. Tax revenues may, however, be reinvested by the government in development plans and services, thereby using energy to support these plans. We save energy, but the government does not. Also in this last case, no real energy savings occur, nor do we know if energy consumption remains stable or is likely to increase.

Only when a decrease of energy use at the individual level is coupled to strategies aimed at lowering the total energy expenditures at the country level will there be decreasing imports or decreasing depletion of local storages. This scenario happens, for instance, when the decreased buying power of the local currency of an oil-importing country or the increased oil costs in the international market force a country to buy less oil and apply energy-conservation strategies, as many European countries did after the 1973 and 1980 oil crises. When such a case occurs, depletion of oil storages in oil-exporting countries is slowed down and some oil remains unused

for a while. The oil price may drop down to levels that other countries are able to afford and thus the use rate will rise again.

What do we learn from these facts? We learn that when a resource is abundant or cheap, it is very likely to be depleted quickly by systems which are well organized for this purpose (Darwinian survival of the fittest). Increased efficiency calls for additional consumption. When a resource is scarce (high cost is a temporary kind of scarcity), use becomes more efficient to delay (not to prevent) final depletion. This trend was clearly recognized in 1922 by A. J. Lotka as a Maximum Power Principle, suggesting that systems self-organize for survival and that under competitive conditions systems prevail when they develop designs that allow for maximum flow of available energy. As environmental conditions change, the response of the system will adapt so that maximum power output can be maintained. In this way, systems tune their thermodynamic performance according to the changing environment.

## III. IS ENERGY A SCARCE RESOURCE?

In 1956, M. K. Hubbert pointed out that all finite resources follow the same trend of exploitation. Exploitation is first very fast due to abundance and availability of easily accessible and high-quality resources. In the progress of the exploitation process, availability declines and production is slowed down. Finally, production declines, due to both lower abundance and lower quality.[1] This trend allows the investigator to forecast a scenario for each resource, based upon estimates of available reserves and exploitation trends. New discoveries of unknown storages will shift the scenario some years into the future, but it cannot ultimately be avoided.

Hubbert also predicted the year 1970 for oil production to peak in the U.S. He likewise predicted the year 2000 for oil production to peak on a planetary scale. He was right in the first case and it seems he was not very far from right in the second prediction. Based upon this rationale and energy analysis data, C. A. S. Hall and C. J. Cleveland and Cleveland and co-workers have predicted that U.S. oil would no longer be a resource soon after the year 2000. Around this time, extraction energy costs would become higher than the actual energy yield, due to increased energy costs for research, deep drilling as well as to the lower quality and accessibility of still available oil storages.

Finally, in 1998 C. Campbell and J. Laherrère predicted the end of cheap oil within the next decade, based upon

[1]In principle, if an unlimited energy supply were available, there would be no material shortages because every substance would become reusable. This would shift the focus to energy resources, further enhancing their importance.

**TABLE I  World Proved Reserves and World Consumption of Fossil Fuels[a]**

| Fossil fuel | World proved reserves[b] ($10^9$ tons of oil equivalent) | Annual world consumption[c] ($10^9$ tons of oil equivalent) | Reserve-to-consumption ratio, R/C, years[d] |
|---|---|---|---|
| Oil | 140 | 3.46 | 40.5 |
| Coal | 490 | 2.13 | 229.9 |
| Natural gas | 134 | 2.06 | 65.2 |
| Total[e] | 764 | 8.53 | 99.8 |

[a] BP Amoco, 2000; based on 1999 estimates.
[b] Proved reserves are defined as those quantities which geological and engineering information indicates with reasonable certainty can be recovered in the future from known reservoirs under existing economic and operating conditions.
[c] Indicates the amount of the total reserve that is used up in a given year.
[d] Ratio of known reserves at the end of 1999 to energy consumption in the same year. Indicates how long the reserve will last at the present consumption rate. A reserve-to-annual production ratio, R/P, can also be used, with negligible differences due to annual oscillations of consumption and production rates.
[e] Estimated, assuming that oil, coal, and natural gas can be substituted for each other. Of course, while this is true for some uses (e.g., use in thermal power plants followed by conversion to electricity), it is not easy for others unless conversion technologies are applied and some energy is lost in conversion (e.g., converting coal to a liquid fuel for transportation use or to hydrogen for making electricity by using fuel cells).

geological and statistical data. It is important to emphasize here that they did not predict the end of oil production but only the end of oil production at low cost, with foreseeable consequences on the overall productive structure of our economies running on fossil fuels. The present fossil resources (known energy storages that can be exploited by means of present technology) and resource-to-consumption ratios are listed in Table I. We are not including nuclear energy in Table I, due to the fact that the increasing environmental concerns and public opinion in most industrialized countries are reducing the use of this technology and the so-called "intrinsically safe nuclear reactor" remains at a research stage of development.

Other authors believe that energy availability is not yet going to be a limiting factor for the stability of developed economies as well as to the growth of still-developing countries due to the possible exploitation of lower quality storages (e.g., oil from tar sands and oil shales, coal supplies) at affordable energy costs. However, they admit that economic costs are very likely to increase. Still available fossil fuels, coupled with energy-conservation strategies, as well as new conversion technologies (among which are coal gasification and gas reforming for hydrogen production, and use of fuel cells) might help delay the end of the fossil-fuel era, allowing for a transition period at higher

costs. The transition time may be long enough to allow for a redesigning of economic, social, and population-growth strategies to come out of the fossil-fuel era in an acceptable manner. It is still under debate whether the new pattern will be characterized by additional growth supported by new energy sources or will be a global and controlled downsizing of our economies, which was anticipated by H. T. Odum as a "prosperous way down."

Other kinds of scarcity, especially economic and environmental factors, are more likely to be strong constraints on energy use. We have already pointed out that an increase of the energy price is a kind of scarcity factor. Indeed, it affects availability of energy for those economies and individuals who are not able to afford higher costs. The economic cost of energy is closely linked to development, creation of jobs, and possible use of other energy sources that could not compete in times of cheap oil and might provide alternatives for future development in times of reduced fossil-fuel use. On the other hand, environmental concerns (related to nuclear waste, the greenhouse effect and global warming potential, need for clean urban air, etc.) affect societal assets in many ways and may alter production patterns when they are properly taken into account. All of these issues call for correct and agreed upon procedures of energy-sources evaluation, energy-conversion processes, energy uses, and energy costs.

## IV. HOW DO WE EVALUATE RESOURCE QUALITY? THE SEARCH FOR AN INTEGRATED ASSESSMENT IN BIOPHYSICAL ANALYSES

The relation between energy and the economy has been deeply investigated by J. Martinez-Alier who often emphasized that thermodynamic analyses cannot be the only bases for policy, nor will they lead to a new theory of economic value. They can, instead, do something that is much more important. In addition to providing a biophysical basis for economic descriptions, they may help to expand the scope of economics, away from a single numeraire or standard of evaluation toward systems thinking and a multicriteria framework. For this to be done, it must be accepted that solutions of different problems may require different methodologies to be dealt with properly. Each approach may only answer a given set of questions and may require the support of other methodologies for a more complete view of system behavior. The correct use and integration of complementary approaches might ease descriptions and enhance understanding of complex systems dynamics, including human economies.

The demand for integration of methodologies has been great during the past decade. Previous studies in the first two decades of energy analyses (the 1970s and 1980s) were very often devoted to assessing and demonstrating the superiority of one approach over others. After introduction of systems thinking, hierarchy theory, nonlinear dynamics, fractal geometry, and complex systems analysis, it is now increasingly clear that different approaches are very often required by the very nature of the problems we deal with to build a set of complementary descriptions that may be used to provide different assessments or different views of changes defined on different space-time scales. Biophysical analyses make it possible to bridge these different perspectives. The values of variables described at a certain level will affect and be affected by the values of variables describing the system at a different level. A clear assessment and call for integration was made at the beginning of the 1990s by the economist C. F. Kiker. Comparing available energy (exergy) analysis and emergy accounting methodologies, he came to the conclusion that "while the backgrounds of available energy and emergy analysts are substantially different, the underlying conceptual bases are sufficiently the same for the two methods to be used jointly. If clear analytical boundaries are specified for human-controlled systems, for natural systems and for systems derived from the interaction of these two types, available energy analysis and emergy analysis can be used together to evaluate the fund of service from these systems." He concluded that "what is needed now are real world examples that actually do this." M. T. Brown and R. Herendeen, in 1998, have compared in the same way the embodied energy approach and the emergy approach, looking for similarities and trying to make a clear assessment of differences and still unresolved questions. M. Giampietro and co-workers, also in 1998, applying concepts of complex systems theory to socioeconomic systems, described as adaptive dissipative holarchies,[2] emphasize that "improvements in efficiency can only be defined at a particular point in time and space (by adopting a quasi–steady state view); whereas improvements in adaptability can only be defined from an evolutionary perspective (by losing track of details). Short-term and long-term perspectives cannot be collapsed into a single description."

P. Frankl and M. Gamberale, in 1998, presented, the Life-Cycle Assessment (LCA) approach as a tool for energy policy "to identify (future) optimal solutions which maximize environmental benefits, . . . to identify priorities

---

[2]" . . . Supersystems (e.g., ecosystems or organisms) consist of subsystems (organisms and cells). They are nested within each other, and from this view are inseparable, since they, though clearly individual, consist of each other," wrote R. V. O'Neill and co-workers in 1986. Open subsystems of higher order systems are called holons, hierarchies of holons are called holarchies. They form a continuum from the cell to the ecosphere.

for orienting research and development activities and identifying best fiscal and market incentive measures." These authors recall that life-cycle analysis is a technique for assessing the environmental aspects and potential impacts associated with a product, by compiling an inventory of relevant inputs and outputs of a system, evaluating the potential environmental impacts associated with those inputs and outputs, and finally interpreting the results of the inventory and impact phases in relation to the goals of the study. They emphasize that "LCA results should not be reduced to a simple overall conclusion, since trade-offs and complexities exist for the systems analyzed at different stages of their life cycle," and that "there is no single method to conduct LCAs . . . LCA is a valuable analytical tool in the decision-making process in conjunction with other tools." It clearly appears that LCA could greatly benefit from integration procedures with other approaches such as exergy and emergy evaluations.

A call for the importance of using integrated methodologies has come from scientists attending the International Workshop" Advances in Energy Studies. Energy Flows in Ecology and Economy" at Porto Venere, Italy, in 1998, to improve definitions of scales and boundaries, methods of analysis, policy initiatives, and future research needs required to generate a balanced evaluation of humanity in the environment. In a final Workshop document, S. Ulgiati and colleagues challenge the scientific community to reframe previous insights and research activity into a systems perspective, enhancing complementarity of approaches, systems thinking, and multi scale, multi criteria evaluation tools. Finally, understanding the mechanism of the environmental support to human wealth and its embodiment in the goods that human societies make and use has been suggested as a research and policy priority to design sustainable interactions of humans and nature by the participants in the Second International Workshop on "Advances in Energy Studies. Exploring Supplies, Constraints, and Strategies," Porto Venere, Italy, held in 2000.

## V. TOWARD AN INTEGRATED EVALUATION APPROACH

Evaluation procedures so far proposed by many authors have been applied to different space-time windows of interest and were aimed at different investigations and policy goals. Many of these have offered valuable insights toward the understanding and description of aspects of resource conversions and use. However, without integration of different points of view, it is very unlikely that biophysical tools can be coupled to economic tools to support environmentally sound development policies.

We must also be aware that nature and economies are self-organizing systems, where processes are linked and affect each other at multiple scales. Investigating only the behavior of a single process and seeking maximization of one parameter (efficiency, production cost, jobs, etc.) is unlikely to provide sufficient insight for policy making. It is of primary importance to develop approaches that focus on the process scale and then to expand our view to the larger scale of the global system in which the process is embedded, to account for the interaction with other upstream, downstream, and parallel processes.

Finally, systems are dynamic and not static. They may appear static when a short time scale is chosen as the window of interest, but their behavior very often follows oscillating patterns. This also applies to societies. Many scientists and policy makers suggest that human societies can grow to a defined state, where resource supply and use are balanced by limiting their rate of depletion to the rate of creation of renewable substitutes. Since the planet as a whole is a self-organizing system, where storages of resources are continuously depleted and replaced at different rates and matter is recycled and organized within a self-organization activity, "sustainability concerns managing and adapting to the frequencies of oscillation of natural capital that perform best. Sustainability may not be the level 'steady-state' of the classical sigmoid curve but the process of adapting to oscillation. The human economic society may be constrained by the thermodynamics that is appropriate for each stage of the global oscillation," according to Odum[3]. Evaluation approaches must be able to account for this aspect too.

### A. Describing Matter Flows

No process description can be provided, at any level, without a clear assessment of matter flows. Either a mechanical transformation or a chemical reaction may be involved and quantifying input and output mass flows is only a preliminary step for a careful evaluation. Without nuclear transformations, matter is conserved according to a fundamental law of nature. Mass cannot be created or destroyed. Whenever we burn 40 kg of gasoline in our engine, we are not only supplying energy to run the car, we are also pushing far more than 40 kg of combustion compounds formed from gasoline into the air. There is a

---

[3]When resources are abundant, systems that are able to maximize resource throughput prevail, no matter their efficiency. Maximum efficiency is not always the best strategy for survival. Only when resources are scarce, systems self-organize to achieve maximum output by increasing efficiency. They therefore choose different thermodynamic strategies depending on external constraints and available resources.

complex set of chemical reactions involving carbon and atmospheric oxygen and nitrogen to yield carbon dioxide and nitrogen oxides. The mass of each individual atomic species is conserved in the process and the total mass of reactants must equal the mass of reaction products. If these changes are carefully accounted for, the process can be described quite well while we make sure that we are not neglecting any output chemical species that could be profitably used or that should be safely disposed of. In addition, when we expand our scale of investigation, we realize that each flow of matter supplied to a process has been extracted and processed elsewhere. In so doing, additional matter is moved from place to place, processed and then disposed of. For instance, nickel is used in the ceramic components of some kinds of fuel cells. The fraction of nickel in pentlandite (nickel ore) is only $5.5 \times 10^{-3}$ g Ni/g ore. To supply 1 g of pure nickel to an industrial process requires that at least 181 g of pentlandite be dug. If the nickel mine is underground, the upper layers of topsoil and clay are also excavated, perhaps a forest is also cut and biotic material degraded. Finally, some process energy in the form of fossil fuels is used up and a huge amount of process water is very often needed as input and formed as output. In addition, combustion also requires oxygen and releases carbon dioxide. Accounting for the amount of biotic and abiotic material involved in the whole chain of steps supporting the investigated process has been suggested as a measure of environmental disturbance by the process itself. We therefore find two main aspects of the material balance. When focusing on the product side, we must make sure that economically and environmentally significant matter flows have not been neglected. When addressing the input side, we must allow for the total mass transfer supporting a process and thereby measure indirectly how the process affects the environment. Both of these two points of view are based on the law of conservation of mass in nonrelativistic systems. Mass demand per unit of product ($g_{in}/g_{out}$) and total input mass supporting a process are suitable indicators. Output/input mass ratios (equal to one in the case of careful accounting[4]) and mass of by-products released per unit of main product ($g_{byprod}/g_{out}$) are also needed. After mass flows have been carefully accounted for, economic, energy, and environmental evaluations are made much easier, as with all of them a good process description is clearly the starting point.

[4]When chemical reactions are involved, all reactants and products must be accounted for. Atmospheric oxygen and nitrogen, for instance, are involved in the combustion reaction of fossil fuels via a large number of elementary reaction steps, with known or estimated reaction rates. Their masses must also be accounted for.

## B. Describing Heat Flows

First-law heat accounting should always be coupled to mass accounting as a required step of any biophysical investigation. The same item may be identified both as a material and heat source. For instance, oil can be measured in kilograms of hydrocarbons or in joules of combustion heat that it can generate. Both measures are useful, depending upon the goal of the evaluation. In plastic polymer production, oil is the substrate to be converted to the final product. Measuring oil in joules when it is more a substrate than a fuel may not be useful. In electricity production, oil is the fuel. Its combustion releases heat to power the turbine and generate electricity. In both cases, some waste heat is released to the environment while the main products are generated. In times of cheap oil, waste-heat flows may be neglected, since the focus is on the main product. In times of decreasing oil availability or increasing costs, there may be uses for waste heat, depending upon its temperature, and we may then redefine waste heat as a by-product for domestic heating or some industrial application. However, total input heat flow must always be equal to total output heat flow for isothermal systems. When heat is produced by combustion to make a product from raw mineral, we have a mass output in the form of the product and an added waste-heat output equal to the input energy used. When the product is a form of energy, such as electricity, the output will be the sum of waste heat and produced electricity.

Environmental as well as economic concerns may motivate us to investigate the consequences of releasing to the environment a resource characterized by a higher temperature than the environmental temperature. To address these opportunities, a careful description and quantification of input and output heat flows is needed.

The result of a first-law energy accounting is two-fold. As in the mass-accounting procedure, we remain with an intensive indicator of efficiency (energy output/energy input or energy input/unit of output) and an extensive accounting of total heat input and total heat output. The ratio of total energy output to total energy input must be equal to one, unless some output flows have been neglected.

## C. Accounting for (User-Side) Resource Quality: The Exergy Approach

Not all forms of energy are equivalent with respect to their ability to produce useful work. While heat is conserved, its ability to support a transformation process must decrease according to the second law of thermodynamics (increasing entropy). This is very often neglected when calculating efficiency based only on input and output heat flows (first-law efficiency) and leads to an avoidable waste of still

usable energy and to erroneous efficiency estimates. The same need for quality assessment applies to every kind of resource supplied or released in a process. The ability of resources to supply useful work or to support a further transformation process must be taken into account and is the basis for inside-the-process optimization procedures, recycling of still usable flows, and downstream allocation of usable resource flows to another process.

The ability of driving a transformation process and, as a special case, producing mechanical work, may be quantified by means of the exergy concept. According to Szargut and co-workers in 1988, exergy is "the amount of work obtainable when some matter is brought to a state of thermodynamic equilibrium with the common components of the natural surroundings by means of reversible processes, involving interaction only with the above-mentioned components of nature." Except for cases such as electricity production by means of nuclear reactors, the gravitational potential of water, or the geothermal potential of deep heat, chemical exergy is the only significant free energy source in most processes. In the procedure of Szargut and co-workers, chemical exergy is calculated as the Gibbs free energy relative to average physical and chemical parameters of the environment. Szargut and co-workers calculated the chemical exergy of a resource, $b_{ch,res}$, relative to common components of nature, by means of the following procedure:

1. The most stable end products of reactions involving the resource under consideration are chosen as reference chemical species "rs" in the environment (air, ocean, or earth crust). These reference species are not in thermodynamic equilibrium, as the biosphere is not a thermodynamic equilibrium system. Reference species are assigned a specific free energy (exergy) $b_{rs}$ due to their molar fraction $z_{rs}$ in the environment by means of the formula $b_{rs} = -RT \ln z_{rs}$, where the reference species are assumed to form an ideal mixture or solution.

2. In accord with the classical procedure of Gibbs, a chemical reaction linking a resource "res" to its reference species is chosen. The chemical exergy difference $\Delta b_{ch}$ between the two states is calculated as $\Delta b_{ch} = \Delta b_f + \sum b_{ch,i}$, where $\Delta b_f$ is the standard chemical exergy of formation of the resource from elements, and $b_{ch,i}$ is the standard chemical exergy of the $i$th element relative to its reference species. Tabulated values of free energy can be used for this purpose.

3. The total chemical exergy of the resource is finally calculated as $b_{ch,res} = \Delta b_f + \sum b_{ch,i} + \sum b_{rs}$. In so doing, tables of exergies per unit mass or per mole (specific exergy) can be constructed for each resource.

By definition, the exergy (ability of doing reversible work) is not conserved in a process: the total exergy of inputs equals the total exergy of outputs (including waste products) plus the exergy losses due to irreversibility. Quantifying the exergy losses due to irreversibility (which depends on deviations from an ideal, reversible case) for a process offers a way to calculate how much of the resource and economic cost of a product can be ascribed to the irreversibility affecting the specific technological device that is used (as in exergoeconomics as described by M. A. Lozano and A. Valero in 1993 and A. Lazzaretto and colleagues in 1998), as well as to figure out possible process improvements and optimization procedures aimed at decreasing exergy losses in the form of waste materials and heat. Exergy losses due to irreversibilities in a process are very often referred to as "destruction of exergy."

Szargut, in 1998, suggested that the "cumulative exergy consumption" required to make a given product could be a measure of the ecological cost of the product itself. Finally, R. U. Ayres and A. Masini,l in 1998, have suggested that "there is a class of environmental impacts arising simply from uncontrolled chemical reactions in the environment initiated by the presence of 'reactive' chemicals. The formation of photochemical smog in the atmosphere is one example . . . ." Exergy may therefore be a useful tool for potential harm evaluation, but requires further investigations for quantitative analyses of this type.

After exergy accounting has been performed, we remain with exergy efficiency indicators ($b_{out}/b_{in}$, $b_{waste}/b_{in}$, $b_{in}$/unit of product) that can be used for optimization and cost-allocation purposes.

# VI. PUTTING PROBLEMS IN A LARGER PERSPECTIVE

Optimizing the performance of a given process requires that many different aspects be taken into account. Some of them, mostly of a technical nature, relate to the local scale at which the process occurs. Other technological, economic, and environmental aspects are likely to affect the dynamics of the larger space and time scales in which the process in embedded. These scale effects require that a careful evaluation be performed of the relation between the process and its "surroundings" so that hidden consequences and possible sources of inefficiencies are clearly identified.

## A. Factors of Scale and System Boundaries

It is apparent that each evaluation (mass, energy, and exergy-flow accounting) can be performed at different space scales. In Fig. 1 we assume that we are evaluating a power plant, with natural gas used to produce electricity. The immediate local scale includes only plant components and buildings. An input mass balance at this
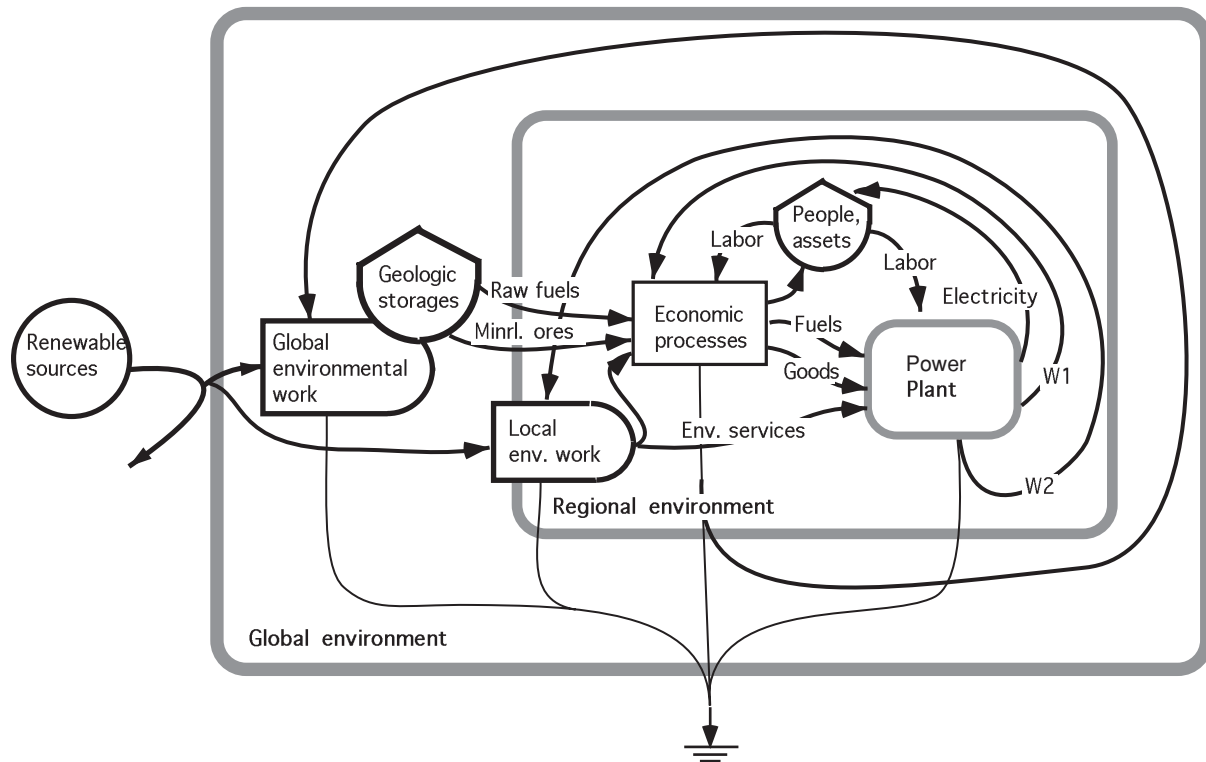
**FIGURE 1** Schematic showing the hierarchy of space and time scales that should be considered when evaluating a production process. W1 is the flow of usable waste materials that can be recycled within the economic process; W2 is the flow of waste materials released to the environment in a form that is not yet completely degraded. Flows to the heat sink represent waste heat and materials that cannot be further degraded. Systems symbols after Odum, 1996.

scale (over the actual plant area for a time of one year) should only include the actual masses of components and building materials discounted over the plant lifetime plus the fuel supplied. The same fraction of building materials from decommissioning plus airborne emissions should be accounted for as mass output. Similarly, an energy balance should only account for input natural gas and output electricity and waste heat. As the scale is expanded to the regional level [whatever the regional area is, it should include the production process for machinery components (boiler, turbine, insulating materials, etc.) and plant building materials like concrete and steel], new mass and energy inputs must be taken into account. For instance, the huge amount of electricity needed for component manufacture from metals contributes heavily to the energy cost, while the amount of fossil fuels needed to produce the electricity remains assigned to the material flow at the input side. If the scale is further expanded, the mass of raw minerals that must be excavated to extract the pure metals for making plant components may contribute heavily to all of the calculated indicators [consider the above examples of nickel ore or of iron, which is 25–30% raw iron ore]. At this larger scale, raw oil used up in the extraction and refining of minerals and the oil

itself must also be accounted for. It is a common and somehow standardized practice to account for embodied large-scale input flows while evaluating the process at the local scale. In so doing the evaluation is more easily done and results are summarized shortly. We do not suggest this procedure, as large-scale processes may hide local-scale performance and optimization needs. We prefer to perform the evaluation at three different scales (local, regional, and global), each one characterized by well-specified processes (respectively, resource final use, components manufacturing and transport, resource extraction and refining), so that inefficiencies at each scale may be easily identified and dealt with. The larger the space scale, the larger the cost in terms of material and energy flows (i.e., the lower the conversion efficiency). If a process evaluation is performed at a small scale, its performance may not be well understood and may be overestimated. Depending upon the goal of the investigation, a small-scale analysis may be sufficient to shed light on the process performance for technical optimization purposes, while a large-scale overview is needed to investigate how the process interacts with other upstream and downstream processes as well as the biosphere as a whole. Defining the system boundary

and making it clear at what time scale an assessment is performed is therefore of paramount importance, even if the scale of the assessment is sometimes implicit in the context of the investigation. It is very important to be aware that a "true" value of net return or other performance indicators does not exist. Each value of a given indicator is "true" at the scale at which it is calculated. When the same value is used at a different scale, it does not necessarily become false. It is, however, out of the right context and inapplicable.

Finally, how is the time scale accounted for in a process evaluation? The simplest case is when we have inputs, whose lifetime exceeds the time window of the analysis. For instance, assets in a farm have a lifetime of about 20–25 years, agricultural machinery never lasts more than about 10 years, while fertilizers and fuels are used up soon after they are supplied to the process. It is therefore easy to transform long-life inputs into annual flows by dividing by their lifetime (in years).

Another and perhaps more important time scale is hidden in the resources used, that is, the time it takes to make a given resource. When we use oil, it is not the extraction time or the combustion time that makes the difference but rather the very long time it took for crude oil formation in the deep storages of the earth crust. It is this time that makes oil a nonrenewable resource. Similar arguments apply to other resources, so that we may distinguish between renewable, slowly renewable, and nonrenewable resources. Turnover time (i.e., the time it takes to replace a resource) is often a good measure of resource renewability. An effort to go beyond the concept of turnover time in resource evaluations is the introduction of emergy-accounting procedures. This approach is dealt with in the following section. It is an attempt to place a value on resources by calculating the environmental work and time it took to supply a given resource to a user process. *Emergy* has been suggested as an aggregated measure of environmental support from present (space scale) and past (time scale) activities of the ecosystem. This important concept is discussed in succeeding sections.

## B. Life-Cycle Assessment

As already pointed out, careful material and energy assessments at different space-time scales are prerequisites for a good description of process dynamics. Assessing material and energy input and output flows is the basis for evaluations of an environmental disturbance generated by withdrawal of resources and release of waste, as well as energy costs associated with resource processing and disposal. Life-cycle assessment (LCA) methodologies are used to assess environmental impacts of a product (or a service) from "cradle to grave" or better "from cradle to cradle," including recycle and reclamation of degraded environmental resources. More than a specific methodology, LCA is a cooperative effort performed by many investigators throughout the world (many working in the industrial sectors) to follow the fate of resources from initial extraction and processing of raw materials to final disposal. Different approaches have been suggested, thousands of papers published, and meetings organized. This effort is converging toward developing standard procedures or at least a common framework that will make procedures and results comparable. SETAC (the International Society for Environmental Toxicology and Chemistry) has created a steering committee to "identify, resolve, and communicate issues regarding LCAs (and to) facilitate, coordinate, and provide guidance for the development and implementation of LCAs." In a series of workshops between 1990 and 1993, SETAC developed a "code of practice" to be adopted as a commonly agreed procedure for reliable LCAs. According to this code, LCA investigations should be divided into four phases: Goal Definition and Scope, Inventory, Impact Assessment, and Improvement Assessment. The SETAC standardization trial has been followed by a robust effort of the International Standard Office in 1997 and 1998 to develop a very detailed investigation procedure for environmental management, based on LCA, and for a comparable quality assessment. The International Standard clearly "recognizes that LCA is still at an early stage of development. Some phases of the LCA technique, such as impact assessment, are still in relative infancy. Considerable work remains to be done and practical experience gained in order to further develop the level of LCA practice." The ISO documents suggest clear and standard procedures for descriptions of data categories, definitions of goals and scope, statements of functions and functional units, assessments of system boundaries, criteria for inclusions of inputs and outputs, data quality requirements, data collections, calculation procedures, validations of data, allocations of flows and releases, reuse and recycling, and reporting of results. Nowhere in the ISO documents is a preference given to a particular approach. LCA is suggested as a standardized (and still to be improved) framework, where most of the methodologies already developed for technical and environmental investigations may be included and usefully contribute. This last statement is of fundamental importance, as it gives LCA a larger role than was recognized in earlier steps toward development. LCA may become a preferred pattern for replacing thousands of evaluations of case studies that are not comparable, since these may have had different boundaries (or did not describe a boundary at all), different approaches, different foci, different calculation procedures, different functional units, different measure units. It is no wonder that environmental and energy assessments have not yet

been accepted as a required practice in resource policy making. Standardization of procedures is a key feature in making significant progress toward policy and economic relevance of environmental assessments.

## C. Accounting for (Donor-Side) Resource Quality: The Emergy Approach

It takes roughly[5] 1.25 g of crude oil to make 1 g of refined oil in a typical refinery process; it takes 3 J of refined oil to supply 1 J of electricity in a typical power plant. Therefore, it takes about 3.8 J of crude oil to supply 1 J of electricity for Carnot-efficiency-limited conversions in current practice. Some additional oil is also invested in machinery and other goods that are components of the refinery and power plant. Oil is also needed to transport components. The oil energy that has been invested in the overall production process is no longer available. It has been used up and is not contained in the final product. The actual energy content or availability (measured as combustion enthalpy, H.H.V., L.H.V., exergy, etc.) of the product differs from the total input oil energy (or exergy) because of losses in many processes leading to the final product. This required total energy in the form of crude oil equivalent is sometimes referred to as "embodied energy" by energy analysts, although it may also be simply referred to as the sum of usable and lost energy in product manufacture. Embodied exergy has been suggested as another useful term.

### 1. From Embodied Exergy to Emergy

While a conservation law (first law) always applies to the total energy content of a resource, such a law is obviously not applicable to the embodied energy (exergy) concept. As defined, the embodied energy (exergy) of a usable resource is a measure of the sum of the product energy (exergy) and the energy (exergy) previously used up to make the product available. The same product may be produced via different production pathways and with different energy (exergy) expenditures depending on the technology used and other factors. The embodied energy (exergy) assigned to an item does not depend only on the initial and final states, but also on boundary conditions that may vary from case to case and are strongly affected by the level of process irreversibility. If we were able to use reversible processes from the raw resources to the final products, we would, of course, obtain the optimal value of the embodied energy (exergy). Unfortunately, this is not the case and we therefore remain with sets of values that may be averaged or given as a range of values. Embodied energy (exergy) algebra is therefore more a "memory algebra" than a

"conservation algebra." It is a process-dependent estimate which may vary significantly for a given product.

However, resources are always available in a limited supply. When a chain of processes is investigated to assess the actual resource investment that is required, we may need to quantify the relation of resources with the environmental work which supports them. In other words, a given resource might require a larger environmental work than others. The use of such a resource means a larger appropriation of environmental support and services, and may be assumed as a measure of sustainability and/or pressure on the environment by the system. As a further development of these ideas, Odum has introduced the concept of *form emergy*[6], that is, "the total amount of exergy *of one kind* that is directly or indirectly required to make a given product or to support a given flow." We may therefore have an oil emergy, a coal emergy, etc., according to the specific goal and scale of the process. In some way, this concept of embodiment supports the idea that something has a value according to what was invested into making it. This way of accounting for required inputs over a hierarchy of levels might be called a "donor system of value," while exergy analysis and economic evaluation are "receiver systems of value" (i.e., something has a value according to its usefulness to the end user).

The focus on crude oil and other fossil fuels (solar energy stored by ecosystem work) as prime energy sources is very recent in human history. Because we need and pay for these energy sources, we worry about possible shortages. Before fossil fuels were used, the focus was on free, direct solar radiation as well as indirect solar energy in the form of rain, wind, and other environmental sources. Everybody was well aware that the availability of direct and indirect solar energy was the main driving force of natural ecosystems and human societies as well. H. T. Odum therefore switched his focus from the interface human societies–fossil storages to the interface human societies–environment, identifying the free environmental work as the source of each kind of resource supporting human activities.

Expanding the scale of the investigation requires the concept of oil emergy to be expanded to *solar emergy*, a measure of the total environmental support to all kinds of processes in the biosphere by means of a new unit, the *solar emergy joule*. This new point of view also accounts for energy concentration processes through a hierarchy of processes that are not under human control and may therefore follow a different optimization pattern than humans do. Human societies (deterministic and product oriented in their strategies) usually maximize efficiency,

---

[5]All numbers are approximate.

[6]The name *emergy* and a valuable effort at developing the emergy nomenclature were provided by D. Scienceman in 1987.

short time-scale return on investment, employment, profit, and one-product output. At the other end, natural processes are stochastic and system-oriented and are assumed to try to maximize the utility of the total flow of resources processed through optimization of efficiencies and feedback reinforcement.

## 2. Solar Emergy—Concepts and Definitions

The *solar emergy* is defined as the sum of all inputs of solar exergy (or equivalent amounts of geothermal or gravitational exergy) directly or indirectly required in a process. Usually these inputs are produced by another process or sequence of processes, by which solar emergy is concentrated and upgraded. In mathematical terms, the solar emergy Em assigned to a product is defined as

$$\text{Em} = \iiint \varepsilon(\lambda, \tau, \sigma)\, d\lambda\, d\tau\, d\sigma,$$

where $\varepsilon(\lambda, \tau, \sigma)$ is the total exergy density of any source at the soil level, depending on the time $\tau$ and land area $\sigma$ over the entire space and time scales involved in the investigated process. In principle, it also depends on the wavelength $\lambda$ of the incoming solar radiation. Sources that are not from solar source (like deep heat and gravitational potential) are expressed as solar equivalent energy by means of suitable transformation coefficients.

The amount of input emergy dissipated (availability used up) per unit output exergy is called *solar transformity*. For investigators who are used to thinking in terms of entropy, the transformity may also be considered as an indicator of entropy production, from which the total entropy produced over the whole process can be in principle calculated. The solar transformity (emergy per unit product) is a measure of the conversion of solar exergy through a hierarchy of processes or levels; it may therefore be considered a "quality" factor which functions as a measure of the intensity of biosphere support for the product under study. The total solar emergy of a product may be expressed as

solar emergy (seJ) = exergy of the product (J)
        × conversion factor (solar transformity in seJ/J).

Solar emergy is usually measured in solar emergy joules (seJ), while the unit for solar transformity is solar emergy joules per joule of product (seJ/J). Sometimes emergy per unit mass of product or emergy per unit of currency is also used (seJ/g, seJ/$, etc.). In so doing, all kinds of flows to a system are expressed in the same unit (seJ of solar emergy) and have a built-in quality factor to account for the conversion of input flows through the biosphere hierarchy. It is useful to recall that emergy is not energy and therefore it is not conserved in the way that energy is. The emergy of a given flow or product is, by definition, a measure of the work of self-organization of the planet in making it. Nature supplies resources by cycling and concentrating matter through interacting and converging patterns. Since there is energy in everything including information and since there are energy transformations in all processes on earth and possibly in the universe as well, all these processes can be regarded as part of an energy hierarchy. All these energy transformations are observed to be connected to others so as to form a network of energy transformations, where many lower components support a few units at higher hierarchical position. A part of this work has been performed in the past (viz., production of fossil fuels) over millions of years. A part is the present work of self-organization of the geobiosphere. The greater the flow that is required, the greater the present and past environmental costs that are exploited to support a given process or product.

According to the process efficiencies along a given pathway, more or less emergy might have been required to reach the same result. The second law of thermodynamics dictates that there is a lower limit below which the product cannot be made. There is also some upper limit above which the process would not be feasible in practice although, in principle, one could invest an infinite amount of fuel in a process and thus have an infinitely high transformity. Average transformities are used whenever the exact origin of a resource or commodity is not known or when it is not calculated separately. It follows that: (1) Transformities are not constants nor are they the same for the same product everywhere, since many different pathways may be chosen to reach the same end state. (2) Emergy is not a point function in the way energy and other thermodynamic state functions are. Its value depends upon space and time convergence, as more emergy is used up (and assigned to the product) over a pathway requiring a higher level of processing. The emergy value has a "memory" of resources invested over all processes leading to a product. While the exergy content of a given resource indicates something that is still available, the emergy assigned to a given item means something that has already been used up and depends upon the processes followed in making the product. This last characteristic is shared with the embodied exergy and cumulative exergy consumption concepts, as it comes out of the different irreversibilities that characterize each pathway. (3) Optimum performance for specified external constraints may be exhibited by systems that have undergone natural selection during a long "trial and error" period and that have therefore self-organized their feedback for maximum power output. Their performance may result in optimum (not necessarily minimum) transformity.

Transformities are a very central concept in emergy accounting and are calculated by means of preliminary studies for the production process of a given good or flow. Values of transformities are available in the scientific literature on emergy. When a large set of transformities is available, other natural and economic processes can be evaluated by calculating input flows, throughput flows, storages within the system, and final products in emergy units. After emergy flows and storages in a process or system have been evaluated, it is possible to calculate a set of indices and ratios that may be suitable for policymaking.

Emergy as defined here provides emergy indicators that expand the evaluation process to the larger space and time scales of the biosphere. While the emergy approach is unlikely to be of practical use in making decisions about the price of food at the grocery store or the way a process should be improved to maximize exergy efficiency at the local scale, its ability of linking local processes to the global dynamics of the biosphere may provide a valuable tool for adapting human driven processes to the oscillations and rates of natural processes. As pointed out in Section V, this may be a useful step toward developing sustainable patterns of human economies.

## VII. CASE STUDY: INDUSTRIAL PRODUCTION OF SOLID OXIDE FUEL CELLS AND HYDROGEN

Mass, energy, exergy and emergy accounting, when used together, provide a valuable tool for multicriteria life-cycle assessment. Data from Tables II to VII may help clarify how system performance may be investigated. The tables refer to electricity production by using solid oxide fuel cells (SOFCs). Fuel cells (FCs) run on hydrogen that can be produced within the cell (internal reforming) if natural gas or other fuels are supplied. Hydrogen can also be supplied directly (from external reforming or water electrolysis). We have investigated both FC production and operating phases and hydrogen production from natural gas via steam reforming and from water via electrolysis to calculate a set of performance parameters that could be used to support decision making about this innovative and developing technology. Figure 2 shows diagrams of the two investigated pathways for hydrogen production.

The questions to be answered relate to energy efficiency, irreversibility of process steps, material demand as a measure of environmental disturbance, and ecological footprint as a measure of environmental support. We need a procedure that differentiates processes to support planning and decision making.

### A. Solid Oxide Fuel Cells

A fuel cell is an electrochemical device for converting the chemical energy of a fuel directly into electricity, without the efficiency losses of conventional combustion processes which are constrained by the Carnot limit. In principle, the conversion efficiency is 100% for FCs, while it is around 35–40% for the Carnot conversion at the temperatures used. In real cases, efficiencies of FCs are lowered by the requirements of other plant components, such as reforming and preheating phases as well as irreversibilities affecting the process.

A fuel and an oxidizing gas are involved. The fuel is hydrogen which is oxidized at the anode according to the reaction $H_2 \rightarrow 2H^+ + 2e^-$. Positive ions migrate from the anode to the cathode through the electrolyte conductor

**TABLE II   Mass Balance for a Solid Oxide Fuel Cell. Production and Use of a Single Cell**[a]

| Product | Amount produced | Oil requirement per unit of product | Material input per unit of product (including oil) | $CO_2$ emissions | $NO_2$ emissions |
|---|---|---|---|---|---|
| Anode | 0.55 g | 369 g/g | 3320 g/g | 1170 g/g | 4.8 g/g |
| Cathode | 0.55 g | 451 g/g | 3850 g/g | 1400 g/g | 5.7 g/g |
| Electrolyte | 1.60 g | 1087 g/g | 9550 g/g | 3460 g/g | 14.2 g/g |
| Interconnector | 25.00 g | 25 g/g | 192 g/g | 79.1 g/g | 8.1 g/g |
| Total cell | 27.70 g | 102 g/g | 16900 g/g | 322 g/g | 1.3 g/g |
| Electricity[b] | 192 kWh | 214 g/kWh | 4250 g/kWh | 578 g/kWh | 1.1 g/kWh |

[a] Ulgiati *et al.*, 2000.

[b] Including the fractions of energy and matter expenditures for the whole plant. The data account for all input flows, from extraction of raw materials to processing and final production of electricity (see the text, Section VII.A). A fraction of inputs for stack and plant manufacturing is also accounted for and assigned to the cell when evaluating the production of electricity.
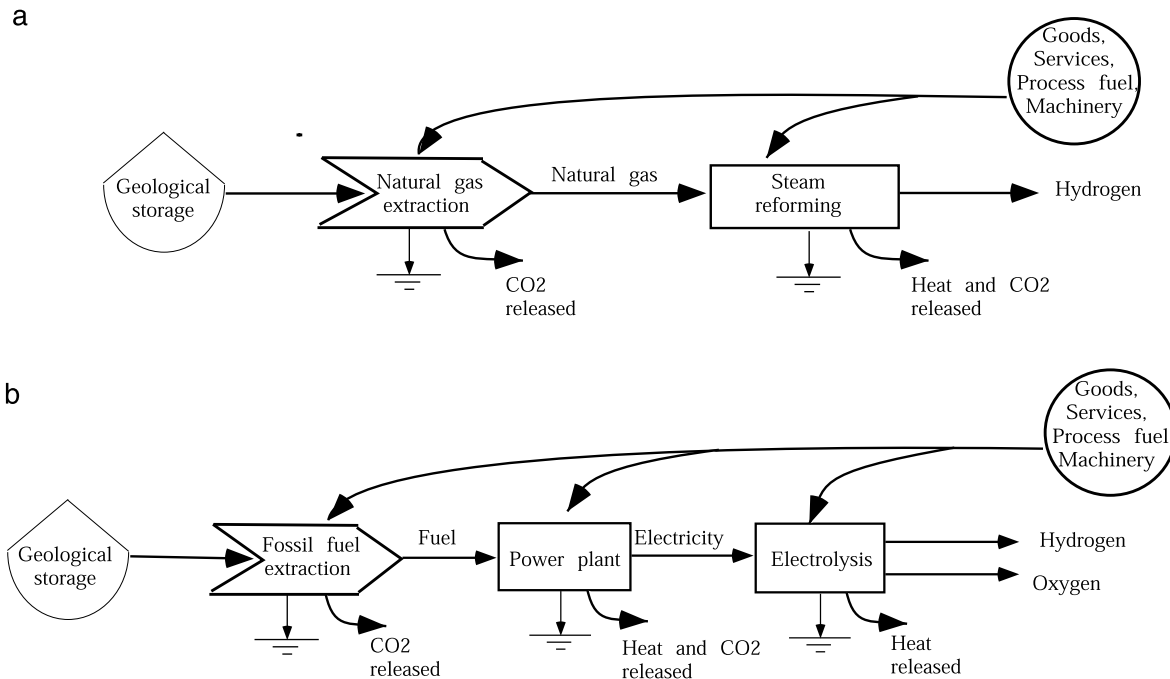
**FIGURE 2** Energy systems diagrams of hydrogen production processes: (a) steam reforming of natural gas; (b) water electrolysis.

of the cell, while electrons migrate through the outside circuit to the cathode where the oxidizing gas is reduced according to the process $O_2 + 4e^- \rightarrow 2O^{2-}$. The overall reaction is $4H^+ + 2O^{2-} \rightarrow 2H_2O$, so that emissions consist mainly of water vapor. Fuel cells also operate on fuels other than hydrogen ($CH_4$, methanol, gasoline, etc.). In this case, preliminary reforming (oxidation) is needed to convert the fuel primarily to hydrogen and carbon dioxide. The electric output of a fuel cell is in the form of current (DC) that is usually converted into alternating current (AC) before use. Our investigation deals with the whole process, from extraction of raw resources to production of electricity in the FC power plant. Reforming of natural gas and preheating of input air and fuel are also accounted for. Local-scale evaluation includes preheating of input gases, reforming of $CH_4$, and conversion of $H_2$ to electricity. Downstream combustion of $CH_4$, CO, and $H_2$ in exhaust gases is also included, as it provides heat to the preheating step and other uses. Regional-scale evaluation includes components manufacturing and transport. Finally, global biosphere-scale assessment includes additional expenses for extraction and refining of raw resources. Performance values calculated at the local scale are always higher than those calculated at larger scales. Performance indicators are also affected by the assumptions made for natural gas production costs. Available energy-cost estimates for natural gas are in the range 1.10–1.40 g of oil equivalent per gram of natural gas and depend on the location and the way natural gas is extracted, refined, and transported. These values affect the energy ratio and the material balance of our case study (the higher the energy cost the lower the performance). Our calculated data refer to an energy cost equal to 1.1 g of oil equivalent per gram of natural gas.

Table II shows an overview of material and energy performance of the cell components as well as of the cell as a whole, including the electricity produced. Electrolyte production appears to be the most oil-intensive component, as well as the most mass-intensive and $CO_2$ releasing component. The opposite is true for the interconnector. Globally, it takes 214 g of crude oil equivalent to produce one kWh of electricity. In addition to this direct and indirect fuel input, it takes about $4 \times 10^3$ g of other material inputs of different nature (minerals, oxygen, etc.) per kWh of electricity delivered. Energy data from Table III show a global efficiency of 40.25%, which is a relatively good performance compared with other fossil-fueled power plants investigated (Table IV). The energy efficiency of the single cell alone is the ratio of delivered electricity (192 kWh) to the energy of the hydrogen input from the reformer ($7.5 \times 10^8$ J), yielding a value of 92% which is very close to the theoretical maximum. This clearly indicates that if hydrogen could be produced in ways different than steam reforming (for instance, electrolysis from excess hydroelectricity or high-temperature catalysis), FCs may become a very effective way of producing electricity. The transformity, that is, the environmental support for the

**TABLE III   Energy and Emergy Accounting for a Solid Oxide Fuel Cell. Production and Use Refer to a Single Cell[a]**

| Product | Amount of item produced | Energy requirement per unit of product (MJ/unit) | Emergy assigned to a unit of product, with services ($\times 10^{12}$) | Emergy assigned to a unit of product, without services ($\times 10^{12}$) |
|---|---|---|---|---|
| Anode | 0.55 g | 15.45 | 1.37 seJ/g | 1.08 seJ/g |
| Cathode | 0.55 g | 18.91 | 1.20 seJ/g | 1.03 seJ/g |
| Electrolyte | 1.60 g | 45.62 | 3.06 seJ/g | 2.77 seJ/g |
| Interconnector | 25.00 g | 1.04 | 0.23 seJ/g | 0.06 seJ/g |
| Total cell | 27.70 g | 4.26 | 0.91 seJ/g | 0.26 seJ/g |
| Electricity[b] | 192 kWh | 8.94 | 0.57 seJ/kWh | 0.48 seJ/kWh |
|  | | Output/input electric efficiency | Transformity, with services ($\times 10^5$) | Transformity, without services ($\times 10^5$) |
|  | | $\eta = 40.25\%$ | 1.59 seJ/J | 1.32 seJ/J |

[a] Ulgiati *et al.*, 2000.

[b] Including the fractions of energy and matter expenditures for the whole plant. The data account for all input flows, from extraction of raw materials to processing and final production of electricity (see the text, Section VII.A). A fraction of inputs for stack and plant manufacturing is also accounted for and assigned to the cell when evaluating the production of electricity.

process, calculated without including labor and services, is $1.32 \times 10^5$ seJ/J, while inclusion of labor and services yields a higher value of $1.59 \times 10^5$ seJ/J (Table III). The difference ($0.27 \times 10^5$ seJ/J, i.e., 17% of the total value) is a measure of the fraction depending upon the labor and services cost in the country, i.e., a measure of the "economic part" of the process. This fraction might be lower in a different economic context with lower labor costs. The

**TABLE IV   Performance of Solid Oxide Fuel Cell System, Compared with Selected Electricity Power Plants[a]**

| Source of electricity | Output/input energy ratio | CO$_2$ emissions per unit of product (g/kWh) | Transformity without including services ($\times 10^4$ seJ/J) |
|---|---|---|---|
| Geothermal energy | 20.76 | 655 | 14.0 |
| Hydroelectric power | 14.02 | 19 | 6.10 |
| Wind | 7.67 | 35 | 5.89 |
| MCFC[b,c] | 0.43 | 561 | 14.1 |
| SOFC[c] | 0.40 | 578 | 13.2 |
| Natural Gas | 0.35 | 766 | 14.5 |
| Oil | 0.30 | 902 | 14.9 |
| Coal | 0.25 | 1084 | 13.0 |

[a] Ulgiati *et al.*, 2000.

[b] Molten Carbonate Fuel Cells.

[c] Including reforming of natural gas. In principle, fuel cells are expected to show the highest overall efficiencies. These are lowered in real plants by the up-stream reforming process, due to losses in the conversion to H$_2$ and the need for preheating of input gases.

emergy input from natural gas accounts for about 76% of the total, while the emergy of technological inputs (the stack of fuel cells and other devices plus process energy) only accounts for about 7%. The technological fraction might change over time, depending on technological improvements. However, the contribution of the fuel (76%) is not likely to change. This means that the value of the calculated transformity can be assumed to be very stable and only a small percentage may change depending on variations in the economy and technology. The small fraction of "technological emergy" might appear negligible when compared with the high input of emergy in the fuel. However, it is not negligible at the regional scale where cell components are manufactured and cells assembled. Improvements at this level would affect the performance and balance of the cell-manufacturing plant, and therefore translate into an advantage to the producer and the regional environment. The results derived from the data in Tables II to IV are the following: (1) Additional effort is needed to decrease the energy and material intensity of electrolyte production. (2) SOFC emissions of CO$_2$ are lower than emissions of other fossil-based power plants. (3) Indirect resource flows supporting labor and services do not appear to be important (emergy) inputs to this technology. (4) At the larger biosphere scale, FC operation affects overall performance more than construction. At the regional scale, construction inputs are important for a given design. At this level, an exergy evaluation may be needed to obtain optimized production patterns. (5) The electric energy efficiency of SOFCs is greater than the efficiencies of other fossil-fuel power plants investigated.

**TABLE V   Performance of a Single SOFC Evaluated at Different Scales, from Local to Global**[a,b]

| Index | Local scale | Regional scale | Global scale |
|---|---|---|---|
| Material input per unit of product (g/kWh) | 3250 | 3390 | 4250 |
| $CO_2$ released per unit of product (g/kWh) | 379 | 409 | 578 |
| $NO_2$ released per unit of product (g/kWh) | 0.11 | 0.23 | 1.07 |
| Energy efficiency, whole FC plant (%) | 54.67 | 51.54 | 40.25 |
| Transformity of electricity produced, without including services (seJ/J) | $1.32 \times 10^5$ | $1.32 \times 10^5$ | $1.32 \times 10^5$ |

[a] Ulgiati *et al.*, 2000.
[b] Including reforming of natural gas. In principle, SOFC cells are expected to show the highest overall efficiencies. These are lowered in real plants by the up-stream reforming process, due to losses in the conversion to $H_2$ and the need for preheating of input gases.

(6) The transformity of electricity from SOFCs (i.e., the environmental support per unit of product energy) is in the same order of magnitude than that for electricity production from conventional fossil-fuel power plants (Table IV) and higher than that from nonfossil sources. This is mostly due to energy inputs for upstream steps of reforming and preheating of input gases. These steps should be carefully studied to improve their efficiencies. However, it should be emphasized that fossil-fuel power plants are optimized by use, while FCs still are at a demonstration or precommercial stage, with large potential for improvement. Table V shows that performance indicators depend on the scale of the investigation as we emphasized in Section VI.A. For the sake of simplicity, we refer only to data at the global scale in Tables II to IV.

Preliminary results from a similar study on molten carbonate fuel cells (MCFCs) compared with SOFCs show a higher output/input energy ratio (42.7%), a higher transformity ($1.41 \times 10^5$ seJ/J) and lower $CO_2$ emission (561 g $CO_2$/kWh) for the whole plant system (Table IV).

## B. Industrial Production of Hydrogen Fuel

Tables VI and VII present some results from an LCA for hydrogen production by means of natural-gas steam-reforming and alkaline electrolysis. The two technological pathways to hydrogen have been investigated by using mass, energy, exergy, and emergy accounting.

Steam-reforming shows quite good energy and exergy conversion efficiencies, 46 and 42%, and a solar transformity of the expected order of magnitude ($1.15 \times 10^5$ seJ/J of hydrogen, labor and services included). Steam-

reforming supplies hydrogen to fuel cells. The previously discussed evaluation of SOFCs has yielded a transformity of about $1.59 \times 10^5$ seJ/J of electricity delivered. If we consider that an SOFC module operates in a two-step process (natural-gas reforming and electrochemical conversion in a cell) a lower transformity for the reforming process alone is perfectly consistent with SOFC data. Alkaline electrolysis shows lower performance due to the fact that the production of electricity is performed by conventional combustion of natural gas at 39% efficiency. This is the bottleneck of the process, whose global efficiency is quite good at 40% due to the better performance of the water-splitting process. The oil-equivalent input per unit of product as well as $CO_2$ emission (other chemicals released are not shown) do not show significant differences between the two processes. Instead, it may appear surprising that there is a huge difference in the total material demand per unit of output, which is four times higher in the steam-reforming process. The reason is the huge water consumption, for both natural gas oxidation and cooling. Water is a cheap resource that can still be obtained at low energy cost. There is no evident reason to design a process that saves water, because no one perceives this problem as crucial. However, if hydrogen is to become the favorite energy carrier for many uses, the shortages of water foreseen for the next decades may become limiting factors in some countries. Water supplies may become a research priority that is more important than the need for raising the already good energy and exergy efficiencies. Finally, alkaline electrolysis requires environmental support that is greater than for steam-reforming or electricity production from SOFCs, as is clearly shown by the higher transformity. This makes alkaline electrolysis a less decidable step toward the production of hydrogen as a fuel.

Our case studies provide both performance data and identification of bottlenecks for the investigated processes. Improvements can be made at different scales by looking at the values of the calculated indicators and designing technological changes where appropriate. It appears from our case studies that none of the methodologies used provides a comprehensive, multiscale assessment. Their combined use leads to synergistic results that are useful in decision making about these nonconventional processes.

## VIII.   A BASIS FOR VALUE

Investigating a system performance is by itself a very difficult task, due to the complexity of the problems that are always involved. When a simplified model is adopted, this is certainly a way to address part of the problem at

**TABLE VI   Energy Evaluation of Hydrogen Production from Natural-Gas Steam-Reforming and from Alkaline Electrolysis[a]**

| Production process | Amount of hydrogen produced (g/yr) | Output/input energy ratio | Exergy efficiency | Solar transformity ($\times 10^5$ seJ/J) |
|---|---|---|---|---|
| Natural-gas steam-reforming | $7.18 \times 10^8$ | 0.46 | 0.42 | 1.15 |
| Alkaline electrolysis[b] | $1.94 \times 10^{11}$ | 0.40 | 0.38 | 1.74 |

[a] Ulgiati *et al*., 2000.

[b] The electricity is produced with in the plant using natural-gas combustion.

the cost of leaving unsolved another part of it. Depending upon the goal of the investigation, this might be a very useful procedure.

Investigators very often run the risk of neglecting the complexity of the problem and taking their model for the reality. As a consequence, they assign a value to a process product according to the results of their simplified investigation. The outcome of this evaluation process is then used in other following evaluations and translated into economic and policy actions. In so doing, the complexity is lost: the reality does not fit the model and the planned policy fails or is inadequate.

Complexity and the consequent difficult task of reaching agreement on a decision can be shown by means of a simple example that recalls the paradoxes already cited in Section II. Assume that I commute every day due to my job and that a friend of mine does the same. The bus takes one and a half hours. We spend half an hour reading our newspapers, then we talk for a while, until we reach our destination. To save money and the environment, we decide to purchase only one newspaper and take turns reading it, one reading the paper, while the other one talks with other passengers or is involved in other reading. We save therefore about 20 newspapers per month, that translates into a savings of about U.S.$7 each. We feel that we are generating a lower pressure on the environment (fewer trees are cut for papermaking, less ink is used, less

garbage is to be disposed of). At first glance, it seems that our approach aimed at saving money and resources works well. We might also decide to purchase two different newspapers and read both of them while traveling, thereby gaining insight into different political opinions. The decision about these two alternatives has no roots in any biophysical analysis; it is just a different approach for a different return on investment. Thus we may get two copies of the same newspaper, two different newspapers, one shared newspaper plus additional goods purchased with the saved money.

At the local scale of our personal life, we find the following consequences: (1) we will have more money available and may decide to purchase a magazine each week; (2) we may invest this money into drinking more coffee at the cafeteria of the train station, which will possibly affect our health and therefore the health care bill at the personal and national levels (the process of coffee-making may have a larger environmental impact than paper-making); (3) as we read at different times, we talk less during travel and we therefore have a reduced exchange of ideas and personal feelings. When we expand the scale to take into account other possible advantages (or disadvantages) generated by our action, we find: (4) the decrease in newspaper sales may cause a decrease of jobs in the sectors of paper and newspaper making and the government will be required to invest money to support this activity; (5) the government

**TABLE VII   Mass-Flow Evaluations for Hydrogen Production from Natural-Gas Steam-Reforming and from Alkaline Electrolysis[a]**

| Production process | Amount of hydrogen produced (g/yr) | Oil requirement per unit of product (g/g $H_2$) | Material input per unit of product (including oil) (g/g $H_2$) | $CO_2$ emissions per unit of product (g/g $H_2$) |
|---|---|---|---|---|
| Natural-gas steam-reforming | $7.18 \times 10^8$ | 7.44 | 196[c] | 24.18 |
| Alkaline electrolysis[b] | $1.94 \times 10^{11}$ | 8.60 | 47[d] | 39.84 |

[a] Ulgiati *et al*., 2000.

[b] The electricity is produced within the plant using natural-gas combustion.

[c] Mostly due to the very high requirement of process and cooling water.

[d] Mostly due to the requirement for process water.

may get less revenue from taxing the newspaper business as well as from taxing a decreased number of workers; (6) a reduced environmental pressure is generated by the downsizing of the newspaper sector, but an increased pressure results from our investment of money in other sectors. We do not know which one of the two alternatives is worse unless a careful evaluation is performed. It is very hard to reach a balance among these different gains and losses. Maybe some of these interests are irreducible to a common optimization pattern. In any case, whatever we do affects the environment in a measure that is partially predictable by means of a careful biophysical analysis.

Let us now assume that we have developed a well-standardized multicriteria, multiscale procedure for the biophysical and environmental evaluation of a process or for the development of the whole economy. We retain many unsolved problems that call for new aspects and values affecting the process of decision making. Let us consider the case of the Maheshwar hydroelectric dam (Narmada Valley, State of Madhya Pradesh, Central India), as described by the Indian writer Arunhati Roy in her 1997 best-seller "The God of Small Things." One of the largest hydroelectric projects in the world will make the village of Sulgaon and 60 other villages disappear, while 40,000 people will have to move somewhere else. A fertile area, filled with cultural and social traditions and a good level of economic development based on agriculture (sugar cane, cotton, cash crops), will become an artificial lake to support the dam turbines. A political and legal controversy is now in progress. What is the problem here? There is no doubt that the people have the right to live in the area where they were born, where they have a home and a job, where they have traditions and social structures, etc. There is no doubt that forcing them to move will cause social disruptions and the collapse of that part of the economy that is supported in this area. On the other hand, there is no doubt that the additional electricity available to the economic system will support new activities, will create jobs and wealth, and will eventually give rise to new villages and social structures. It appears that the diverse projections are not reducible to an agreed upon strategy. No biophysical theories of value appear to be able to determine the process of decision making. The problem is not only conversion efficiency or environmental degradation, nor is it a purely economic cost-benefit evaluation.

There is no way that an absolute value can be put on an item or a strategy for a given approach because the concept of value is scale- and goal-dependent. Of course, values assigned to an item may fit the goals and scale of validity of each approach very well. There is no consensus among scientists on the whole set of approaches that is needed for a complete evaluation. This is not necessarily bad, as it leads to cultural diversity. As for natural systems, diversity is a fundamental prerequisite for stability and progress.

The commonly accepted economic belief is that capital and labor form the real basis of wealth. It does not assign any special value to resources and services that are supplied for free by nature and therefore are not accounted for by market mechanisms. This belief has been a source of huge environmental degradation and is being gradually abandoned by concerned environmental economists. As previously pointed out, most evaluation systems are based on utility of what is received from an energy transformation process. Thus, fossil fuels are evaluated on the basis of the heat that will be received when they are burned. Economic evaluation is very often based on the willingness to pay for perceived utility. An opposite view of value in the biosphere could be based on what is put into something rather than what is received. In other words, the more energy, time, and materials that are "invested" in something, the greater its value. As a consequence of this last statement, the source of resources and the way they are used should always be carefully assessed and form the basis of a new economic view that assigns an additional value, complementary to utility values, to something such as natural services that have no market by definition.

We started this paper observing that both physical and economic aspects are involved in the process of running a car. We have stressed many aspects involved in the trial for developing an acceptable unique theory of value that is able to account for all aspects of a technological or economic process. Should we keep trying, by again getting involved in this never-ending debate? In our opinion, more fruitful research would deal with the nonreducibility of different theories of value while searching for integration and synergistic patterns.

On the basis of this discussion, it can be concluded that: (1) any evaluation process that is not supported by a preliminary integrated biophysical assessment is not reliable or usable. Biophysical theories provide the scientific basis for choices by including productivity, climate, efficiency, pollution, and environmental degradation scenarios. (2) Biophysical assessments based on a single approach or numeraire may have research value but are unlikely to contribute to effective policy action. (3) After an integrated biophysical assessment has been performed (i.e., the source of natural value is recognized), other values (ethical, social, monetary, political, etc.) should be integrated. The process of decision making will have to take all of these into account, each fitting a different scale and with different goals.

## IX. CONCLUSIONS

A consensus is being developed on the urgent need for integrated and dynamic assessment tools. These are needed both at the larger scale of the economy and at the local

scale of technological processes. Among the latter, energy-conversion processes may be dealt with by using integrated approaches, as is shown for the case studies investigated in this article. The results show the feasibilities and advantages of each energy process and provide comprehensive assessments for governments, industrial corporations, and the public, thereby leading to economically and environmentally sound policies.

It may appear surprising that we have emphasized in this paper the need for biophysical evaluation procedures as a fundamental and useful support tool in decision making and, at the same time, we have not supported the possibility of policy making when relying only on biophysical theories of value. In our opinion, the two statements are not contradictory. As repeatedly emphasized in this article, the applicability of any evaluation depends on the space-time scale of the investigation and its utility depends on the (scale-related) goal of the study. When dealing with policy actions involving human societies and short time scales, values other than biophysical assessments are very likely to affect governmental choices and trends, even if there is no doubt that in the long run human societies, as well as any other self-organizing systems in the biosphere, will be constrained by the availability of energy and resources. Limits to growth have very often been claimed since the publication of the Meadows' report in 1972, which dealt with the consequences of finite resource bases. Integrated biophysical tools are needed to identify which constraints affect the development of our societies as a whole. They are also helpful in understanding how the uses of new technological tools or developments are affected by these constraints. Biophysical evaluation procedures may provide answers to local and short-time scale problems, while providing estimates of planetary and long-run consequences of our actions.

## SEE ALSO THE FOLLOWING ARTICLES

BIOENERGETICS • CARBON CYCLE • ENERGY EFFICIENCY COMPARISONS AMONG COUNTRIES • ENERGY RESOURCES AND RESERVES • FOSSIL FUEL POWER STATIONS • FUEL CELLS, APPLICATIONS IN STATIONARY POWER SYSTEMS • GREENHOUSE WARMING RESEARCH • HYDROELECTRIC POWER STATIONS • PETROLEUM GEOLOGY • THERMODYNAMICS • TRANSPORTATION APPLICATIONS FOR FUEL CELLS • WASTE-TO-ENERGY SYSTEMS

## BIBLIOGRAPHY

Atkins, P. W. (1994). "Physical Chemistry; 5th ed.," Oxford University Press, Oxford, U.K.

Ayres, R. U., and Masini, A. (1998). Waste exergy as a measure of potential harm. *In* "Advances in Energy Studies. Energy Flows in Ecology and Economy" (S. Ulgiati, M. T. Brown, M. Giampietro, R. A. Herendeen, and K. Mayumi, eds.), pp. 113–128. Musis, Rome, Italy.

BP Amoco (2000). Statistical review of world energy. June 2000.

Brown, M. T., and Herendeen, R. (1996). Embodied energy analysis and emergy analysis: A comparative view. *Ecol. Econ.* **19,** 219–235.

Brown, M. T., and Ulgiati, S. (1999). Emergy evaluation of the biosphere and natural capital. *Ambio* **28**(6), 486–493.

Boustead, I., and Hancock, G. F. (1979). "Handbook of Industrial Energy Analysis," Wiley, New York.

Campbell, C., and Laherrère, J. (1998). *The end of cheap oil. Sci. Am.* March 1998, 60–65.

Cleveland, C. J., Hall, C. A. S., and Kaufman, R. (1984). Energy and the U.S. economy: A biophysical perspective. *Science* **225,** 890–897.

Foran, B., and Crane, D. (1998). The OzECCO embodied energy model of Australia's physical economy. *In* "Advances in Energy Studies. Energy Flows in Ecology and Economy" (S. Ulgiati, M. T. Brown, M. Giampietro, R. A. Herendeen, and K. Mayumi, eds.), pp. 579–596. Musis, Rome, Italy.

Frankl, P., and Gamberale, M. (1998). The methodology of life-cycle assessment (LCA) and its application to the energy sector. *In* "Advances in Energy Studies. Energy Flows in Ecology and Economy" (S. Ulgiati, M. T. Brown, M. Giampietro, R. A. Herendeen, and K. Mayumi, eds.), pp. 241–256. Musis, Rome, Italy.

Giampietro, M., Mayumi, K., and Pastore, G. (1998). A dynamic model of socio-economic systems described as adaptive dissipative holarchies. *In* "Advances in Energy Studies. Energy Flows in Ecology and Economy" (S. Ulgiati, M. T. Brown, M. Giampietro, R. A. Herendeen, and K. Mayumi, eds.), pp. 167–190. Musis, Rome, Italy.

Hall, C. A. S., and Cleveland, C. J. (1981). Petroleum drilling and production in the United States: Yield per effort and net energy analysis. *Science* **211,** 576–579.

Herendeen, R. A. (1998). Embodied energy, embodied everything . . . now what? *In* "Advances in Energy Studies. Energy Flows in Ecology and Economy" (S. Ulgiati, M. T. Brown, M. Giampietro, R. A. Herendeen, and K. Mayumi, eds.), pp. 13–48. Musis, Rome, Italy.

Hinterberger, F., and Stiller, H. (1998). Energy and material flows. *In* "Advances in Energy Studies. Energy Flows in Ecology and Economy" (S. Ulgiati, M. T. Brown, M. Giampietro, R. A. Herendeen, and K. Mayumi, eds.), pp. 275–286. Musis, Rome, Italy.

Hubbert, M. K. (1956). Nuclear energy and the fossil fuels. *In* "Drilling and Production Practices," pp. 7–25. American Petroleum Institute, New York.

Hubbert, M. K. (1968). Energy resources. *In* "Resources and Man," pp. 157–242. National Academy of Sciences. W. H. Freeman, San Francisco.

ISO. (1997). Environmental management—Life-cycle assessment—Principles and framework (ISO 14040:1997). International Standard Organization, June 1997, pp. 16. Brussels.

ISO. (1998). Environmental management—Life-cycle assessment—Goal and scope definition and inventory analysis (ISO 14041:1998). International Standard Organization, October 1998, pp. 27. Brussels.

Jarach, M. (1985). Sui valori di equivalenza per l'analisi e il bilancio energetici in agricoltura. *Rivista di Ingegneria Agraria* **2,** 102–114.

Jørgensen, S. E. (1992). "Integration of ecosystem theories: A pattern," Kluwer, Dordrecht and Boston.

Kiker, C. F. (1992). Emergy and available energy analyses: Are these consistent means for delineating natural and artifactal capital? Paper presented at the Second Meeting of the International Society for Ecological Economics, 3–6 August 1992, Stockholm, Sweden.

Lazzaretto, A., Macor, A., Mirandola, A., and Stoppato, A. (1998). Potentialities and limits of exergoeconomic methods in the design, analysis and diagnosis of exergy conversion plants. *In* "Advances in Energy Studies. Energy Flows in Ecology and Economy" (S. Ulgiati,

M. T. Brown, M. Giampietro, R. A. Herendeen, and K. Mayumi, eds.), pp. 515–530. Musis, Rome, Italy.

Lozano, M. A., and Valero, A. (1993). Theory of the exergetic cost. *Energy* **18**(9), 939–960.

Martinez-Alier, J. (1987). "Ecological Economic," Basil Blackwell, New York.

Odum, H. T. (1983). Maximum power and efficiency: A rebuttal. *Ecol. Model*. **20,** 71–82.

Odum, H. T. (1988). Self organization, transformity and information. *Science* **242,** 1132–1139.

Odum, H. T. (1994). The emergy of natural capital. *In* "Investing in Natural Capital" (A. M. Jansson, M. Hammer, C. Folke, and R. Costanza, eds.), pp. 200–212. Island Press, Covelo, CA.

Odum, H. T. (1996). "Environmental Accounting. Emergy and Environmental Decision Making," Wiley, New York.

Odum, H. T., and Pinkerton, R. C. (1955). Time's speed regulator: The optimum efficiency for maximum power output in physical and biological systems. *Am. Sci*. **43,** 331–343.

O'Neill, R. V., DeAngelis, D. L., Waide, J. B., and Allen, T. F. H. (1986). "A Hierarchical Concept of Ecosystems," Princeton University Press, Princeton.

Penner, S. S. (2000). Policy issues in providing energy supplies for the 21st century and beyond. Manuscript presented at the Second International Workshop "Advances in Energy Studies. Exploring Supplies, Constraints, and Strategies," Porto Venere, Italy, May 23–27, 2000. Book of Proceedings, 2001, pp. 367–377.

Scienceman, D. (1987). Energy and Emergy. *In* "Environmental Economics" (G. Pillet and T. Murota, eds.), pp. 257–276. Roland Leimgruber, Geneva, Switzerland.

Smil, V. (1991). "General Energetics. Energy in the Biosphere and Civilization," Wiley, New York.

Szargut, J. (1998). Exergy analysis of thermal processes. Ecological cost. *In* "Advances in Energy Studies. Energy Flows in Ecology and Economy" (S. Ulgiati, M. T. Brown, M. Giampietro, R. A. Herendeen, and K. Mayumi, eds.), Musis, Rome, Italy.

Szargut, J., Morris, D. R., and Steward, F. R. (1988). "Exergy Analysis of Thermal, Chemical, and Metallurgical Processes," Hemisphere, London.

Ulgiati, S., and Brown, M. T. (1998). Monitoring patterns of sustainability in natural and man-made ecosystems. *Ecol. Model*. **108,** 23–36.

Ulgiati, S., Brown, M. T., Bastianoni, S., and Marchettini, N. (1995). Emergy-based indices and ratios to evaluate the sustainable use of resources. *Ecol. Eng*. **5,** 519–531.

Ulgiati, S., Brown, M. T., Giampietro, M., Herendeen, R. A., and Mayumi, K. (eds.) (1998). "Advances in Energy Studies. Energy Flows in Ecology and Economy," Musis, Rome, Italy.

Ulgiati, S., Bargigli, S., Raugei, M., and Tabacco, A. M. (2001). Life-cycle assessment and environmental impact of advanced energy systems. *In* "Advances in Energy Studies. Exploring Supplies, Constraints, and Strategies" (S. Ulgiati, M. T. Brown, M. Giampietro, R. A. Herendeen, and K. Mayumi, eds.), Palombo, Rome, Italy. pp. 29–42.
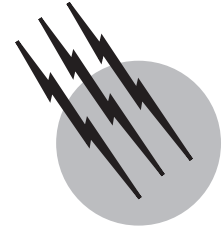
# Energy Resources and Reserves

## S. S. Penner

*University of California, San Diego*

## GLOSSARY

**Fossil fuels** Coals, petroleums, natural gases, oils from shales and tar sands, methane hydrates, and any other supplies from which hydrocarbons for energy applications may be extracted.

**Fuels for nuclear breeder reactors** These include U-238 and Th-232, which may be converted to fissile isotopes (e.g., U-233, U-235, Pu-239, and Pu-241) as the result of neutron capture.

**Fuels for (nuclear) fission reactors** The naturally occurring fissile isotope U-235, as well as Pu-239 and Pu-241 produced by neutron capture in U-238 and U-233 from Th-232.

**Fuels for (nuclear) fusion reactors** Deuterium-2 (which occurs in water as HDO or $D_2O$) and Li-6, which constitutes about 7.4% of naturally occurring lithium.

**Nonrenewable resources (nonrenewables)** Resources located on the planet with estimable times for exhaustion at allowable costs and use rates.

**Renewable resources (renewables)** Usually defined as extraterrestrial energy supplies such as solar resources, but some authors include energy supplies of such types and magnitudes that they will be available for the estimated duration of human habitation on the planet.

**Reserves** Energy supplies which are immediately usable at or very close to current prices.

**Resources** The totality of energy supplies of specified types that include reserves and may become usable in time at competitive prices with improved technologies.

**AN ASSESSMENT** of energy reserves and resources is properly made in the context of needs for an acceptable standard of living for all humans. This approach may be quantified by beginning with a discussion of current energy use in a developed country such as the United States, noting ascertainable benefits from high energy-use levels, estimating the possible impacts of a vigorously pursued,

stringent energy-conservation program, and then defining human needs for the estimated world population by 2050 on the assumption that population growth will level off after this time. This is an egalitarian approach based on documented benefits to humans of high levels of energy use while asserting the right of all people everywhere to enjoy the good life which technology has brought to the beneficiaries of large energy supplies at acceptable costs in a small number of developed countries. Of special interest are reserves and resources with a time to exhaustion that is commensurate with the very long time span which we optimistically foresee for human existence on our planet.

# I. INTRODUCTION

## A. Historical Perspective of Current Energy-Use Levels in the United States

Per capita energy consumption in the United States corresponds to about 12 kW of thermal power used continuously. This high level of continuous power use may be appreciated by examining two examples of rapid power consumption and the associated energy use.

Top power consumption for an elite athlete exceeds 2 kW (Smil, 1999, p. 90), whereas running a 100-m dash in 10 sec requires a power level of 1.3 kW for a typical world-class sprinter (Smil, 1999, p. XV, Table 5). Thus, the power expenditure of a world-class sprinter, if maintained *continuously*, would be about a factor of 10 smaller than that enjoyed by an average American.

Another striking example of how rapidly we consume energy to attain the good life we take for granted may be given by asking how many Frenchmen the disciplinarian Finance Minister Jean Baptiste Colbert of King Louis XIV (1643–1715) would have had to motivate in expending *all of their usable energy* for the King's benefit to allow Louis XIV to luxuriate at the energy-expenditure level we have reached in the United States on the average. A worker consuming 3000 cal/day and converting all of the food energy input to useful work would have a power output[1] of

$$(3000 \text{ cal/p-day}) \times (3.73 \times 10^{-7} \text{ hp-hr}/0.24 \text{ cal})$$

$$\times (1 \text{ day}/24 \text{ hr}) = 2 \times 10^{-4} \text{ hp/p},$$

where p stands for person and hp for horsepower. At this power-output level per person, Colbert's requirement becomes

---

[1]The numerator and denominator in the fractions contained within the middle set of parentheses both equal 1 J.

$$(12 \text{ kW}) \times (3.73 \times 10^{-7} \text{ hp-hr}/2.78$$

$$\times 10^{-7} \text{ kW-hr})/(2 \times 10^{-4} \text{ hp/p}) = 80{,}000 \text{ p}.$$

Thus, Colbert needed a veritable army of 80,000 Frenchmen who would convert all of their 3000-cal/day food consumption into useful work for the king with 100% efficiency in order to provide him with the energy-use opportunities enjoyed on the average by the people of the United States.

## B. Is a High Level of Energy Consumption Necessary or Desirable for Human Welfare?

There are many definitions for the good life or desirable standard of living. Demands for a good life include low infant mortality at birth, good health, long life expectancy, high per capita income levels to allow purchase of the many goods and services we take for granted in the developed part of the world, a clean and attractive environment, good educational opportunities, the needed time for and access to major cultural events, etc. There has been a strong historical correlation everywhere between per capita income and energy use on the one hand and between all of the desired amenities and income on the other, i.e. the level of energy consumption is a direct determinant of the standard of living defined in terms of the specified benefits. Two notable examples are reductions in infant mortality from about 150 per 1000 for undeveloped countries with low energy-use-rate levels to fewer than 15 per 1000 in the United States and a corresponding increase in life expectancy from about 40 years to nearly double this number. The gradual introduction of energy efficiency improvements may briefly disrupt historical trends, but has always been overwhelmed over the longer term by new discoveries providing new amenities and hence escalating energy consumption.

We shall ignore the historical record and assume that we can achieve significant energy conservation while maintaining our high standard of living. This desirable goal will be achieved through accelerated inventions and continuing emphasis on the desirability of minimizing energy use in order to reduce fossil-fuel consumption because we must reduce the rate of growth of atmospheric greenhouse-gas concentrations and also decelerate the rate of accumulation of long-lived radioactive isotopes produced in nuclear fission reactors. Whereas the more ambitious projections of energy-use reductions top out at about 30% for constant income levels, we shall use a reduction of 40% as the optimistic result of concentrated worldwide energy-conservation efforts. In summary, we assume that successful energy-conservation measures will allow us to attain the desired standard of living for 10 billion people by the year 2050 at 60% of the late-20th-century per capita use levels in the United States.

During 1999, total U.S. energy use was approximately $1 \times 10^{17}$ Btu/yr, or

$$(1 \times 10^{17} \text{ Btu/yr}) \times (10^3 \text{ J/Btu}) \times (1 \text{ EJ}/10^{18} \text{ J})$$
$$= 10^2 \text{ EJ/yr,}$$

where $1 \text{ EJ} = 1 \text{ exajoule} = 10^{18}$ J and Btu and J stand for the British thermal unit and the joule, respectively. By the year 2050, there will be around 10 billion people on the earth. If all of these people were to enjoy 60% of the U.S. per capita energy consumption used during 1999, then annual worldwide energy use would be

$$0.6 \times 10^{10} \text{ p} \times (10^2 \text{ EJ/yr})/3 \times 10^8 \text{ p} = 2 \times 10^3 \text{ EJ/yr.}$$

In summary, 10 billion people will hopefully enjoy the good life enjoyed by U.S. citizens at the end of the 20th century provided they each consume 60% of the U.S. per capita average during the year 1999. At that time, worldwide energy use will reach $2 \times 10^3$ EJ/y.

### C. Choice of a Common Unit to Compare the Magnitudes of Resources and Reserves

In order to gauge the relative importance of nonrenewable and renewable energy supplies over the long term, it is desirable to list all of their magnitudes in terms of a common unit. We choose here the cgs unit of energy or erg (1 erg = $10^{-25}$ EJ). There are large uncertainties in resource estimates, as well as variabilities in the heats of combustion of fossil fuels bearing the same name (such as coals, petroleum, tar oils, and shale oils). As an approximation, the following conversion factors are used here for energy deposits: one standard cubic foot of natural gas (1 SCF or 1 ft$^3$ of NG) = $(10^3 \text{ Btu/ft}^3) \times (10^{10} \text{ erg/Btu})$ = $10^{13}$ erg/ft$^3$; one barrel (bbl or B) of oil = $5.8 \times 10^6$ Btu = $5.8 \times 10^{16}$ erg; one metric ton of coal (1 mt of coal; $\approx 1.1$ short ton, 1 t, of coal) = $28 \times 10^6$ Btu = $2.8 \times 10^{17}$ erg; 1 mt of oil = $4.3 \times 10^7$ Btu $\approx 7.3$ bbl $\approx 4.3 \times 10^{17}$ erg.

For *nonrenewable resources,*[2] which are discussed in Sections II and III, the listed entries for the total available energies represent estimates of values that have been identified in some manner as probably available and, over the long term and after sufficient research and development has been carried out, actually usable as reserves provided resource recovery can be implemented at acceptable costs. The estimated size of the nonrenewable resource base may

---

[2]The distinction between *nonrenewable* and *renewable* resources is tenuous, with the latter generally referring to energy resources that are the result of extraterrestial inputs (e.g. solar energy in its many manifestations) but often including also the utilization of geophysical features of the earth such as its rotation about an axis or heat flow from the core to the surface. Some authors mean by renewables those resources which will be usable as long as the earth remains suitable for human habitation.

increase with time as improved exploration and assessments are performed. *Renewable resources* are discussed in Section IV.

## II. FOSSIL-FUEL RESOURCES

As used here and by most other authors, the terms *resources* and *reserves* have greatly different meanings. *Resources* are potential supplies if no attention is paid to commercial readiness at currently paid market prices and generally include also supplies for which recovery and utilization techniques remain to be developed. *Reserves*, on the other hand, are supplies that may be utilized with known technologies at or very near current market prices. In the normal course of events, resource recoveries precede reserve classification, and reclassification as reserves may require new technologies or at least improvements of existing, tested, or developing technologies.

Estimations of resources and reserves tend to be grossly oversimplified. For this reason, the use of a single significant figure is often appropriate. An entire paragraph of assumptions and constraints may be needed to justify a numerical estimate. To illustrate this point, we begin with the potentially largest of the fossil-fuel supplies, namely, shale oil. A classical worldwide assessment was published by Duncan and Swanson (1965). The supplies were labeled as falling into the following categories: known resources, possible extensions of known resources, undiscovered and unappraised resources, and total resources. The shales were further subdivided according to shale grade in terms of the number of gallons (ga) of oil which could be recovered from one short ton (t) of the raw material with subdivisions according to shale grades of 5–10, 10–25 and 25–100 ga/t. Nothing is said about shale grades below 5 ga/t, of which there are enormous deposits. Needless to say, the greater the shale grade, the easier and more cost-effective is the recovery. As an example, a total resource estimate of potentially usable oil shale may then be made by including all categories of deposits and shale grades down to 10–25 ga/t. The estimate derived from the data of Duncan and Swanson (1965) becomes $344,900 \times 10^9$ bbl of oil or, using $5.8 \times 10^6$ Btu/bbl,

$$3.45 \times 10^{14} \text{ bbl} \times (5.8 \times 10^6 \text{ Btu/bbl})$$
$$\times (10^7 \text{ erg}/10^{-3} \text{ Btu}) = 2 \times 10^{31} \text{ erg.}$$

If we decided to broaden the resource base by including shale grades down to 5–10 ga/t, the preceding resource estimate would be increased by

$$1.75 \times 10^{15} \text{ bbl} \times (5.8 \times 10^6 \text{ Btu/bbl})$$
$$\times (10^7 \text{ erg}/10^{-3} \text{ Btu}) = 10 \times 10^{31} \text{ erg.}$$

Problems similar to those for **shale oil** arise for **coal resources** (for which estimates customarily depend on the thickness of the coal seams and on the depth below the surface where the coal is recovered), as well as for **oils from tar sands**, **petroleum**, and **natural gas**.

Averitt (1969) published estimates of the total original coal resources of the world and of regions of the world on the basis of (a) resources determined by mapping and exploration and (b) probable additional resources in unmapped and unexplored areas for 12-foot-thick or thicker coal beds and (in nearly all cases) less than 4000 feet below the surface of the earth. The total estimate was $16{,}830 \times 10^9$ t $\times$ (0.909 mt/t) $\times$ ($2.8 \times 10^{17}$ erg/mt) $= 4 \times 10^{30}$ erg. Of this total, the North American contribution was about one-fourth or $1 \times 10^{30}$ erg. The coal resources are predominantly bituminous and subbituminous coals with perhaps 20% representing low-rank (i.e. low carbon content) **lignite** and **peat** and less than 1% high-rank (nearly 100% C content) **anthracite**. Since the heat of combustion we use for coal applies to bituminous coals and generally decreases with rank (or carbon content), we may be overestimating the resource base by as much as 30%. Worldwide resource estimates were discussed in 1972 (Cornell, 1972) with L. Ryman contributing values for the ultimately recoverable **resources of NG and oil**. For NG, the estimate was about $1.5 \times 10^{16}$ SCF $= 1.5 \times 10^{29}$ erg and for oil it was about $4 \times 10^{12}$ bbl $= 2.3 \times 10^{29}$ erg. All of these estimates do not refer to resources **regardless of cost**, but rather to **resources that are not unlikely to become reserves over time**. The resource base for natural gas may be greatly expanded in the future if deep-lying hydrates are included (see Section V.B).

The worldwide base for **oil from tar sands** has not been well determined. For the Athabasca tar sands in Alberta, Canada, where commercial recovery has been pursued for more than 25 years, a resource value of $625 \times 10^9$ bbl was estimated by Spragins (1967) and corresponds to $3.6 \times 10^{28}$ erg; a 1981 resource estimate for Athabasca was $8 \times 10^{28}$ erg.

We may summarize this qualitative overview by the statement that, over a long period of time, **shale oil** could contribute substantially more than $2 \times 10^{31}$ **erg**, **coal** more than $4 \times 10^{30}$ **erg**, **NG** and **oil** each more than $2 \times 10^{29}$ **erg**, and worldwide **oil from tar sands** probably considerably more than $2 \times 10^{29}$ erg. Thus, the combined fossil-fuel resource base certainly exceeds $2 \times 10^{31}$ erg by a substantial margin.

It is interesting to compare the fossil-fuel resource estimates with past fossil-fuel extraction. According to Smil (1999, p. 135), total global extraction of fossil fuels (in the form of coal, oil, and natural gas) from 1850 to 1990

was $3 \times 10^2$ EJ or ($3 \times 10^2 \times 10^{18}$ J) $\times$ ($10^7$ erg/J) $= 3 \times 10^{27}$ erg, which equals about 1.5% of the resource estimates for the least abundant fossil fuels. Total global electricity generation from 1900 to 1990 produced (Smil, 1999, p. 135) $1.3 \times 10^4$ terrawatt-hours (TWh$_e$) $= 1.3 \times 10^{16}$ Wh$_e$, which corresponds approximately to $1.3 \times 10^{13}$ kWh$_e \times 4 \times$ ($10^7$ erg/$2.78 \times 10^{-7}$ kWh$_e$), or a fossil-fuel input of about $1.9 \times 10^{27}$ erg, where the subscript e identifies electrical energy and we have assumed that four units of thermal energy were required per unit of electrical energy produced during the first century of electricity generation, i.e. the historical average efficiency for electricity generation has been taken to be 25%. Worldwide coal production from 1800 to 1990 (Smil, 1999, p. 137) was $4.8 \times 10^3$ Mt (1 Mt $= 10^6$ t) $= 4.8 \times 10^9$ t $\times$ ($2.78 \times 10^{16}$ erg/t) $= 1.3 \times 10^{27}$ erg. The conclusion to be reached from this historical overview is that more than 98% of estimated fossil-fuel resources of all types remained unused as of the end of the 20th century.

## III. RESOURCES FOR NUCLEAR FISSION, BREEDER, AND FUSION REACTORS

We shall estimate the resources for nuclear fission, breeder, and fusion reactors by using geological data. In breeder reactors, the "fertile" isotopes U-238 and Th-232 are converted to the "fissile" isotopes U-233, U-235, Pu-239, and Pu-241 as the result of neutron capture. Thorium is a very widely distributed element[3] and does not represent a limiting supply when used in breeder reactors with uranium. For this reason, the following discussion is restricted to uranium resources for fission and breeder reactors.

### A. Fuels for Nuclear Fission Reactors

Fission reactors consume the only naturally occurring fissile isotope, namely, U-235. Fission occurs as the result of absorption of slow (thermal) neutrons. Uranium ores contain the following isotope distribution: $6 \times 10^{-5}$ of U-234, $7.11 \times 10^{-3}$ of U-235, and 0.99283 of U-238. Chattanooga shale is a typical deposit; it was formed 33–29 million years ago and is widely distributed in Illinois, Indiana, Kentucky, Ohio, and Tennessee (Swanson, 1960). The Gassaway member of this shale deposit is about 16 feet thick.

It is appropriate to estimate the ultimate resources for fission reactors by following the procedure that is custom-

---

[3]A 1966 assessment (IAEA, 1966) placed U.S. resources at $2.26 \times 10^7$ t for costs of \$50–100 and at $2.72 \times 10^9$ t for costs of \$100–500.

arily applied to fusion reactors, i.e. to estimate the total resource on the planet without regard to costs or near-term availability. Seawater (Best and Driscoll, 1980, 1981) contains $3.3 \times 10^{-9}$ part by weight of uranium (U). Hence, the $1.5 \times 10^{18}$ m$^3$ of water on the earth contains $1.5 \times 10^{18}$ m$^3$ of water $\times$ ($10^3$ kg of water/m$^3$ of water) $\times$ ($3.3 \times 10^{-9}$ kg of U/kg of water) $\times$ ($7.11 \times 10^{-3}$ kg of U-235/kg of U) $\times$ ($10^7$ erg/$1.22 \times 10^{-14}$ kg of U-235) $= 4 \times 10^{31}$ erg, where we use the estimate that the fission-energy equivalent of $10^7$ erg is derived from $1.22 \times 10^{-14}$ kg of U-235.

The uranium content of the earth's crust has been estimated to be $6.5 \times 10^{13}$ mt (Koczy, 1954). For use in a fission reactor, the uranium content of the crust has an energy content of ($6.5 \times 10^{13}$ mt of U) $\times$ ($7.11 \times 10^{-3}$ mt of U-235/mt of U) $\times$ ($10^3$ kg of U-235/mt of U-235) $\times$ ($10^7$ erg/$1.22 \times 10^{-14}$ kg of U-235) $= 4 \times 10^{35}$ erg.

The *current availability* of uranium is generally specified in terms of availability at a specified price. A 1983 DOE estimate for the United States was 433,400 t of U at \$80/kgU, corresponding to the sum of the probable potential (251,500 t), possible potential (98,800 t), and speculative potential (83,100 t) resources; at \$260/kg U, the corresponding values were $725,700 + 323,800 + 250,700 = 1,300,200$ t of U. Uranium prices declined from nearly \$110/kg in 1980 to about \$40/kg by mid-1984 and remained at about this level during the next 4 years. Needless to say, the current availability in the United States and elsewhere of uranium resources even at the highest prices is smaller by many orders of magnitude than the ultimately available resources. A similar remark applies to resources outside of the United States. Thus, the total Western world resources (not including the centrally planned economies) were estimated in 1987 to exceed the U.S. resources by factors of about two to four. In 1981, the United States was the world's dominant uranium producer, with 33.5% of total Western production; by 1986, the U.S. share had declined to 14.3%, while Canada had become the dominant producer with 31.6% of total production. In 1986, South African (at 12.5%) and Australian (at 11.2%) production followed U.S. production closely.
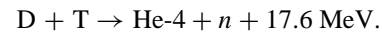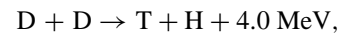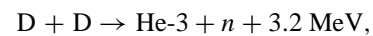
## B. Fuels for Nuclear Breeder Reactors

When burned in a breeder reactor, the energy content of naturally occurring uranium (Cohen, 1983) is approximately (1 MW-day/g of U) $\times$ (24 hr/day) $\times$ [$10^3$ kW/MW] $\times$ ($10^7$ erg/$2.78 \times 10^{-7}$ kWh) $= 8.6 \times 10^{17}$ erg/g of U. It follows that for use in breeder reactors, the uranium contents of the oceans have an energy value of $1.5 \times 10^{18}$ m$^3$ of water $\times$ (1 mt of water/m$^3$ of water) $\times$ ($3.3 \times 10^{-9}$ mt of U/mt of water) $\times$ ($8.6 \times 10^{17}$ erg/g of U) $\times$ ($10^6$ g of U/mt

of U) $= 4 \times 10^{33}$ erg. The uranium content of the earth's crust has been estimated as $6.5 \times 10^{13}$ mt. The energy equivalent for use in a breeder reactor of this resource is ($6.5 \times 10^{13}$ mt of U) $\times$ ($10^6$ g of U/mt of U) $\times$ ($8.6 \times 10^{17}$ erg/g of U) $= 5.6 \times 10^{37}$ erg; at the estimated year 2050 consumption rate of $2 \times 10^{28}$ erg/year (see Table I), this energy source would last for $2.8 \times 10^9$ years or a large fraction of the remaining life on the planet provided both world populations and per capita energy use become stabilized at the estimated level for the year 2050.

## C. Fuels for Nuclear Fusion Reactors

The sequence of nuclear fusion reactions of deuterium to produce helium with atomic weights of 3 and 4 is

$$D + D \rightarrow \text{He-3} + n + 3.2 \text{ MeV},$$

$$D + D \rightarrow T + H + 4.0 \text{ MeV},$$

$$D + T \rightarrow \text{He-4} + n + 17.6 \text{ MeV}.$$

Thus the fusion of five deuterium atoms (D), including fusion with the reaction intermediate tritium (T), releases 24.8 MeV of energy or 4.96 MeV per D atom, where 1 J $= 6.24 \times 10^{12}$ MeV $= 0.949 \times 10^{-3}$ Btu. Since 1 m$^3$ of liquid water contains 34.4 g of D, fusion of the deuterium contained in 1 m$^3$ of liquid water releases (4.96 MeV/D atom) $\times$ [($6.023 \times 10^{23}/2$) D atoms/g of D] $\times$ (34.4 g of D/m$^3$ of H$_2$O) $= (5.14 \times 10^{25}$ MeV/m$^3$) $\times$ ($0.949 \times 10^{-3}$ Btu/$6.24 \times 10^{12}$ MeV) $= 7.8 \times 10^9$ Btu/m$^3$ of H$_2$O. Since the total amount of water on the surface of the earth is about $1.5 \times 10^{18}$ m$^3$, the resource estimate for planetary fusion energy from deuterium becomes ($7.8 \times 10^9$ Btu/m$^3$) $\times$ ($1.5 \times 10^{18}$ m$^3$) $\times$ ($10^7$ erg/$0.949 \times 10^{-3}$ Btu) $= 1.2 \times 10^{38}$ erg.

Lithium occurs in water at 1 part in $10^7$ with 7.42% of Li being Li-6; mineable deposits of Li$_2$O are $5 \times 10^6 - 10 \times 10^6$ mt. At a sufficiently high neutron flux, Li-6 undergoes the fission reaction

$$\text{Li-6} + n \rightarrow \text{He} + T + 4.8 \text{ MeV},$$
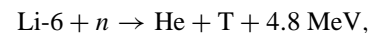
which may be followed by tritium fusion with deuterium to release 17.6 MeV, thus leading to the overall reaction

$$\text{Li-6} + D + n \rightarrow 2\text{He-4} + n + 22.4 \text{ MeV}.$$

This energy release must be reduced by 4.96 MeV/D atom since the fusion of deuterium and lithium decreases the supply of deuterium that can fuse with itself, i.e. the net incremental fusion energy becomes

**TABLE I  Worldwide Energy Resources without Consideration of Utilization Efficiencies or Technical Readiness**[a]

| "Nonrenewable" energy sources | Energy (EJ) ($1\ \text{EJ} = 10^{18}\ \text{J} = 10^{25}\ \text{erg}$) | Years of availability at $2 \times 10^3$ EJ/yr |
|---|---|---|
| Fusion energy from D in $1.5 \times 10^{18}$ m$^3$ of water | $1 \times 10^{13}$ | $5 \times 10^9$ |
| Uranium in the earth's crust used in breeder reactors | $6 \times 10^{12}$ | $3 \times 10^9$ |
| Uranium in the earth's crust used in fission reactors | $4 \times 10^{10}$ | $2 \times 10^7$ |
| Fusion of lithium contained in seawater | $3 \times 10^9$ | $1.5 \times 10^6$ |
| Uranium in seawater used in breeder reactors | $4 \times 10^8$ | $2 \times 10^5$ |
| Uranium in seawater used in fission reactors | $4 \times 10^6$ | $2 \times 10^3$ |
| Worldwide fossil-fuel resources (shale oils, coals, oils from tar sands, petroleum, natural gas, natural gas liquids) | $>2 \times 10^6$ | $>1 \times 10^3$ |
| U.S. fossil-fuel resources | $\sim 1 \times 10^5$ | $\sim 5 \times 10$ |
| Fusion of lithium in terrestrial Li$_2$O deposits | $8 \times 10^4$ | $4 \times 10$ |
| Hydrothermal energy stored to a depth of 10 km worldwide | $4 \times 10^4$ | $2 \times 10$ |
| U.S. uranium used in water-moderated fission reactors | $2 \times 10^4$ | $1 \times 10$ |
| Hydrothermal energy stored to a depth of 3 km worldwide | $8 \times 10^3$ | 4 |

| "Renewable" energy sources | Power (EJ/yr) | Multiple of need at $2 \times 10^3$ EJ/yr |
|---|---|---|
| Solar energy at the "outer boundary" of the atmosphere[b] | $5 \times 10^6$ | $2.5 \times 10^3$ |
| Worldwide wind energy | $2 \times 10^6$ | $1 \times 10^3$ |
| Ocean thermal energy conversion (OTEC)[c] | $3.78 \times 10^5$ | 190 |
| Geothermal energy (outward flow of heat from the earth's core) | $8 \times 10^2$ | $4 \times 10^{-1}$ |
| Tidal energy | $1 \times 10^2$ | $5 \times 10^{-2}$ |
| Hydroelectric energy worldwide | $9 \times 10^1$ | $4.5 \times 10^{-2}$ |
| Commercial, worldwide energy use in 1995 | $4 \times 10^2$ | — |
| Stabilized annual worldwide energy demand as of 2050 with conservation for $10 \times 10^9$ people and allowance of energy use to establish an acceptable standard of living for all[d] | $2 \times 10^3$ | — |

[a] See the text for sources of these values.

[b] This is the total solar input which may be recovered in part as biomass, wind energy, OTEC, direct water photolysis, photovoltaic energy, etc.

[c] A 1964 assessment is $2 \times 10^3$ EJ for the Gulf Stream alone. If 0.1% of the solar input power is used for tropical waters with accessible temperature differences $>22°$C (area $\approx 60 \times 10^6$ km$^2$), the continuous power production is $12 \times 10^6$ GW$_e$ year $= 378 \times 10^3$ EJ/year.

[d] This long-range, steady-state value refers to $10 \times 10^9$ people (now expected by the year 2050) with each person using 60% of the U.S. *per capita* consumption of 1998, i.e. each person enjoying the average 1998 U.S. standard of living if conservation efforts lead to a 40% energy-use reduction over the long term.

$$\frac{17.44\ \text{MeV}}{\text{Li-6 atom}} \times \frac{10^7\ \text{erg}}{6.24 \times 10^{12}\ \text{MeV}} \times \frac{1\ \text{g of Li}}{10^7\ \text{g of H}_2\text{O}}$$

$$\times \frac{7.42 \times 10^{-2}\ \text{g of Li-6}}{\text{g of Li}} \times \frac{10^6\ \text{g of H}_2\text{O}}{\text{m}^3\ \text{of water}}$$

$$\times \frac{6.023 \times 10^{23}\ \text{atoms of Li-6}}{6.94\ \text{g of Li-6}} \times 1.5$$

$$\times 10^{18}\ \text{m}^3 \text{of water} = 2.7 \times 10^{34}\ \text{erg}.$$

Li-7 may also breed with T, but with lower efficiency than Li-6.

Mineable deposits of Li$_2$O have been estimated (Penner, 1978) to be as much as $11 \times 10^7$ mt of which the lithium weight is about $4.64 \times 10^6$ mt, while that of Li-6 is equal to $3.44 \times 10^5$ mt. The incremental fusion energy (after

subtracting that of the deuterium used in the process) then becomes about

$$(3.44 \times 10^5\ \text{mt of Li-6}) \times \frac{17.44\ \text{MeV}}{\text{atom of Li-6}}$$

$$\times \frac{6.023 \times 10^{23}\ \text{atoms of Li-6}}{6.94\ \text{g of Li-6}} \times \frac{10^3\ \text{kg of Li-6}}{\text{mt of Li-6}}$$

$$\times \frac{10^3\ \text{g of Li-6}}{\text{kg of Li-6}} \times \frac{10^7\ \text{erg}}{6.24 \times 10^{12}\ \text{MeV}} = 8.3 \times 10^{29}\ \text{erg}.$$

## D. Concluding Observations on Resources for Breeder and Fusion Reactors

It is apparent that the resource bases for breeder and fusion reactors are so vast that the ascertainable amounts of both

of these materials, when used in currently contemplated systems, could last until the white dwarf that is our sun cools and becomes a red giant of vastly increased radius (increased by about a factor of 10), at which time the rivers and oceans of our planet will boil away, thereby causing a fiery demise for all of the earth's inhabitants.

While the resource estimates are not greatly different, the status of technological development (Zebroski *et al.,* 1998) is not comparable for fusion and breeder reactors. For fusion reactors, large-scale demonstrations have now been deferred by three or more decades because of the lack of adequate funding for R&D on the scale needed for rapid commercialization. Although large-scale breeder reactors represent a demonstrated technology, both technology improvement and commercialization have been suspended in nearly all countries where breeders were being developed because they are not currently cost-competitive with fossil-fuel electricity generators or with conventional fission-nuclear reactors.

The procedure used for resource estimations of the "nonrenewable" breeder and fusion fuels makes these estimates, in common with those of *renewable* resource estimates considered in the following section, unachievable limiting values in practice because costs for the recovery of resources must escalate as the source concentrations are reduced through recovery.

## IV. RENEWABLE RESOURCES

For *renewable resources*, the listed values represent limits associated with defined physical parameters such as the solar constant and area of the earth's surface (for solar energy), estimates of the total energy that could possibly be extracted (Palmén and Newton, 1969) from atmospheric winds (for wind energy), the known geothermal heat flow for the entire earth (for geothermal energy recovery from this source), known water flows in all of the rivers of the planet (for hydroelectric energy utilization), and total tidal friction as determined by the gravitational pulls of the moon and sun on the earth's oceans, which lead to differences in tidal range (for tidal energy). These limits are absolute values that cannot increase with time or be augmented in magnitude through improved technology. What can increase with time is the fraction of the total resource base that may be usable at acceptable costs. The renewable resource base may also increase through identification and utilization of physical effects that have not been included in the listed resource base (e.g., collection of solar energy in geostationary orbits, followed by conversion to microwave energy of the electric energy generated in photovoltaic cells used for collection, before transmission of the remaining collected energy to the sur-

face of the earth; or using the differences in salinities at the mouths of rivers between waters in rivers and estuaries where they are discharged into the oceans; or an as yet unidentified scheme to utilize the rotation of the earth in a manner not involving planetary winds; or cost-effective utilization of wave energy, etc.).

The magnitudes of entries for renewable resources will now be justified by utilizing known physical phenomena.

## A. Solar Energy Input to the Surface of the Earth

The *solar constant* $Q_0$ is defined as the radiation power density at the top of the earth's atmosphere which faces the sun directly. Its value is close to the easily remembered estimate 2 cal/min-cm$^2$; it is about 3% greater than the average value during the summer in the northern hemisphere and about 3% smaller during the winter in the northern hemisphere because the earth's orbit around the sun is not exactly circular. Using 2 cal/min-cm$^2$, we may estimate a reasonable average radiant power density for the earth as a whole by proceeding as follows.

We first multiply the approximate value 2 cal/min-cm$^2$ by factors of unity in order to convert the solar constant to the power density in watts per square meter (W/m$^2$), viz., $Q_0 = 2.0$ (cal/min-cm$^2$) $\times$ ($10^4$ cm$^2$/m$^2$) $\times$ (1 J/0.24 cal) $\times$ (1 min/60 sec) $\times$ (1 W-sec/J) or $Q_0 = 1389$ W/m$^2$. The approximate average *solar radiation power density at the top of the earth's atmosphere* is obtained from the solar constant by dividing by 4 to allow for variations of seasons and times of day around the globe. The resulting *global average radiant solar power density* is then about 345 W/m$^2$. Since the mean radius of the earth is $r_e = 6.371 \times 10^6$ m and its surface area is approximately $4\pi r_e^2 = 5.10 \times 10^{14}$ m$^2$, the *radiant power incident on the entire surface of the earth* becomes (345 W/m$^2$) $\times$ ($5.10 \times 10^{14}$ m)$^2$ $= 1.77 \times 10^{17}$ W $\times$ ($10^7$ erg/1 W-sec) $\times$ ($3.154 \times 10^7$ sec/year) $\approx 5 \times 10^{31}$ erg/year, where we have slightly reduced the calculated value of $5.6 \times 10^{31}$ erg/year and neglected the small difference between the radius of the earth and the radial distance to the top of the atmosphere from the center of earth. This widely used numerical value is itself only a fair approximation to the actual value because of the many idealizations made in the calculations (e.g., the earth is spherical, $Q_0$ is exactly 2 cal/min-cm$^2$ rather than a slightly smaller value, etc.).

It should be noted that local solar radiant power inputs are variable not only because of seasonal and diurnal changes but also because of variability of optical properties of the atmosphere (changes in water vapor concentration, clouds, and particulate concentrations) and of surface reflectivities of the earth. A careful evaluation of

these effects is very complex and requires use of the complete equations for radiant energy transfer with absorption, scattering, and reflection from variable atmospheres and highly variable surfaces.

The number $5 \times 10^{31}$ erg/year provides a reasonable resource estimate for a discussion of solar energy availability over the entire planetary surface.

## B. Wind, Geothermal, Hydroelectric, and Tidal Energy

The numerical value listed for the worldwide wind-energy potential (Palmén and Newton, 1969) results from the estimate that $2 \times (10 \times 10^{10}$ MW) are available if the average winds are the same in the northern and southern hemispheres, which corresponds to $2 \times 10 \times 10^{10}$ MW $\times (10^6$ W/MW) $\times (10^7$ erg/$2.78 \times 10^{-4}$ W-hr) $\times (8.76 \times 10^3$ hr/year) $= 6 \times 10^{31}$ W/yr. The listed value for the outward flow of heat from the core of the earth equals $8 \times 10^{17}$ Btu/yr $= 8 \times 10^{27}$ erg/year and includes heat flow to the large areas covered by the earth's oceans; at locations where this geothermal heat flow is usable, the conversion efficiency to electricity is not likely to exceed about 25%. The limiting worldwide potential for hydroelectric energy is $2.86 \times 10^6$ MW $\times [10^6$ W/MW] $\times (10^7$ erg/$2.78 \times 10^{-4}$ W-hr) $\times (8.76 \times 10^3$ hr/yr) $= 9 \times 10^{26}$ erg/year; the fossil-fuel equivalent for this resource base as a source of electric energy is about three times larger or about $3 \times 10^{27}$ erg/year. Based on the increase in the length of the day during the last 100 years, the total tidal energy has been estimated to be about $3 \times 10^6$ MW $= 3 \times 10^{12}$ W $\times (10^7$ erg/$2.78 \times 10^{-4}$ W-hr) $\times (8.76 \times 10^3$ hr/year) $= 1 \times 10^{27}$ erg/year. The achieved conversion efficiencies for tidal to electrical energy are in the range of 10–20%.

## C. Ocean Thermal Energy Conversion [OTEC]

A total assessment of the worldwide potential for energy recovery using OTEC does not appear to have been made. That the magnitude of this resource is very large is certain in view of a 1964 estimate of the potential for electricity generation using only the existing thermal gradients in the Gulf Stream (Anderson and Anderson, 1964, cited in Anderson, 1966). The estimate is $1.82 \times 10^{14}$ kW$_e$h/yr. Assuming 33% conversion efficiency to electricity, the thermal fossil-energy equivalent becomes

$$(1.82 \times 10^{14} \text{ kW}_e\text{h/yr}) \times (3 \text{ kW/kW}_e) \times (10^7 \text{ erg}/2.78$$
$$\times 10^{-7} \text{ kWh}) = 1.96 \times 10^{28} \text{ erg/yr}$$

The worldwide potential (see the article in 18.05 G by W. H. Avery) is probably at least 20 times greater, i.e.

$>4 \times 10^{29}$ erg/year, and hence may represent a complete, sustainable solution to the world's energy needs at the estimated steady-state requirement beyond 2050 for 5% recovery of the worldwide ocean thermal energy conversion potential.

## V. RECENT ESTIMATES OF FOSSIL-FUEL RESERVES AND RESOURCES LIKELY TO BECOME USABLE OVER THE NEAR TERM

Estimations of reserves should be more accurate than those of resources since they presumably refer to currently available supplies at acceptable costs. Nevertheless, listed tabulations have built-in inaccuracies arising from diverse causes such as the following: (1) different reserve holders (e.g., international and domestic oil companies, governmental agencies, coal companies, drillers for NG, etc.) may not use compatible evaluation procedures; (2) national representatives of oil interests may overestimate reserves in order to obtain increased leverage in dealing with monopolistic consortia such as OPEC and thereby influence approval of increased recovery and sales; and (3) companies may find it expedient to underestimate reserves for the purpose of augmented drilling-right approvals or to leverage assets and stock prices over time. It is therefore not surprising to find diverse estimates from "authoritative" sources.

## A. Oil Reserves

In a *Scientific American* article addressed to a general audience, Campbell and Laherrère (1998) express doubts about official reserve reports from six OPEC countries (Saudi Arabia, Iran, UAE, Iraq, Kuwait, and Venezuela), which show an increase in oil reserves from about $370 \times 10^9$ to $710 \times 10^9$ bbl between 1980 and 1997, on the grounds that the increases were reported without major discoveries of new fields.[4] This type of argument is reinforced by noting a decline in announcements of newly discovered oil fields since 1980, which has been accompanied by depletions of aging fields. Omitted from this type of analysis are likely beneficial impacts from enhanced

---

[4]Developed countries typically maintain petroleum stocks at levels corresponding to about 1 month of consumption, with somewhat smaller levels in Canada and Italy and somewhat larger stocks in Germany and Japan. These low ratios of petroleum stock to annual consumption make the OECD countries highly sensitive to short-term supply disruptions. Petroleum production in non-OPEC countries was smaller than in OPEC countries during the period 1973–1977, slightly exceeded OPEC production during 1978, and exceeded it by substantial amounts in later years. Since 1995, OPEC production has grown more rapidly than non-OPEC production and may well again supply more than 50% of totals by 2010 if present trends continue.

oil recovery and the possibility that the accounting used in 1980 may have been low in order to justify high oil prices. Without alternative oil supplies or new discoveries, the authors expect OPEC's supplies to reach 30% of world supplies (as in the 1970s) during the year 2000 or 2001 and the 50% level by 2010, thereby suggesting the possibility of arbitrarily set increased oil prices within a decade.

A recent staff report of *The Wall Street Journal* (Cooper, 1999) refers to a reservoir off Louisiana at Eugene Island (E.G. 330) and describes output peaking at about 15,000 bbl/day following initiation of recovery after 1973. Production slowed to about 4000 bbl/day by 1989 but then began to recover and reached 13,000 bbl/day in 1999. As a result, reserve estimates have changed from 60 to 400 million bbl for E.G. 330. The conjecture is that the reservoir is being refilled from a deep source miles below the surface of the earth. During the last 20 years, the reserve estimates for the Middle East have doubled, while active exploitation was in progress and no major new discoveries were announced. A theory to account for the observations, namely, that superheated methane was carrying crude oil upward from greater depths, was apparently not substantiated by a drilling test.[5] Although more research is clearly needed to explain the augmented oil recovery achieved at E.G. 330 and perhaps also the growing estimates for Middle East reserves, the suggestion that deep oil reservoirs of great magnitude remain to be tapped is not unreasonable, especially in view of recent successes with deep-sea drilling off the coast of West Africa.

In the same issue of *Scientific American*, Anderson (1998) discusses the likelihood of augmented recovery above the conventional 25–35% of oil in place as the result of underground imaging, gas flooding, steerable drilling, and discoveries of giant deposits off the coast of West Africa.

In its 1998 Annual Report to shareholders, Exxon shows a bar graph for each of the years 1994–1998 labeled "proved reserves replacement," which exceeds 100% for each of the 5 years and includes an excess of about 120% for 1997, i.e., the company's reserve estimates have been steadily increasing. Major oil and gas resource additions are shown for Alaska, the Gulf of Mexico, in Europe (for the North Sea off the coasts of England and Germany and near northeastern Spain), and offshore West Africa near Angola and also for Papua New Guinea. These resource

additions are defined as "new discoveries and acquisitions of proved reserves and other resources that will likely be developed." Areas with the largest new resource additions are the Commonwealth of Independent States (former Soviet Union) and Africa, closely followed by Europe. It is not credible to interpret data of this type as pointing to near-term oil-reserve depletions.

Another company perspective on oil reserves is provided by the British Petroleum Company's "Statistical Reviews of World Energy," which show that generally rising consumptions of oil and natural gas since the end of World War II have been accompanied by increasing reserve estimates: world production ($3.3616 \times 10^9$ mt) and consumption ($3.3128 \times 10^9$ mt) of oil are in balance, as are those for other fossil fuels. In 1996, the yearly reserve-to-production ($R/P$) ratio for the world as a whole was given as 42.2, with 8.2 for Europe, 15.7 for the Asia–Pacific Region, 18.1 for North America, about 25 for both Africa and the CIS, 36.1 for South and Central America, and 93.1 for the Middle East. These values do not suggest strong price escalation over the near term. Price determinants through market changes are discussed by Hannesson (1998), as well as declines in production rates from reservoirs. An interesting associated topic is the intelligent management of fossil-fuel wealth through partial revenue contributions to a permanent fund for the purpose of leveling the income of a country over time as oil revenues decline (Hannesson, 1998, Chapter 7).

Assertions of declining oil stocks accompanied by rapidly rising prices have a substantial history of being faulty, in spite of the fact that announcements of huge new oil discoveries often turn out to represent overly optimistic reserve estimates. Echoing statements from Exxon and British Petroleum, wells in waters 3300–4600 feet deep off the coast of West Africa are being hailed as showing "excellent" results (George, 1998). According to Exxon (1998–1999), some of these new fields are believed to contain reserves of $1 \times 10^9$ bbl or more (which are referred to as "elephants" by geologists).

In 1995, the U.S. Geological Survey (USGS) performed a U.S. assessment of oil and gas *resources*. Measured (proven) *reserves* in BBO (billions of barrels of oil) were 20.2, augmented by (a) 2.1 BBO of continuous-type accumulations in sandstones, shales, and chalks, (b) 60.0 BBO expected reserve growth in conventional fields, and (c) expected, but as yet undiscovered conventional resources of 30.3 BBO, thus yielding *total resources* of 112.6 BBO. Similarly, in TCFG (trillions of cubic feet of natural gas), the estimates were (a) measured *reserves* of 135.1 TCFG, (b) continuous-type accumulations in coal beds of 49.9 TCFC, (c) continuous-type accumulations in sandstones, shales, and chalks of 308.1 TCFG, (d) expected reserve growth in conventional fields of 322.0 TCFG, and

---

[5]In this connection, mention should be made of a hypothesis by Gold (1999), who describes a biosphere under pressure with temperatures around $100°$C at depths near 7 km that exceeds in contents the surface biosphere where we live. There "chemilithotropic" reactions involving primordially stored hydrogen and hydrocarbons lead to coal and petroleum formations that continuously augment the availability of fossil-fuel resources. Some supporting evidence has been found in bringing to the surface downhole materials from 6.7-km depths.

(e) expected, but as yet undiscovered conventional resources of 258.7 TCFG, for a total of 1073.8 TCFG. Finally, for natural gas liquids (NGL), the estimates in BBO were (a) measured reserves of 6.6, (b) continuous-type accumulations in sandstones, shales, and chalks of 2.1, (c) expected reserve growth in conventional fields of 13.4, and (d) expected, but as yet undiscovered conventional resources of 7.2, for a *total resource* estimate of 29.3 BBO. These fossil-fuel *reserves* and *soon-to-be-usable resources* have a combined energy content of approximately $(112.6 \times 10^9$ bbl of oil $\times 5.8 \times 10^{16}$ erg/bbl$) + (1.074 \times 10^3 \times 10^{12}$ feet$^3$ of NG $\times 10^{13}$ erg/foot$^3] + ( 29.3 \times 10^9$ bbl of oil equivalent $\times 5.8 \times 10^{16}$ erg/bbl of oil equivalent$) = 6.5 \times 10^{27}$ erg as oil $+ 1.1 \times 10^{28}$ erg from NG $+ 1.7 \times 10^{27}$ erg from NGL $\approx 1.9 \times 10^{28}$ erg total.

A year 2000 estimate by the USGS (2000; for a brief summary, see Science, 2000) shows a very substantial increase in oil resources that should become usable over the near term. A conclusion is that worldwide peak conventional oil production is more than 25 years in the future as compared with a prior estimate of worldwide peak oil production during the first decade of the 21st century. The large supply increases are very unevenly distributed over the globe with most occurring in the Middle East and the CIS. The absence of large new discoveries in North America may reflect the absence of significant recent exploration in this region.

## 1. Oil Recovery

Recent U.S. oil production of 6.4 MBO/day involves mostly primary recovery (which is accomplished by using only the energy stored in the reservoir), yielding 17% of the oil in place, while secondary recovery may contribute 51% and tertiary recovery (or enhanced oil recovery, EOR) an additional 12% (US DOE, 1996). Secondary recovery may be accomplished with water flooding. Tertiary recovery or EOR (using thermal, gas-miscible, chemical, or microbiological techniques) could potentially augment U.S. production by 50 BBO. However, current high costs have limited the use of EOR. The U.S. DOE program on Production and Reservoir Extension is aimed at achieving lower cost secondary and tertiary oil recovery from known and previously worked reservoirs.

## 2. The U.S. Strategic Petroleum Reserve

The Strategic Petroleum Reserve (SPR) was established in 1975 as a buffer against possible oil embargoes of the type that occurred during 1973–1974 (US Congress, 1998). The SPR holds currently about 564 MBO (million barrels of oil), well below the initially planned level of 1000 MBO. Since its establishment, a sufficient number of tests have been performed to ensure that filling and emptying (drawdown) procedures for reserve additions and subtractions function properly. As of 1998, the SPR had become a depository for the sale of oil in order to generate revenue for the government, while additions were curtailed to reduce federal spending, on the assumption that the likelihood of supply interruptions was low because OPEC's share of U.S. imports had been reduced to about 30% while the unregulated oil markets operated efficiently with respect to price allocation. For reasons that are unclear, a consensus has emerged that an SPR containing around 500 MBO is sufficient, even though this level of supply is used up in the United States in little more than 30 days. A review of management procedures used for the SPR may suggest that reserve additions have been made preferentially when oil prices were high while reserve drawdowns occurred occasionally when oil prices were low. In view of the privatization of the U.S. Navy's Elk Hill petroleum reserve, the SPR may be said to provide now about the same level of domestic security as was available prior to establishment of the SPR and sale of the Navy reserve.

The total input into the SPR through 1995 was 612.8 MBO, with 41.9% or 256.7 MBO from Mexico, 147.3 MBO (24.0%) from the North Sea, 48.1 MBO (7.8%) from U.S. production, 27.1 MBO (4.4%) from Saudi Arabia, 23.7 MBO (3.9%) from Libya, 20.0 MBO (3.3%) from Iran, and lesser amounts from other countries. Drawdown during the Gulf War reduced SPR holdings to 568.5 MBO, whereas subsequent purchases led to refilling to 591.6 MBO.

The drawdown and distribution capability was designed to be 4.3 MBO/day for 90 days. Gas seepage and sale or transfer of 69 MBO from the Week Island site, which was decommissioned during 1996, reduced the drawdown capacity to 3.7 MBO/day by the spring of 1998.

We refer to the U.S. Congress (1998) for some discussion of political decisions that have determined drawdown, refilling, and sale of SPR oil.

## B. Natural Gas Reserves and Methane Hydrate Resources

The reserve-to-production ratio for natural gas (NG) was larger than that for oil in 1996.[6] The world total was 62.2; it was 11.8 for North America, 18.6 for Europe, 40.1 for the Asia-Pacific region, 70.2 for South and Central America, 80.1 for the CIS, and more than 100 for Africa and the Middle East. NG supply estimates, like those for conventional oil, were greatly increased in the year 2000 USGS revisions.

---

[6]As cited in the British Petroleum Company's "Statistical Reviews of World Energy" for that year.

Resource estimates for methane hydrates are highly uncertain but clearly much greater than the known resources of NG. Examples are the following (Mielke, 1998): (1) The world's total gas hydrate accumulations range from $1 \times 10^5$ to $3 \times 10^8$ trillion feet$^3$ of NG, i.e. from $1 \times 10^{30}$ to $3 \times 10^{33}$ erg. These energy estimates range from several times the known NG reserves to those of uranium from seawater used in breeder reactors, with the larger value sufficient to supply $10 \times 10^9$ people at the estimated year 2050 energy demand for 150,000 years. (2) The USGS 1995 appraisal of NG for the United States is $3 \times 10^5$ trillion feet$^3$, or $3 \times 10^{30}$ erg, which exceeds the known U.S. *reserves* of NG by about a factor of 140. Seafloor stability and safe recovery of hydrates are open issues.

## C. Reserves and Resources of Coals

Based on 1978 technology, the world's coal reserves (Penner, 1987, p. 627) were estimated to equal about $3 \times 10^{12}$ bbl of oil-equivalent (bbl$_e$) $= 1.7 \times 10^{29}$ erg, with the U.S. share of 28% exceeding that of Europe (20%), the CIS (17%), and China (16%). Estimates generally include coals of all ranks (from peat, lignite, subbituminous and bituminous coals, to anthracite).

### 1. U.S. Coal Resources and Coal Use

U.S. coal resources are being reassessed by the USGS with initial emphasis on (a) the Appalachian and Illinois basins, (b) the Gulf Coast, (c) the northern Rocky Mountains and Great Plains, and (d) the Rocky Mountains–Colorado Plateau (USGS, 1996). Even before completion of this study, it is apparent that U.S. coal reserves will not soon be exhausted. About one billion short tons (t) of coal were produced for the first time in the United States in 1994 after a continuous sequence of production rises extending for a period of time longer than a century.

Consumption during 1994 in the United States was $930 \times 10^6$ t, of which electric utilities used $817 \times 10^6$ t and $71 \times 10^6$ t were exported. Nonutility use was mostly confined to steel production.

### 2. Environmental Issues in Coal Use

Concerns have often been expressed about coal utilization as a source of toxic and also of radioactive materials since coals contain uranium and thorium at concentration levels of one to four parts per million (ppm), with uranium at levels of 20 ppm rare and thorium as high as 20 ppm very rare (USGS, 1997). Coal combustion releases both U and Th from the coal matrix. As a result, more than 99.5% of these elements ends up in the 10% of solid ash which is formed on the average, i.e., the concentrations of radioac-

tive U and Th in the ash are approximately 10 times those in the original coals. The resulting levels of radioactive elements are comparable to or lower than those normally found in granitic and phosphate rocks and lower than those of Chattanooga shales (10-85 ppm of U). The radioactive combustion products impact the environment when they become dispersed in air and water or in flyash from coals, which is incorporated in such commercial products as concrete building blocks. Upper-bound estimates suggest possible radioactive dose enhancement up to about 3% in residences, corresponding to roughly 10% of the dose from indoor radon that is normally present and contributes about 65% of the total radioactive load. Of the man-made radiation sources, X-rays for diagnostic purposes are the largest contributors at about 11% of the total exposure. About 75% of the annual production of flyash ends up in impoundments, landfills, and abandoned mines or quarries. Standard tests to ensure contamination levels below 100 times those of potable water for leachability of toxic flyash constituents such as arsenic, selenium, lead, and mercury show that the actual amounts of dissolved constituents justify classification of flyash as a nonhazardous solid waste. The U.S. Environmental Protection Agency has established radioactivity standards for drinking water at 5 picocuries per liter for dissolved radium; standards for U and radon are to be established soon. The leachability of radioactive elements increases greatly for strongly alkaline (pH $\gtrsim 8$) or strongly acidic (pH $\lesssim 4$) water-based solutions. Further studies are required, especially for past flyash disposals in leaky landfills. Under current regulations for flyash disposal, leaching of radioactive elements or compounds should not pose a significant environmental hazard.

## D. Oil Reserves for Oil Recovery from Tar Sands

Of the very large Athabasca tar-sand resources, about $300 \times 10^9$ bbl $= 1.7 \times 10^{28}$ erg are recoverable with current technology (George, 1998) and may therefore be classified as reserves. This reserve assessment is expected to increase greatly with oil price escalation above about \$20/bbl.

## E. Shale-Oil Reserves

Although shale-oil resources are enormous and limited commercial recovery has been in progress during most of the 20th century (e.g., in China and Estonia), a major U.S. development utilizing *in situ* recovery during the 1970s and 1980s (Penner and Icerman, 1984, pp. 10–67) was terminated because of a decline in oil prices. It is unlikely that this program will be reactivated before *stable* oil

prices rise above about \$30/bbl, at which time substantial reserve estimates may be demonstrated for shale oil.

## VI. THE KYOTO PROTOCOL

The Kyoto Protocol was signed by the Clinton Administration in November 1998 but has not yet been ratified by the U.S. Senate. It obligates the United States to reduce carbon emissions 7% below 1990 levels during the 2008-2012 time period.

Workers at the Energy Information Agency (EIA) of the U.S. Department of Energy have considered six projections to 2012 for U.S. carbon emissions, which range from an upper bound representing the "reference case" or business as usual to the protocol-mandated reduction levels. Year 2012 U.S. carbon-emission levels range from about $1.25 \times 10^9$ to $1.87 \times 10^9$ mt.

The carbon reduction scenarios have large associated costs. The year 2010 U.S. gross domestic product is estimated to be reduced by about \$400 billion dollars to meet protocol requirements by 2010 without interdicted emissions trading with less developed countries. Emissions reductions may be achieved in a number of ways, e.g., by imposing a carbon emission tax, by extensive fuel switching from coal to NG for electricity generation and from oil to natural gas for transportation, by curtailing gasoline use through oil-price hikes, etc. According to the EIA, meeting protocol requirements will increase the average year 2010 household energy costs by \$1740.

It is unlikely that appreciable carbon mitigation can be accomplished as early as 2010 through implementation of such new energy-efficient technologies as fuel-cell-powered vehicles. Probably the most cost-effective procedure for reducing carbon emissions is a large-scale change to nuclear power generation, perhaps to levels exceeding 50%, which were reached some years ago in France, Japan, and some other developed countries.

## VII. POLICY CONSIDERATIONS

We have seen that there are abundant low-cost energy supplies, especially in the form of fossil fuels and fissionable uranium, which could be used to supply the world's energy needs for many years to come. Unfortunately, this is not the end of story. The use of any energy reserve or resource for human well-being is accompanied not only by an increase in global entropy but also by a variety of detrimental environmental effects. Among these, public attention has been focused primarily on possibly disastrous long-range climate changes associated with the very well documented relentless escalation of the atmospheric $CO_2$ concentration since the beginning of large-scale anthropogenic conversion of gaseous, liquid, and solid fossil fuels to combustion products, of which benign water and perhaps not so benign $CO_2$ are the chief constituents.

### A. Mitigation of the Greenhouse Effect

The increased gaseous water production associated with fossil fuel use is of no special environmental consequence for three reasons: (1) The resulting increase in the atmospheric water concentration is negligibly small compared with the preexisting steady-state atmospheric water concentrations, (b) consequently, the changes in atmospheric heating produced by any increases in the dominant greenhouse gas, water, are also negligibly small, and (c) finally, the residence time of the added water in the atmosphere is very short (on the order of days to weeks) and therefore water cannot accumulate over very long periods of time. Unfortunately, the story is reversed for anthropogenic $CO_2$ addition to the atmosphere because (a) the added amounts of $CO_2$ are not negligibly small compared with preexisting levels (manmade contributions have already raised preexisting $CO_2$ levels by upward of 25%), (b) the added $CO_2$ increases the greenhouse effect measurably because long-wavelength absorption by the added $CO_2$ augments atmospheric heating, and (c) the residence time of the added $CO_2$ is of the order of 100–200 years and may therefore be responsible for long-term atmospheric heating changes as the result of absorption of long-wavelength radiation emitted from the planetary surface. The inevitable conclusions derived from these physical effects are that whereas hydrogen would be an ideal fuel, fossil fuels may cause unacceptable long-term climate changes because of $CO_2$ emissions to the atmosphere. Hence, we must face proposals to eliminate fossil-fuel use without $CO_2$ mitigation over the long term.

### B. Elimination of Long-Lived Radioactive Products

The second supply source that is now in large-scale use involves nuclear fission reactors. With these, we also face difficult environmental issues, resulting from (a) inadequately designed, maintained, or improperly operating reactors leading to calamitous accidents (e.g., Chernobyl, Three-Mile Island, Tokaimura, and elsewhere) and (b) the certain production of long-lived radioactive isotopes which will require safe storage for very long periods of time, exceeding hundreds of thousands of years in some cases. Whereas the accident issues will be controllable by constructing passively safe nuclear reactors (i.e. by constructing reactors that cannot cause calamitous accidents as the result of human errors, which are designs that will not produce runaway heating or radioactive emissions because of the judicious use of physical phenomena that will inhibit runaway calamities as the result of human

**TABLE II   A Global, Large-Scale Energy Supply Trade-off Matrix (1999)[a]**

| Large-scale energy source | Overriding consideration in source selection | | | |
|---|---|---|---|---|
| | Low current cost | No net $CO_2$ emission | No long-lived radioactive products | Low risk because of established large-scale technology |
| Fossil fuels (all) | 5–10 | 2 | 10 | 9–10 |
| Coals | 10 | 2 | 10 | 10 |
| Oils | 8–10 | 2 | 10 | 10 |
| Natural gases | 9–10 | 2 | 10 | 10 |
| Shale oils | 5–10 | 2 | 10 | 9–10 |
| Oils from tar sands | 7–10 | 2 | 10 | 9–10 |
| Solar technologies (all) | 2–10 | 3–10 | 10 | 2–10 |
| Wind energy | 4–10 | 9–10 | 10 | 7–9 |
| Photovoltaics | 2–10 | 2–10 | 10 | 2–10 |
| Biomass | 2–10 | 7–10 | 10 | 2–10 |
| Hydrogen production | 2 | 3–10 | 10 | 2 |
| Ocean thermal energy production | 4–8 | 4–10 | 10 | 4–8 |
| Hydroelectric power (partial supply) | 10 | 10 | 10 | 10 |
| Hydrothermal energy (partial supply) | 10 | 10 | 10 | 10 |
| Hot dry-rock geothermal energy (partial supply) | 2–8 | 10 | 10 | 2–8 |
| Nuclear fission reactors (all) | 10 | 9–10 | 2 | 10 |
| Nuclear breeder reactors (all) | 7 | 9–10 | 2 | 9 |
| Fusion reactors | 2 | 9–10 | 7–9 | 2 |

[a] 10, Very high probability of meeting goals; 1, very low probability of meeting goals.

errors), long-term safe storage of highly radioactive reactor products remains an inevitable component of the "*Faustian bargain*" (to quote Alvin Weinberg) which accompanies our use of nuclear fission reactors. Proponents of nuclear fission reactors consider long-term, safe storage in stable geological strata to provide a proper technical response to this issue. Opponents of nuclear power continue to believe that we are incapable of coping with the required long-term safe storage issues, especially because their use would, in principle, facilitate access to nuclear weapons materials and thereby invite weapons proliferation.

The preceding arguments lead to the conclusion that we must change to renewable resources, provided their use is not so costly as to cause economic collapse for the growing world population, which appears inevitably reaching around 10 billion people in about 50 years, only a few years longer than the historical time scale for replacing an established energy technology by a new energy technology. After more than 25 years of investments in and subsidies for renewable technologies, the absence of significant progress makes it appear likely that this program will not lead to acceptable, *cost-effective*, *large-scale* replacements of fossil and nuclear energy supply systems over the near term.

The preceding considerations are summarized in Table II in terms of numerical estimates reflecting the author's assessments of the current state of technological development on a scale from 1 to 10, with 10 representing a *currently competitive* technology, 2 a research agenda aimed at creating a competitive technology, and numbers between 2 and 10 representing subjective assessments of how close we have come thus far in creating replacements for currently competitive technologies on *a large scale* and *at acceptable costs*.

While one may view Table II as a proper basis for rational progress, it is important to remember that experts in other countries may find our approach for defining a new energy road map to the future either uninteresting or irrelevant to their needs and preferences. With an inevitable emphasis on *low-cost supplies* for *local well-being*, Table III for China, India, France, Norway, and OPEC members reflects *local political preferences* and goals rather than the entries in Table II.

## C. Research and Development Agenda for Future Energy Security

If a political decision is made to eliminate man-made, long-lived radioactive products from the environment, no

**TABLE III   Local Energy Supply Trade-off Matrices for China, India, France, Norway, and OPEC Members[a]**

| Country | Energy source | Overriding consideration in source selection | | | |
| | | Low local cost | Export value | Established technology | Environmental opposition |
|---|---|---|---|---|---|
| China | Fossil fuels (all) | 8–10 | 8–10 | 10 | Minimal |
| | Coals | 10 | 10 | 10 | Minimal |
| | Oils | 10 | 10 | 10 | Minimal |
| | Natural gases | 10 | 10 | 10 | Minimal |
| | Shale oils | 8 | 10 | 10 | Minimal |
| | Oils from tar sands | 8 | 10 | 10 | Minimal |
| | Nuclear fission reactors | 10 | 10 | 10 | Minimal |
| | Nuclear breeder reactors | 7 | 7 | 9 | Minimal |
| | Hydroelectric power | 10 | 10 | 10 | Minimal |
| India | Coal | 10 | 10 | 10 | Minimal |
| | Fuelwood | 10 | 1 | 10 | Moderate |
| | Hydroelectric power | 8–10 | 1–10 | 10 | Moderate to strong |
| | Biomass (large scale) | 5–10 | 1–3 | 7 | Moderate |
| | Nuclear fission reactors | 10 | 10 | 10 | Minimal |
| | Nuclear breeder reactors | 7 | 10 | 9 | Minimal |
| France | Nuclear fission reactors | 10 | 10 | 10 | Low |
| | Nuclear breeder ractors | 7 | 10 | 9 | Low |
| Norway | Oils | 10 | 10 | 10 | Minimal |
| | Hydroelectric power | 10 | 10 | 10 | Minimal |
| | Natural gases | 10 | 10 | 10 | Minimal |
| OPEC members | Oils | 10 | 10 | 10 | Minimal |

[a] 10, Very high probability of meeting goals; 1, very low probability of meeting goals.

special supply problems will arise for many years because we can continue and augment the use of low-cost fossil fuels. The practical agenda will stress improved energy utilization efficiencies, while R&D emphasis will remain on $CO_2$ sequestration at the source and on a variety of solar technologies and the development of fusion reactors (see Table IV). Unfortunately, our present perspective must assign low probabilities of success to all of the identified programs if the focus is on low-cost, large-scale supply sources by about the year 2025.

If our political choice becomes elimination of $CO_2$ emissions to the atmosphere, or disposal in the deep sea or stable caverns of the enormous amounts of $CO_2$ generated from fossil-fuel burning, then we see again no disastrous near-term or long-term calamities since the nuclear fission options remain for known, relatively low-cost supply. In this environment, near-term construction emphasis must properly shift to the use of passively safe nuclear fission and breeder reactors at somewhat elevated costs in order to ensure the availability over the very long term of adequate energy supplies for an increasingly affluent and growing world population. While the R&D program on solar and

fusion reactors continues as before, greatly augmented effort should be directed also at the commercialization of $CO_2$ sequestration with the hope of retaining some of the fossil-fuel supply sources, especially during the transition period to large-scale use of nuclear breeder technologies (see Table IV).

What should be done if both fossil-fuel and nuclear fission reactors are interdicted? As Table IV shows, we are now left without usable, low-cost, large-scale energy supply sources and only the promise of better things to come if we can succeed in the cost-effective commercialization of large-scale solar or fusion technologies. During the foreseeable future, this will be unacceptable to people everywhere and will not be followed by their elected officials because the economic consequences would be calamitous and stable political structures could not survive.

### D. The Preferred Research and Development Agenda for the Future

There is little doubt that the best energy technology for the world at large is the ability to produce low-cost $H_2$ on

**TABLE IV   Selected Research and Development Agenda**

| Goal | Energy source | Agenda | Probability of success at acceptable cost[a] |
|---|---|---|---|
| Early elimination of $CO_2$ emissions to the atmosphere | Fossil fuels (all) | $CO_2$ capture at the source and utilization for the production of salable chemicals; $CO_2$ capture and separation at the source prior to long-term storage in the sea or elsewhere | 2–4 <br> 2–4 |
| | Solar technologies (all) | Direct $H_2$ production from water (e.g., by water splitting in low-temperature thermal cycles, direct water photolysis, etc.); large-scale development of energy-from-the-sea technologies followed by hydrogen or other fuel production; genetically altered biomass to improve crop yields for energy use; solar energy collection in space orbits followed by conversion and transmission to earth stations; large-scale use of wind power or photovoltaics | 2 <br> 6–8 <br> 2 <br> 2 <br> 2 |
| | Nuclear fission reactors | Implement large-scale use of passively safe nuclear fission reactors | 10 |
| | Nuclear breeder reactors | Implement large-scale use of passively safe nuclear breeder reactors | 10 |
| | Fusion reactors | Implement large-scale use of systems | 2 |
| Early cessation of the production of long-lived radioactive isotopes | Fossil fuels (all) | Improve utilization efficiencies | 10 |
| | Solar technologies (all) | As above | |
| | Fusion reactors | Implement large-scale use of systems | 2 |
| No $CO_2$ emissions and no nuclear fission reactors | Solar technologies (all) | As above | |
| | Fusion reactors | Implement large-scale use of systems | 2 |

[a] 10, High probability of success; 1, low probability of success.

the required global scale, i.e. the preferred goal is what enthusiasts have long labeled "the hydrogen economy." It matters little how the hydrogen is ultimately produced. A viable fusion energy supply is, of course, completely compatible with the hydrogen economy since water splitting by using electrical energy is a long-established technology. Similarly, the use of solar energy to produce $H_2$ in low-temperature thermal cycles or by the direct photolysis of water constitute priority methods for $H_2$ production, as is also $NH_3$ synthesis in an ocean thermal energy conversion system followed by conversion to $H_2$ after transmission to user locations, etc. The emphasis on combining new energy supply options with large-scale hydrogen production should become the worldwide priority R&D program.
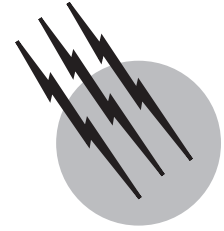
## SEE ALSO THE FOLLOWING ARTICLES

ENERGY EFFICIENCY COMPARISONS BETWEEN COUNTRIES • ENERGY FLOWS IN ECOLOGY AND IN THE ECONOMY • FOSSIL FUEL POWER STATIONS • HYDROELECTRIC POWER STATIONS • NUCLEAR POWER REACTORS • OCEAN THERMAL ENERGY CONVERSION • PETROLEUM GEOLOGY • RENEWABLE ENERGY FROM BIOMASS • SOLAR THERMAL POWER STATIONS • TIDAL POWER SYSTEMS • WASTE-TO-ENERGY SYSTEMS • WIND POWER SYSTEMS

## BIBLIOGRAPHY

Anderson, J. H., and Anderson, J. H., Jr. (1964). "Power from the sun, via the sea," Paper presented at the Susquehanna Section of the ASME, November 1964.

Anderson, J. H., and Anderson, J. H., Jr. (April 1966). "Thermal power from sea water: 100,000 kW at 3 mills/kWh," *Mech. Eng*. **88,** 41–46.

Anderson, R. N. (1998). "Oil production in the 21st century," *Sci. Am.* **1998**(March), 86–91.

Averitt, P. (1969). "Coal Resources of the United States," U.S. Geological Survey Bulletin 1275, Washington, DC.

Best, F. R., and Driscoll, M. J. (1980). *Trans. Am. Nucl. Soc*. **34,** 380.

Best, F. R., and Driscoll, M. J. (1981). *Trans. Am. Nucl. Soc*. **38,** 200.

Campbell, C. J., and Laherrère, J. H. (1998). "The end of cheap oil," *Sci. Am*. **1998**(March), 78–83.

Cohen, B. L. (1983). "Breeder reactors: A renewable energy source," *Am. J. Phys*. **51,** 1.

Cooper, C. (1999). "It's no crude joke: This oil field grows even as it's tapped," *Wall Street Journal*, Friday, April 16, p. A1.

Cornell (1972). "Cornell University Workshop on Energy and the Environment," US Government Printing Office, Washington, DC.

Duncan, D. C., and Swanson, V. E. (1965). "Organic-Rich Shales of the United States and World Land Areas," US Geological Survey Circular 523, Washington, DC.

Exxon (1998–1999). "The Lamp," Exxon Corporation, Shareholder Services, Irvine, TX.

George, R. L. (1998). "Mining for Oil," *Sci. Am*. **1998**(March), 84–85.

Gold, T. (1999). "The Deep Hot Biosphere," Springer-Verlag, New York.

Hannesson, R. (1998). "Petroleum Economics: Issues and Strategies of Oil and Natural Gas Production, II," Quorum Books, Westport, CT.

IAEA. (1966). "Utilization of thorium in power reactors," In Technical Report Series No. 52, pp. 10–11, International Atomic Energy Agency, Vienna.

Koczy, F. F. (1954). *In* "Nuclear Geology" (H. Faul, ed.), Wiley, New York.

Mielke, J. E. (1998). "Methane Hydrates: Energy Prospect or Natural Hazard?" Congressional Research Services Report, The Committee for the National Institute for the Environment, Washington, DC.

Palmén, E. H., and Newton, C. W. (1969). "Atmospheric Circulation Systems: Their Structure and Physical Interpretations," Academic Press, New York.

Penner, S. S. (ed.). (1978). "Lithium Needs and Resources," Pergamon Press, Oxford.

Penner, S. S. (ed.). (1987). "Coal Gasification," Pergamon Press, Oxford [reprinted from *Energy–The Int. J*. **12**(8/9), 623–903 (1987)].

Penner, S. S., and Icerman, L. (1984). "Energy," Vol. II, "Non-Nuclear Energy Technologies," 2nd ed., Addison-Wesley, Reading, MA.

Science. (2000). "USGS Optimistic on World Oil Prospects," *Science* **289,** 237; review by R. A. Kerr.

Smil, V. (1999). "Energies," MIT Press, Cambridge, MA.

Spragins, F. K. (1967). "Mining at Athabasca—A new approach to oil production," *J. Petroleum Technol*. **19,** 1337–1343.

Swanson, V. E. (1960). "Oil yield and uranium content of beach shales," *In* "Uranium and Carbonaceous Rocks," US Geological Survey Paper 356, Washington, DC.

US Congress. (1998). "Strategic Petroleum Reserve," Congressional Research Service (CRS) Report 8705, December 7, 1998.

US DOE. (1996). US Department of Energy Internet data.

USGS. (1996). USGS Fact Sheet FS-157-96, US Geological Survey, Washington, DC.

USGS. (1997). USGS Fact Sheet FS-163-97, US Geological Survey, Washington, DC.

USGS. (2000). Web site http://greenwood.cr.usgs.gov/energy/World/Energy/DDS-60.

Zebroski, E. L., Zaleski, C.-P., Simnad, M., Baker, C. S., and Penner, S. S. (eds.). (1998). "Advanced Nuclear Reactors—Current Developments and Future Prospects," Elsevier, Oxford [reprinted from *Energy–The Int. J.* **23,** 521–694].

# Enhanced Oil Recovery

**Ronald E. Terry**
*Brigham Young University*

## GLOSSARY

**Capillary pressure** Difference in the pressure between two fluids measured at the interface between the two fluids.

**Interfacial tension** Measure of the ease with which a new interface between two fluids can be made.

**Miscible** When two or more phases become single phased.

**Mobility** Measure of the ease with which a fluid moves through porous media.

**Permeability** Measure of the capacity of a porous medium to conduct a fluid.

**Polymer** Large-molecular-weight chemical used to thicken a solution.

**Primary production** Production of oil using only the natural energy of the formation.

**Reservoir** Volume of underground porous media usually containing rock, oil, gas, and water.

**Residual oil** Amount of oil remaining in a reservoir after primary and secondary production.

**Secondary production** Production of oil when gas, water, or both are injected into the formation and the injected fluid immiscibly displaces the oil.

**Surfactant** Molecule that is made up of both hydrophilic and hydrophobic entities and can reduce the interfacial tension between two fluids.

**Sweep efficiency** Measure of how evenly a fluid has moved through the available flow volume in a porous medium.

**Viscosity** Property of a fluid that is a measure of its resistance to flow.

**ENHANCED OIL RECOVERY** refers to the process of producing liquid hydrocarbons by methods other than the conventional use of reservoir energy and reservoir repressurizing schemes with gas or water. On the average, conventional production methods will produce from a reservoir about 30% of the initial oil in place. The remaining oil, nearly 70% of the initial resource, is a large and attractive target for enhanced oil recovery methods.

# I. INTRODUCTION

## A. Classification of Hydrocarbon Production

The initial production of hydrocarbons from an underground reservoir is accomplished by the use of natural reservoir energy. This type of production is termed primary production. Sources of natural reservoir energy that lead to primary production include the swelling of reservoir fluids, the release of solution gas as the reservoir pressure declines, nearby communicating aquifers, and gravity drainage. When the natural reservoir energy has been depleted, it becomes necessary to augment the natural energy with an external source. This is usually accomplished by the injection of fluids, either a natural gas or water. The use of this injection scheme is called a secondary recovery operation. When water injection is the secondary recovery process, the process is referred to as waterflooding. The main purpose of either a natural gas or a water injection process is to repressurize the reservoir and then to maintain the reservoir at a high pressure. Hence, the term *pressure maintenance* is sometimes used to describe a secondary recovery process.

When gas is used as the pressure maintenance agent, it is usually injected into a zone of free gas (i.e., a gas cap) to maximize recovery by gravity drainage. The injected gas is usually natural gas produced from the reservoir in question. This, of course, defers the sale of that gas until the secondary operation is completed and the gas can be recovered by depletion. Other gases, such as $N_2$, can be injected to maintain reservoir pressure. This allows the natural gas to be sold as it is produced.

Waterflooding recovers oil by the water's moving through the reservoir as a bank of fluid and "pushing" oil ahead of it. The recovery efficiency of a waterflood is largely a function of the sweep efficiency of the flood and the ratio of the oil and water viscosities.

Sweep efficiency is a measure of how well the water has come in contact with the available pore space in the oil-bearing zone. Gross heterogeneities in the rock matrix lead to low sweep efficiencies. Fractures, high-permeability streaks, and faults are examples of gross heterogeneities. Homogeneous rock formations provide the optimum setting for high sweep efficiencies.

When an injected water is much less viscous than the oil it is meant to displace, the water could begin to finger, or channel, through the reservoir. This is referred to as viscous fingering and leads to significant bypassing of residual oil and lower flooding efficiencies. This bypassing of residual oil is an important issue in applying enhanced oil recovery techniques as well as in waterflooding.

Tertiary recovery processes were developed for application in situations in which secondary processes had become ineffective. However, the same tertiary processes were also considered for reservoir applications for which secondary recovery techniques were not used because of low recovery potential. In the latter case, the name *tertiary* is a misnomer. For most reservoirs, it is advantageous to begin a secondary or a tertiary process concurrent with primary production. For these applications, the term *enhanced oil recovery (EOR)* was introduced and has become popular in referring to, in general, any recovery process that enhances the recovery of oil beyond what primary and secondary production would normally be expected to yield.

Enhanced oil recovery processes can be classified into four categories:

1. Miscible flooding processes
2. Chemical flooding processes
3. Thermal flooding processes
4. Microbial flooding processes

The category of miscible displacement includes single-contact and multiple-contact miscible processes. Chemical processes are polymer, micellar–polymer, and alkaline flooding. Thermal processes include hot water, steam cycling, steam drive, and *in situ* combustion. In general, thermal processes are applicable in reservoirs containing heavy crude oils, whereas chemical and miscible displacement processes are used in reservoirs containing light crude oils. Microbial processes use microorganisms to assist in oil recovery.

## B. Hydrocarbon Reserves and Potential of Enhanced Recovery

In the United States, the remaining producible reserve is estimated to be 21 billion barrels. Of this 21 billion, currently implemented EOR projects are expected to recover 3 billion barrels. A 1998 report in the *Oil and Gas Journal* listed a production of 759,653 barrels of oil per day (b/d) from EOR projects in the United States. This amount represented about 12% of the total U.S. oil production.

A somewhat dated but highly informative study conducted by the U.S. National Petroleum Council (NPC) and published in 1984 determined that, with current EOR technology, an estimated 14.5 billion barrels of oil could be produced in the United States over a 30-yr period. This amount includes the 3 billion barrels that are expected to be produced from current EOR projects. The 14.5-billion-barrel figure was derived from a series of assumptions and subsequent model predictions. Included in the assumptions was an oil base price of $30 per barrel in constant 1983 U.S. dollars. The ultimate oil recovery was projected to be very sensitive to oil price, as shown in Table I.

**TABLE I   Ultimate Oil Recovery from Enhanced Oil Recovery Methods as a Function of Oil Price[a]**

| Oil price per bbl (1983 U.S. dollars) | Ultimate recovery (billions of bbl) |
|---|---|
| 20 | 7.4 |
| 30 | 14.5 |
| 40 | 17.5 |
| 50 | 19.0 |

[a] bbl, barrel(s).

The NPC study also attempted to predict what the ultimate recoveries could reach if technological advancements were made in EOR processes. A potential 13–15 billion barrels of oil could be added to the figures in Table I if research activities led to improvements in EOR technology.

Interest in EOR activity is increasing outside the United States. In the same *Oil and Gas Journal* report that was referenced earlier, the worldwide production from EOR projects was listed as 2.3 million b/d. The largest project in the world at the time of the 1998 *Oil and Gas Journal* report was a steam drive in the Duri field in Indonesia. The Duri steam drive was producing about 310,000 b/d in the first quarter of 1998. Total estimated recovery from the Duri field is expected to be nearly 2 billion barrels of oil. Canada continues to report significant EOR production of 400,000 b/d, while China and Russia have increased their production to 280,000 and 200,000 b/d, respectively.

## II. FUNDAMENTALS OF FLUID PRODUCTION

### A. Overall Recovery Efficiency

The overall recovery efficiency $E$ of any fluid displacement process is given by the product of the macroscopic, or volumetric, displacement efficiency $E_v$ and the microscopic displacement efficiency $E_d$:

$$E = E_v E_d.$$

The macroscopic displacement efficiency is a measure of how well the displacing fluid has come in contact with the oil-bearing parts of the reservoir. The microscopic displacement efficiency is a measure of how well the displacing fluid mobilizes the residual oil once the fluid has come in contact with the oil.

The macroscopic displacement efficiency is made up of two other terms, the areal, $E_s$, and vertical, $E_i$, sweep efficiencies:

$$E_v = E_s E_i.$$

### B. Microscopic Displacement Efficiency

The microscopic displacement efficiency is affected by the following factors: interfacial and surface tension forces, wettability, capillary pressure, and relative permeability.

When a drop of one immiscible fluid is immersed in another fluid and comes to rest on a solid surface, the surface area of the drop will take a minimum value due to the forces acting at the fluid–fluid and rock–fluid interfaces. The forces per unit length acting at the fluid–fluid and rock–fluid interfaces are referred to as interfacial tensions. The interfacial tension between two fluids represents the amount of work required to create a new unit of surface area at the interface. The interfacial tension can also be thought of as a measure of the immiscibility of two fluids. Typical values of oil–brine interfacial tensions are on the order of 20–30 dyn/cm. When certain chemical agents are added to an oil–brine system, it is possible to reduce the interfacial tension by several orders of magnitude.

The tendency for a solid to prefer one fluid over another is called wettability. Wettability is a function of the chemical composition of both the fluids and the rock. Surfaces can be either oil-wet or water-wet, depending on the chemical composition of the fluids. The degree to which a rock is either oil-wet or water-wet is strongly affected by the adsorption or desorption of constituents in the oil phase. Large, polar compounds in the oil phase can absorb onto the solid surface; this leaves an oil film that may alter the wettability of the surface.

The concept of wettability leads to another significant factor in the recovery of residual oil. This factor is capillary pressure. To illustrate capillary pressure, let us consider a capillary tube that contains both oil and brine, the oil having a lower density than that of the brine. The pressure in the oil phase immediately above the oil–brine interface in the capillary tube will be slightly greater than the pressure in the water phase just below the interface. This difference in pressure is called the capillary pressure $P_c$ of the system. The greater pressure will always occur in the nonwetting phase. An expression relating the contact angle $\theta$, the radius $r$ of the capillary, the oil–brine interfacial tension $\gamma_{wo}$, and the capillary pressure $P_c$ is given in Eq. (1):

$$P_c = (2\gamma_{wo} \cos\theta)/r. \tag{1}$$

This equation suggests that the capillary pressure in a porous medium is a function of the chemical composition of the rock and fluids, the pore size distribution, and the saturation of the fluids in the pores. Capillary pressures have also been found to be a function of the saturation history, although this dependence is not reflected in Eq. (1). Because of this, different values will be obtained during the drainage process (i.e., displacing the wetting phase

with the nonwetting phase) than those obtained during the imbibition process (i.e., displacing the nonwetting phase with the wetting phase). This historesis phenomenon is exhibited in all rock–fluid systems and is an important consideration in the mathematical modeling of EOR processes.

It has been shown that the pressure required to force a nonwetting phase through a small capillary can be very large. For instance, the pressure drop required to force an oil drop through a tapering constriction that has a forward radius of 0.00062 cm, a rearward radius of 0.0015 cm, a contact angle of 0°, and an interfacial tension of 25 dyn/cm is 0.68 psi. If the oil drop were 0.01 cm long, a pressure gradient of 2073 psi/ft would be required to move the drop through the constriction. This is an enormous pressure gradient, which cannot be achieved in practice. Typical pressure gradients obtained in reservoir systems are of the order of 1–2 psi/ft.

Another factor affecting the microscopic displacement efficiency is the fact that two or more fluids are usually flowing in an EOR process. When two or more fluid phases are present, the saturation of one phase affects the permeability of the other(s), and relative permeabilities have to be considered. Figure 1 is an example of a set of relative permeability curves plotted against the wetting phase saturation (water in this case).

The relative permeability to oil, $K_{ro}$, is plotted on the left vertical axis, whereas the relative permeability to water, $K_{rw}$, is plotted on the right. The curve for $K_{ro}$ goes to zero at $S_{or}$, the residual oil saturation. Once the oil saturation has been reduced to this point in a pore space by a waterflood, no more oil will flow, since $K_{ro}$ is zero. Similarly, at saturations below the irreducible water saturations, $S_{wr}$
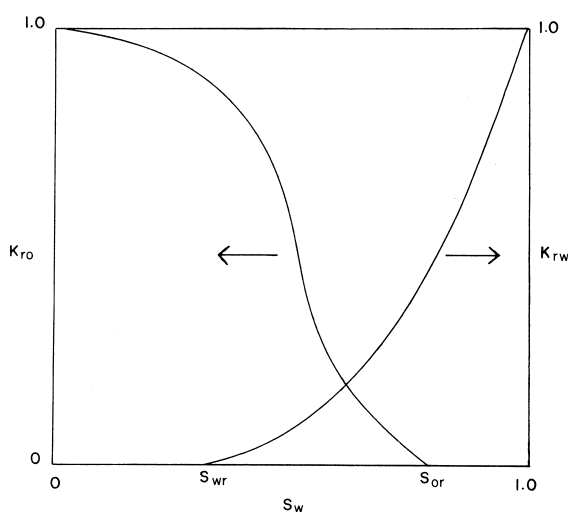


**FIGURE 1** Typical water–oil relative permeability curves for a porous medium.

in the figure, water will not flow, since $K_{rw}$ is zero. The curves are strong functions of wettability and do exhibit a historesis effect (especially for the nonwetting phase permeability).

## C. Macroscopic Displacement Efficiency

Factors that affect the macroscopic displacement efficiency are the following: heterogeneities and anisotropy, the mobility of the displacing fluids compared with the mobility of the displaced fluids, the physical arrangement of injection and production wells, and the type of rock matrix in which the oil exists.

Heterogeneities and anisotropy of an oil-bearing formation have a significant effect on the macroscopic displacement efficiency. The movement of fluids through the reservoir will not be uniform if there are large variations in such properties as porosity, permeability, and clay content. Limestone formations generally have wide fluctuations in porosity and permeability. Also, many formations have a system of microfractures or large macrofractures. Any time a fracture occurs in a reservoir, fluids will be inclined to travel through the fracture because of the high permeability of the fracture. This may lead to substantial bypassing of residual oil. The bypassing of residual oil by injected fluids is a major reason for the failure of many pilot EOR projects. Much research is being conducted on how to improve the sweep efficiency of injected fluids.

Mobility is a relative measure of how easily a fluid moves through porous media. The apparent mobility is defined as the ratio of effective permeability to fluid viscosity. Since the effective permeability is a function of fluid saturations, several apparent mobilities can be defined. The mobility ratio $M$ is a measure of the relative apparent mobilities in a displacement process and is given by Eq. (2):

$$M = \frac{\text{mobility of displacing phase}}{\text{mobility of displaced phase}}. \qquad (2)$$

When a fluid is being injected into a porous medium containing both the injected fluid and a second fluid, the apparent mobility of the displacing phase is usually measured at the average displacing phase saturation as the displacing phase just begins to break through at the production site. The apparent mobility of the nondisplacing phase is measured at the displacing phase saturation that occurs just before the beginning of the injection of the displacing phase. Sweep efficiencies are a strong function of the mobility ratio. The phenomenon called viscous fingering can take place if the mobility of the displacing phase is much greater than the mobility of the displaced phase.

The arrangement of injection and production wells depends primarily on the geology of the formation and the

size (areal extent) of the reservoir. When an operator is considering an EOR project for a given reservoir, he or she will have the option of using the existing well arrangement or drilling new wells in different locations. If the operator opts to use the existing well arrangement, there may be a need to consider converting production wells to injection wells or vice versa. This will require analysis of tubing and other factors to determine whether the existing equipment can withstand the properties of the chemicals or thermal energy to be injected. An operator should also recognize that when a production well is converted to an injection well, the production capacity of the reservoir has been reduced. Often this decision can lead to major cost items in the overall project and should involve a great deal of consideration. Knowledge of any directional permeability effects and other heterogeneities can aid in the consideration of well arrangements. The presence of faults, fractures, and high-permeability streaks could dictate the shutting in of a well near one of these heterogeneities. Directional permeability trends could lead to a poor sweep efficiency in a developed pattern and might suggest that the pattern be altered in one direction or that a different pattern be used.

Sandstone formations are characterized by a more uniform pore geometry than that of limestones. Limestones have large holes (vugs) and can have significant fractures, which are often connected. Limestone formations are associated with connate water that can have high levels of divalent ions such as $Ca^{2+}$ and $Mg^{2+}$. Vugular porosity and high divalent ion content in their connate waters hinder the application of EOR processes in limestone reservoirs. Conversely, a sandstone formation can be composed of such small sand grain sizes and be so tightly packed that fluids will not readily flow through the formation.

## D. Capillary Number Correlation

In a water-wet system, during the early stages of a waterflood, the brine exists as a film around the sand grains and the oil fills the remaining pore space. At an intermediate time during the flood, the oil saturation has been decreased and exists partly as a continuous phase in some pore channels but as discontinuous droplets in other channels. At the end of the flood, when the oil has been reduced to residual oil saturation $S_{or}$, the oil exists primarily as a discontinuous phase of droplets or globules that have been isolated and trapped by the displacing brine.

The waterflooding of oil in an oil-wet system yields a different fluid distribution at $S_{or}$. Early in the waterflood, the brine forms continuous flow paths through the center portions of some of the pore channels. The brine enters more and more of the pore channels as the waterflood progresses. At residual oil saturation, the brine has entered

a sufficient number of pore channels to shut off the oil flow. The residual oil exists as a film around the sand grains. In the smaller flow channels this film may occupy the entire void space.

The mobilization of the residual oil saturation in a water-wet system requires that the discontinuous globules be connected to form a flow channel. In an oil-wet porous medium, the film of oil around the sand grains needs to be displaced to large pore channels and be connected in a continuous phase before it can be mobilized. The mobilization of oil is governed by the viscous forces (pressure gradients) and the interfacial tension forces that exist in the sand grain–oil–water system.

There have been several investigations of the effects of viscous forces and interfacial tension forces on the trapping and mobilization of residual oil. From these studies correlations between a dimensionless parameter called the capillary number $N_{vc}$ and the fraction of oil recovered have been developed. The capillary number is the ratio of viscous forces to interfacial tension forces and is defined by Eq. (3):

$$N_{vc} = V\mu_w/\gamma_{ow} = K_0\,\Delta P/\phi\gamma_{ow}L, \qquad (3)$$

where $V$ is the Darcy velocity, $\mu_w$ the viscosity of displacing fluid, $\gamma_{ow}$ the interfacial tension between displaced and displacing fluid, $K_0$ the effective permeability to displaced phase, $\phi$ the porosity, and $\Delta P/L$ the pressure drop associated with Darcy velocity.

Figure 2 is a schematic representation of the capillary number correlation. The correlation suggests that a capillary number greater than $10^{-5}$ is necessary for the mobilization of unconnected oil droplets. The capillary number
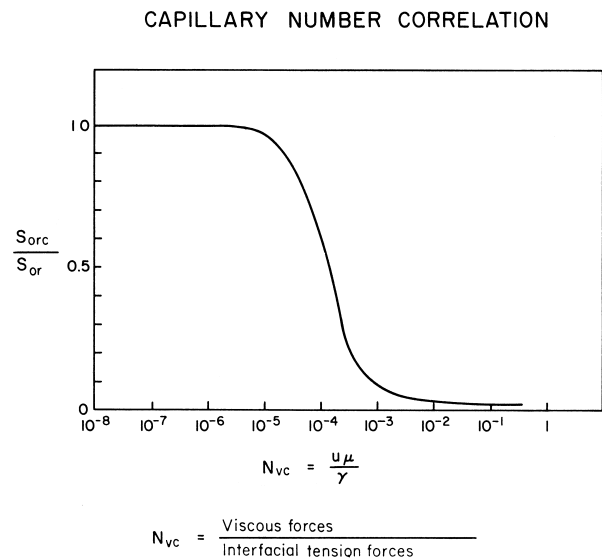
CAPILLARY NUMBER CORRELATION



**FIGURE 2**  Capillary number correlation.

increases as the viscous forces increase or as the interfacial tension forces decrease.

## III. MISCIBLE FLOODING

### A. Introduction

In the previous section, it was noted that the microscopic displacement efficiency is largely a function of interfacial forces acting between the oil, rock, and displacing fluid. If the interfacial tension between the trapped oil and the displacing fluid could be lowered to $10^{-2}$ to $10^{-3}$ dyn/cm, the oil droplets could be deformed and could squeeze through the pore constrictions. A miscible process is one in which the interfacial tension is zero; that is, the displacing fluid and the residual oil mix to form one phase. If the interfacial tension is zero, then the capillary number $N_{vc}$ becomes infinite and the microscopic displacement efficiency is maximized.

Figure 3 is a schematic of a miscible process. Fluid A is injected into the formation and mixes with the crude oil, which forms an oil bank. A mixing zone develops between fluid A and the oil bank and will grow due to dispersion. Fluid A is followed by fluid B, which is miscible with fluid A but not generally miscible with the oil and which is much cheaper than fluid A. A mixing zone will also be created at the fluid A–fluid B interface. It is important that the amount of fluid A that is injected be large enough that the two mixing zones do not come in contact. If the front of the fluid A–fluid B mixing zone reaches the rear of the fluid A–oil mixing zone, viscous fingering of fluid B through the oil could occur. Nevertheless, the volume of fluid A must be kept small to avoid large injected-chemical costs.

Consider a miscible process with $n$-decane as the residual oil, propane as fluid A, and methane as fluid B. The
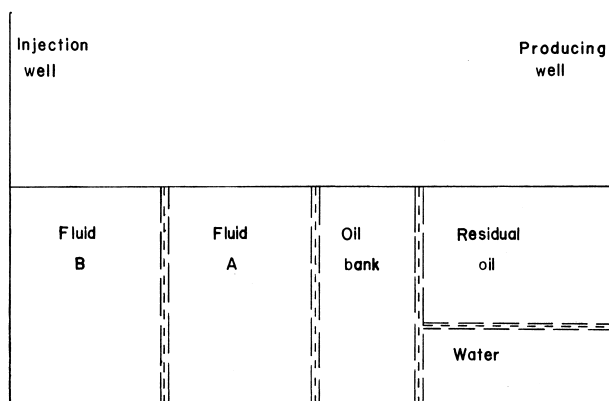


**FIGURE 3** Schematic of an enhanced oil recovery process requiring the injection of two fluids.
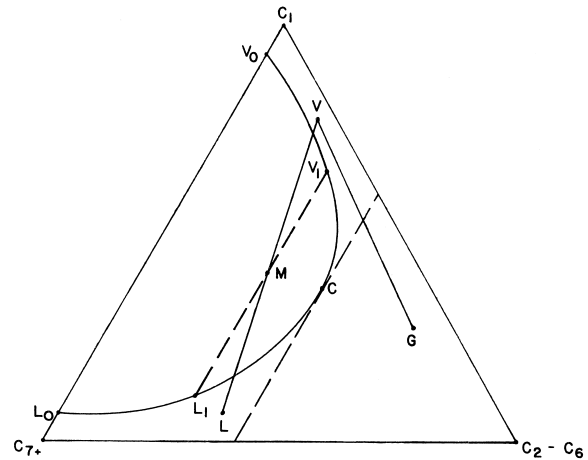


**FIGURE 4** Ternary diagram illustrating typical hydrocarbon phase behavior at constant temperature and pressure.

system pressure and temperature are 2000 pounds per square inch absolute (psia) and 100°F, respectively. At these conditions both the $n$-decane and the propane exist as liquids and are therefore miscible in all proportions. The system temperature and pressure indicate that any mixture of methane and propane would be in the gas state; therefore, the methane and propane would be miscible in all proportions. However, the methane and $n$-decane would not be miscible for similar reasons. If the pressure were reduced to 1000 psia and the temperature held constant, the propane and $n$-decane would again be miscible. However, mixtures of methane and propane could be located in a two-phase region and would not lend themselves to a miscible displacement. Note that in this example the propane appears to act as a liquid when it is in the presence of $n$-decane and as a gas when it is in contact with methane. It is this unique capacity of propane and other intermediate gases that leads to the miscible process.

There are, in general, two types of miscible processes. One is referred to as the single-contact miscible process and involves such injection fluids as liquefied petroleum gases (LPGs) and alcohols. The injected fluids are miscible with residual oil immediately on contact. The second type is the multiple-contact, or dynamic, miscible process. The injected fluids in this case are usually methane, inert fluids, or an enriched methane gas supplemented with a $C_2$–$C_6$ fraction. The injected fluid and oil are usually not miscible on first contact but rely on a process of chemical exchange between phases to achieve miscibility.

### B. Single-Contact Miscible Processes

The phase behavior of hydrocarbon systems can be described with the use of ternary diagrams such as Fig. 4. Researchers have shown that crude oil phase behavior can

be represented reasonably well by three fractions of the crude. One fraction is methane ($C_1$). A second fraction is a mixture of ethane through hexane ($C_2$–$C_6$). The third fraction is the remaining hydrocarbon species lumped together and called $C_{7+}$. Figure 4 illustrates the ternary phase diagram for a typical hydrocarbon system with these three pseudocomponents at the corners of the triangle. There are one-phase and two-phase regions in the diagram. The one-phase region may be vapor or liquid (to the left of the dashed line through the critical point $C$) or gas (to the right of the dashed line through the critical point). A gas could be mixed with either a liquid or a vapor in appropriate percentages and yield a miscible system. However, when liquid is mixed with a vapor, often the result is a composition in the two-phase region. A mixing process is represented on a ternary diagram as a straight line. For example, if compositions $V$ and $G$ are mixed in appropriate proportions, the resulting mixture would fall on the line $VG$. If compositions $V$ and $L$ are mixed, the resulting overall composition $M$ would fall on the line $VL$ but the mixture would yield two phases. If two phases are formed, their compositions, $V_1$ and $L_1$, would be given by a tie line extended through the point $M$ to the phase envelope. The part of the phase boundary on the phase envelope from the critical point $C$ to point $V_0$ is the dew point curve. The phase boundary from $C$ to $L_0$ is the bubble point curve. The entire bubble point–dew point curve is referred to as the binodal curve.

The oil–LPG–dry gas system will be used to illustrate the behavior of the first-contact miscible process on a ternary diagram. Figure 5 is a ternary diagram with the points $O$, $P$, and $V$ representing the oil, LPG, and dry gas, respectively. The oil and LPG are miscible in all proportions. A mixing zone at the oil–LPG interface will grow
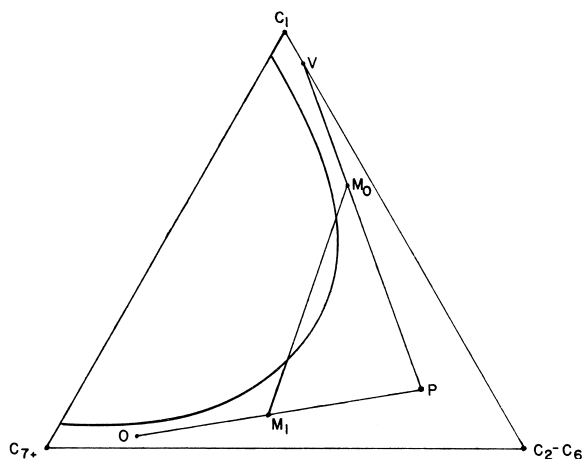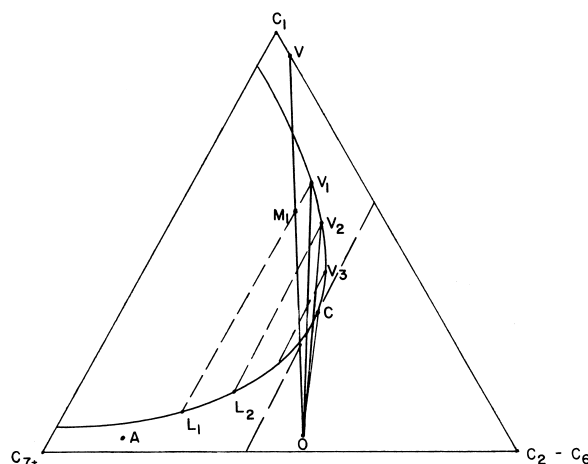


**FIGURE 6** Ternary diagram illustrating the multicontact dry gas miscible process.

as the front advances through the reservoir. At the rear of the LPG slug, the dry gas and LPG are miscible and a mixing zone will also be created at this interface. If the dry gas–LPG mixing zone overtakes the LPG–oil mixing zone, miscibility will be maintained unless the contact of the two zones yields mixtures inside the two-phase region (see line $M_0M_1$, Fig. 5).

Reservoir pressures sufficient to achieve miscibility are required. This limits the application of LPG processes to reservoirs having pressures at least of the order of 1500 psia. Reservoirs with pressures less than this might be amenable to alcohol flooding, another first-contact miscible process, since alcohols tend to be soluble with both oil and water (the drive fluid in this case). The two main problems with alcohols are that they are expensive and they become diluted with connate water during a flooding process, which reduces the miscibility with the oil. Alcohols that have been considered are in the $C_1$–$C_4$ range.

## C. Multiple-Contact Miscible Processes

Multiple-contact, or dynamic, miscible processes do not require the oil and displacing fluid to be miscible immediately on contact but rely on chemical exchange between the two phases for miscibility to be achieved. Figure 6 illustrates the high-pressure (lean-gas) vaporizing process. The temperature and pressure are constant throughout the diagram at reservoir conditions. A vapor denoted by $V$ in Fig. 6, consisting mainly of methane and a small percentage of intermediates, will serve as the injection fluid. The oil composition is given by the point $O$. The following sequence of steps occurs in the development of miscibility:

1. The injection fluid $V$ comes in contact with crude oil $O$. They mix, and the resulting overall composition is



**FIGURE 5** Ternary diagram illustrating the single-contact miscible process.

given by $M_1$. Since $M_1$ is located in the two-phase region, a liquid phase $L_1$ and a vapor phase $V_1$ will form with the compositions given by the intersections of a tie line through $M_1$ with the bubble point and dew point curves, respectively.

2. The liquid $L_1$ has been created from the original oil $O$ by vaporization of some of the light components. Since the oil $O$ was at its residual saturation and was immobile due to $K_{ro}$'s being zero, when a portion of the oil is extracted, the volume, and hence the saturation, will decrease and the oil will remain immobile. The vapor phase, since $K_{rg}$ is greater than zero, will move away from the oil and be displaced downstream.

3. The vapor $V_1$ will come in contact with fresh crude oil $O$, and again the mixing process will occur. The overall composition will yield two phases, $V_2$ and $L_2$. The liquid again remains immobile and the vapor moves downstream, where it comes in contact with more fresh crude.

4. The process is repeated with the vapor phase composition moving along the dew point curve, $V_1$–$V_2$–$V_3$, and so on, until the critical point, $c$, is reached. At this point, the process becomes miscible. In the real case, because of reservoir and fluid property heterogeneities and dispersion, there may be a breaking down and a reestablishment of miscibility.

Behind the miscible front, the vapor phase composition continually changes along the dew point curve. This leads to partial condensing of the vapor phase with the resulting condensate being immobile, but the amount of liquid formed will be quite small. The liquid phase, behind the miscible front, continually changes in composition along the bubble point. When all the extractable components have been removed from the liquid, a small amount of liquid will be left, which will also remain immobile. There will be these two quantities of liquid that will remain immobile and will not be recovered by the miscible process. In practice, operators have reported that the vapor front travels anywhere from 20 to 40 ft from the well bore before miscibility is achieved.

The high-pressure vaporizing process requires a crude oil with significant percentages of intermediate compounds. It is these intermediates that are vaporized and added to the injection fluid to form a vapor that will eventually be miscible with the crude oil. This requirement of intermediates means that the oil composition must lie to the right of a tie line extended through the critical point on the binodal curve (see Fig. 6). A composition lying to the left, such as denoted by point $A$, will not contain sufficient intermediates for miscibility to develop. This is due to the fact that the richest vapor in intermediates that can be formed will be on a tie line extended through point $A$. Clearly, this vapor will not be miscible with crude oil $A$.

As pressure is reduced, the two-phase region increases. It is desirable, of course, to keep the two-phase region minimal in size. As a rule, pressures of the order of 3000 psia or greater and an oil with an American Petroleum Institute (API) gravity greater than 35 are required for miscibility in the high-pressure vaporizing process.

The enriched-gas-condensing process is a second type of dynamic miscible process (Fig. 7). As in the high-pressure vaporizing process, where chemical exchange of intermediates is required for miscibility, miscibility is developed during a process of exchange of intermediates with the injection fluid and the residual oil. In this case, however, the intermediates are condensed from the injection fluid to yield a "new" oil, which becomes miscible with the "old" oil and the injected fluid. The following steps occur in the process (the sequence of steps is similar to those described for the high-pressure vaporizing process but contain some significant differences):

1. An injection fluid $G$ rich in intermediates mixes with residual oil $O$.

2. The mixture, given by the overall composition $M_1$ separates into a vapor phase $V_1$ and a liquid phase $L_1$.

3. The vapor moves ahead of the liquid that remains immobile. The remaining liquid $L_1$ then comes in contact with fresh injection fluid $G$. Another equilibrium occurs, and phases having compositions $V_2$ and $L_2$ are formed.

4. The process is repeated until a liquid is formed from one of the equilibration steps that is miscible with $G$. Miscibility is then said to have been achieved.

Ahead of the miscible front, the oil continually changes in composition along the bubble point curve. In contrast with the high-pressure vaporizing process, there is the potential for no residual oil to be left behind in the reservoir as
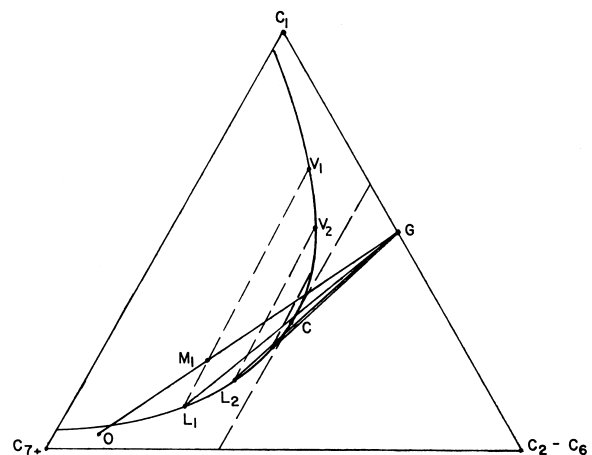


**FIGURE 7**   Ternary diagram illustrating the multicontact enriched-gas-condensing miscible process.

long as there is a sufficient amount of *G* injected to supply the condensing intermediates. The enriched gas process may be applied to reservoirs containing crude oils with only small quantities of intermediates. Reservoir pressures are usually in the range 2000–3000 psia.

The intermediates are expensive, so usually a dry gas is injected after a sufficient slug of enriched gas has been injected.

### D. Inert Gas Injection Processes

The use of inert gases, in particular carbon dioxide, $CO_2$, and nitrogen, $N_2$, as injected fluids in miscible processes has become extremely popular. The ternary diagram representation of the process with $CO_2$ or $N_2$ is exactly the same as that for the high-pressure vaporizing process except that either $CO_2$ or $N_2$ becomes a component and methane is lumped with the intermediates. Typically the one-phase region is largest for $CO_2$, with $N_2$ and dry gas having about the same one-phase size. The larger the one-phase region, the more readily miscibility will be achieved. Miscibility pressures are lower for $CO_2$, usually in the neighborhood of 1200–1500 psia, whereas $N_2$ and dry gas yield much higher miscibility pressures (i.e., 3000 psia or more).

The capacity of $CO_2$ to vaporize hydrocarbons is much greater than that of natural gas. It has been shown that $CO_2$ vaporizes hydrocarbons primarily in the gasoline and gas–oil range. This capacity of $CO_2$ to extract hydrocarbons is the primary reason for the use of $CO_2$ as an oil recovery agent. It is also the reason $CO_2$ requires lower miscibility pressures than those required by natural gas. The presence of other diluent gases such as $N_2$, methane, or flue gas with the $CO_2$ will raise the miscibility pressure. The multiple-contact mechanism works nearly the same with a diluent gas added to the $CO_2$ as it does for pure $CO_2$. Frequently an application of the $CO_2$ process in the field will tolerate higher miscibility pressures than what pure $CO_2$ would require. If this is the case, the operator can dilute the $CO_2$ with other available gas, which will raise the miscibility pressure but also reduce the $CO_2$ requirements.

The pressure at which miscibility is achieved is best determined by conducting a series of displacement experiments in a long, slim tube. A plot of oil recovery versus flooding pressure is made, and the minimum miscibility pressure is determined from the plot. There have been several attempts to correlate miscibility pressures for $CO_2$–oil systems. An early study showed that there is a correlation among the API gravity of the crude, the temperature of the reservoir, and the minimum miscibility pressure. A second study revealed that the molecular weight of the $C_{5+}$ fraction was a better variable to use than the API gravity. In further investigations, it has been found that the minimum miscibility pressure is the

pressure that will achieve a minimum density in the $CO_2$ phase.

### E. Problems in Applying the Miscible Process

Because of differences in density and viscosity between the injected fluid and the reservoir fluid(s), the miscible process often suffers from poor mobility. Viscous fingering and gravity override frequently occur. The simultaneous injection of a miscible agent and a brine was suggested in order to take advantage of the high microscopic displacement efficiency of the miscible process and the high macroscopic displacement efficiency of a waterflood. The improvement was not as good as hoped for since the miscible agent and brine tended to separate due to density differences, with the miscible agent flowing along the top of the porous medium and the brine along the bottom.

Several variations of the simultaneous injection scheme have been suggested and researched. They typically involve the injection of a miscible agent followed by brine or the alternating of miscible agent–brine injection. The latter variation has been named the WAG (water alternate gas) process and has become the most popular. A balance between amounts of injected water and gas must be achieved. Too much gas will lead to viscous fingering and gravity override of the gas, whereas too much water could lead to the trapping of reservoir oil by the water. The addition of foam-generating substances to the brine phase has been suggested as a way to aid in reducing the mobility of the gas phase. Research is continuing in this area.

Operational problems involving miscible processes include transportation of the miscible flooding agent, corrosion of equipment and tubing, and separation and recycling of the miscible flooding agent.

## IV. CHEMICAL FLOODING

### A. Introduction

Chemical flooding relies on the addition of one or more chemical compounds to an injected fluid either to reduce the interfacial tension between the reservoir oil and the injected fluid or to improve the sweep efficiency of the injected fluid.

There are three general methods in chemical flooding technology. The first is polymer flooding, in which a large macromolecule is used to increase the displacing fluid viscosity. This leads to improved sweep efficiency in the reservoir. The second and third methods, micellar–polymer and alkaline flooding, make use of chemicals that reduce the interfacial tension between an oil and a displacing fluid.

## B. Polymer Processes

The addition of large-molecular-weight molecules called polymers to an injected water can often increase the effectiveness of a conventional waterflood. Polymers are usually added to the water in concentrations ranging from 250 to 2000 parts per million (ppm). A polymer solution is more viscous than a brine without polymer. In a flooding application, the increased viscosity will alter the mobility ratio between the injected fluid and the reservoir oil. The improved mobility ratio will lead to better vertical and areal sweep efficiencies and thus higher oil recoveries. Polymers have also been used to alter gross permeability variations in some reservoirs. In this application, polymers form a gel-like material by cross-linking with other chemical species. The polymer gel sets up in large permeability streaks and diverts the flow of any injected fluid to a different location.

Two general types of polymers have been used. These are synthetically produced polyacrylamides and biologically produced polysaccharides. Polyacrylamides are long molecules with a small effective diameter. Thus, they are susceptible to mechanical shear. High rates of flow through valves will sometimes break the polymer into smaller entities and reduce the viscosity of the solution. A reduction in viscosity can also occur as the polymer solution tries to squeeze through the pore openings on the sand face of the injection well. A carefully designed injection scheme is necessary. Polyacrylamides are also sensitive to salt. Large salt concentrations (i.e., greater than 1–2 wt %) tend to make the polymer molecules curl up and lose their viscosity-building effect.

Polysaccharides are less susceptible to both mechanical shear and salt. Since they are produced biologically, care must be taken to prevent biological degradation in the reservoir. As a rule, polysaccharides are more expensive than polyacrylamides.

Polymer flooding has not been successful in high-temperature reservoirs. Neither polymer type has exhibited sufficiently long-term stability above 160°F in moderate-salinity or heavy-salinity brines.

Polymer flooding has the best application in moderately heterogeneous reservoirs and reservoirs containing oils with viscosities less than 100 centipoise (cP). In the United States, there has been a significant increase in the number of active polymer projects since 1978. The projects involve reservoirs having widely differing properties, that is, permeabilities ranging from 20 to 2000 millidarcies (mD), *in situ* oil viscosities of up to 100 cP, and reservoir temperatures of up to 200°F.

Since the use of polymers does not affect the microscopic displacement efficiency, the improvement in oil recovery will be due to an improved sweep efficiency over

what is obtained during a conventional waterflood. Typical oil recoveries from polymer flooding applications are in the range of 1–5% of the initial oil in place. It has been found that operators are more likely to have a successful polymer flood if they start the process early in the producing life of the reservoir.

## C. Micellar–Polymer Processes

The micellar–polymer process uses a surfactant to lower the interfacial tension between the injected fluid and the reservoir oil. A surfactant is a surface-active agent that contains a hydrophobic ("dislikes" water) part to the molecule and a hydrophilic ("likes" water) part. The surfactant migrates to the interface between the oil and water phases and helps make the two phases more miscible. Interfacial tensions can be reduced from ∼30 dyn/cm, found in typical waterflooding applications, to $10^{-4}$ dyn/cm with the addition of as little as 0.1–5.0 wt % surfactant to water–oil systems. Soaps and detergents used in the cleaning industry are surfactants. The same principles involved in washing soiled linen or greasy hands are used in "washing" residual oil off rock formations. As the interfacial tension between an oil phase and a water phase is reduced, the capacity of the aqueous phase to displace the trapped oil phase from the pores of the rock matrix increases. The reduction of interfacial tension results in a shifting of the relative permeability curves such that the oil will flow more readily at lower oil saturations.

When surfactants are mixed above a critical saturation in a water–oil system, the result is a stable mixture called a micellar solution. The micellar solution is made up of structures called microemulsions, which are homogeneous, transparent, and stable to phase separation. They can exist in several shapes, which depend on the concentrations of surfactant, oil, water, and other constituents. Spherical microemulsions have typical size ranges from $10^{-6}$ to $10^{-4}$ mm. A microemulsion consists of external and internal phases sandwiched around one or more layers of surfactant molecules. The external phase can be either aqueous or hydrocarbon in nature, as can the internal phase.

Solutions of microemulsions are known by several other names in the petroleum literature. These include surfactant solutions and soluble oils.

Figure 3 can be used to represent the micellar–polymer process. A certain volume of the micellar or surfactant solution fluid A is injected into the reservoir. The surfactant solution is then followed by a polymer solution, fluid B, to provide a mobility buffer between the surfactant solution and a drive water, which is used to push the entire system through the reservoir. The polymer solution is designed to prevent viscous fingering of the drive

water through the surfactant solution as it starts to build up an oil bank ahead of it. As the surfactant solution moves through the reservoir, surfactant molecules are retained on the rock surface due to the process of adsorption. Often a preflush is injected ahead of the surfactant to precondition the reservoir and reduce the loss of surfactants to adsorption. This preflush contains sacrificial agents such as sodium tripolyphosphate.

There are, in general, two types of micellar–polymer processes. The first uses a low-concentration surfactant solution ($<2.5$ wt %) but a large injected volume (up to 50% pore volume). The second involves a high-concentration surfactant solution (5–12 wt %) and a small injected volume (5–15% pore volume). Either type of process has the potential of achieving low interfacial tensions with a wide variety of brine–crude oil systems. Both have been used in pilot field trials with moderate technical success.

Whether the low-concentration or the high-concentration system is selected, the system is made up of several components. The multicomponent facet leads to an optimization problem, since many different combinations could be chosen. Because of this, a detailed laboratory screening procedure is usually undertaken. The screening procedure typically involves three types of tests: (1) phase behavior studies, (2) interfacial tension studies, and (3) oil displacement studies.

Phase behavior studies are typically conducted in small (up to 100 mL) vials in order to determine what type, if any, of microemulsion is formed with a given micellar–crude oil system. The salinity of the micellar solution is usually varied around the salt concentration of the field brine where the process will be applied. Besides the microemulsion type, other factors examined could be oil uptake into the microemulsion, ease with which the oil and aqueous phases mix, viscosity of the microemulsion, and phase stability of the microemulsion.

Interfacial tension studies are conducted with various concentrations of micellar solution components to determine optimal concentration ranges. Measurements are usually made with the spinning drop, pendent drop, or sessile drop techniques.

The oil displacement studies are the final step in the screening procedure. They are usually conducted in two or more types of porous media. Often initial screening experiments are conducted in unconsolidated sand packs and then in Berea sandstone. The last step in the sequence is to conduct the oil displacement experiments in actual cored samples of reservoir rock. Frequently, actual core samples are placed end to end in order to obtain a core of reasonable length since the individual core samples are typically only 5–7 in. long.

If the oil recoveries from the oil displacement tests warrant further study of the process, the next step is

usually a small field pilot study involving anywhere from 1 to 10 acres.

The micellar–polymer process has been applied in several pilot projects and one large field-scale project. The results have not been very encouraging. The process has demonstrated that it can be a technical success, but the economics of the process has been either marginal or poor in nearly every application.

## D. Alkaline Processes

When an alkaline solution is mixed with certain crude oils, surfactant molecules are formed. When the formation of surfactant molecules occurs *in situ*, the interfacial tension between the brine and oil phases could be reduced. The reduction of interfacial tension causes the microscopic displacement efficiency to increase, which thereby increases oil recovery.

Alkaline substances that have been used include sodium hydroxide, sodium orthosilicate, sodium metasilicate, sodium carbonate, ammonia, and ammonium hydroxide. Sodium hydroxide has been the most popular. Sodium orthosilicate has some advantages in brines with a high divalent ion content.

There are optimum concentrations of alkaline and salt and optimum pH where the interfacial tension values experience a minimum. Finding these requires a screening procedure similar to the one discussed above for the micellar–polymer process. When the interfacial tension is lowered to a point where the capillary number is greater than $10^{-5}$, oil can be mobilized and displaced.

Several mechanisms have been identified that aid oil recovery in the alkaline process. These include the following: lowering of interfacial tension, emulsification of oil, and wettability changes in the rock formation. All three mechanisms can affect the microscopic displacement efficiency, and emulsification can also affect the macroscopic displacement efficiency. If a wettability change is desired, a high (2.0–5.0 wt %) concentration of alkaline should be used. Otherwise, concentrations of the order of 0.5–2.0 wt % of alkaline are used.

The emulsification mechanism has been suggested to work by either of two methods. The first is by forming an emulsion, which becomes mobile and later trapped in downstream pores. The emulsion "blocks" the pores, which thereby diverts flow and increases sweep efficiency. The second mechanism is by again forming an emulsion, which becomes mobile and carries oil droplets that it has entrained to downstream production sites.

The wettability changes that sometimes occur with the use of alkaline affect relative permeability characteristics, which in turn affect mobility and sweep efficiencies.

Mobility control is an important consideration in the alkaline process as it is in all EOR processes. Often, it is

necessary to include polymer in the alkaline solution in order to reduce the tendency of viscous fingering to take place.

Not all crude oils are amenable to alkaline flooding. The surfactant molecules are formed with the heavier, acidic components of the crude oil. Tests have been designed to determine the susceptibility of a given crude oil to alkaline flooding. One of these tests involves titrating the oil with potassium hydroxide (KOH). An acid number is found by determining the number of milligrams of KOH required to neutralize 1 g of oil. The higher the acid number, the more reactive the oil will be and the more readily it will form surfactants. An acid number larger than ~0.2 mg KOH suggests a potential for alkaline flooding.

In general, alkaline projects have been inexpensive to conduct. However, recoveries have not been large in the past field pilots.

### E. Problems in Applying Chemical Processes

The main technical problems associated with chemical processes include the following: (1) screening chemicals to optimize the microscopic displacement efficiency, (2) making contact with the oil in the reservoir, and (3) maintaining good mobility in order to lessen the effects of viscous fingering. The requirements for screening of chemicals vary with the type of process. Obviously, as the number of components increases, the more complicated the screening procedure becomes. The chemicals must also be able to tolerate the environment they are placed in. High temperature and salinity may limit the chemicals that could be used.

The major problem experienced in the field to date in chemical flooding processes has been the inability to make contact with residual oil. Laboratory screening procedures have developed micellar–polymer systems that have displacement efficiencies approaching 100% when sand packs or uniform consolidated sandstones are used as the porous medium. When the same micellar–polymer system is applied in an actual reservoir rock sample, however, the efficiencies are usually lowered significantly. This is due to the heterogeneities in the reservoir samples. When the process is applied to the reservoir, the efficiencies become even worse. Research is being conducted on methods to reduce the effect of the rock heterogeneities and to improve the displacement efficiencies.

Mobility research is also being conducted to improve displacement sweep efficiencies. If good mobility is not maintained, the displacing fluid front will not be effective in making contact with residual oil.

Operational problems involve treating the water used to make up the chemical systems, mixing the chemicals to maintain proper chemical compositions, plugging the formation with particular chemicals such as polymers,

dealing with the consumption of chemicals due to adsorption and mechanical shear and other processing steps, and creating emulsions in the production facilities.

## V. THERMAL FLOODING

### A. Introduction

Primary and secondary production from reservoirs containing heavy, low-gravity crude oils is usually a small fraction of the initial oil in place. This is due to the fact that these types of oils are very thick and viscous and as a result do not migrate readily to producing wells. Figure 8 shows a typical relationship between the viscosity of a heavy, viscous crude oil and temperature. As can be seen, for certain crude oils, viscosities decrease by orders of magnitude with an increase in temperature of 100–200°F. This suggests that if the temperature of a crude oil in the reservoir can be raised by 100–200°F over the normal reservoir temperature, the oil viscosity will be reduced significantly and will flow much more easily to a producing well. The temperature of a reservoir can be raised by injecting a hot fluid or by generating thermal energy *in situ* by combusting the oil. Hot water or steam can be injected as the hot fluid. Three types of processes will be discussed in this section: steam cycling, steam drive, and *in situ* combustion. In addition to the lowering of the crude oil viscosity, there are other mechanisms by which these three processes recover oil. These mechanisms will also be discussed.

Most of the oil that has been produced by EOR methods to date has been as a result of thermal processes. There is
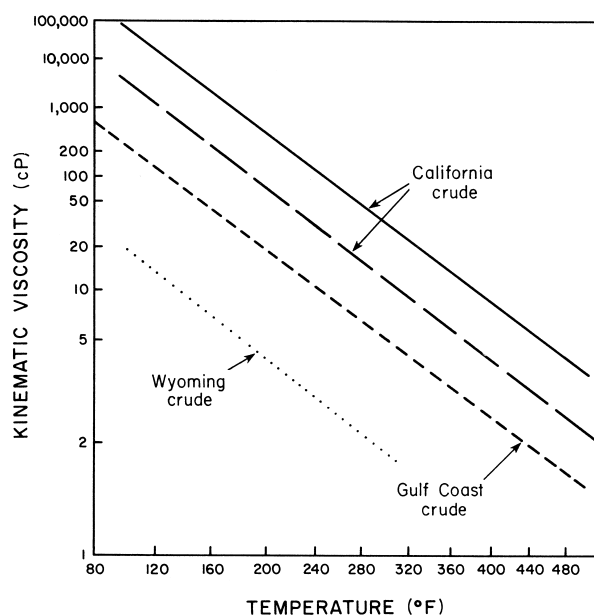
**FIGURE 8** Typical viscosity–temperature relationships for several crude oils.

a practical reason for this, as well as several technical reasons. In order to produce more than 1–2% of the initial oil in place from a heavy-oil reservoir, operators had to employ thermal methods. As a result, thermal methods were investigated much earlier than either miscible or chemical methods, and the resulting technology was developed much more rapidly.

## B. Steam Stimulation Processes

The steam stimulation process was discovered by accident in the Mene Grande Tar Sands, Venezuela, in 1959. During a steam injection trial, it was decided to relieve the pressure from the injection well by backflowing the well. When this was done, a very high oil production rate was observed. Since this discovery, many fields have been placed on steam stimulation.

The steam stimulation process, also known as the steam huff and puff, steam soak, or cyclic steam injection, begins with the injection of 5000–15,000 bbl of high-quality steam. This can take a period of days to weeks to accomplish. The well is then shut in, and the steam is allowed to soak the area around the injection well. This soak period is fairly short, usually from 1 to 5 days. The injection well is then placed on production. The length of the production period is dictated by the oil production rate but can last from several months to a year or more. The cycle is repeated as many times as is economically feasible. The oil production will decrease with each new cycle.

Mechanisms of oil recovery due to this process include (1) reduction of flow resistance near the well bore by reducing the crude oil viscosity and (2) enhancement of the solution gas drive mechanism by decreasing the gas solubility in an oil as temperature increases.

Often, in heavy-oil reservoirs, the steam stimulation process is applied to develop injectivity around an injection well. Once injectivity has been established, the steam stimulation process is converted to a continuous steam drive process.

The oil recoveries obtained from steam stimulation processes are much smaller than the oil recoveries that could be obtained from a steam drive. However, it should be apparent that the steam stimulation process is much less expensive to operate. The cyclic steam stimulation process is the most common thermal recovery technique. Recoveries of additional oil have ranged from 0.21 to 5.0 bbl of oil per barrel of steam injected.

## C. Steam Drive Processes

The steam drive process is much like a conventional waterflood. Once a pattern arrangement is established, steam is injected into several injection wells while the oil is produced from other wells. This is different from the steam stimulation process, whereby the oil is produced from the same well into which the steam is injected. As the steam is injected into the formation, the thermal energy is used to heat the reservoir oil. Unfortunately, the energy also heats the entire environment such as formation rock and water. Some energy is also lost to the underburden and overburden. Once the oil viscosity is reduced by the increased temperature, the oil can flow more readily to the producing wells. The steam moves through the reservoir and comes in contact with cold oil, rock, and water. As the steam comes in contact with the cold environment, it condenses and a hot water bank is formed. This hot water bank acts as a waterflood and pushes additional oil to the producing wells.

Several mechanisms have been identified that are responsible for the production of oil from a steam drive. These include thermal expansion of the crude oil, viscosity reduction of the crude oil, changes in surface forces as the reservoir temperature increases, and steam distillation of the lighter portions of the crude oil.

Steam applications have been limited to shallow reservoirs because as the steam is injected it loses heat energy in the well bore. If the well is very deep, all the steam will be converted to liquid water. Recently, interest has been shown in downhole steam generation; research to develop an economical system is continuing in this area.

Steam drives have been applied in many pilot and field-scale projects with very good success. Oil recoveries have ranged from 0.3 to 0.6 bbl of oil per barrel of steam injected.

## D. *In Situ* Combustion

Early attempts at *in situ* combustion involved what is referred to as the forward dry combustion process. The crude oil was ignited downhole, and then a stream of air or oxygen-enriched air was injected in the well where the combustion was originated. The flame front was then propagated through the reservoir. Large portions of heat energy were lost to the overburden and underburden with this process. To reduce the heat losses, researchers devised a reverse combustion process. In reverse combustion, the oil is ignited as in forward combustion but the airstream is injected in a different well. The air is then "pushed" through the flame front as the flame front moves in the opposite direction. Researchers found the process to work in the laboratory, but when it was tried in the field on a pilot scale, it was never successful. What they found was that the flame would be shut off because there was no oxygen supply and that where the oxygen was being injected, the oil would self-ignite. The whole process would then revert to a forward combustion process.

When the reverse combustion process failed, a new technique called the forward wet combustion process was introduced. This process begins as a forward dry

combustion does, but once the flame front is established, the oxygen stream is replaced by water. As the water comes in contact with the hot zone left by the combustion front, it flashes to steam, using energy that otherwise would have been wasted. The steam moves through the reservoir and aids the displacement of oil. The wet combustion process has become the primary method of conducting combustion projects.

Not all crude oils are amenable to the combustion process. For the combustion process to function properly, the crude oil has to have enough heavy components to serve as the fuel source for the combustion. Usually this requires an oil of low API gravity. As the heavy components in the oil are combusted, lighter components as well as flue gases are formed. These gases are produced with the oil and raise the effective API gravity of the produced oil.

The number of *in situ* combustion projects has decreased since 1980. Environmental and other operational problems have proved to be more than what some operators want to deal with.

### E. Problems in Applying Thermal Processes

The main technical problems associated with thermal techniques are poor sweep efficiencies, loss of heat energy to unproductive zones underground, and poor injectivity of steam or air. Poor sweep efficiencies are due to the density differences between the injected fluids and the reservoir crude oils. The lighter steam or air tends to rise to the top of the formation and bypass large portions of crude oil. Data have been reported from field projects in which coring operations have revealed significant differences in residual oil saturations in the top and bottom parts of the swept formation. Research is being conducted on methods of reducing the tendency for the injected fluids to override the reservoir oil. Techniques involving foams are being employed.

Large heat losses continue to be associated with thermal processes. The wet combustion process has lowered these losses for the higher-temperature combustion techniques, but the losses are severe enough in many applications to prohibit the combustion process. The losses are not as large with the steam processes because they involve smaller temperatures. The development of a feasible downhole generator will significantly reduce the losses associated with steam injection processes.

The poor injectivity found in thermal processes is largely a result of the nature of the reservoir crudes. Operators have applied fracture technology in connection with the injection of fluids in thermal processes. This has helped in many reservoirs.

Operational problems include the following: the formation of emulsions, the corrosion of injection and production tubing and facilities, and the creation of adverse

effects on the environment. When emulsions are formed with heavy crude oil, they are very difficult to break. Operators need to be prepared for this. In the high-temperature environments created in the combustion processes and when water and stack gases mix in the production wells and facilities, corrosion becomes a serious problem. Special well liners are often required. Stack gases also pose environmental concerns in both steam and combustion applications. Stack gases are formed when steam is generated by either coal- or oil-fired generators and, of course, during the combustion process as the crude is burned.

## VI. MICROBIAL FLOODING

### A. Introduction

Microbial enhanced oil recovery (MEOR) flooding involves the injection of microorganisms that react with reservoir fluids to assist in the production of residual oil. The U.S. National Institute for Petroleum & Energy Research (NIPER) maintains a database of field projects that have used microbial technology. There has been significant research conducted on MEOR, but few pilot projects are currently being conducted. The *Oil and Gas Journal*'s 1998 survey reported just 1 ongoing project in the United States related to this technology out of 199 projects. However, China reported several pilot projects that indicated mild success with MEOR.

There are two general types of MEOR processes—those in which microorganisms react with reservoir fluids to generate surfactants and those in which microorganisms react with reservoir fluids to generate polymers. Both processes are discussed below along with a few concluding comments regarding the problems in applying them. The success of MEOR processes will be highly affected by reservoir characteristics. MEOR systems can be designed for reservoirs that have either a high or a low degree of channeling. Therefore, MEOR applications require a thorough knowledge of the reservoir. Mineral content of the reservoir brine will also affect the growth of microorganisms.

### B. Microbial Processes

Microorganisms can be reacted with reservoir fluids to generate either surfactants or polymers in the reservoir. Once either the surfactant or the polymer is produced, the processes of mobilizing and recovering residual oil become similar to those discussed with regard to chemical flooding.

Most pilot projects have involved near-well treatments of stripper wells in an application of the huff and puff process discussed with regard to thermal flooding. A solution of microorganisms is injected along with a nutrient, usually molasses. When the solution of microorganisms has been designed to react with the oil to form polymers,

the injected solution will enter high-permeability zones and react to form the polymers that will then act as a permeability-reducing agent. When oil is produced during the huff stage, oil from lower permeability zones will be produced. Conversely, the solution of microorganisms can be designed to react with the residual crude oil to form a surfactant. The surfactant lowers the interfacial tension of the brine-water system, which thereby mobilizes the residual oil. The oil is then produced in the huff part of the process.

The reaction of the microorganisms with the reservoir fluids may also produce gases, such as $CO_2$, $N_2$, $H_2$, and $CH_4$. The production of these gases will result in an increase in reservoir pressure, which will thereby enhance the reservoir energy.

### C. Problems in Applying Microbial Processes

Since microorganisms can be reacted to form either polymers or surfactants, a knowledge of the reservoir characteristics is critical. If the reservoir is fairly Heterogeneous, then it would be desirable to generate polymers *in situ* that could be used to divert fluid flow from high-to low-permeability channels. If the reservoir has low injectivity, then using microorganisms that produced polymers could be very damaging to the flow of fluids near the well bore. Hence, a thorough knowledge of the reservoir characteristics, particularly those immediately around the well bore, is extremely important.

Reservoir brines can inhibit the growth of the microorganisms. Therefore, some simple compatibility tests can result in useful information as to the viability of the process. These can be simple test-tube experiments in which reservoir fluids and/or rock are placed in microorganism–nutrient solutions and growth and metabolite production of the microorganisms are monitored.

MEOR processes have been applied in reservoirs with brines up to less than 100,000 ppm, rock permeabilities greater than 75 mD, and depths less than 6800 ft. This depth corresponds to a temperature of about 75°C. Most MEOR projects have been performed with light crude oils having API gravities between 30 and 40. These should be considered "rule of thumb" criteria. The most important consideration in selecting a microorganism–reservoir system is to conduct compatibility tests to make sure that microorganism growth can be achieved.

## VII. SCREENING CRITERIA FOR EOR PROCESSES

### A. Introduction

A large number of variables are associated with a given oil reservoir, for instance, pressure and temperature, crude oil type and viscosity, and the nature of the rock matrix and connate water. Because of these variables, not every type of EOR process can be applied to every reservoir. An initial screening procedure would quickly eliminate some EOR processes from consideration in particular reservoir applications. This screening procedure involves the analysis of both crude oil and reservoir properties. This section presents screening criteria for each of the general types of processes previously discussed in this article except microbial flooding. (A discussion of MEOR screening criteria was presented in Section VI). It should be recognized that these screening criteria are only guidelines. If a particular reservoir–crude oil application appears to be on a borderline between two different processes, it may be necessary to consider both processes. Once the number of processes has been reduced to one or two, a detailed economic analysis will have to be conducted.

Some general considerations can be discussed before the individual process screening criteria are presented. First, detailed geological study is usually desirable, since operators have found that unexpected reservoir heterogeneities have led to the failure of many EOR field projects. Reservoirs that are found to be highly faulted or fractured typically yield poor recoveries from EOR processes. Second, some general comments pertaining to economics can be made. When an operator is considering EOR in particular applications, candidate reservoirs should contain sufficient recoverable oil and be large enough for the project to be potentially profitable. Also, deep reservoirs could involve large drilling and completion expenses if new wells are to be drilled.

### B. Screening Criteria

Table II contains the screening criteria that have been compiled from the literature for the miscible, chemical, and thermal techniques.

The miscible process requirements are characterized by a low-viscosity crude oil and a thin reservoir. A low-viscosity oil will usually contain enough of the intermediate-range components for the multicontact miscible process to be established. The requirement of a thin reservoir reduces the possibility that gravity override will occur and yields a more even sweep efficiency.

In general, the chemical processes require reservoir temperatures of less than 200°F, a sandstone reservoir, and enough permeability to allow sufficient injectivity. The chemical processes will work on oils that are more viscous than what the miscible processes require, but the oils cannot be so viscous that adverse mobility ratios are encountered. Limitations are set on temperature and rock type so that chemical consumption can be controlled to reasonable values. High temperatures will degrade most of the chemicals that are currently being used in the industry.

**TABLE II  Screening Criteria for Enhanced Oil Recovery Processes**

| Process | Oil gravity (°API) | Oil viscosity (cP) | Oil saturation (%) | Formation type | Net thickness (ft) | Average permeability (mD) | Depth (ft) | Temp (°F) |
|---|---|---|---|---|---|---|---|---|
| Miscible | | | | | | | | |
| Hydrocarbon | >35 | <10 | >30 | Sandstone or carbonate | 15–25 | —[a] | >4500 | —[a] |
| Carbon dioxide | >25 | <12 | >30 | Sandstone or carbonate | 15–25 | —[a] | >2000 | —[a] |
| Nitrogen | >35 | <10 | >30 | Sandstone or carbonate | 15–25 | —[a] | >4500 | —[a] |
| Chemical | | | | | | | | |
| Polymer | >25 | 5–125 | —[b] | Sandstone preferred | —[a] | >20 | <9000 | <200 |
| Surfactant–polymer | >15 | 20–30 | >30 | Sandstone preferred | >10 | >20 | <9000 | <200 |
| Alkaline | 13–35 | <200 | —[b] | Sandstone preferred | —[a] | >20 | <9000 | <200 |
| Thermal | | | | | | | | |
| Steamflooding | >10 | >20 | >40–50 | Sand or sandstone with high porosity | >10 | >50 | 500–5000 | —[a] |
| Combustion | 10–40 | <1000 | >40–50 | Sand or sandstone with high porosity | >10 | >50 | >500 | —[a] |

[a] Not critical but should be compatible.
[b] Ten percent mobile oil above waterflood residual oil.

In applying the thermal methods, it is critical to have a large oil saturation. This is especially pertinent to the steamflooding process, because much of the produced oil will be used on the surface as the source of fuel to fire the steam generators. In the combustion process, crude oil is used as fuel to compress the airstream on the surface. The reservoir should be of significant thickness in order to minimize heat loss to the surroundings.

## VIII. SUMMARY

The recovery of nearly two-thirds of all the oil that has been discovered to date is an attractive target for EOR processes. The application of EOR technology to existing fields could significantly increase the world's proven reserves. Several technical improvements will have to be made, however, before EOR processes are widely implemented. The current economic climate will also have to improve because many of the processes are either marginally economical or not economical at all. Steamflooding and polymer processes are currently economically viable, with produced oil prices per barrel within a range of $20 to $28. In comparison, the $CO_2$ process is more costly, $26–$39. The micellar–polymer process is even more expensive, at $35–$46. (These prices are in 1985 U.S. dollars.) An aggressive research program is needed to assist in making EOR processes more technologically and economically sound.

Enhanced oil recovery technology should be considered early in the producing life of a reservoir. Many of the processes depend on the establishment of an oil bank in order for the process to be successful. When oil saturations are high, the oil bank in easier to form. It is crucial for engineers to understand the potential of EOR and the way EOR can be applied to a particular reservoir.

## SEE ALSO THE FOLLOWING ARTICLES

FLUID DYNAMICS • MINING ENGINEERING • PETROLEUM GEOLOGY • PETROLEUM REFINING • POLYMER PROCESSING • ROCK MECHANICS

## BIBLIOGRAPHY

Bryant, R. S. (1991). "MEOR screening criteria fit 27% of U.S. oil reservoirs," *Oil and Gas J.* (Apr. 15).

"EOR oil production up slightly," (1998). *Oil and Gas J.* (Apr. 20).

Gogerty, W. B. (1983). *J. Petrol. Technol.* (Sept. and Oct., in two parts).

Lake, L. W. (1989). "Enhanced Oil Recovery," Prentice-Hall, Englewood Cliffs, N.J.

National Petroleum Council (1984). "Enhanced Oil Recovery," U.S. Govt. Printing Office, Washington, D.C.

Prats, M. (1982). "Thermal Recovery" (Soc. Pet. Eng. Monograph Ser.), Vol. 7, Soc. Pet. Eng. of AIME, Dallas.

Stalkup, F. I. (1983). "Miscible Displacement" (Soc. Pet. Eng. Monograph Ser.), Vol. 8, Soc. Pet. Eng. of AIME, Dallas.

van Poolen, H. K., and Associates, Inc. (1980). "Enhanced Oil Recovery," PennWell Books, Tulsa, Okla.

# Fossil Fuel Power Stations— Coal Utilization

## L. Douglas Smoot
## Larry L. Baxter

*Brigham Young University*

## GLOSSARY

**Ash** The oxidized product of inorganic material found in coal.

**ASTM** American Society for Testing and Materials, one of several organizations promulgating standards for fuel analysis and classification.

**BACT** Best available control technology.

**CAAA** Clean Air Act amendments of 1990 in the United States, which introduced substantial new controls for nitrogen oxides, sulfur-containing compounds, and, potentially, many metals emissions by power plants.

**Char** Residual solid from coal particle devolatilization.

**Char oxidation** Surface reaction of char with oxygen.

**Clean Air Act** A federal law governing air quality.

**Coal devolatilization** Thermal decomposition of coal to release volatiles and tars.

**Coal rank** A systematic method to classify coals based on measurement of volatile matter, fixed carbon and heating value.

**Comprehensive code** Computerized numerical model for describing flow and combustion processes.

**Criteria pollutants** Pollutants for which federal law governs emissions; (nitrogen oxides, sulfur-containing compounds, CO, and particulates are the most significant criteria pollutants for coal-based power systems).

**daf** Dry, ash-free, a commonly cited basis for many coal analyses.

**Fouling and slagging** Accumulation of residual ash on inner surfaces of power plant boilers, more generically referred to as ash deposition.

**Greenhouse gases** Gases (e.g., $CO_2$, $CH_4$, $N_2O$, halocarbons) that contribute to atmospheric warming.

**Inorganic material** The portion of coal that contributes

to ash or is not a fundamental portion of the organic matrix.

**ISO** International Organization for Standards, an international association of national/regional standards organizations that publishes standards for fuel characterization and classification.

**Mineral matter** The portion of the inorganic material in coal that occurs in the form of identifiable mineral structures.

**NAAQS** National Ambient Air Quality Standard.

**NSPS** New source performance standard; regulatory standards applied to new power stations.

**PSD** Prevention of Significant Deterioration.

**SCR** Selective catalytic reduction of nitrogen oxides by $NH_3$ over a generally vanadium-based catalytic surface.

**SNCR** Selective noncatalytic conversion of nitrogen oxides by $NH_3$ via gas-phase reactions.

**COAL** is the world's most abundant fossil fuel and is the energy source most commonly used for generation of electrical energy. World use of electrical energy represents about 35% of the total world consumption of energy and this share is increasing. Even so, coal presents formidable challenges to the environment. This article outlines the commercial technology used for generation of electricity and mentions developing technologies. Properties of coals and their reaction processes are discussed and advantages and challenges in the use of coal are enumerated. New analysis methods for coal-fired electrical power generators are noted and future directions are indicated.

## I. COAL USE IN POWER STATIONS

Most of the world's coal production of over 5 billion short tons a year is used to help generate nearly 14 trillion kWhr of electricity, with especially significant use in the United States, China, Japan, Russia, Canada, and Germany. Coal generates about 56% of the world's electrical power.

World generation of electricity is growing at a more rapid rate than total world energy consumption. The use of fossil fuels (i.e., coal, natural gas, liquified natural gas, petroleum) dominates total energy and electric power production, accounting for 85–86% of the total for both the United States and the world. Petroleum provides the largest fraction of the U.S. (39%) and the world (37%) energy requirements, while coal is the most commonly used fuel for electrical power generation in the United States (52% of total kWhr) and the world. Approximately 87% of all coal consumed in the United States is for gener-

ation of electricity. The U.S. production of coal has increased by 55% in the past 20 years, during the time that U.S. production of natural gas and petroleum has held steady or declined. During this same period, world coal production has increased 39% (with little increase in the past decade). Proven world recoverable reserves of coal represent 72% of all fossil fuel reserves, with petroleum and natural gas each representing about 14% of the total reserve. The world depends increasingly on electrical energy, while coal represents the most common fuel of choice for electric power generation and the world's most abundant fossil fuel reserve. With electricity generation from nuclear energy currently declining and with hydroelectric power increasing more slowly than electric power from fossil fuels, coal may continue to be the world's primary energy source for electricity generation. Yet coal is not without substantial technical challenges, which include mining, pollutant emissions, and global warming concerns.

## II. POWER STATION TECHNOLOGIES

### A. Power Station Systems and Conversion Cycles

Fuel conversion technologies of current commercial significance include cyclone boilers, pulverized coal boilers, combustion grates, and fluidized beds. Similar technologies, generally operated at less extreme conditions of pressure and temperature, provide steam and hot water in regions with high demand for district or process heating. The most important advanced technologies include gasifiers, high-temperature furnaces and fuel cells. Some of the general characteristics of these technologies are given in Table I, and characteristics of advanced systems are shown in Table II. The advantages and challenges of the various technologies occur in part from the thermodynamic cycles involved in these heat engines.

### 1. Carnot Cycle

The most efficient means of converting thermal energy into work or electricity is the Carnot cycle. The Carnot cycle represents an idealized heat engine that is unachievable in practical systems but that sets a useful standard for heat engine performance and comparisons. The Carnot efficiency is given by

$$\eta = 1 - \frac{T_{\text{lo}}}{T_{\text{hi}}}, \tag{1}$$

where $T_{\text{hi}}$ and $T_{\text{lo}}$ are the peak and exhaust temperatures, respectively, of the working fluid (expressed as absolute

**TABLE I   Figures of Merit for Coal-Based Power Generation Systems[a]**

| | Pulverized coal boilers | | | | Fluid beds | | | |
|---|---|---|---|---|---|---|---|---|
| | Total | Wall-fired | Tangential | Cyclones | Total | Circulating | Bubbling | Grate-fired |
| Capital cost ($/kW) | 950 | 950 | 950 | 850 | 1350 | 1450 | 1250 | 1250 |
| Peak furnace temperature (K) | 2000 | 2000 | 2000 | 2400 | 1700 | 1700 | 1700 | 1700 |
| Energy/fuel efficiency (%) | 38/100 | 38/100 | 38/100 | 38/100 | 37/99 | 38/99 | 36/98 | 32/95 |
| Fuel flexibility | Moderate | Moderate | Moderate | Excellent | Excellent | Excellent | Excellent | Excellent |
| Average/max size ($MW_e$) | 250/1000 | 250/1000 | 250/1000 | 600/1000 | 30/250 | 30/250 | 30/250 | 10/150 |
| Pollutant emissions | Moderate | Moderate | Moderate | High | Low | Low | Low | Moderate |
| Installed capacity (%) | | | | | | | | |
|     World | 82 | 43 | 39 | 2 | 8 | 4 | 5 | 1 |
|     North America | 80 | 43 | 38 | 8 | 1 | <1 | <1 | <1 |
|     Western Europe | 80 | 43 | 37 | 0 | 10 | 4 | 6 | 8 |
|     Asia | 84 | 43 | 41 | 0 | 8 | 4 | 4 | 8 |
|     Australia | 88 | 43 | 45 | 0 | 3 | 1 | 2 | 2 |

[a] Values are illustrative of typical trends but vary significantly on case-by-case bases.

temperatures). The strong dependence of efficiency on temperature is clear from this expression.

## 2. Rankine Cycle

Nearly all steam-based power boilers use the Rankine cycle to generate electricity. A simple Rankine cycle is based on pressurizing water with a pump, heating the pressurized water beyond its boiling point to produce superheated steam in the boiler, expanding the steam through a steam turbine, cooling the residual steam to regenerate water, and returning the water to the pump to start through the cycle again. Commercial boilers operate at both supercritical (>221.2 bar) and subcritical pressures, with supercritical boilers exhibiting higher efficiencies and subcritical boilers encountering less severe pressure requirements.

The efficiency of the Rankine cycle can be improved by returning the steam to the boiler after partial steam expansion in a process referred to as *reheat*. Reheat cycles add significant complexity to the turbine, the boiler, and the controls but, at large scale, the increased complexity and cost can be justified by the increase in efficiency of a few percent. Other significant factors affecting overall

**TABLE II   Figures of Merit for Coal-Based Advanced Power Generation Systems**

| | Gasifiers | Fuel cells | High-temperature furnaces |
|---|---|---|---|
| Capital cost ($/kW) | 2000 | 2500 | 1100 |
| Energy/fuel efficiency (%) | 50/85 | 45/100 | 40/100 |
| Fuel flexibility | Moderate | Very poor | Poor |
| Average/max size ($MW_e$) | 10/200 | <0.2/15 | 250/1000 |

efficiency in the Rankine cycle are parasitic losses (soot-blowing, pumps, fans, etc.) and nonidealities in pumps, turbines, etc. In theory, Rankine efficiencies of about 60% are achievable. Actual plant efficiencies for coal-fired systems range from approximately 40% (on a higher heating value or gross calorific value basis) for large-scale systems to less than 20% for small-scale systems. Mathematical expressions for the efficiency of the Rankine cycle are more cumbersome than those for the Carnot efficiency, but the strong reliance on peak temperature is common to all of them.

## 3. Brayton (or Joule) Cycle

Essentially all gas turbines are based on the Brayton cycle, which is sometimes referred to as a Joule cycle. In this cycle, fuel and air are pressurized, burned, pass through a gas turbine, and exhausted. The exhaust gases are generally used to preheat the fuel or air. Modern gas turbines operate at significantly higher temperatures than steam turbines. The turbine blades present the most critical materials issues, and blades are thus cooled by internal circulation or transpiration. The Brayton cycle is of little current commercial significance in coal-fired power plants but is a major consideration in the most promising proposed schemes for the next generation of coal-driven heat engines. Gas turbine efficiencies can be as high as 60% and gas turbines enjoy a significant turndown ratio with little loss in efficiency.

## B. Power Station Technologies

The Rankine cycle forms the basis of power generation from nearly all current coal-fired power stations. The

Brayton cycle offers efficiency and other advantages and is incorporated into many future coal-based technologies, generally in combination with the Rankine cycle (so-called combined cycle systems). The Brayton cycle is also the basis of gas- and oil-fired turbine technologies. Some of the distinguishing characteristics of the major coal conversion technologies and the leading advanced technologies are summarized here.

Most of the energy associated with the efficiency losses compared with ideal systems is available as heat, but not as work or electricity. Somewhat over half of the total energy potentially available from fuel is not converted to electricity/work because of these losses. In many countries (including the United States) the available heat from these losses is rarely used for industrial processes or district heating. Recovering this heat would require collocation of power stations with industrial parks or communities and installation of district heating or process heating systems in the communities or collocated industries. These changes are fraught with social and economic challenges. Nevertheless, these energy losses represent major inefficiencies in the overall energy budgets in many countries.

### 1. Pulverized-Coal Boilers

The most widely implemented coal-based technology for power generation is the pulverized-coal (pc) boiler. These boilers provide 88% of the total coal-based electric capacity in the United States and similar large fractions of electric capacity abroad. A typical pulverized coal boiler is shown in Fig. 1. Steps in the process are (1) mining of the coal from near-surface or deep mines, (2) transporting crushed coal several centimeters in size to the power sta-
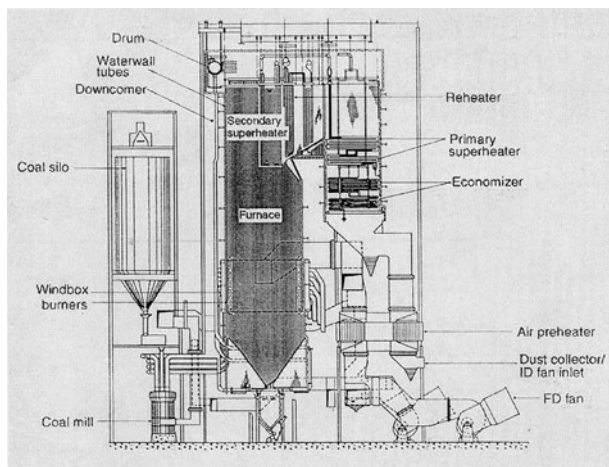
tion, (3) pulverizing the coal to a powder predominantly less than 100 $\mu$m in size, (4) pneumatically transporting this coal dust with combustion air into a large combustion chamber with steel walls cooled by water flowing in wall tubes, (5) ignition and combustion of the coal dust and air to create high temperatures, which heats the pressurized cooling water to high temperatures, (6) expansion of the high-pressure water through large steam turbines to create electricity, (7) cleaning of the combustion off-gases to remove gaseous pollutants, (8) marketing or disposal of the noncombustible ash, and (9) quality control, cooling, and recycling of the water.

Fuel preparation for a pc boiler occurs in a mill, where coal is typically reduced to 70% through a 200-mesh (74-$\mu$m) screen. Figure 2 shows a particle-size distribution from a bowl-mill grinder typically used in power plants. A typical utility specification for particle size distribution of the coal is that 70% of the mass should pass through a 200-mesh screen, which is equivalent to 70% less than 74 $\mu$m. Sometimes the top size is limited to about 120 $\mu$m. The pulverized fuel is pneumatically transported to burner levels (commonly one level fed by each mill) with each level made up of a series of burners.
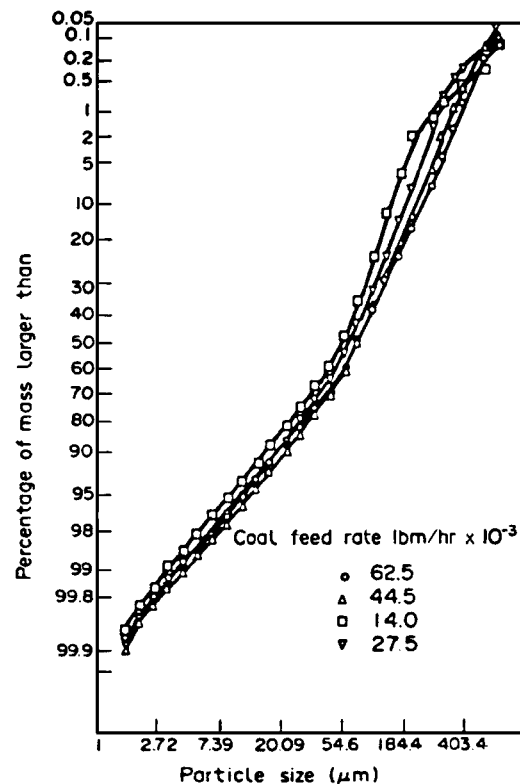


**FIGURE 1** Tangentially fired boiler. [Reproduced with permission from Singer, J. G. (ed.), (1981). "Combustion: Fossil Fuel Systems," 3rd Ed., Combustion Engineering, Windsor, CT.



**FIGURE 2** Measured pulverized-coal size distribution for a full-scale exhauster-type mill operating at a steady state. [Reproduced with permission from Beér, J. M., Chomiak, J., Smoot, L. D. (1984). *Progress Energy Combustion Sci.* **10,** 229–272.

Two basic types of burners define the two major designs of pc boilers: wall and corner burners. Wall burners are generally configured as a series of annular flows, with the pulverized coal entering near the center and air, often swirled, entering through one or more outer flows. Wall burners are located on either the front or the back wall, or both. Corner burners comprise a stack of individual ducts in or near the boiler corners that individually contain air or a coal suspended in air. The direction of the air and coal streams is biased either clockwise or counterclockwise, creating a large vortex that fills the furnace. The burner tilt and yaw (angle made in the horizontal plane) can be adjusted as operational parameters, although in practice the yaw is generally fixed.

The furnace exit is generally defined as the location at which the flue gas enters in-flow heat exchangers. Such heat exchangers are first encountered either in the form of platens hanging in the top of the furnace or in the form of tube bundles over the furnace nose. Beginning at the in-flow heat exchangers, the gas exits the furnace and enters the convection pass. The highest temperature tube bundles are called superheaters, which are the last heat exchangers through which steam passes before entering the turbine. The steam temperature and pressure entering the turbine are two of the most closely controlled set points in a boiler. The steam temperature range is 500–650°C, with 540°C being a common value. Pressures vary from subcritical to supercritical ($>221$ bar), with subcritical boilers being most common and generally requiring less maintenance, and supercritical boilers having higher efficiencies. Steam turbines in large-scale power plants are split into two or three sections. Steam exits the high-pressure turbine after a partial expansion, where it flows through another set of heat exchangers called reheaters before returning to the intermediate-pressure turbine.

### 2. Gasifiers

Coal gasification is a process in which coal is converted to a low-grade gas; it can be regarded in many ways as fuel-rich combustion. Oxygen-blown coal gasifiers operate at overall equivalence ratios (fuel to oxidizer ratio normalized by fuel to oxidizer ratio required to fully oxidize fuel) greater than two. Commercial or near-commercial systems are available that are either air-blown or oxygen-blown. Different designs use suspension firing, slurry firing, or fluid beds. They can be slagging or dry-bottom systems. They operate at either atmospheric or high pressure. The most commercially significant systems for coal are entrained-flow, high-temperature, high-pressure systems. Gasification offers potential environmental and efficiency advantages over conventional coal combustion. Efficiency benefits derive from the potential combustion of the gasi-

fied coal, commonly referred to as producer gas, in a gas turbine. A Brayton cycle for a gas turbine is a more efficient heat engine than the Rankine cycle used in a boiler. If the gases are introduced into the turbine at their peak gasification temperatures and if a conventional Rankine cycle is used to burn the residual char and for the low-grade heat, overall thermal efficiencies as high as 60% (higher-heating-value basis) can be achieved. Gas-phase sulfur forms $H_2S$ in fuel-rich systems, while gas-phase nitrogen forms HCN and $NH_3$, all of which are much easier to scrub from combustion gases than are their fully oxidized counterparts ($SO_2$ and nitrogen oxides, $NO_x$). Cleaning coal-derived gases to the standards demanded by gas turbines (in particular, residual ash, alkali, and sulfur) is a technical challenge.

Other commercial systems include cyclone boilers, fluidized-bed boilers, grate-based furnaces, and fuel cells.

## C. Fuel Considerations

A boiler converts chemical energy stored in fuels into steam that is used to produce electricity or for process use. The theoretical and practical limitations of this process determine the ultimate technology and systems making up the overall power plant. Carbon, hydrogen and oxygen are responsible for the majority of the calorific value of the fuel.

A convenient illustration of the relationships between many fuels is based on the carbon, hydrogen, and oxygen content and is referred to as a coalification diagram when restricted to coal (Fig. 3). Natural gas has a hydrogen-to-carbon ratio of about 3.6 and an oxygen-to-carbon ratio of near 0, but is not conveniently illustrated at the scale of this diagram. As is seen, major solid fuels fall into consistent regions of the diagram.

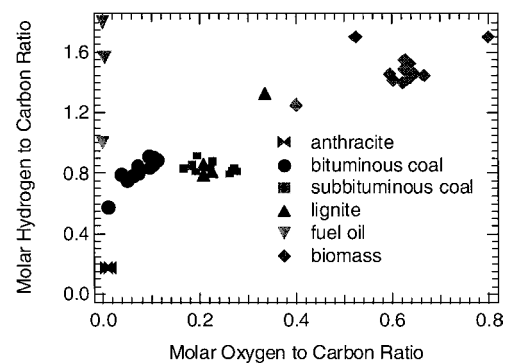The remaining fuel components are typically nitrogen, sulfur, and inorganic impurities. It is ultimately the



**FIGURE 3** Coalification diagram indicating the relationships among a variety of solid fuels in terms of their chemical composition.

impurities in the fuel and process materials that drive engineering decisions regarding power station technologies and operating conditions. The contributions most directly associated with carbon, hydrogen, and oxygen such as peak flame temperature have less influence on performance than idealized theoretical analyses might suggest. Fuel impurities and limitations of current materials place the greatest restrictions on boiler design and operation. With low-grade fuels these restrictions can be substantial.

Coal is commonly classified by rank, with the major classifications being, in order of decreasing rank, anthracite, bituminous coal, subbituminous coal, and lignite. The data in Fig. 3 illustrate how fuel properties vary with rank and how they relate to each other. Coals most commonly used for steam generation have molecular hydrogen-to-carbon ratios ranging from 0.7 to 0.9 and molecular oxygen-to-carbon ratios ranging from near 0 to about 0.2.

While coal rank correlates with oxygen and hydrogen content, coal rank is determined principally by heating value except for the highest rank coals and anthracites, which are distinguished on the basis of fixed carbon contents (according to ASTM standards). Heating value is the amount of energy released from a fuel during combustion and is discussed later in this article.

## III. COAL PROPERTIES AND PREPARATION

### A. Coal Classification

A common means of characterizing the broad range of coals in use is by rank, moisture, inorganic material, and heating value, which ranges from about 9 to 35 MJ/kg (4000 to about 15,000 Btu/lb). Moisture and inorganic material percentages typically range from 1% to 30% and from 5% to over 50%, respectively. Table III summarizes several properties of various coals. The heating values are conveniently cited for scientific purposes on dry, ash-free bases, but for commercial purposes they are typically moist, inorganic-matter-free values. In the United States, coals are ranked primarily on a heating-value basis up to a value of 33 MJ/kg (14,000 Btu/lb). Coals with heating values higher than 33 MJ/kg are ranked based on volatile matter/fixed carbon contents. Coals range in rank from lignites to low-volatile, bituminous coals and anthracites, with the bulk of the coals representing either subbituminous, low-sulfur coals or high-volatile bituminous coals.

Figure 3 illustrates the relationship of anthracite, bituminous and subbituminous coals, lignite, fuel oil, and biomass in terms of the atomic hydrogen-to-carbon (H:C) and oxygen-to-carbon (O:C) ratios. This type of diagram indicates some chemical structure, combustion, and inorganic aspects of coals and other solid fuels. For example, increasing the H:C or O:C ratio implies decreasing aromaticity of the fuel. Increasing the O:C ratio implies increasing hydroxyl, carboxyl, ether, and ketone functional groups in the fuel. Both the aromaticity and the oxygen-containing functional groups influence the modes of occurrence of inorganic material in fuel and inorganic transformations during combustion.

### B. Inorganic Matter

One type of inorganic material in coal is inherent inorganic material that can be volatile at combustion temperatures and includes some of the alkali and alkaline earth metals, most notably sodium and potassium. Anthracites typically lose less than 10% of their mass by pyrolysis. Bituminous and subbituminous coals lose between 5% and 65% of their mass by this process. Lignites, peats, and biomass can lose over 90% of their mass in this first stage of combustion. The large quantities of gases or tars leaving coals, lignites, and biomass fuels can convectively carry inorganic material out of the fuel, even if the inorganic material itself is nonvolatile. The combination of high oxygen content and high organic volatile matter in subbituminous coals indicates a potential for creating large amounts of inorganic vapors during combustion.

The second class of inorganic material in solid fuels includes material that is added to the fuel from extraneous sources. Geologic processes and mining techniques contribute much of this material to coal. This adventitious material is often particulate in nature (in contrast to the atomically dispersed material) and is the dominant contributor to fly ash particles larger than about 10 $\mu$m. Examples include silica, pyrite, calcite, kaolinite, illite, and other silicates.

Typically, a boiler designed for combustion of subbituminous coals is about 40% taller and 20% wider and deeper than a similarly rated boiler designed for bituminous coal. This increase in size is required even though the low-rank fuel combusts notably faster than the high-rank fuel. The primary reason for the increase in size is to accommodate differences in the behavior of the inorganic material associated with these two coals, specifically the properties of the ash deposits that they form. The increased surface area in the boiler for the low-rank fuel allows the operator to manage this ash behavior without compromising boiler lifetime or damaging boiler components. In the past, boilers largely burned coals supplied by local or regional mines under long-term contracts. A combination of environmental pressures and deregulation has created much larger variation in the coals used by most current boilers and has led to widespread fuel blending.

**TABLE III   Major Fuel Properties as Determined by Standardized Analyses for Typical Samples of Each Major Coal Rank**

|  | Beulah | Black Thunder | Pittsburgh #8 | Pocahontas #3 |
|---|---|---|---|---|
| Nominal rank | Lignite | Subbituminous | Bituminous | Low Volatile Bituminous |
| Proximate analysis |  |  |  |  |
| Moisture (as received) | 26.89 | 21.3 | 1.02 | 0.62 |
| Ash (dry basis) | 13.86 | 6.46 | 10.68 | 4.51 |
| Volatile matter (dry basis) | 42.78 | 54.26 | 40.16 | 18.49 |
| Fixed carbon (dry basis) | 43.36 | 39.28 | 49.16 | 77 |
| Ultimate analysis (daf basis) |  |  |  |  |
| Carbon | 70.49 | 74.73 | 80.06 | 91.65 |
| Hydrogen | 4.75 | 5.4 | 5.63 | 4.46 |
| Nitrogen | 1.22 | 1 | 1.39 | 1.31 |
| Sulfur | 2.14 | 0.51 | 5.35 | 0.79 |
| Oxygen (by difference) | 21.36 | 18.27 | 7.57 | 1.62 |
| Chlorine | 0.05 | 0.08 | 0 | 0.17 |
| Ash chemistry/elemental ash (% of ash) |  |  |  |  |
| $SiO_2$ | 21.23 | 30.67 | 41.7 | 37.92 |
| $Al_2O_3$ | 13.97 | 16.46 | 20.66 | 24.16 |
| $TiO_2$ | 0.42 | 1.29 | 0.9 | 1.14 |
| $Fe_2O_3$ | 12.25 | 5.1 | 29.24 | 17.14 |
| CaO | 16.36 | 21.32 | 2.08 | 7.67 |
| MgO | 4.46 | 4.8 | 0.79 | 2.4 |
| $K_2O$ | 0.22 | 0.35 | 1.74 | 1.84 |
| $Na_2O$ | 6.5 | 1.43 | 0.4 | 0.83 |
| $SO_3$ | 24.6 | 17.67 | 2.35 | 6.79 |
| $P_2O_5$ | 0 | 0.92 | 0.15 | 0.1 |
| Heating value (MJ/kg, daf basis) | 23.4 | 29.8 | 32.2 | 35.0 |
| Form of sulfur (% of daf fuel) |  |  |  |  |
| Sulfatic | 0.16 | 0 | 0.15 | 0 |
| Pyritic | 0.47 | 0.08 | 1.81 | 0.21 |
| Organic | 1.18 | 0.4 | 2.82 | 0.54 |
| Ash fusion temperature (°C) |  |  |  |  |
| Reducing atmosphere |  |  |  |  |
| Initial deformation | 774 | 1161 | 1047 | 1191 |
| Hemispherical | 1636 | 1178 | 1082 | 1249 |
| Spherical | 1650 | 1189 | 1179 | 1278 |
| Fluid | 1659 | 1210 | 1222 | 1331 |
| Oxidizing atmosphere |  |  |  |  |
| Initial deformation | 1714 | 1179 | 1337 | 1297 |
| Hemispherical | 1753 | 1200 | 1372 | 1313 |
| Spherical | 1756 | 1209 | 1381 | 1337 |
| Fluid | 1769 | 1243 | 1389 | 1364 |

In the past, long-term experience with a limited number of coals partially compensated for an inability to quantitatively anticipate the behavior of the inorganic material in these coals. Current fuel purchasing trends require much more adept boiler operation.

In general, ash behavior depends on fuel properties, boiler design, and boiler operation. Industrial practice has led to the development of a wide variety of indices, nearly all of which are based on fuel properties, for predicting ash behavior. These include arithmetic combinations of elemental mass fractions in the ash, fusion temperatures, and slag viscosity-related measurements. It is on the basis of these parameters that the coal-fired electrical power industry has developed a large, stable, reliable suite of boilers. However, there is a critical need for improved accuracy, given the combination of the effect of ash

on boiler performance and current market trends leading to increased variation and generally decreased quality of coals used in most boilers.

## C. Standardized Analysis Methods

A large number of coal analyses have been standardized by ASTM, ISO, and similar organizations around the world. The most commonly cited and generally useful among these include energy-content, proximate, ultimate, ash-chemistry, and ash-fusion-temperature analyses. The specific procedures by which these analyses are performed differ slightly from organization to organization, but generally they are more similar than they are different. The following briefly discusses the general objectives of the analyses, some typical results, and a few easily misinterpretable features of these analyses.

### 1. Coal Energy Content and Flame Temperatures

Energy content represents possibly the single most relevant property of a fuel. The standard measure of a fuel's energy content is the heating value or calorific value, which typically ranges from 23 to 35 MJ/kg (10,000–15,000 Btu/lb) on a dry, ash-free basis. Since some low-rank coals contain large quantities of both ash and moisture, as-received values may be as low as 13 MJ/kg (6000 Btu/lb) or sometimes lower in exceptional cases. The practical importance of this parameter for combustion technology is obvious. The heating value is more commonly referred to as calorific value in countries that use SI units in commerce. Two different heating values/calorific values are commonly cited. The higher heating value (gross calorific value) assumes that water forms a liquid product after combustion, whereas the lower heating value (net calorific value) assumes that steam vapor is the final form of the water, the difference between the two being the heat of vaporization of the water formed during combustion and contained in the fuel. Most computations in the United States and Australia are based on higher heating value or gross calorific value, whereas most of the rest of the world bases calculations on net calorific value. Performance statistics such as efficiencies are consequently not directly comparable among various countries.

All standardized heating value analyses measure the heat released from the fuel when burned with enough air to fully oxidize the fuel (carbon forms carbon dioxide, hydrogen forms water, etc.), normalized by the mass of the fuel, *not the mass of all combustion products*. This value determines the amount of fuel required to release a given amount of heat during combustion. It does not directly indicate the peak temperatures of the resulting flames. These temperatures play an important role in some
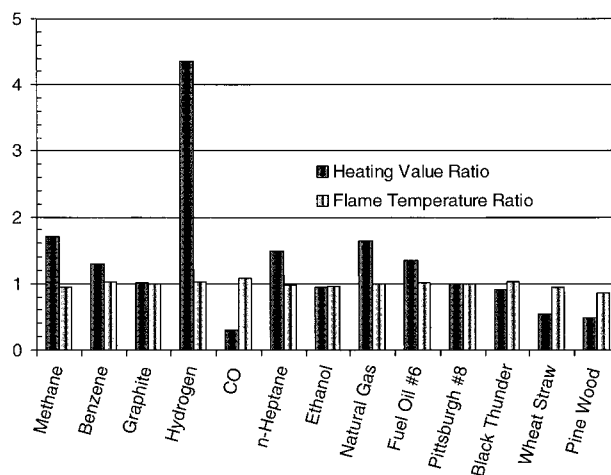


**FIGURE 4** Variation of heating value and adiabatic flame temperature with fuel type. All values are normalized by those for bituminous coal.

theoretical analyses of boiler performance such as the heat engine cycles discussed previously.

A useful figure of merit for most fuels is the adiabatic flame temperature. The adiabatic flame temperature defines the temperature of the products of combustion after all chemical reactions have reached equilibrium and when no heat is allowed to escape (or enter) the combustor. Each fuel has a unique adiabatic flame temperature for a given amount of air. As the ratio of fuel to air is varied, the adiabatic flame temperature varies. The highest value generally occurs at an air-to-fuel ratio with slightly less air than is required to convert all of the carbon of $CO_2$ and hydrogen to $H_2O$.

The adiabatic flame temperature does not correlate well with heating value over a broad range of fuels. Figure 4 illustrates both heating values and adiabatic flame temperatures for a variety of fuels. All values are normalized to those for a high-volatile bituminous coal and all calculations assume dry fuels. The lack of correlation between heating value and flame temperature is clear. A comparison of hydrogen and carbon monoxide vividly illustrates this point. The heating values for these two fuels differ by a factor of about 14, whereas the adiabatic flame temperatures differ by less than 6%, with that of CO exceeding that of hydrogen. The wide disparity in heating values but similar adiabatic flame temperatures arises primarily from the varying oxygen content of the fuels. Fuels that contain significant oxygen require less air to completely burn, so the heat release is distributed within less mass. The net effect of less heat generation and less mass to absorb that heat is a flame temperature that does not significantly vary from fuel to fuel, presuming dry fuels.

Despite the dominant role that flame temperature plays in theoretical analysis of cycle efficiencies, peak flame

temperature has little role in practical systems. The efficiency of virtually all combustion systems is limited ultimately by materials, not flame temperatures. Most dry fuels have peak adiabatic flame temperatures that slightly exceed 2000°C (assuming air as an oxidizer). However, no commercially viable boiler materials can withstand such temperatures as either structural members or pressure parts. Therefore, the peak temperatures achieved in gas and steam turbines are determined by materials constraints, not fuel constraints. Furthermore, simple theoretical cycles such as the Carnot cycle assume ideal heat transfer and fluid flow that are not achievable in practice, further limiting the efficiency of real systems. A consequence of these limitations is that most coal combustion systems convert less than 40% of the energy in the fuel (as measured by gross calorific value or higher heating value) to electricity. The departure from ideal behavior arises, in order of significance, from a lack of suitable high-temperature materials, nonidealities in fluid flow and heat transfer, and process-related losses of energy for auxiliary equipment. Steam turbines typically operate with peak steam temperatures of 500–600°C (900–1100°F). Gas turbines typically operate with peak temperatures as high as 1425°C (2600°F), with the high-temperature limit depending on sophisticated blade-cooling systems. There are many examples of lower performance systems, especially in small systems and those that are optimized for steam rather than power production.

## 2. Proximate and Ultimate Analyses

The proximate analysis indicates the moisture, ash, and volatiles content of a fuel. The volatile yield is sensitive to heating rate and peak temperature. The yield experienced during the relatively slow and cool proximate analysis is lower than that experienced by coal in many combustors by about 30%. However, the relative volatile yields of many coals are approximately indicated by the proximate analysis.

The ultimate analysis apportions the composition of the organic portion of the fuel among the elements C, H, N, and S. The remaining mass is generally assumed to be oxygen. Some details of this analysis are commonly misunderstood with respect to sulfur, chorine, and ash. Ash typically includes some sulfate. The sulfur content reported in the ultimate analysis is most commonly based on a sample heated to 1350°C in an oxidizing environment, well above the proximate ashing temperature of about 750–815°C. Often, 20% or more of the ash from western coals is in the form of sulfates or other forms represented as $SO_3$. When proximate analysis ash is included with the ultimate analysis, the sum of ash, carbon, hydrogen, nitrogen, sulfur, and oxygen will exceed 100%. The excess arises from a portion of the sulfur being counted twice, once in the ash

and once in the ultimate analysis for sulfur. Since oxygen is normally determined by difference, this confusion potentially biases both the sulfur and oxygen values. To avoid such confusion, ash should be included with the proximate analysis and the ultimate analysis should be reported on an ash-free basis. A significant fraction of the sulfur in high-rank U.S. coals derives from pyrite; an accurate estimate of the amount of pyrite in the fuel can be obtained from relatively simple, standardized forms of sulfur analyses.

Chlorine analyses generally are performed separately from ultimate analyses. Essentially all of the chlorine in coals is released during the ultimate analysis process. Corrections to the oxygen content should be made to account for the presence of chlorine in the evolved gases if oxygen is determined by difference. U.S. coals typically have less than 0.15% chlorine, with 1% chlorine being an extreme upper bound. Oxygen content, on the other hand, ranges from 4% to 20% depending strongly on rank. Even small amounts of chlorine can have large impacts on inorganic transformation and deposition. Chlorine is much more significant in some international coals and in many fuels fired in conjunction with coal.

## 3. Ash Chemistry and Ash Fusion Temperatures

The most uniquely suited standardized analyses for ash deposition include ash chemistry and ash fusion temperature. The total ash content from proximate analysis and ash composition provide the fuel information that goes into the majority of common empirical indices of ash behavior, along with ash fusion temperature. The ash chemistry analysis typically reports the ash elemental composition on an oxide basis. This does not mean that all of the species exist as oxides in the fuel (which they do not). It is a convenient method of checking the consistency of the data. The sum of the oxides should be about the same as the total ash content. The analysis is fundamentally an elemental analysis with no distinction of the chemical speciation of the inorganic species.

## D. Advanced Analysis Methods

A variety of advanced techniques exists for determining the species composition of the fuel, which is vital to improved predictive methods for coal, and especially ash behavior, in boilers. X-ray diffraction, performed on low-temperature ash, is the most widely used technique for qualitatively identifying the presence of minerals in their crystalline form in concentrations of a few weight percent or greater. Thermal analytical techniques such as differential thermal analysis (DTA) and thermogravimetric analysis (TGA) have been used as a signature analysis based on changes in physical properties with temperature.

Microanalytical techniques, including scanning electron microscopy (SEM) equipped with energy dispersive microscopy (TEM), resolve particles to submicrometer scales. The techniques may be applied to low-temperature ash as well as raw coal and provide visual images of the morphology of the inorganic structure. Extended X-ray analysis (XANE) is another signature analytical technique sometimes applied to coal and petroleum coke. Mössbauer spectroscopy has been widely used for characterizing the inorganic forms of iron in coal as well as in slags and deposits. Computer-controlled scanning electron microscopy (CCSEM) is a commonly used technique that measures the chemical composition and size of individual inorganic grains in coal particles. It forms the basis of several models of ash deposition during coal combustion. Chemical fractionation is a nonstandardized but relatively accessible technique for characterization of the modes of occurrence of both mineral and nonmineral inorganic material in coal. These analyses, particularly the last two, are becoming increasingly important for predicting ash behavior in boilers. Detailed descriptions of inorganic material in coal can be inferred from chemical fractionation, as illustrated in Fig. 5.

## IV. COAL REACTION PROCESSES

### A. Reaction Types

Coal particle reactions are an essential aspect of all coal combustion processes. These reactions have a direct impact on the formation of fine particles, nitrogen-

**TABLE IV  Typical Process Characteristics for Various Coal-Related Combustion Technologies**[a]

| Process | Particle size ($\mu$m) | Flame temperature (K) | Typical residence time (sec) |
|---|---|---|---|
| Entrained flow | 1–100 | 1900–2000 | <1 |
| Fluidized bed | 1500–6000 | 1000–1450 | 10–500 |
| Stoker/fixed bed | 10,000–50,000 | <2000 | 500–5000 |

[a] Adapted from Table 3.1, *in* Smoot, L. Douglas, and Smith, Philip J. (1985). "Coal Combustion and Gasification," Plenum Press, New York.

and sulfur-containing species, and other pollutants. They also dominate heat transfer and fluid dynamic processes. Table IV summarizes typical residence times, flame temperatures, and particle sizes for three generic process types: entrained flow, fluidized bed, and fixed bed.

A schematic diagram of a reacting coal particle is given in Fig. 6. The diagram suggests that the particle, at any time in the reaction process, may be composed of moisture, (raw) coal, char, inorganic matter, and ash. Ash represents the condensed-phase product of inorganic matter transformations in a manner similar to char representing the condensed-phase product of coal pyrolysis. The particle may be reacting in either oxidizing or reducing gases, depending on boiler design and operation and the location of the particle. Commonly, the center of the particle is in a reducing environment while the surface is oxidizing. Reactions of the organic portion of the coal proceed by two primary processes: devolatilization and char oxidation.
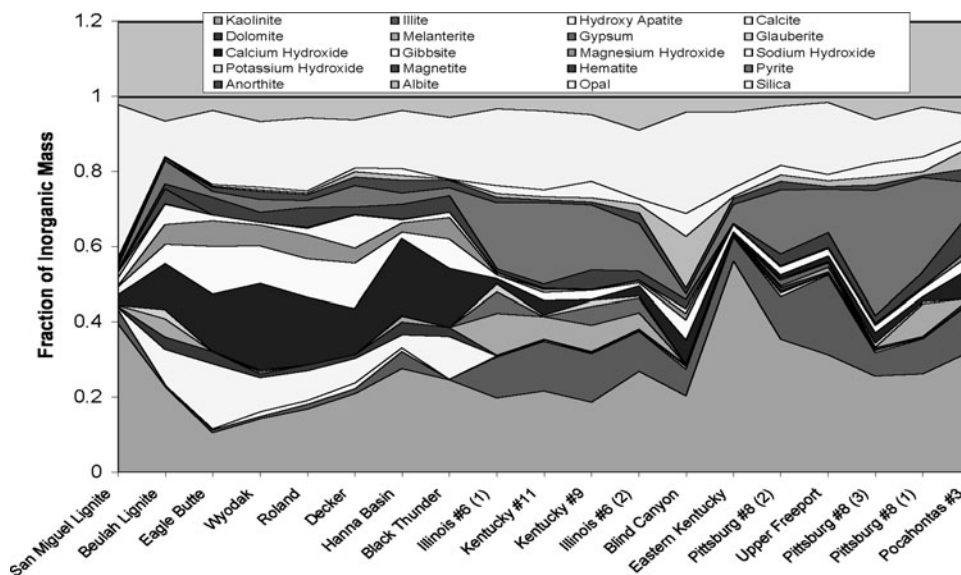


**FIGURE 5**  Major forms of inorganic material in a suite of U.S. coals, arranged in rank order and by mineral class. Species are presented from bottom to top in the order indicated in the legend.
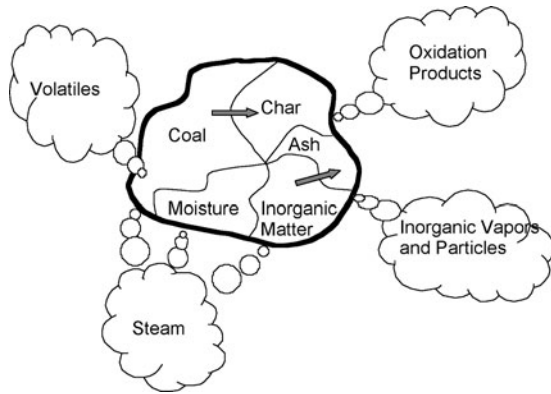
**FIGURE 6** Schematic diagram of coal particle reactions during combustion.



**FIGURE 7** Comparison of predicted and measured tar and total volatiles yields for a wide range of coals. Carbon content is used to illustrate coal rank. Predictions were made with a coal devolatilization network model. Solid and dashed lines represent correlation of the measured total volatiles and tar yields, respectively. [Reproduced with permission from Fletcher, T. H., Solum, M. S., Grant, D. M., Pugmire, R. J. (1992). "The Chemical Structure of Chars in Transition from Devolatilization to Combustion," *Energy and Fuels* **6,** 643–650.]

### 1. Coal Devolatilization

Coal devolatilization or pyrolysis dominates the early stages of combustion. This process involves the homogeneous, solid-phase thermal decomposition of coal chemical structure, producing gaseous and solid products. The gaseous products include light gases ($CO, CO_2, CH_4, H_2O,$ $H_2$, etc.) and heavy hydrocarbon gases that condense at room temperature. The latter are called *tars*. The general, but not universal, convention is to call this process *devolatilization* when it occurs under unspecified conditions or combinations of reducing and oxidizing conditions. When it occurs under neutral or reducing conditions, it is sometimes called *pyrolysis*. Devolatilization generates the gases that form visible flames in wood (and coal) fires.

This part of the reaction cycle occurs as the raw coal is heated and thermally decomposes. The particle may soften and undergo internal transformation. Moisture present in the coal evolves early as the temperature rises. As the temperature continues to increase, gases and heavy tarry substances are emitted as the coal chemical structure thermally decomposes. The extent of devolatilization can vary from a few percent up to 70–80% of the total particle weight and can take place in a few milliseconds to several minutes or longer, depending on coal size and type and temperature. Devolatilization typically occurs in a few milliseconds and accounts for about 60% of the organic mass loss in pulverized coal processes. The residual mass, enriched in carbon and inorganic material and depleted in oxygen and hydrogen, is called *char*. The char particle often has many cracks or holes made by escaping gases, may have swelled to a larger size than the original coal particle, and can have high porosity. The nature of the char is dependent on the original coal type and size and the conditions of devolatilization.

Five principal devolatilization phases have been identified. The first phase is moisture vaporization near the boiling point of water. The second phase (627–675 K)
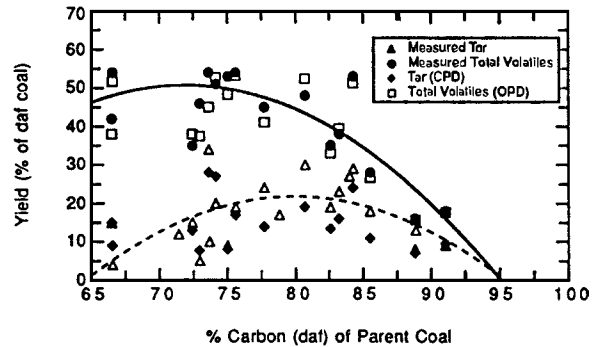
is associated with the initial evolution of carbon dioxide and carbon monoxide and a small amount of tar. The third phase (800–1000 K) is the evolution of chemically formed water and carbon oxides. The fourth phase (1000–1200 K) involves the final evolution of tar, hydrogen, and hydrocarbon gases. The fifth phase is the formation of carbon oxides at high temperatures. Structural properties of the coal strongly affect devolatilization mechanisms and rates. Yield of volatiles depends on external factors such as peak temperature, heating rate, pressure, and particle size as well as coal type and chemical structure. Proportions of gases and tars vary widely, as shown in Fig. 7. Rates and amounts of weight loss during devolatilization differ among coal types, with much more liquid and tar products for the bituminous coals and more gaseous products (including $H_2O$) from lower rank coals.

Weight loss is not significantly size dependent for particles up to about 400 $\mu$m in diameter. However, very large particles behave differently from finely pulverized coal. Larger particles heat less rapidly and less uniformly, so that a single temperature cannot be used to characterize the entire particle. The internal char surface provides sites where secondary reactions occur. Devolatilization products generated near the center of a particle must migrate to the outside to escape. During this migration, they may crack, condense, or polymerize with some carbon deposition and smaller yield of volatiles. Devolatilization rates and yields consequently decrease with increasing particle size.

### 2. Char Oxidation

Char oxidation is an inherently heterogeneous, gas–solid reaction with the gas composition dominating the rates

and yields. When this reaction occurs with oxygen it is generally called *oxidation*. When it occurs with $CO_2$, $H_2O$, or other gases it can be called *gasification*. Gasification is sometimes used to describe the combination of devolatilization/pyrolysis reactions and heterogeneous reactions with the char under overall reducing conditions.

The residual char particles can be oxidized or burned away by direct contact with oxygen at sufficiently high temperature. This reaction of the char and oxygen is heterogeneous, with gaseous oxygen diffusing toward and into the particle, adsorbing and reacting on the particle surface, and desorbing as either CO or $CO_2$. For small particles at high temperature, CO is the dominant surface product. This heterogeneous process is often much slower than the devolatilization process, requiring seconds for small particles to several minutes or more for larger particles. These rates vary with coal type, temperature, pressure, char characteristics (e.g., size, surface area, porosity), presence of inorganic impurities, and oxidizer concentration. At atmospheric pressure, gasification reactions of char with $CO_2$ and $H_2O$ are much slower than oxidation.

These two processes (i.e., devolatilization and char oxidation) may take place simultaneously, especially at very high heating rates. If devolatilization takes place in an oxidizing environment (e.g., air), then the fuel-rich gaseous and tar products react further in the gas phase to produce high temperatures in the vicinity of the coal particles.

## B. Rate Processes and Equations

### 1. Coal Devolatilization

The earliest coal devolatilization models were simple single-step, single-equation expressions in which coal was conceived to form gases and char with a specified yield, as follows:

$$\frac{dv}{dt} = k(v_\infty - v), \qquad (2)$$

$$k = A \exp(-E/RT), \qquad (3)$$

where $v$ is the mass of volatiles emitted, $t$ is time, $k$ is the reaction rate coefficient, $v_\infty$ is the ultimate volatiles yield, $T$ is temperature, and $A$ and $E$ are respectively the preexponential factor and the activation energy that describe the reaction rate coefficient.

This single-step process lacks the flexibility required to describe much of the experimental data and nonisothermal effects. More sophisticated but still essentially empirical models describe devolatilization with two parallel, first-order reactions, by a first-order reaction with a statistical distribution of activation energy, and by various combina-
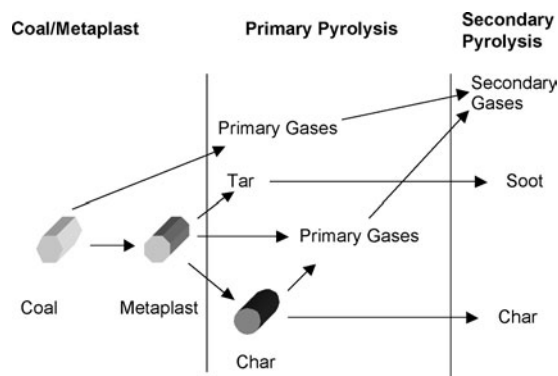


**FIGURE 8** A schematic illustration of the major steps during coal devolatilization (pyrolysis).

tions of parallel and sequential first-order reactions. The most successful of these empirical models describe the dependence of volatile yields on temperature, heating rate, and pressure with some amount of differentiation of the species produced by devolatilization (tars, light gases, and chars).

The organic properties and structural characteristics of coal have a large impact on its devolatilization behavior. Fundamental devolatilization models base predictions on measurable organic properties of the coal. For example, devolatilization of higher rank, melting and swelling coals has been described by the following two-step process, as illustrated in Fig. 8. In the first step, coal undergoes a reduction of hydrogen bonding, and the macromolecular network softens to partially form a metaplast containing liquid coal components. Tars are evaporated from the generated metaplast, and the mobile phase is released during this stage. Not all of the macromolecular structure is depolymerized into metaplast. Low-rank coals produce very little metaplast. In the second step, further bond breaking leads to the evolution of tars and gases and repolymerization of coal fragments from char. Competition between bond scission and cross-linking reactions controls the rate of tar evolution and the properties of char. Gas formation is correlated with the composition of the coal. Large amounts of char and primary gases are generated directly from the coal macromolecular phase, especially for low-rank coals. Tar formation is viewed as a combined depolymerization and vaporization process. The structural evolution of char during pyrolysis suggests that functional groups are released from the aromatic clusters, with cross-linking occurring among the aromatic clusters. At the end of devolatilization, the carbon skeletal structures of chars from a variety of coals become very similar chemically. However, the physical structure of chars varies significantly depending on the type and rank of the coal from which they are derived.

Coal network models have more recently been developed to relate the structures of coal to devolatilization processes and rate measurements. Model parameters have been obtained from several optical or spectral measurements of coals and from rate measurements. Initially, physical melting occurs due to weakening of hydrogen bonds in the network, and the mobile phase is released as early tar. Further scission of covalent bonds in the network generates fragments corresponding to the main evolution of aromatic tars. Depending on coal rank, oxygen structures in coal such as carboxyl and hydroxyl groups thermally decompose early in the devolatilization process, generating radical groups that rapidly cross-link the coal network structure. Tar precursors are then not allowed to form and evolve. Thus, early cross-linking is empirically correlated with the evolution of $CO_2$.

Coal structure is typically modeled with four basic groups: aromatic nuclei, labile bridges, char or stable bridges, and peripheral groups. These structural parameters have been determined from many advanced coal characterization techniques. Competition occurs in the evolving coal structure between the cleavage of labile bonds and the formation of stable char links. With increasing temperature, the labile bonds are rapidly consumed and the decomposition of methyl and methoxy groups forms radicals that rapidly cross-link the structure into char. With further increases in temperature, methane and hydrogen are evolved from the char structure. It has been shown that two moieties, aryl–alkyl ethers and hydroaromatic structures, can be used to correlate weight loss and tar yield.

Network models use lattice statistics to quantitatively describe the thermal breakup of the coal macromolecular structure and interpret the interrelationships in the lattice characteristics of coal, tar, and char. The capability of one network model to predict total volatiles and char release for a large number of coals of various rank without the use of any fitting parameters was illustrated in Fig. 7. Char is formed from charring reactions within the network and from cross-linking of nonvolatile metaplast.

### 2. Coal Volatiles Ignition and Combustion

Many products are produced during devolatilization of coal, including tars and hydrocarbon liquids, hydrocarbon gases, $CO_2$, $CO$, $H_2$, $H_2O$, and HCN. These products react with oxygen in the vicinity of the char particles, increasing temperature and depleting oxygen. This complex reaction process is important to control soot formation, flame stability, and char ignition. Rate processes include the following: volatiles release, tar condensation and repolymerization in char pores, hydrocarbon cloud evolution through small pores from the moving particle, cracking of the hydrocarbons to smaller hydrocarbon fragments with local production of soot, condensation of gaseous hydrocarbons and agglomeration of sooty particles, macromixing of the devolatilizing coal particles with oxygen, micromixing of the volatiles cloud and oxygen, oxidation of the gaseous species to combustion products, production of nitrogen oxides and sulfur oxides by reaction of devolatilized products with oxygen, and heat transfer from the reacting fluids to the char particles.

At high combustion temperatures, up to 70% or more of the reactive coal mass can be consumed through this process. Volatiles combustion in practical systems is complicated by turbulent mixing of fuel and oxidizer, soot formation and radiation, and near-burner fluid dynamics. Such systems usually exhibit volatiles cloud combustion, rather than single-particle combustion.

### 3. Char Oxidation

The rate of char oxidation is determined by a combination of chemical kinetics and mass transport. Heterogeneous char oxidation can proceed simultaneously with or after vaporization and devolatilization, depending on reaction conditions. The time required for char and coke combustion is substantially greater than that required for devolatilization. There is also much emphasis on reducing the amount of carbon in the ash, which means that the burnout of the char or coke needs to be more effective. The physical structure of the char, such as pore structure, surface area, particle size, and inorganic content, controls the reaction processes of the char. However, the structure of char is determined by the devolatilization process as well as by the parent fuel. Models of char combustion can include the following processes: (1) diffusion of gaseous reactants and products through the boundary layer surrounding the particle, (2) diffusion of reactants within the porous structure of the particle, (3) conduction of heat through the particle, (4) adsorption of oxygen, (5) chemical reaction of the oxidizer on the particle surface, (6) desorption of products (e.g., CO) from the solid surface, (7) the homogeneous reaction of carbon monoxide with oxidizer within the pores and in the boundary layer, and (8) the evolution of particle and pores during the reactions.

Global models of heterogeneous reaction typically include only processes 1 and 5 described above as resistances in the reaction rate expression. The resistance due to film diffusion is usually calculated through the use of diffusional coefficients corrected for the mass leaving the particle surface (i.e., blowing factor). The effective diffusivity of the gases into the particle and the portion of the particle that participates in the reaction can be estimated by using methods that have been published.

The char reacts with oxygen to form CO primarily, and $CO_2$, depending on conditions. The measured oxidation rate is correlated through reaction rate expressions such as

$$r_1 = \eta \xi k_x T_p^m p_{O_2,s}^n = \eta \xi k_x R_g^n T_p^{m-n} C_{O_2}^n, \qquad (4)$$

where $T_p$ is the particle temperature, $k_x$ is the kinetic rate coefficient, which varies exponentially with surface temperature, and $p_{O_2,s}$ is the oxygen partial pressure at the solid surface. The parameter $\xi$ is the ratio of active surface area to external surface area and can be used to incorporate intrinsic kinetics in the model. The parameter $\eta$ is the fraction of the surface area that is available for reaction. As the ash fraction of the char particle increases, rates of reaction of the entire particle based on total surface area generally decrease. The particle diameter changes as the particle burns and is typically modeled as a shrinking sphere with constant density, a constant-diameter sphere with changing porosity, or some intermediate relationship. Some chars fragment during combustion, a process most pronounced in high-rank, initially large coal particles. Since the oxygen partial pressure is not known at the particle surface, the molecular diffusion equation is used in conjunction with the oxidation equation, Eq. (4), to solve for reaction rate, assuming quasi-steady state as the char particle reduces in size and/or mass. Other complications during char oxidation include possible changes in reactivity during reaction because of char annealing and graphitization, effects of impurities on surface reaction rates, and the ratio of $CO/CO_2$ formed on the char surface. More recent char oxidation models include oxygen adsorption and product desorption steps through the classic Langmuir expression, which differs in form substantially from the empirical expression above.

Descriptions of pore development and structure require microscopic models of the particle. These models include intrinsic kinetics and pore structural changes during burnoff. Three of the most popular microscopic models are a random capillary pore model, one in which the pores are considered spherical vesicles connected by cylindrical micropores, and one in which the pores have a treelike structure. These models allow for pore growth and coalescence in their respective fashions and provide estimates of reactive surface area. Parameters required for these models are obtained from experimental measurements of the various chars.

Other microscopic models have been developed that are based on percolation or other char transformation theories. These microscopic models currently serve more as scientific descriptors than practical tools.
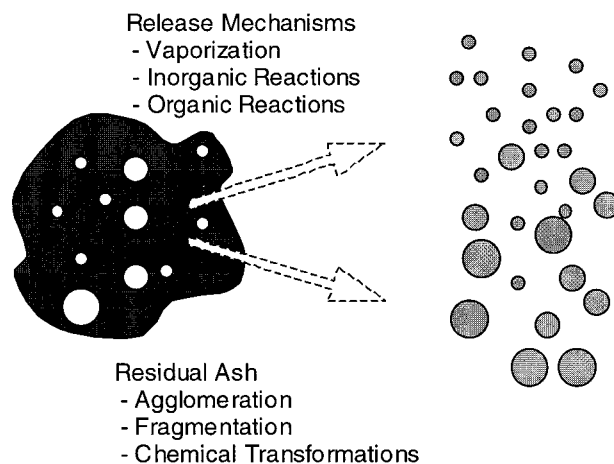


Release Mechanisms
- Vaporization
- Inorganic Reactions
- Organic Reactions

Residual Ash
- Agglomeration
- Fragmentation
- Chemical Transformations

**FIGURE 9** Schematic illustration of the fate of inorganic material in coals during combustion.

## C. Fate of Inorganic Material

### 1. Ash Formation

Figure 9 schematically illustrates the fate of inorganic material in coal leading to the production of fly ash. Inorganic grains may be imbedded in the particle or may be extraneous to the particle itself. The fate of this second class of inorganic material differs substantially from that of the inherent material. The minerals undergo chemical reactions and phase changes determined by their thermochemistry as well as interacting with other inorganic components and the organic material. Components of the minerals may be released from the fuel by either thermal decomposition or vaporization during combustion.

The transformations are divided into two types: release mechanisms and the fate of the residual ash. Release mechanisms are indicated as vaporization, thermal or chemical disintegration of the inorganic material (inorganic reaction), or convection during rapid devolatilization or other organic reactions. These mechanisms tend to produce small ($<0.8\ \mu$m) particles or vapors. The residual ash may undergo fragmentation either as an inorganic grain or in conjunction with fragmentation of burning char particles, may coalesce with some or all of the remaining inorganic material, and may undergo significant chemical or physical transformations. This material tends to produce larger ash particles. Depending on the type of inorganic material and the combustion conditions, the ash produced during combustion is composed of varying amounts of vapor, fume ($<1$-$\mu$m-diameter particles), and larger particles.

Many inorganic materials occur as hydrates in coal, all of which dehydrate through endothermic reactions, generally at temperatures from $100°C$ to $200°C$. Sulfates typically thermally decompose as they are heated,

as do carbonates and sulfides. In cases such as most of the alkaline-earth-containing materials, these salts decompose prior to melting and appreciable vaporization, while in cases such as most of the alkali-containing material, the salts melt, partially vaporize, and react further in the gas phase. Chlorides and hydroxides are more stable than the salts with bivalent anions and generally do not completely decompose or vaporize. At flame temperatures, these species represent the most stable form of the alkali metals and of chlorine and hydroxyl. Phosphates and silicates undergo transformations when heated but generally do not completely decompose or vaporize. Their melting behavior depends on their composition and structure, and they can generally be grouped into high- and low-temperature-melting materials. High-temperature-melting materials include phosphates, kaolinite, and their decomposition products. Low-temperature-melting materials include illite, salts, and alkali-containing species. Mathematical models have recently been developed to predict the decomposition and reaction rates of the inorganic species in coal.

### 2. Ash Deposition

A great deal of information is available on rates and mechanisms of ash deposition. Five major mechanisms of deposit formation include (1) inertial transport including impaction and sticking, (2) eddy impaction, (3) thermophoresis, (4) condensation, and (5) chemical reaction. Rates of inertial impaction in most environments are well established, and inertial impaction commonly accounts for most of the accumulated mass. Rates of eddy impaction are less well known. The capture efficiency, a measure of the propensity of material to stick to a surface upon impaction, is far less well established. The rates of thermophoretic deposition are reasonably well established when local temperature gradients and the functional form of the thermophoretic force on a particle (or the thermophoretic velocity) are known. Condensation rates can be predicted reasonably well, given accurate vapor pressure and concentration data. The accuracy to which rates of chemical reaction are known is often inadequate, especially those involving sulfation and alkali adsorption in silicates.

### 3. Deposit Properties

Properties of deposits that are important to boiler behavior and management include emittance and absorbance, thermal conductivity, strength, tenacity and thermal shock resistance, viscosity, and porosity. Methods for measurement of each of these properties are available, while techniques for prediction or correlation of these properties are in various states of development. Figure 10 illustrates the
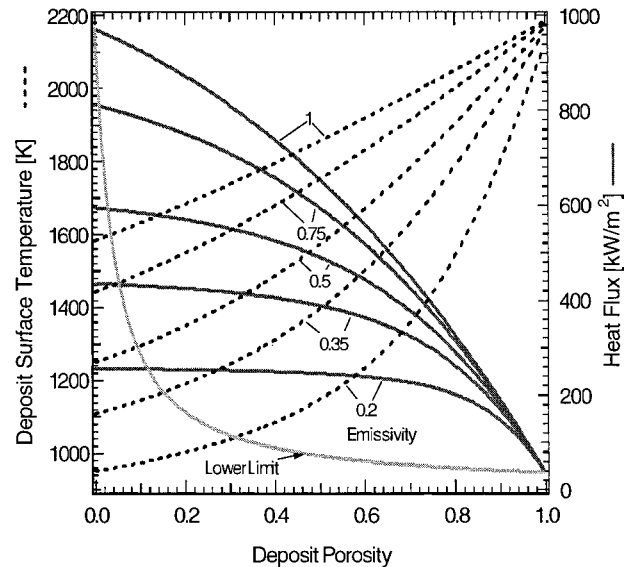


**FIGURE 10**  Deposit surface temperature and heat flux as a function of porosity and emissivity for a specified set of assumptions.

dramatic impact of deposit thermal conductivity and emissivity on a boiler's deposit surface temperature and heat flux.

## V. COAL-GENERATED POLLUTANTS

### A. Criteria Pollutants

#### 1. Pollutant Species

The most challenging problem associated with coal consumption is the control of pollutants. Increasing power demand will lead to increased production and use of fossil fuels with the potential for increased pollutant emissions.

In the United States, burning of coal accounts for 93% of the $SO_2$ emissions from electric generating units and 85% of the nitrogen oxides ($NO_x$) with total emissions of about 14 million tons of $SO_2$ and nearly 9 million tons of $NO_x$ per year (figures for 1997). Total $SO_2$ emissions from electric power generation declined by 16% from 1989 to 1997, while total $NO_x$ emissions increased by about 3%. During the same period, electric power generation increased by 22%.

Concern for the atmospheric quality can be dated to several early episodes when coal began to replace wood as the primary source of domestic heating and industrial fuel. By the mid-20th century, public support and government resolve were strong enough to regulate air pollutant emissions from stationary combustors.

The combustion of coal produces both primary and secondary pollutants. Primary pollutants include all species

of the combustor exhaust gases that are considered contaminates to the environment. The major primary pollutants include carbon monoxide (CO), hydrocarbon fragments, sulfur-containing compounds ($SO_x$), nitrogen oxides ($NO_x$), particulate material, and various trace metals. Secondary pollutants are defined as environmentally detrimental species that are formed in the atmosphere from precursor combustion emissions. The list of secondary pollutants includes particulate matter and aerosols that accumulate in the size range of 0.1–10 $\mu$m in diameter, $NO_2$, $O_3$, other photochemical oxidants, and acid vapors. The connection between combustion-generated pollutants and airborne toxins, acid rain, visibility degradation, and stratospheric ozone depletion is well established. The detrimental impact of these contaminants on ecosystems in the biosphere has been the impetus for stricter standards around the world.

Studies have shown that anthropogenic (i.e., human-generated) emissions of sulfur-containing pollutants decreased in the United States during the 1970s and early 1980s, whereas nitrogen oxides continued to increase. In 1985, stationary fuel combustion in electric utilities and industry accounted for approximately 50% of all $NO_x$ emissions and more than 90% of all $SO_x$ emissions. The increase in $NO_x$ emissions can be correlated with the construction of many coal-fired, power-generating facilities throughout the U.S. midwest and west. Coal naturally contains nitrogen in the range 0.5–1.5 wt%, while sulfur typically varies from 0.5 to 10 wt%, with 1–2 wt% being more common.

## 2. Air-Quality and Emissions Regulations

The National Ambient Air Quality Standards (NAAQS) set by the United States specify the maximum allowable concentrations for the following *criteria pollutants*: CO, hydrocarbons, particulate matter, $NO_2$, and $SO_2$ (see Table V). The last three are of most concern in the burning of coal. The NAAQS define Primary Standards, which are intended to protect public health with an adequate margin of safety. Secondary Standards are also specified by the NAAQS, which are intended to protect public welfare (e.g., soils, vegetation, wildlife). Other countries have similar, sometimes more stringent, requirements, as shown in Table V. Regions are classified as either in "attainment" or "nonattainment" depending on whether the ambient level of pollutant concentrations exceeds or does not exceed the standards established by the regulations. While these standards specify the maximum allowable ambient concentrations for key pollutants, they do not effectively control point-source emissions rates.

Two common provisions dictate the type and level of pollutant abatement required by existing and new combustion facilities in the United States. These include the New Source Performance Standards (NSPS) and the Prevention of Significant Deterioration (PSD) regulations. NSPS establish nationwide emission standards for new and upgraded facilities regardless of source location or regional air quality. The NSPS are set at levels that reflect the degree of control achievable through the application of the best system of continuous emission reduction that has been adequately demonstrated for each category of sources. Under PSD regulations, the permitting of most new facilities requires the installation of best available control technology (BACT) for the control of nitrogen and sulfur oxides. BACT is defined as "an emissions limitation based on the maximum degree of reduction for each pollutant . . . which the reviewing authority, on a case-by-case basis, taking into account energy, environmental, and economic impacts and other costs, determines is achievable." A BACT analysis must be performed for each new, modified, or reconstructed combustion source.

The U.S. Clean Air Act and its 1990 amendment provide the framework for controlling major sources, including stationary combustors. The regulations establish NSPS which apply to units of more than 73 $MW_e$ ($250 \times 10^6$ Btu hr$^{-1}$) of heat input and are directed toward the control of $SO_2$, particulates, and $NO_x$.

The 1980 Clean Air Act regulation of $SO_2$ concentrations for gaseous and liquid fuels is 96 mg/$10^6$ J (0.20 lb/$10^6$ Btu). The standard for coal-fired unit emissions is somewhat complicated. A unique maximum emission rate and a unique minimum reduction of potential (uncontrolled) emissions is based on the sulfur content and heating value of the coal. For high-sulfur coals ($>1.5\%$ sulfur content), a 90% reduction of uncontrolled $SO_2$ emissions is required with a maximum allowable $SO_2$ emission rate of approximately 520 mg/$10^6$ J (1.2 lb/$10^6$ Btu). For low-sulfur coals, sometimes called compliance coals, the standard requires at least a 70% reduction in uncontrolled $SO_2$ emissions and limits the emission rate to 287 mg/$10^6$ J (0.60 lb/$10^6$ Btu).

The 1990 Clean Air Act amendment (CAAA) further reduced $SO_2$ emissions to 50% of their 1980 levels by the year 2000, and caps $SO_2$ emissions at that level. The new standards not only affect new sources, but also require a constant total emissions allowed from all electric utility generation units of $8.1 \times 10^6$ metric tons of $SO_2$ per year. The Act contains provisions for trading of emission allowances among units in order to achieve the most economical application of control technologies.

In Germany, $NO_x$ emissions limits for new coal-fired boilers are almost half the currently prescribed level in the United States. Japan has placed nearly as stringent limits on $NO_x$ emissions. However, under CAAA, by 2003–04, emissions limits in the United States will drop by more than half. Strict limits on sulfur oxides have also been set in Japan, where owners of $SO_2$ emitters pay a proportional

**TABLE V  Ambient Air Quality Standards or Recommendations for Combustion Pollutants in Various Countries[a]**

|  | Country/organization | Concentration | Time |
|---|---|---|---|
| $NO_x$ (as $NO_2$) | Germany | 0.05 ppm | 2–12 mo |
|  | Japan | 0.04–0.06 ppm | 24 hr |
|  | United States | 0.05 ppm[b,c,d] | Annual arithmetic mean |
|  | Former USSR | 0.05 | 24 hr |
| $SO_2$ | Germany | 0.06 ppm | 24 hr |
|  | United States | 0.03 ppm[b,c] | Annual arithmetic mean |
|  |  | 0.14 ppm[b,c] | 24 hr |
|  |  | 0.5 ppm[b,d] | 3 hr |
|  | Former USSR | 0.02 ppm | 24 hr |
|  | World Health Organization | 38–37 ppb[e] | 24 hr |
|  |  | 15–23 ppb[e] | Annual arithmetic mean |
| CO | Canada | 13 ppm | 8 hr |
|  | Germany | 26 ppm | 0.5 hr |
|  | Japan | 20 ppm | 8 hr |
|  | World Health Organization | 25 ppm[f] | 24 hr |
|  | United States | 9.0 ppm[c] | 8 hr |
|  |  | 35 ppm[d] | 1 hr |
|  | Former USSR | 1.3 ppm | 24 hr |
| NMHC[g] | Canada | 0.24 ppm | — |
|  | United States | 0.24 ppm[c,d] | Average from 6 to 9 A.M. |
| Particulate matter | Japan | 200 $\mu$g m$^{-3}$ | 1 hr |
|  | United States | 150 $\mu$g m$^{-3\,h}$ | 24 hr |
|  |  | 50 $\mu$g m$^{-3\,h}$ | Annual arithmetic mean |
|  | World Health Organization | 60–90 $\mu$g m$^{-3}$ | Annual arithmetic mean |

[a] From Smoot, L. D., (Ed.) (1993). "Fundamentals of Coal Combustion," Table 6.1, Elsevier Science Publishers, New York.

[b] NAAQS in effect in 1984.

[c] NAAQS primary standard.

[d] NAAQS secondary standard.

[e] Based on observation of community populations exposed simultaneously to mixtures of $SO_2$ and "smoke," that is, particulate matter.

[f] Based on recommended limit in blood of 4% carboxyhemoglobin (COHb) or less.

[g] Nonmethane hydrocarbons; expressed as ppm carbon (ppmC).

[h] Primary and Secondary Standards effective July 1, 1987. Includes only particles having an equivalent spherical diameter of 10 Φm or less. Referred to as the PM10 standard.

release levy. Some European communities have low limits on nitrogen and sulfur oxide emissions. These strict standards have been the impetus for developing and implementing BACT, which includes flue gas treatment for both nitrogen and sulfur oxides. In fact, Japan and Germany are the first to use catalyst beds in the flue duct system to control nitrogen oxide emissions. Similar technologies are widely expected to be deployed in the United States in the near future as power plants prepare for Phase II reductions of $NO_x$ required by the 1990 CAAA.

In the United States, the 1990 CAAA called on the Environmental Protection Agency (EPA) to analyze the health impacts of over 180 specific chemicals and all chemicals containing any of 12 hazardous metals. Mercury, arsenic, and cadmium are among the metals listed and are also of specific concern for coal power plants. The Act required the EPA to submit a report to Congress on the health hazards of power plant (and other industrial) hazardous air pollutant emissions and to describe possible control strategies. The EPA must then regulate those hazardous air pollutants found to be significant in power plant stack exhausts. Currently, utilities in the United States are not required to control the emission of hazardous toxins. Small particles, mercury, arsenic, and cadmium are among the emissions of greatest concern from power plants.

## 3. $NO_x$ Formation and Reduction

A detailed understanding of nitrogen oxide formation during the combustion of fossil fuels has been achieved by

years of research. Most of the nitrogen oxides emitted to the atmosphere by combustion systems are in the form of nitric oxide (NO), with smaller fractions appearing as nitrogen dioxide ($NO_2$) and nitrous oxide ($N_2O$). Other nitrogenous pollutants include ammonia and hydrogen cyanide.

Nitric oxide is formed in flames predominantly by three mechanisms: thermal $NO_x$ (the fixation of atmospheric molecular nitrogen by reaction with oxygen atoms at high temperatures), fuel $NO_x$ (the formation of $NO_x$ from nitrogen contained in the fuel), and prompt $NO_x$ (the attack of a hydrocarbon radical on molecular nitrogen, producing $NO_x$ precursors). $N_2O$ can be formed by a number of paths in gaseous and coal-laden reactors and is more prevalent at lower temperatures. $N_2O$ levels in coal combustors are generally less than 5 ppm (except for fluidized-bed systems). $NO_2$ is typically significant only in combustion sources such as fuel-lean gas turbines. Neither $N_2O$ nor $NO_2$ reaches high concentrations in pulverized coal flames or in the unquenched effluent of most gas combustors. Hence, NO is normally the most significant nitrogen oxide species emitted from coal-fired furnaces.

It is generally agreed that the reaction network for thermal NO is qualitatively described by the modified Zel'dovich mechanism:

$$N_2 + O \longleftrightarrow NO + N, \qquad (5)$$

$$O_2 + N \longleftrightarrow NO + O, \qquad (6)$$

$$N + OH \longleftrightarrow NO + H, \qquad (7)$$
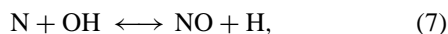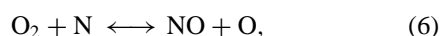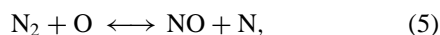
and where O atom concentrations are often assumed to be in equilibrium.

The first step is rate-limiting and requires high temperatures to be effective due to the high activation energy barrier with the designation *Thermal* NO. The third step is important in fuel-rich environments.

Prompt NO occurs by the collision and fast reaction of hydrocarbons with molecular nitrogen in fuel-rich flames. This mechanism accounts for rates of NO formation in the early flame region that are much greater than the rates of formation predicted by the thermal NO mechanism just described. This mechanism is much more significant in fuel-rich hydrocarbon flames. Two such reactions are believed to be particularly significant:

$$CH + N_2 \rightarrow HCN + N, \qquad (8)$$

$$C + N_2 \rightarrow CN + N. \qquad (9)$$
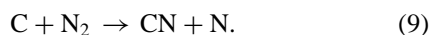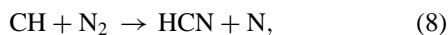
The cyanide species is further oxidized to NO.

Fuel NO typically accounts for 75–95% of the total $NO_x$ accumulation in coal flames. Nitrogen in coal is predominantly found as a heteroatom in aromatic rings. Nitrogen primarily evolves from tars as HCN, with much

lower concentrations of $NH_3$. The conversion of fuel nitrogen to HCN is independent of the chemical nature of the initial fuel nitrogen (assuming that it is in-ring aromatic nitrogen) and is not rate limiting. Once the fuel nitrogen has been converted to HCN, it rapidly decays to $NH_i$ ($i = 0, 1, 2, 3$), which reacts to form NO and $N_2$ as illustrated by the following simplified global reaction scheme, though the process is far more complicated:

$$\text{fuel-N} \longrightarrow HCN \longrightarrow NH_i \overset{\overset{NO}{\nearrow} N_2}{\underset{\underset{O_2}{\searrow} NO}{}} . \qquad (10)$$
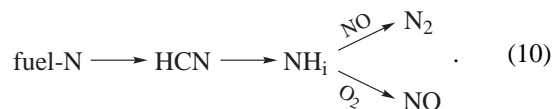
In fuel-rich combustion systems, NO can also be reduced by hydrocarbon radicals leading to the formation of HCN and then molecular $N_2$. Based on such observations an $NO_x$ abatement technology termed *reburning* has been developed by injecting and burning secondary fuel in the postburner flame to convert NO back to HCN. It is generally accepted that light hydrocarbon gases are the most effective staging fuels, although pulverized coal, particularly lignite, has been shown to work.

In a typical coal flame, char retains a significant fraction (30–60%) of the initial nitrogen. NO formation during char reactions often accounts for 20–30% or more of the total NO formed.

Heterogeneous destruction of gas-phase NO is significant during fuel-lean char burnout, but homogeneous decay reactions appear to dominate NO removal in fuel-rich zones of coal combustion. Soot formation and subsequent conversion of soot nitrogen to nitrogen pollutants represents yet another way to produce $NO_x$ from coal burning. The mechanism of NO destruction by soot particles is also well documented, with results indicating that NO removal rates on carbon black particles are much greater than heterogeneous destruction of NO by char.

Commercially significant $NO_x$ abatement technologies include installation of low-$NO_x$ burners, fuel/air staging, reburning, nonselective catalytic reduction (NSCR), and selective catalytic reduction (SCR). These are listed in approximate order of their implementation. $NO_x$ reductions as high as 70% can be expected by incorporating all of the noncatalytic processes, and nearly all power stations currently do so. However, the next phase of $NO_x$ regulations incorporated in the CAAA require many power plants to reduce $NO_x$ emissions further than is practical using burner and simple boiler modifications, and SCR represents one of the few commercially demonstrated technologies capable of such high reduction. Therefore, SCR systems are expected to become much more prevalent in the United States in the near future. Such systems are already installed in Germany and Japan, where $NO_x$ emission limits are more stringent than in the United States.

### 4. Sulfur Oxides Formation and Capture

Sulfur oxides are formed in stationary combustors from the sulfur entering with the coal. The sulfur content of coal typically varies from 0.5 to 10 wt%, depending on rank and the environmental conditions during the geological coal formation process. Sulfur oxides emissions can be reduced substantially by simply converting from high- to low-sulfur coal. Such coal switching or partial switching (blending) is responsible for the large shift in coal production from the eastern and midwestern coal seams in the United States to the low-sulfur coals predominantly found in Wyoming and Montana. For certain coals, the sulfur content can be reduced by removing some of the sulfur from the coal through washing or magnetic separation before burning it.

There are three forms of sulfur in coal: (1) organic sulfur such as thiophene, sulfides, and thiols, (2) pyritic sulfur ($FeS_2$), and (3) sulfates such as calcium or iron salts. Both organic and inorganic sulfur phases undergo significant chemical transformations during coal devolatilization and combustion. Sulfur-containing pollutants formed by burning coal include $SO_2$, $SO_3$, $H_2S$, COS, and $CS_2$. Under normal boiler operating conditions, with overall excess oxygen being used, virtually all of the sulfur is oxidized to $SO_2$, with small quantities of $SO_3$. The $SO_2$ is the thermodynamically favored product at high temperatures. At the stack temperatures, $SO_3$ is the most stable species, but kinetic limitations prevent significant amounts of $SO_3$ from forming, leading to $SO_x$ emissions being dominated by $SO_2$.

Natural minerals such as limestone can be pulverized and added directly to the furnace to capture the sulfur. The limestone calcines to form the calcium oxide with formation of pores in the particle:

$$CaCO_3 \longleftrightarrow CaO + CO_2. \qquad (11)$$
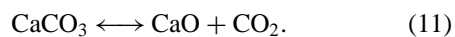
The sulfur dioxide penetrates the pores and reacts with the calcium oxide to form solid calcuim sulfate that can be removed with the ash. Dolomite ($CaCO_3 \cdot MgCO_3$) and hydrated lime [$Ca(OH)_2$] are also used as sorbents. Sulfur scrubbers based on a variety of chemical reactions have become more common since 1990. Such systems produce by-products with some commercial value, such as elemental sulfur, sulfuric acid, and gypsum. Scrubbers have added benefits of removing some $NO_x$, mercury, arsenic, and other pollutants that either currently or in the future may fall under formal regulation. However, scrubbers add about 25% to the capital and operating costs of a power station, leading most power stations to switch to low-sulfur coals rather than build scrubbers.

## B. Non-Criteria Pollutants

Coal contains significant amounts of many trace elements. Most trace elements are associated with the inorganic portion of the coal, although some are also chemically and physically bound with the organic material. Trace element concentrations vary significantly among coal seams and even among mines within the same seam. During coal combustion, these trace elements become partitioned in various effluent streams, depending on their chemical behavior. Classification of trace element emissions has been partitioned into three general groups.

- *Group 1*. Elements concentrated in coarse residues, such as bottom ash or slag, or partitioned between coarse residues and particulates that are trapped by particulate control systems (e.g., Ba, Ce, Cs, Mg, Mn, and Th).
- *Group 2*. Elements concentrated in particulates compared with coarse slag or ash and tend to be enriched in fine particulate material that may escape particle control systems (e.g., As, Cd, Cu, Pb, Se, and Zn).
- *Group 3*. Elements that volatilize easily during combustion, are concentrated in the gas phase, and deplete in solid phases (e.g., Br, Hg, and I).

Emission of elements found in group 1 is controlled by standard particle handling systems. Because group 2 elements are concentrated in fine particles, the release of these elements depends largely on the efficiency of the gas-cleaning system. The emission of volatile trace elements, such as those in group 3, can be controlled by combustion conditions or downstream processing. For example, trace element release may be lower from a fluidized-bed coal furnace than from entrained-flow systems because of the lower temperatures in fluid beds. Furthermore, wet scrubbers used for flue gas desulfurization allow volatile elements to condense and therefore provide an effective method of reducing emissions of certain trace elements. Trace elements that are of particular concern from coal combustion are Hg, As, Cd, and Pb.

## C. Greenhouse Gas Emissions

Greenhouse gas emissions represent a major and complex concern for the future use of fossil energy in general and coal in particular. Greenhouse gases are atmospheric components that are generally transparent in the visible region but absorb longer wavelength radiation, thereby contributing to a higher atmospheric temperature. They include water vapor ($H_2O$), carbon dioxide ($CO_2$), nitrous oxide ($N_2O$), methane ($CH_4$), hydrofluorocarbons

(HFCs), perfluorocarbons (PFCs), and sulfur hexafluoride ($SF_6$). Water vapor plays a large role in the greenhouse effect and, in the form of clouds, introduces much of the uncertainty in global circulation models. However, water is rarely discussed in terms of anthropogenic sources because of its very high background concentration and vital role in weather and precipitation. Most analyses, and this discussion, are presented on a water-vapor-free basis. Because of differences in the absorption efficiency of these gases and wide differences in their existing concentrations, they vary widely in potency as greenhouse gases. For example, methane is about 21 times as potent as carbon dioxide as a greenhouse gas. However, $CO_2$ is by far the largest potential contributor to emissions due to much higher concentration in both the atmosphere and in net emissions from most sources. In globally averaged numbers, $CO_2$ accounts for 80–90% of the total anthropogenic greenhouse gas effect, $CH_4$ between 5% to 10%, and NO slightly less than $CH_4$, with all other gases making up the remaining few percent. Therefore, greenhouse gas emissions are often almost synonymous with carbon dioxide and methane emissions.

The complexity of the issue can be partially reduced by analyzing its aspects, starting at the most certain and progressing to the less certain but more important. There is no doubt that the average concentration of greenhouse gases in the earth's atmosphere increased during the 20th century. There is also no doubt that much of this increase is anthropogenic as opposed to natural fluctuations. The large increase in fossil energy utilization during the same period is responsible for much of this increase. There is debate among scientists as to whether these global atmospheric chemistry changes are driving a measurable increase in the average world temperature, i.e., global warming, although a consensus seems to be growing that measurable warming is occurring. There is considerable uncertainty regarding the current and long-term magnitude of this change and its impact on global climates, national and international economies, and global and regional environments. The most radical predicted changes from credible sources suggest very large impacts on both environments and economies, with potentially massive destruction of property and loss of life associated with changes in sea levels, regional precipitation and drought, crop yields, and ecosystems. Others predict more modest but still measurable changes. Most scientists acknowledge that such changes are difficult to predict with confidence. However, the possible consequences are such that some proactive measures may be warranted. The difficulty and magnitude of lifestyle change required to affect a decrease or even to significantly slow the increase in global $CO_2$ concentrations is not widely acknowledged.

Coal figures prominently in this discussion. Coal contributes a disproportionately high amount of $CO_2$ from power production. For example, in the United States about 54% of the electric power production is derived from coal, whereas about 88% of the $CO_2$ emissions from electric power production are derived from coal. Furthermore, the largest increases in $CO_2$ emissions in the future are anticipated to arise in association with increased industrialization of large population centers, specifically China and India. Both countries have large coal reserves and are actively developing them to support their improving economies. If the per capita energy consumption of such countries becomes comparable to that of the United States or western Europe, or even if it becomes a significant fraction of it, the impact on greenhouse gas emissions will increase significantly.

Technologies exist that effectively reduce greenhouse gas emissions. Most are renewable energy options (solar, wind, biomass, and hydrogen-based power, among others), which generally do not produce a net change in greenhouse gases. Another effective and proven, but highly unpopular and nonrenewable option is nuclear power. Those familiar with the practical implementation of renewable systems recognize the enormous change in infrastructure required to implement these at large enough scale to impact greenhouse gas emissions. Most renewable options are more expensive, less proven, and less convenient than fossil energy systems. Many suggest that the increased costs of renewable or sustainable power production are more accurate indicators of the true cost of power production than are current market costs based mainly on fossil energy.

Government programs are actively exploring increased efficiency, renewable energies, energy conservation, and other systems to reduce greenhouse gases. Nevertheless, several prominent nations, including the United States, have yet to commit to specific reduction targets such as those described in the Kyoto Protocol and many nations that have committed to these targets are well behind schedule in meeting these commitments. Greenhouse gas emissions may well become one of the major issues in coal utilization in the coming years.

## VI. NUMERICAL SIMULATION OF COAL BOILERS AND FURNACES

### A. Foundations

Development and application of comprehensive, multidimensional, computational combustion models for coal-fired power stations is increasing across the world. While once confined to specialized research computer codes,

these combustion models are now commercially available. Simulations made with such computer codes offer substantial potential for use in analyzing, designing, retrofitting, and optimizing the performance of coal combustion during power generation.

Development of fossil-fuel combustion technology in the past was largely empirical in nature, being based primarily on years of accumulated experience in the operations of utility furnaces and on data obtained from subscale and full-scale test facilities. Empirically based experience and data have limited applicability, however, when considering changes in process parameters, such as evaluating firing strategies for improving combustion efficiencies or mitigating pollutant formation. Large-scale furnace data are typically limited to effluent measurements. Subscale data provide valuable information and insights into controlling phenomena and are less expensive to obtain. However, the results of subscale tests can be difficult to extrapolate to large-scale systems because of the complex nature of turbulent, reactive flow processes in furnace diffusion flames. Combustion modeling technology bridges the gap between subscale testing, which tends to be phenomenological in nature, and the operation of large-scale furnaces typically used for power generation by providing information about the combustion processes that experimental data alone cannot practically provide.

When modeling pulverized coal combustion, the framework for the gas-phase solution approach is most rigorously based on computational fluid dynamics (CFD) using numerical solutions of multidimensional differential equations describing mass, energy, and momentum transport. Source terms and transport coefficients in the CFD models are derived from particle models, gaseous turbulent combustion models, radiation models, and similar coupled descriptions. Information available from model predictions includes temperature, composition, and velocity for gases and particles, ash/slag accumulation, and so forth. Most models of this type provide spatial variation of such quantities, while some also provide variation with time.

Comprehensive combustion models offer many advantages in characterizing combustion processes that can effectively complement experimental tests. Code predictions are typically less expensive and take less time than experimental programs, and such models often provide additional information that cannot always be measured. Computational modeling thereby becomes a cost-effective, complementary tool to testing in designing, retrofitting, analyzing, and optimizing the performance of coal-fired power stations. Typical objectives of these applications have been to determine key information required for program planning, study the effects of various firing alternatives or different fuels on process performance, or aid in process design and optimization. Other objectives include trend analysis, retrofitting, identification of test variables, or scaling of measurements. Examples include reducing carbon loss, optimization of low-$NO_x$ burners, definition of regions of particle impact on walls of furnaces, effects of changes in fuel feedstock quality, effects of changes in burner configuration, or impacts of larger and smaller coal particle sizes.

Computer capacity for running these codes is also becoming much more acceptable due to improved numerical methods and more advanced computer hardware. Three-dimensional code solutions for full-scale furnaces utilizing hundreds of thousands of grid nodes are being obtained on advanced, high-powered engineering workstations within a few hours to a few days.

While combustion modeling technology offers great potential, a major challenge in the use of such technology is establishing confidence that such models adequately characterize, both qualitatively and quantitatively, the combustion processes of interest. This is typically accomplished by making comparisons of code predictions with experimental data measured from flames in reactors that embody the pertinent aspects of the turbulent combustion of coal. Consequently, data from a range of different-size facilities are necessary to validate the adequacy of code predictions and establish code reliability in simulating the behavior of industrial furnaces. Such detailed data also provide new insights into combustion processes and strategies.

## B. Equations and Solutions

The significant physical and chemical phenomena that are typically required to classify a combustion model as *comprehensive* include (see Fig. 11) (1) gaseous, turbulent fluid mechanics with heat transfer, (2) gaseous, turbulent combustion, (3) radiative energy transport, (4) multiphase, turbulent fluid mechanics, (5) water vaporization from particles, (6) particle devolatilization, (7) particle oxidation, (8) soot formation, (9) $NO_x$ and $SO_x$ pollutant formation and distribution, and (10) fouling/slagging behavior. The equations that constitute the mathematical model of reacting gaseous and particulate turbulent fluid flow with heat transfer make use of the fundamental transport equations for mass, momentum, and energy. For the gas phase, the equations are generally cast into fixed-coordinate (i.e., Eulerian) form:

$$\frac{\partial \bar{\rho}\tilde{u}\phi}{\partial x} + \frac{\partial \bar{\rho}\tilde{v}\phi}{\partial y} + \frac{\partial \bar{\rho}\tilde{w}\phi}{\partial z} - \frac{\partial}{\partial x}\left(\Gamma_\phi \frac{\partial \phi}{\partial x}\right)$$
$$- \frac{\partial}{\partial y}\left(\Gamma_\phi \frac{\partial \phi}{\partial y}\right) - \frac{\partial}{\partial z}\left(\Gamma_\phi \frac{\partial \phi}{\partial z}\right) = S_\phi, \qquad (12)$$

where $\phi$ refers to the conserved scalar (mass, momentum, enthalpy), $x$, $y$, and $z$ represent the three coordinate
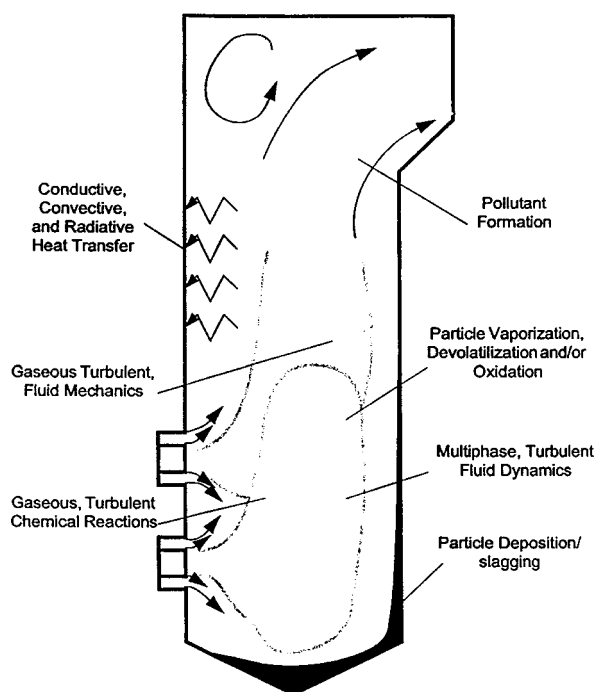
**FIGURE 11** Major physical and chemical mechanisms in a pulverized-coal combustion process. [Reproduced with permission from Eaton, A. M., Smoot, L. D., Hill, S. C., Eatough, C. N. (1999). "Components, Formulations, Solutions, Evaluation and Application of Comprehensive Combustion Models," *Progress Energy Combustion Sci.* **25,** 387–436.]

**TABLE VI   Variables Typically Considered in Comprehensive Simulations of Coal Boilers[a]**

| Reactor parameters | Independent variables |
|---|---|
| Configuration (e.g., shape, upflow, downflow) | Physical coordinates $(x, y, z)$ $(x, r, \theta)$ |
| Inlet configurations (e.g., location, presence of quarl) | Time $(t)$ |
| Inlet locations | Dependent variables |
| Dimensions | Gas species composition |
| Wall properties | Gas temperature |
| Wall thickness | Gas velocity |
| Wall temperature | Pressure |
| Boundary/initial conditions | Turbulent kinetic energy |
| Gas velocity | Turbulent energy dissipation rate |
| Gas composition | Mixture fraction (mean and variance) |
| Gas temperature | Particle composition |
| Gas mass flow-rate | Particle temperature |
| Pressure | Particle velocity |
| Particle velocity | Particle diameter |
| Particle composition | Gas density |
| Particle temperature | Gas viscosity |
| Particle size distribution | |
| Particle loading | |
| Particle bulk density | |

[a] Adapted from Smoot, L. Douglas, and Kramer, Stephen, K. (1995). "Combustion Modeling (Table 2)" *Energy Technology and the Environment*, **2,** (Bisio, A. and Boots, S. eds.) John Wiley and Sons, pp. 863–897.

directions (in a Cartesian coordianate system), $u$, $v$, and $w$ are component velocities, $\rho$ is the density, $\Gamma_\phi$ is the turbulent transport coefficient, and $S_\phi$ is the source term. Substitutions of algebraic approximations for the derivatives in such equations produces solvable equations that describe three-dimensional systems. To complete the model formulation, mathematical expressions are required for (1) gaseous reactions, including pollutant formation, (2) effects of turbulence on chemical reaction rates, (3) coal particle reaction rates, (4) radiative heat transfer, (5) temperature and composition dependence of physical properties, (6) particulate flow within the gas continuum, (7) behavior of inorganic matter, and (8) initial and boundary conditions. Recent reviews discuss these components in detail including code solution methods.

A list of typical dependent and independent variables for a furnace simulation is shown in Table VI. Coal particles typically are treated in a procedure that couples the continuum (Eulerian or fixed reference fame) gas-phase grid and the discrete (Lagrangian or moving reference frame) particles. Numerical solutions are repeated until the gas flow field is converged for the computed particle source terms, radiative fluxes, and gaseous reactions. Lagrangian particle trajectories are then calculated. After solving all particle-class trajectories, the new source terms are compared with the old values and the entire procedure is repeated until convergence is obtained. Use of comprehensive combustion models typically requires a technical specialist with appropriate training and experience.

## C. Code Evaluation—Field Tests

Data for validating pulverized coal combustion predictions requires accurate information for the reactor parameters shown in Table VI. Data measured in the combustion chamber typically include (1) locally measured values of the gaseous flow field velocity, temperature, and species composition, (2) coal particle burnout, number density, velocity, temperature, and composition, and (3) wall temperatures and heat fluxes. Evaluation should include comparisons with measurements from a wide variety of combustors and furnaces that range in scale from very small laboratory combustors (0.01–0.5 MW) and industrial furnaces (1–10 MW) to large utility boilers (up to 1000 MW).

Laboratory-scale data are very important to comprehensive model validation for several reasons. First, the nonintrusive, laser-based instrumentation needed to characterize the turbulent behavior in a flame is used more conveniently in smaller laboratory reactors. Second, a

wider range and better control of reactor operating conditions are possible for a smaller facility. Third, the entire reactor volume can be traversed with small incremental steps, so that much more detailed datasets can be obtained. Other advantages of small laboratory facilities are relatively low operating cost, flexibility and accessibility, and ability to control and to define carefully the boundary and inlet conditions. They can be large enough to give sufficient spatial resolution and to create a near-burner furnace environment, and small enough to utilize the advanced measurement techniques which are essential to providing accurate and complete data for model evaluation and for detailed understanding of combustion processes.

In large-scale facilities, detailed measurements are difficult to make, inlet conditions are often not well defined, operating costs for obtaining data can be high, and the facility may have instrumentation limitations. For example, detailed model evaluation measurements of species and temperature within or across large flames cannot easily be made, although such spatially resolved profile data are often best for comparing flame characteristics, near-field burner performance, and jet mixing behavior with predictions. Such measurements are essential for the evaluation of comprehensive combustion models. These detailed measurements provide important information concerning flame response to parametric variation. Effluent measurements (measurements at the outlet of the system or process) can be useful, particularly when effects of key system variables, such as excess air percentage, firing rate, or burner tilt angle, are measured. In some facilities, effluent data are the only kind available because of furnace size or access constraints, or because that was the only objective for making the measurements. Three-dimensional data at the utility furnace scale obtained with the specific intent of evaluating and validating model predictions have been collected at large-scale 85-$MW_e$ and 160-$MW_e$ corner-fired boilers of one U.S. utility.

## VII. THE FUTURE

Energy production from coal will face at least three major challenges in the future: Legislative reduction of allowable pollutant emissions, effects of deregulation of the power industry, and consideration of greenhouse gas emissions. The coal community has always faced pressure to reduce pollutant production; however, constraints relating to deregulation and greenhouse gas emissions are relatively new.

Emissions from coal-fired power plants have decreased substantially in the last decades due to improved understanding of pollutant formation mechanisms and combustion science. $NO_x$ is routinely reduced by 50–70% in major power stations, well over 99% of the particulate matter is filtered from the stack gases, and low-sulfur coals or sulfur scrubbers reduce $SO_2$ emissions by 40–90% compared to uncontrolled emissions. However, newer and tighter pollutant regulations portend continued pressure to reduce emissions. In particular, $NO_x$ emissions have not been reduced nearly as far as can be accomplished by advanced but proven technologies or as far as is necessary to prevent environmental harm such as acid rain, ozone reactions, or smog. Impending environmental standards will require installation of treatment systems such as selective catalytic reduction (SCR) of $NO_x$. SCR systems are established technology but have not been widely implemented in commercial operation in the United States due to the success of alternative and generally cheaper burner or boiler modifications. Germany and Japan are currently implementing SCR systems and the coming years should see large increases in the number of SCR installations or similarly effective $NO_x$ controls on commercial coal-fired power plants in most countries. Catalyst lifetime extension and catalyst regeneration are not as well understood as are SCR $NO_x$ reduction mechanisms, and these issues may play a prominent role in near-future plant operation. Proposed particle cleanup regulations targeted at particles less than 2.5 $\mu$m in size and mercury, arsenic, and other heavy or toxic metal regulations will also be major pollutant-related issues in the near future.

Many world power markets are deregulating, with power producers now able to compete for customers throughout the grid. As one consequence of this deregulation, power production cost is likely to become increasingly more important than production reliability or consistency. This has the potential to fundamentally change the way in which power plants operate and are managed. The intended consequence of deregulation is an overall decrease in the cost of power, but some regions with inadequate local production capacity have experienced rapid increases in cost and dramatic decreases in reliability during their early efforts to deregulate power systems. Regions with adequate supply from competing interests have made smoother transitions to deregulated systems. Issues associated with greenhouse gases have been discussed above. If nuclear power remains a socially or technically unacceptable form of power generation, it is difficult to foresee the availability of sufficient capacity for inexpensive power generation without coal continuing to play a major role.

To accomplish substantial reductions in greenhouse gas emissions, the efficiency of power production from coal will necessarily increase, the cost of power will increase as more renewable energies are developed, significant conservation of energy will be practiced, or methods of controlling/capturing $CO_2$ emissions will be deployed. A combination of these efforts is likely to be most effective. Efficiency increases as high as 50% (on paper) are achievable through advanced cycles such as
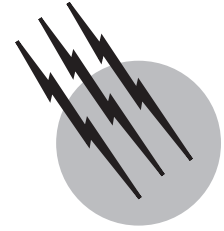
integrated gasification combined cycles and fuel cells. However, capital costs associated with these technologies may be twice those of conventional power systems and technical risks are very high. Since approximately half of the cost of coal-derived power is associated with capital investment, there is likely no amount of fuel savings that could economically compensate for a doubling of capital costs. Global warming concerns may provide the motivation for conservation, sequestration, and advanced power systems that are otherwise difficult to justify on economic terms alone.

## SEE ALSO THE FOLLOWING ARTICLES

BIOMASS UTILIZATION, LIMITS OF • ENERGY FLOWS IN ECOLOGY AND IN THE ECONOMY • ENERGY RESOURCES AND RESERVES • GREENHOUSE EFFECT AND CLIMATE DATA • POLLUTION, AIR • POLLUTION, ENVIRONMENTAL • RENEWABLE ENERGY FROM BIOMASS • THERMODYNAMICS

## BIBLIOGRAPHY

Abbas, T., Costen, P. G., and Lockwood, F. C. (1996). "Solid fuel utilization: From coal to biomass," *In* "Twenty-Sixth Symposium (International) on Combustion," pp. 3041–3058, Combustion Institute, Pittsburgh, PA.

Bartok, W., and Sarofim, A. F. (eds.) (1991). "Fossil Fuel Combustion: A Source Book," Wiley, New York.

Baxter, L. L. (1993). "Ash deposition during biomass and coal combustion: A mechanistic approach," Biomass Bioenergy **4**(2), 85–102.

Baxter, L. L., and DeSollar, R. W. (1993). "A mechanistic description of ash deposition during pulverized coal combustion: Predictions compared to observations," *Fuel* **72**(10), 1411–1418.

Baxter, L. L., and DeSollar, R. W. (eds.) (1996). "Application of Advanced Technology to Ash-Related Problems in Boilers," Plenum Press, New York.

Cengel, Y. A., and Boles, M. A. (1998). "Thermodynamics: An Engineering Approach," 3rd ed., McGraw-Hill, New York.

EIA. (1999). "Annual Energy Review 1998," Energy Information Administration, U.S. Department of Energy, Washington, DC.

Gupta, R. P., Wall, T. F., and Baxter, L. L. (eds). (1999). "Impact of Mineral Impurities in Solid Fuel Combustion," Plenum Press, New York.

Hart, R. H. (1998). "Structure, properties, and reactivity of solid fuels," *In* "Twenty-Seventh Symposium (International) on Combustion," pp. 2887–2904, Combustion Institute, Pittsburgh, PA.

Hinds, W. C. (1999). "Aerosol Technology: Properties, Behavior and Measurements of Airborne Particles," 2nd ed., Wiley, New York.

Leckner, B. (1996). "Fluidized bed combustion: Achievements and problems," *In* "Twenty-Sixth Symposium on Combustion," pp. 3231–3241, Combustion Institute, Pittsburgh, PA.

Niksa, S. (1996). "Coal Combustion Modeling," IEA PR/31, IEA Coal Research, London.

Patankar, S. V. (1980). "Numerical Heat Transfer and Fluid Flow," Hemisphere, Washington, DC.

Rosner, D. E. (2000). "Transport Processes in Chemically Reacting Flow Systems," Dover, New York.

Seinfeld, J. H., and Pandis, S. N. (1998). "Atmospheric Chemistry and Physics," Wiley, New York.

Smith, K. L., Smoot, L. D., Fletcher, T. H., and Pugmire, R. J. (1994). "The Structure and Reaction Processes of Coal," Plenum Press, New York.

Smoot, L. D. (ed.) (1993). "The Fundamentals of Coal Combustion," Elsevier, Amsterdam.

Smoot, L. D. (1998). "International research centers' activities in coal combustion," *Prog. Energy Combustion Sci.* **24,** 409–501.

Solomon, P. R., and Fletcher, T. H. (1994). "Impact of coal pyrolysis on combustion," *In* "Twenty-Fifth Symposium (International) on Combustion," pp. 463–474, Combustion Institute, Pittsburgh, PA.

Tree, D. R., Black, D. L., Rigby, J. R., McQuay, M. Q., and Webb, B. W. (1998). "Experimental measurements in the BYU controlled profile reactor," *Prog. Energy Combustion Sci.* **24,** 355–384.

Wall, T. F. (1992). "Mineral matter transformations and ash deposition in pulverised coal combustion," *In* "Twenty-Fourth Symposium (International) on Combustion," Combustion Institute, Pittsburgh, PA.

Willeke, K., and Baron, P. A. (eds.). (1993). "Aerosol Measurement: Principles, Techniques and Applications," Wiley, New York.

Williams, A., Pourkashanian, M., Jones, J. M., and Skorupska, N. (2000). "Combustion and Gasification of Coal," Taylor and Francis, New York.

# Gas Hydrate in the Ocean Environment

## William P. Dillon
*U.S. Geological Survey*

## GLOSSARY

**Biogenic methane** Methane formed by microbial processes.

**BSR** An abbreviation for the so-called "bottom simulating reflection." A reflection recorded in seismic reflection profiles that results from an acoustic velocity contrast produced by the decrease in sound speed caused primarily by the presence of gas trapped beneath the gas hydrate stability zone. BSRs provide a remotely sensed indication of the presence of gas hydrate.

**Dissociation** The breakdown of gas hydrate into its components, water and gas, commonly resulting from some combination of pressure reduction and temperature increase.

**Gas hydrate or clathrate** A solid crystalline structure formed of cages of water molecules; a cage commonly encloses and is supported by a gas molecule.

**Gas hydrate stability zone** The region within deep ocean water and seafloor sediments where gas hydrate is stable. Gas hydrate commonly occurs in the zone from the seafloor down to some depth, perhaps several hundred meters below the seafloor, where temperature/pressure changes result in conditions where gas hydrate is unstable.

**Geothermal gradient** The rate at which temperature increases downward into the solid Earth.

**Phase boundary** The interface in temperature/pressure space where a phase change occurs. In the case of gas hydrate, the phase change is from stability of solid gas hydrate to stability of gas + water.

**Seismic reflection profile** An image of the structure below the seafloor along a ship trackline that is created by using sound pulses and recording the time of return from various reflecting layers.

**Thermogenic methane** Methane formed from organic matter by heating at sediment depths exceeding about 1 km.

**A GAS HYDRATE**, also known as a gas clathrate, is a gas-bearing, icelike material. It occurs in abundance in marine sediments and stores immense amounts of methane, with major implications for future energy resources and global climate change. Furthermore, gas hydrate controls some of the physical properties of sedimentary deposits and thereby influences seafloor stability.

## I. THE NATURE OF GAS HYDRATE

Gas hydrate or gas clathrate is a crystalline solid; its building blocks consist of a gas molecule surrounded by a cage of water molecules. The water molecules are connected by hydrogen bonds, but the gas molecule is not bonded; rather it is simply trapped within the cage and held by Van der Waal forces (the term "clathrate" refers to a cage or lattice-like structure). Because the material is formed of hydrogen-bonded water molecules, it looks like ice and is very similar to ice, although of different crystallographic form, having its crystal structure stabilized by the guest gas molecules. A conceptual model of a gas hydrate cage is shown in Fig. 1, containing a methane molecule. This is the most common gas hydrate structure, which is classified as structure I. Structure I gas hydrate has the smallest cavity sizes and is comfortable holding methane molecules or other gas molecules of similar size, such as nitrogen, oxygen, carbon dioxide, and hydrogen sulfide. The structure of the clathrate cages varies depending on the size of the guest molecule, and some molecules are too small (smaller than argon) or too big to exist in any of the gas hydrate structures. If larger gas molecules are available, at least two other gas hydrate structures are known. These are struc-
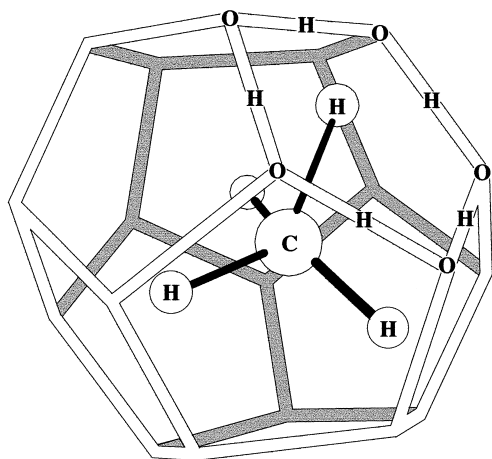


**FIGURE 1** Diagrammatic concept of a structure I gas hydrate cage holding a methane molecule. H, O, and C represent hydrogen, oxygen, and carbon atoms. Double lines, filled or unfilled, represent chemical bonds.

ture II, holding molecules up to the size of propane and isobutane, and structure H, holding even larger molecules like isopentane and neohexane. Actually, more than one cavity size exists in a structure, and thus, a mixture of gases can be accommodated. Because the gas molecules are held within a crystal lattice, they may be held closer together than in the gaseous state, at least at lower pressures, so gas hydrate can act as a concentrator of gas. For example, a unit volume of methane hydrate at one atmosphere (and $0°C$) can hold 163 volumes of methane gas at the same conditions.

Many gas hydrates are stable in the deep ocean conditions, but methane hydrate is by far the dominant type, making up >99% of gas hydrate in the ocean floor. The methane is almost entirely derived from microbial methanogenesis, predominantly through the process of carbon dioxide reduction. In some areas, such as the Gulf of Mexico, gas hydrates are also created by other thermogenically formed hydrocarbon gases and other clathrate-forming gases such as hydrogen sulfide and carbon dioxide. Such gases escape from sediments at depth, rise along faults, and form gas hydrate at or just below the seafloor, but on a worldwide basis these are of minor volumetric importance compared to methane hydrate. Methane hydrate exists in several forms in marine sediments. In coarse-grained sediments it often forms as disseminated grains and pore fillings, whereas in finer silt/clay deposits it commonly appears as nodules and veins. Gas hydrate also is observed as surface crusts on the seafloor.

## II. ENVIRONMENT OF STABILITY

Gas hydrate forms wherever appropriate physical conditions exist—moderately low temperature and moderately high pressure—and the materials are present—gas near saturation and water. These conditions are found in the deep sea commonly at water depths greater than about 500 m or somewhat shallower depths (about 300 m) in the Arctic, where bottom-water temperature is colder. Gas hydrate also occurs beneath permafrost on land in arctic conditions, but, by far, most natural gas hydrate is stored in ocean floor deposits. A simplified phase diagram is shown in Fig. 2A, in which pressure has been converted to water depth in the ocean (thus, pressure increases downward in the diagram). The heavy line in Fig. 2A is the phase boundary, separating conditions in the temperature/pressure field where methane hydrate is stable to the left of the curve (hatched area) from conditions where it is not. In Fig. 2B, some typical conditions of pressure and temperature in the deep ocean were chosen to define the region where methane hydrate is stable. The phase boundary indicated is the same as in Fig. 2A, so methane hydrate is stable
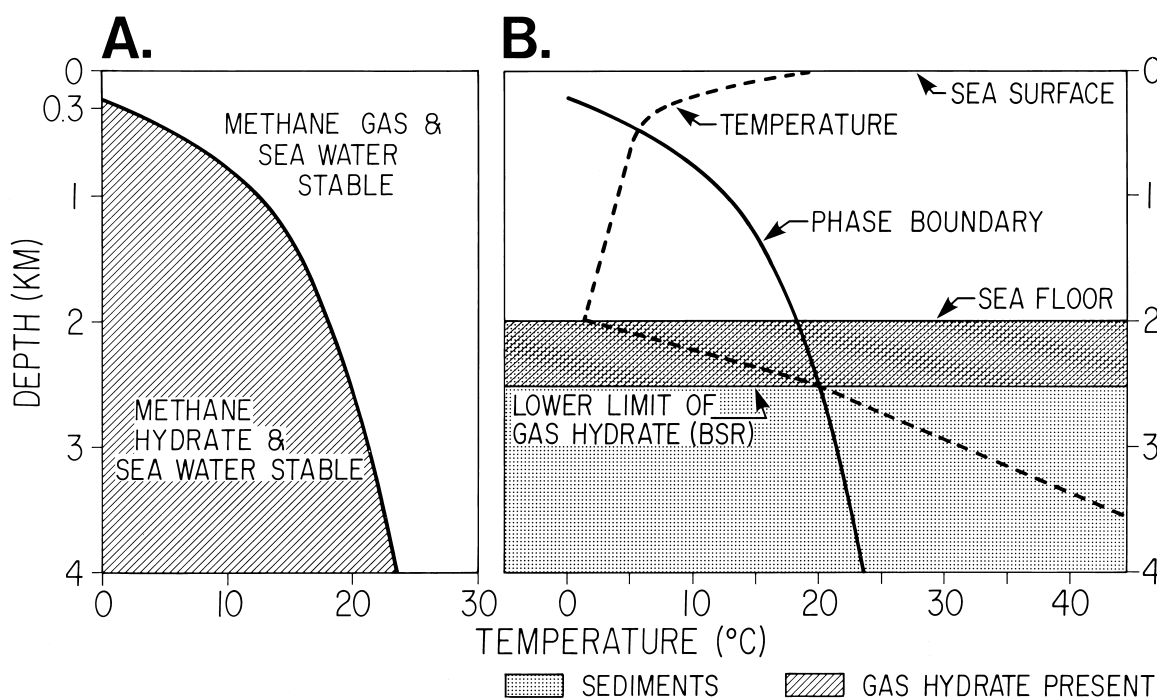
**FIGURE 2**  (A) Phase boundary of methane hydrate in the ocean (solid line). The pressure axis has been converted to depth into the ocean, so pressure increases downward. (B) The same phase boundary as shown in (A) with a seafloor inserted at 2 km depth and a typical temperature curve (dashed line). Hatched region shows the vertical extent of the gas hydrate stability zone under these assumed conditions.

at locations in the pressure/temperature field where conditions plot to the left of the phase boundary. The dashed line shows how temperature conditions typically vary with depth in the deep ocean and underlying sediments. In this case typical western North Atlantic Ocean thermal conditions were chosen and a seafloor at 2 km water depth (Fig. 2B) was selected. Near the ocean surface, temperatures are too warm and pressures too low for methane hydrate to be stable. Moving down through the water column, temperature decreases and an inflection in the temperature curve is reached, known as the main thermocline, which separates the warm surface water from the deeper cold waters. At about 500 m, the temperature and phase boundary curves cross; from there downward, temperatures are cold enough and pressures high enough for methane hydrate to be stable in the ocean. This intersection would occur at a shallower depth in colder, arctic waters.

If methane is sufficiently concentrated (near saturation), gas hydrate will form. However, like ice, the density of crystalline methane hydrate is less than that of water (about 0.9), so if such hydrate formed in the water (e.g., at methane seeps) it would float upward and would dissociate when it crossed the depth where the curves intersect. However, if the gas hydrate formed within sediments, it would be bound in place. Minimum temperature occurs

at the seafloor (Fig. 2). Downward through the sediments, the temperature rises along the geothermal gradient toward the hot center of the Earth. At the point where the curve of conditions in the sediments (dashed line) crosses the phase boundary, one reaches the bottom of the zone where methane hydrate is stable.

The precise location of the base of the gas hydrate stability zone under known pressure/temperature conditions varies somewhat depending on several factors, most important of which is gas chemistry. In places where the gas is not pure methane, for example, in the Gulf of Mexico, at a pressure equivalent to 2.5 km water depth, the base of the gas hydrate stability zone will occur at about 21°C for pure methane, but at 23°C for a typical mixture of approximately 93% methane, 4% ethane, 1% propane, and some smaller amounts of higher hydrocarbons. At the same pressure (2.5 km water depth) but for a possible mixture of about 62% methane, 9% ethane, 23% propane, plus some higher hydrocarbons, the phase limit will be at 28°C. These differences will cause major shifts in depth to the base of the gas hydrate stability zone as would be implied by Fig. 2B. Such mixtures of gases essentially make the formation of gas hydrate easier and therefore can result in the formation of gas hydrate near the seafloor at shallower depths (lower pressures) than for methane hydrate
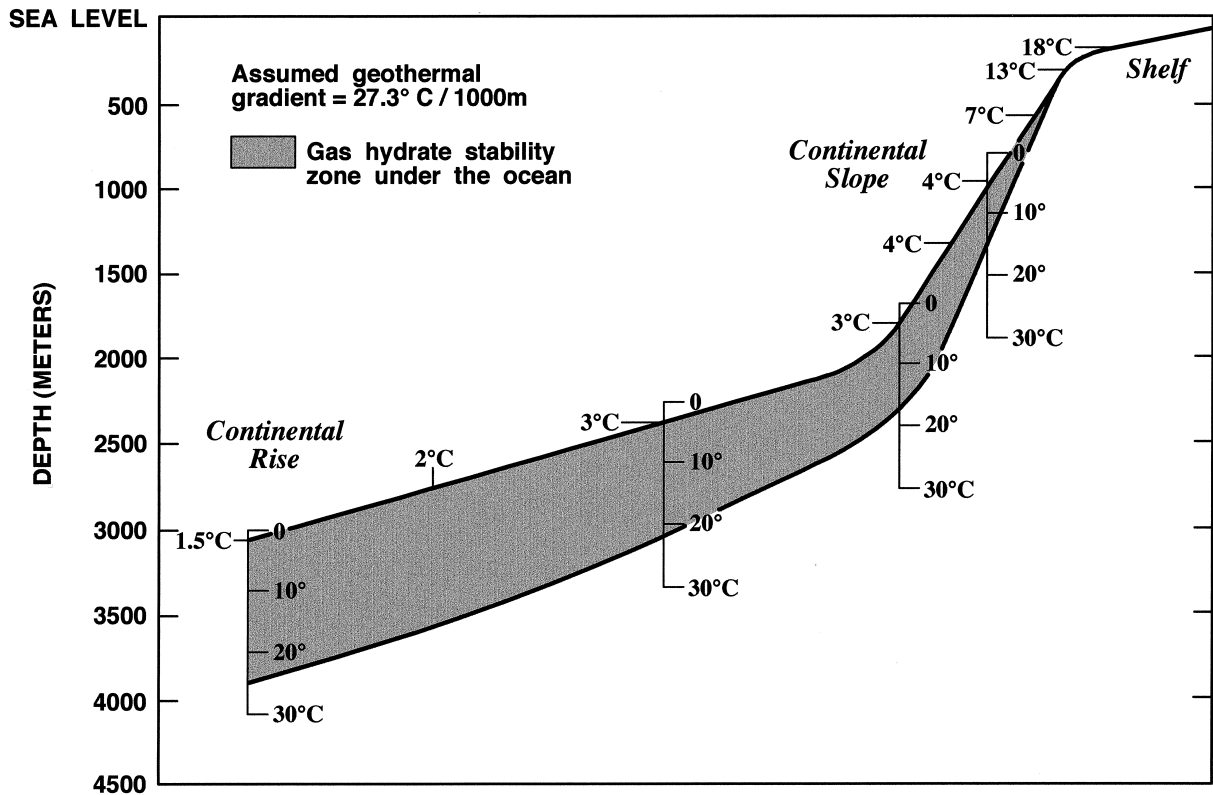
**FIGURE 3** Assumed thickness of the gas hydrate stability zone for changing water depth (pressure) and a constant geothermal gradient. Factors that change the chemistry (changes in gas composition, salinity variations, etc.) or that affect the thermal structure (landslides, salt diapirs) can complicate this simple picture. [From Kvenvolden, K. A., and Barnard, L. A. (1982). Hydrates of natural gas in continental margins, *In* "Studies in Continental Margin Geology" (J. S. Watkins, and C. L. Drake, eds.), Am Assoc. Petroleum Geologists Memoir No. 34, pp. 631–640.]

at equal temperatures. Below the base of the gas hydrate stability zone (500 m in our example in Fig. 2B), methane and water will be stable and methane hydrate will not be found.

The geothermal gradient tends to be quite uniform across broad regions where sediments do not vary. Thus, for a given water depth, the sub-bottom depth to the base of the gas hydrate stability zone will be quite constant. However, because a change in water depth causes change in pressure, we anticipate that the base of the gas hydrate stability zone will extend further below the seafloor as water depth increases (Fig. 3).

## III. PRESENCE IN NATURE

Two basic issues that concern us about gas hydrate are: (1) where does it exist and (2) how much is present? These questions have turned out to be very difficult to answer in a precise manner because gas hydrate exists beneath the ocean floor in conditions of temperature and pres-

sure where human beings cannot survive, and if gas hydrate is transported from the sea bottom to normal Earth-surface conditions, it dissociates. Thus, drilled samples cannot be depended upon to provide accurate estimates of the amount of gas hydrate present, as would be the case with most minerals. Even the heat and changes in chemistry (methane saturation, salinity, etc.) introduced by the drilling process, including the effect of circulating drilling fluids, affect the gas hydrate, independent of the changes brought about by moving a sample to the surface. Gas hydrate has been identified in nature generally from drilling data or by using remotely sensed indications from seismic reflection profiles.

### A. Identifying Gas Hydrate in Nature

#### 1. Measuring Gas Hydrate in Wells

Drilled samples of gas hydrate have survived the trip to the surface from hydrate accumulations below the seafloor, despite the transfer out of the stability field, just because the dissociation of gas hydrate is fairly slow. Such samples

have been preserved, at least temporarily, by returning them to the pressure/temperature conditions where they formed or, more commonly, by keeping them at surface pressure, but at the ultra-cold temperature of liquid nitrogen. Such preserved samples are valuable for many studies, but certainly they do not provide quantitatively accurate indications of the amount or distribution of gas hydrate or its relation to sediment fabric that existed in the undisturbed seafloor sediments.

Several indirect approaches have been used to gain indication of the amount and/or distribution of gas hydrate in the sediments. Some are as obvious as making temperature measurements along a core to indicate where gas hydrate existed, because the dissociation of gas hydrate, being an endothermic reaction, will leave cold spots.

The dissociation of gas hydrate leaves another indirect marker of its former existence, because when gas hydrate forms it extracts pure water to form the clathrate structure, excluding all salts as brine. Therefore, when hydrate dissociates in a core, the interstitial water becomes much fresher, and the amount of gas hydrate present before dissociation can be calculated. An example is shown in Fig. 4 from Ocean Drilling Program hole 997 off the South Carolina coast. Measured chloride content is shown in the left panel. The values near zero depth (depth at the seafloor) represent seawater chloride concentration (chlorinity). It is assumed that a smooth curve of chlorinity vs depth, following the main trend of data points (solid curve), rep-

resents the undisturbed (predrilling) chlorinity and that spikes of low chlorinity to the left of the curve represent the result of hydrate dissociation. The base of hydrate stability here is at 450 m below the seafloor, and the top of significant gas hydrate concentrations is apparently about 200 m. The estimation of the base curve is still a controversial issue in using this method. The second panel shows the base curve straightened out (dashed line) so that the plot is considered to represent the chlorinity anomaly that results from dissociation of previously existing hydrate. The third panel shows the calculated results in terms of proportion of sediment that had been occupied by gas hydrate.

A third approach that is important for identifying and quantifying gas hydrate in wells is the downhole logging methods in which sensors are pulled through the hole and measurements are made. These measurements can give porosity information about the sediments, using gamma-rays or neutron flux, and provide indications about gas hydrate concentration, using electrical resistivity or acoustic velocity measurements.

## 2. Sensing Gas Hydrate in Seismic Reflection Profiles

Beyond the issue of precisely quantifying gas hydrate in the few spots on the seafloor where wells have been drilled, we would like to have a technique of remotely sensing gas hydrate without drilling a hole and by some means that
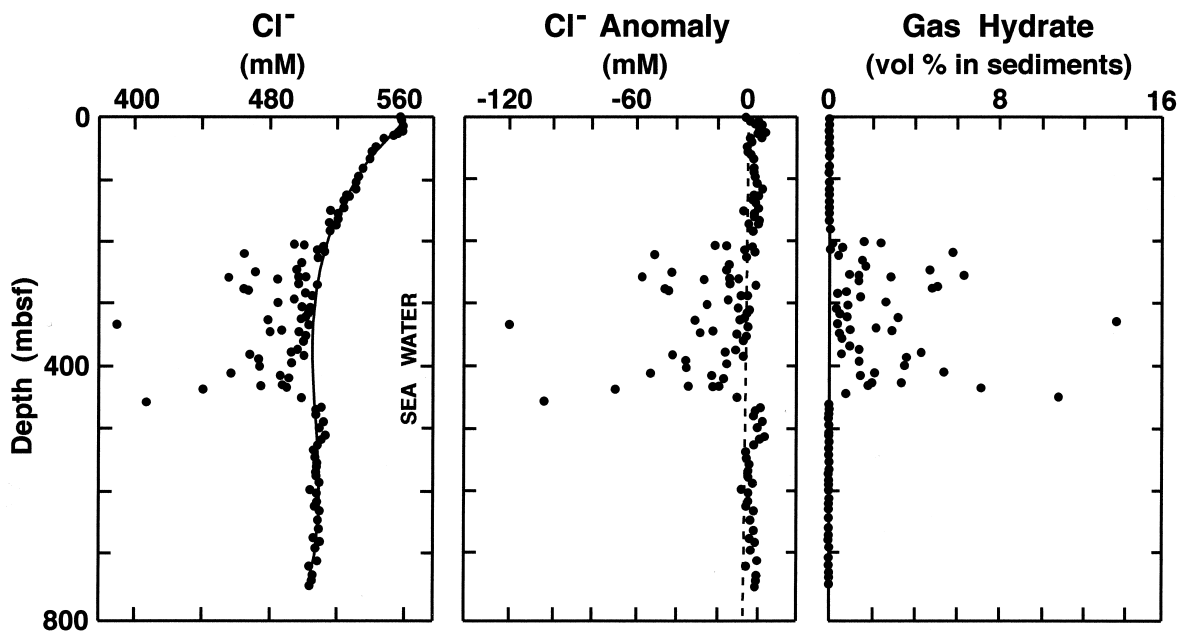


**FIGURE 4**   Chloride content, chloride anomaly, and volume of gas hydrate calculated from this information for Ocean Drilling Program hole 997, Blake Ridge, off the southeastern United States. [From Paull, C. K., Matsumoto, R., and Wallace, P. J. *et al.* (1996). "Proceedings of the ODP Initial Reports," Vol. 164, Figure 47, p. 317, College Station, TX.]

would allow us to survey large areas. The only effective approaches that seem to fulfill these requirements are the use of acoustic imaging of the near-bottom sediments. Most commonly, this means using seismic reflection profiling to image the sediments in the zone of gas hydrate stability just below the seafloor. A seismic reflection profile is created by steaming a ship along a line and triggering a sound source in the water every few seconds. The sound source for gas hydrate studies can be a fairly small one, as the need is to provide penetration only down to the base of the gas hydrate stability zone, which is unlikely to be more than 1000 m below the seafloor. Use of a small seismic source, which provides higher frequency sound, is also consistent with obtaining the highest possible resolution. The sound source is often an "airgun," which is a small high-pressure air chamber (commonly 100–160 in.[3] volume, which is equivalent to 1.6–2.6 liters), from which the air can be abruptly dumped by an electrical signal. Hydrophones trailed in the water behind the ship receive the echoes of each sound pulse from the seafloor and below, and the echoes are recorded on the ship for computer processing into a profile. In a sense, any seismic profile represents a cross section through the seabed, but keep in mind that the image is created by reflections of sound from reflectors formed by density and acoustic velocity changes that are inherent to sedimentary layers and gas hydrate accumulations. The vertical axis is imaged in the travel time of sound.

Fortunately, the base of the the gas hydrate stability zone is often easy to detect in seismic reflection profiles. Free

gas bubbles commonly accumulate, trapped just beneath the base of the gas hydrate stability zone, where free gas is stable and gas hydrate will not exist. Presence of bubbles in intergranular spaces reduces the acoustic velocity of the sediment markedly. Conversely, in the gas hydrate stability zone the velocity is increased slightly by the presence of gas hydrate, which in the pure state has twice the velocity of typical deep sea sediments, and furthermore, bubbles generally cannot be present because any free gas would convert to gas hydrate as long as water is present. A large velocity contrast generates a strong echo when an acoustic pulse impinges on it. Thus, we can receive an image of the base of the gas hydrate stability zone in a seismic reflection profile. The base of gas hydrate stability, as disclosed by this reflection, generally occurs at an approximately uniform sub-bottom depth throughout a restricted area because it is controlled by the temperature. Thermal gradients across an area tend to be consistent, so isothermal surfaces have consistent depth. Hence, the reflection from the base of the gas hydrate stability zone roughly parallels the seafloor in seismic profiles and has become known as the "Bottom Simulating Reflection" (BSR) (Fig. 5). The BSR is a sure sign that gas exists trapped beneath the base of the gas hydrate stability zone and strongly implies that gas hydrate is present, because free gas, which has a tendency to rise, exists just below and in contact with the zone where gas would be converted to gas hydrate. The coincidence in depth of the BSR to the theoretical, extrapolated pressure/temperature conditions that define the base of gas hydrate stability and
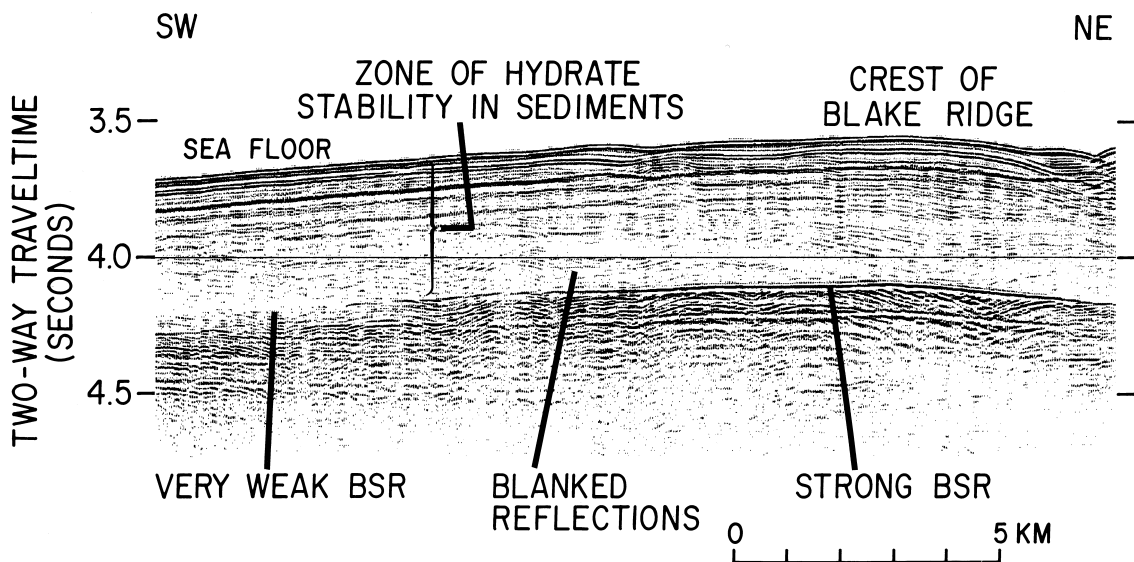


**FIGURE 5** Seismic reflection profile near the crest of the Blake Ridge off South Carolina in the southeastern United States. Note the strong reflection marked BSR that defines the base of the gas hydrate stability zone. The vertical axis is in two-way travel time of sound, which varies with sound velocity in the medium; thus, this axis is a variable scale with respect to distance.

**TABLE I   Global Organic Carbon Distribution in Giga-tons (Gt = 10¹⁵ g)[a]**

| Gas hydrate (methane) | 10,000 | Land | |
|---|---|---|---|
| | | Biota | 830 |
| | | Detritus | 60 |
| Fossil fuels | 5,000 | Soil | 1400 |
| | | Peat | 500 |
| Atmosphere (methane) | 3.6 | | |
| | | Ocean | |
| | | Marine biota | 3 |
| Dispersed carbon | 20,000,000 | Dissolved | 980 |

[a] Data from Kvenvolden (1988).

the sampling of gas hydrate above BSRs and gas below give confidence that this seismic indication of the base of gas hydrates is dependable.

A second significant seismic characteristic of gas hydrate cementation, called "blanking," is also displayed in Fig. 5; blanking is the reduction of the amplitude (weakening) of seismic reflections, which appears to be caused by the presence of gas hydrate. Many observations of blanking in nature have been associated with gas hydrate accumulations, but the reality of blanking has been questioned because an undisputed physical model for it has not yet been identified.

## B. Estimated Quantities of Gas Hydrate Worldwide

Many estimates of the global amount of gas hydrate have been made. Keith Kvenvolden of the U.S. Geological

Survey recently did a survey of these and looked at their variability over time. In 1980 the estimates covered a very broad range, but since 1988 the range of disagreement has been considerably reduced, even though the methods used by ten different sets of workers have varied widely. From 1988 to 1997 the estimates of worldwide methane content of gas hydrate range from 1 to $46 \times 10^{15}$ m³, with a "consensus value" of $21 \times 10^{15}$ m³.

How does this amount of methane compare to other stores of carbon in the natural environment? Table I attempts to answer that question by comparing the amount of organic carbon in various reservoirs on Earth. Obviously, by far, the largest amount of organic carbon exists dispersed in the rocks and sediments. However, consider the first three numbers (Table I). The organic carbon in gas hydrate is estimated to be twice as much as in all fossil fuels on Earth (including coal). Organic carbon in gas hydrate is also estimated to be about 3000 times the amount in the atmosphere, and here, we are directly comparing methane to methane. As methane is a very powerful greenhouse gas, this reservoir of methane, which might be released to the atmosphere, may have significant implications.

## C. Environments of Gas Hydrate Concentration

The presence of gas hydrate, identified on the basis of drilling and seismic reflection profiling, has been reported all around the world, but most commonly around the edges of the continents (Fig. 6). Methane accumulates in
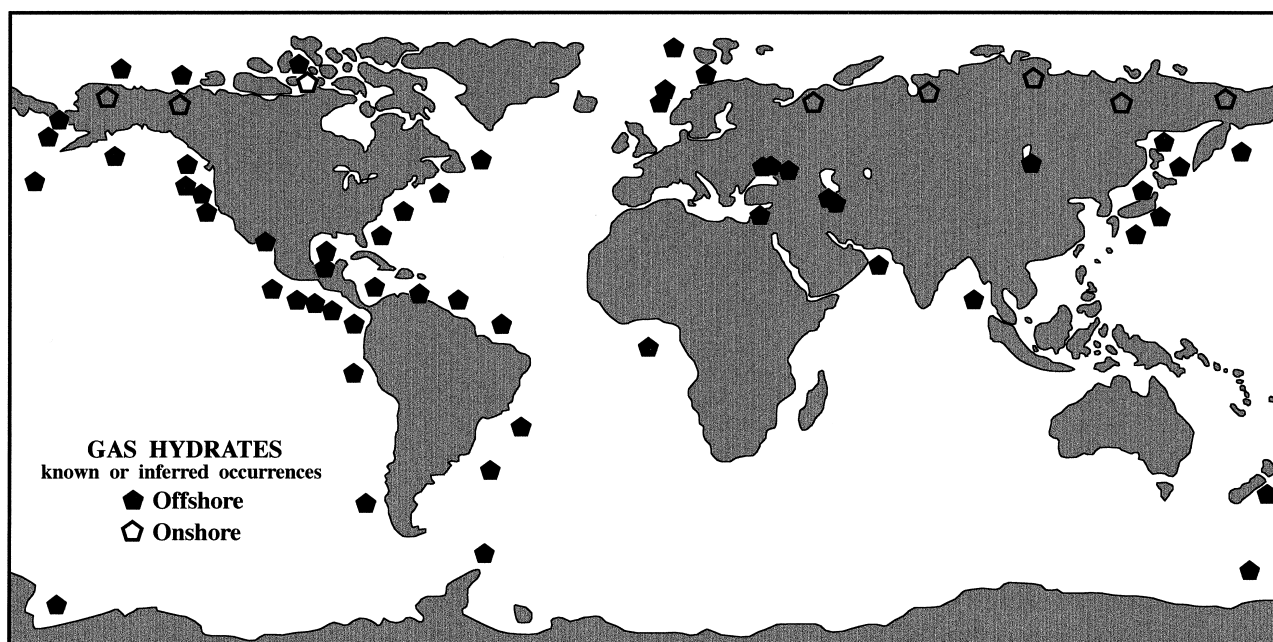


**FIGURE 6** Global distribution of confirmed or inferred gas hydrate sites, 1997 (courtesy of James Booth, U.S. Geological Survey). This information represents our very limited knowledge. Gas hydrate probably is present in essentially all continental margins.

continental margin sediments probably for two reasons. First, the margins of the oceans are where the flux of organic carbon to the seafloor is greatest because oceanic biological productivity is highest there and organic detritus from the continents also collects to some extent. Second, the continental margins are where sedimentation rates are fastest. The rapid accumulation of sediment serves to cover and seal the organic material before it is oxidized, allowing the microorganisms in the sediments to use it as food and form the methane that becomes incorporated into gas hydrate. In addition to sites in the marine environment, the map (Fig. 6) also shows some sites where gas hydrate has been reported associated with permafrost in the Arctic, the one location where it has been found in fresh water in Lake Baikal, Siberia, and in intermediate salinity (between fresh and oceanic salinities) in the Caspian Sea.

Gas hydrate concentration commonly increases downward through the gas hydrate stability zone in marine sediments, based on the evidence of well and seismic reflection profiling data. For example, note the gas hydrate distribution described by chlorinity in Fig. 4 and the increase in blanking downward toward the BSR (base of the gas hydrate stability zone) in Fig. 5. The increase in concentration toward the base of the gas hydrate stability zone suggests that methane is being supplied below the base of the gas hydrate stability zone by some process and that the gas is probably being trapped there before entering the lower part of the gas hydrate stability zone. The presence of trapped gas just beneath the base of the gas hydrate stability zone is demonstrated by the common occurrence of BSRs and has been observed in drilling. This methane may rise from below, either because it is being generated by microbes at greater depth in the sediment or because of solubility considerations. Alternatively, it may be recycled from above. Many continental margin settings have ongoing sediment deposition. As the seafloor builds up, the geothermal gradient tends to remain constant, so the isothermal surfaces must rise with the accreting seafloor. A sediment grain or bit of gas hydrate in the shallow sediments effectively sees the gas hydrate stability zone migrate upward past it as the seafloor builds up. Eventually, that bit of gas hydrate ends up far enough below the seafloor that it is beneath the base of the gas hydrate stability zone and thus outside the range of gas hydrate stability. Then the gas hydrate dissociates and releases its methane, which will tend to rise through the sediments because of the low density of gas and accumulate at the base of the gas hydrate stability zone, ultimately to work its way up into the gas hydrate stability zone, where it forms more gas hydrate.

## IV. EFFECT ON SEAFLOOR SLIDES AND SLUMPS

When gas hydrate forms or dissociates within seafloor sediments, this has major effects on sediment strength. Dissociation converts a solid material, which sometimes at high concentrations acts as a cement, to gas and water. Only one triaxial shear strength test of a natural gas hydrate-bearing sediment has been made at *in situ* conditions (at the U.S. Geological Survey, Gas Hydrate And Sediment Test Laboratory Instrument—GHASTLI; Booth, Winters, and Dillon, 1999), but this confirmed the tremendous increase in strength that was expected from gas hydrate in high concentrations. However, in most locations in natural seafloor sediments, hydrate concentrations are fairly low (<5%), so the cementing effect is minimal. The effect of dissociation of gas hydrate on sediment strength can still be immense, though, because when hydrate breaks down at shallower depths, the products of breakdown, gas + water, occupy greater volume than the gas hydrate they were derived from, and thus, the dissociation will increase the internal pressure in the pore space. Such pressures are called "overpressures," which are pressures greater than the column of water plus sediment above the spot. Generation of overpressures weakens the sediment significantly and is likely to initiate sediment slides and slumps on continental slopes and rises.

Dissociation of gas hydrate that leads to weakening of sediments can result from pressure reduction or warming (Fig. 2). Obviously, warming will occur if ocean bottom waters warm up. However, gas hydrate will only dissociate at its phase boundary. In Fig. 2B, for example, the gas hydrate at the seafloor exists at pressure/temperature conditions some considerable distance within its zone of stability, so a few degrees of warming will not cause it to dissociate. The deep limit of gas hydrate stability at conditions shown in Fig. 2B is about 500 m below the seafloor. If bottom water became abruptly warmer, the warming front would have to propagate downward through the sediment to the depth where the gas hydrate is at the stability limit, which might take hundreds or thousands of years. The change would have to occur as a conductive heat flow, as downward flow of water that could transfer (advect) heat is extremely limited in ocean sediments. Obviously, the place where warming of bottom water will have a rapid influence is where the base of the gas hydrate stability zone is very close to the seafloor (see Fig. 3).

Atmospheric warming is presently occurring. Global surface air temperature has probably increased by roughly $0.8°C$ over the last century. This warming is probably being transferred to the ocean in a manner comparable to chemical tracers that have been observed, which means

that warmed surface water can be expected to circulate down to depths of the shallower gas hydrates in several tens of years. In specific cases in the Gulf of Mexico, where warm bottom currents sometimes sweep through the region, and off northern California, active dissociation of gas hydrate at the seafloor has been observed (evidence is the absence of previously observed gas hydrate and release of methane bubbles). Some of this activity has been related to identifiable water temperature changes, so present atmospheric warming may be leading to hydrate dissociation.

Dissociation of oceanic gas hydrate by pressure reduction clearly must also occur because a lowering of sea level must cause an instantaneous reduction of pressure at the seafloor and down into the sediments. Pressure is dependent on the weight of a column of water and sediment above a spot, and if that changes there is no delay in changing pressure at all depths, as there is in changing temperature. Thus, a lowering of sea level, as occurred during buildup of continental icecaps that removed water from the ocean, will immediately reduce the pressure at the base of the gas hydrate stability zone and cause gas hydrate to dissociate. The most recent glacial sea level lowering of about 120 m ended about 15,000 years ago and must have caused significant dissociation of gas hydrate.

The result of hydrate dissociation on a continental slope is diagrammed in Fig. 7. The initial hydrate breakdown releases a slide mass. When the slide takes place, the removal of sediment reduces the load on the sediment that was be-

low it and, thus, creates another pressure reduction that may cause further gas hydrate dissociation, resulting in cascading submarine slides. Circumstantial evidence that gas hydrate processes and submarine landslides are related is provided by a map of the shallow limit of gas hydrate stability compared to the tops of known landslide scars on the U.S. Atlantic margin that is shown in Fig. 8. Clustering of slope failures within the zone of gas hydrate stability, just below its upper limit, is suggestive of a relationship comparable to that diagrammed in Fig. 7. Even on relatively flat slopes where slides are not triggered, evidence for buildup of pressure due to gas hydrate dissociation; mobilization of gas + water-rich sediments; and escape of methane, water, and sediment has been interpreted. This evidence consists of apparent blowout structures (craters with chaotically disturbed sediments beneath, sometimes with associated ejecta) in the Gulf of Mexico and the U.S. Atlantic margin.

The dissociation of gas hydrate, resulting in gas and water release and sediment deformation, can also be a safety issue in petroleum exploration and production. Throughout the world, oil drilling is moving to deeper water where gas hydrate can be anticipated. In the marine environment, gas hydrate commonly is concentrated near the base of the gas hydrate stability zone. If the gas hydrate is dissociated by warm drilling fluid, gas can enter the wellbore and circulating fluid. This reduces the density of drilling muds, which are placed in the hole to maintain high pressure at depth, and the reduced pressure encourages further dissociation of hydrate. Muds have foamed
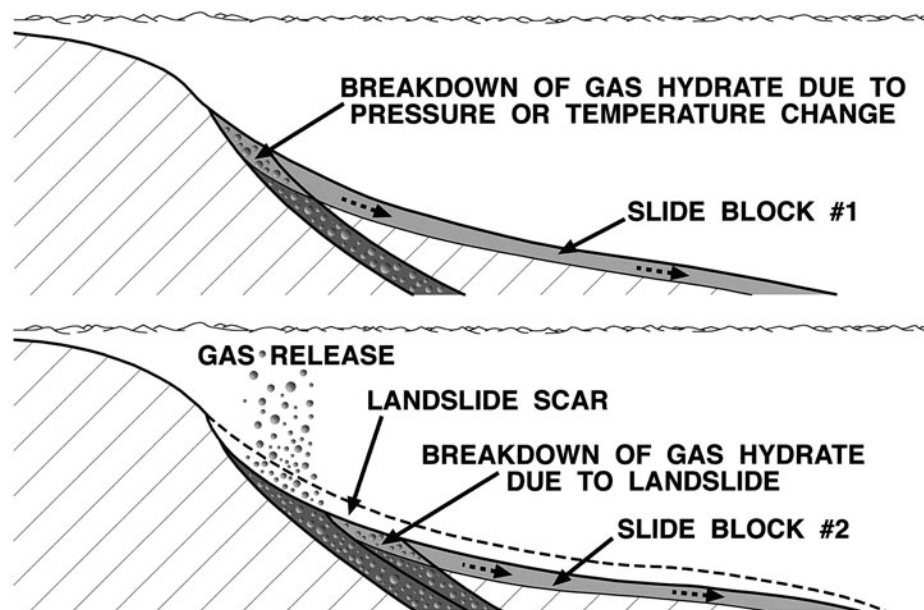


**FIGURE 7** Conceptual drawing of the triggering of a landslide and release of methane due to gas hydrate dissociation.

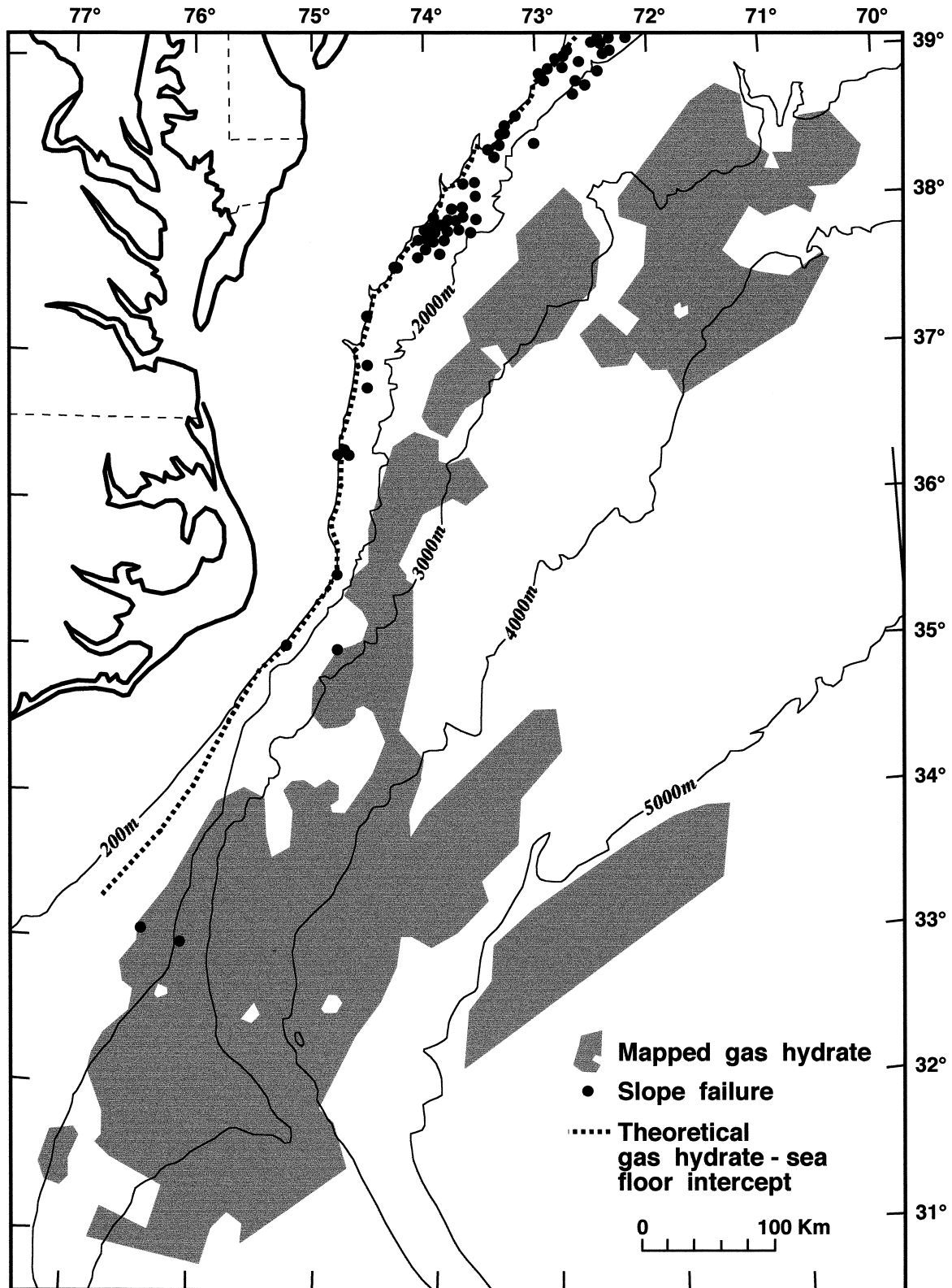**FIGURE 8**  Theoretical shallow limit of gas hydrate stability (dashed line) and distribution of mapped slope failures on the Atlantic continental margin of the United States. [From Booth, J. S., Winters, W. J., and Dillon, W. P. (1994). "Circumstantial evidence of gas hydrate and slope failure associations on the United States Atlantic continental margin." *Ann. N.Y. Acad. Sci.* **715,** 487–489].

and even been completely lost from the hole in gas hydrate areas. However, anticipation of the problems allows proper engineering of the drilling, and such issues can be avoided by controlling temperatures and pressures. During the production phase, flow of warm fluids from below the gas hydrate stability zone upward through the conductor pipe could warm adjacent sediments, which could create an expanding dissociation front around the pipe, possibly resulting in such problems as subsidence around the conductor pipe, escape of gas, loss of support for the platform foundation, or erosion by rafting away of sediment with newly formed gas hydrate at the sediment surface. The same considerations of dissociation by warming and loss of support can affect pipelines. Again, control is possible as long as the situation is recognized in advance.

## V. EFFECT ON CLIMATE

The gas hydrate reservoir in the ocean sediments has significant implications for climate because of the vast amount of methane situated there and the strong greenhouse warming potential of methane in the atmosphere. It has been suggested that ancient climate shifts, such as the Late Paleocene Thermal Maximum, might have been caused by the release of methane associated with gas hydrate. A reasonably conservative estimate (see Table I) suggests that there is roughly 3000 times as much methane in the gas hydrate reservoir as there is in the present atmosphere. Methane absorbs energy at wavelengths that are different from other greenhouse gases, so a small addition of methane can have important effects. If a mass of methane is released into the atmosphere, it will have an immediate greenhouse impact that will slowly decrease as the methane is oxidized to carbon dioxide in the air. The global warming potential of methane is calculated to be 56 times by weight greater than carbon dioxide over a 20-year period. That is, a unit mass of methane introduced into the atmosphere would have 56 times the warming effect of an identical mass of carbon dioxide over that time period. Because of chemical reactions (oxidation) in the atmosphere, this factor decreases over time; for example, the global warming potential factor is 21 for a 100-year time period. Oxidation of methane can also occur in the ocean water when methane or methane hydrate is released from the seafloor. The methane that reaches the atmosphere can be gas released by dissociation of gas hydrate and/or gas that escapes from traps beneath the gas hydrate seal. The slides and collapses, which were discussed in the previous section, are probably the mechanisms that allow gas to escape from the gas hydrate/sediment system to the ocean/atmosphere system.

In the long term, methane from the gas hydrate reservoir might have had a stabilizing influence on global climate. When the Earth cools at the beginning of an ice age, expansion of continental glaciers binds ocean water in vast continental ice sheets and thus causes sea level to drop. Lowering of sea level would reduce pressure on seafloor hydrates, which would cause hydrate dissociation and gas release, possibly in association with seafloor slides and collapses. Such release of methane would increase the greenhouse effect, and so global cooling might trigger a response that would result in general warming. Thus, speculatively, gas hydrate might be part of a great negative feedback mechanism leading to stabilization of Earth temperatures during glacial periods. Currently, researchers are just beginning to analyze the potential effects of this huge reservoir of methane on the global environment and to consider many hypotheses.

## VI. ENERGY RESOURCE

The very large amount of gas hydrate that probably exists in nature suggests the idea that methane hydrate may represent a significant energy resource. However, most of the reports of gas hydrate in marine sediments are only indications that gas hydrate exists at some place (Fig. 6). Almost every natural resource that we extract for human use, including petroleum, is taken from the unusual sites where there are natural high concentrations. Much of the large volume of gas hydrate may be dispersed material and, therefore, may have little significance for the extraction of methane from hydrate as an energy resource. But even if only a small fraction of the estimated gas hydrate exists in extractable concentrations, the resource could be extremely important. The goals for the future use of methane hydrate as an energy resource are to be able to predict where methane hydrate concentrations exist and develop methods to safely extract the gas.

### A. Concentration

Most of the gas hydrate in the world (commonly stated as 99%) is formed of biogenic methane. Therefore, understanding the processes that can cause this gas, which is formed by bacteria dispersed through the sediment, to become concentrated into a hydrate accumulation is a critical issue. A small amount of gas hydrate is formed from thermogenic gas rising from great depths along fault channelways that apparently are kept warm (thus outside the gas hydrate stability field) as a result of circulation of warm fluids. The gas-bearing warm fluids reach the ocean floor, where temperature abruptly drops and gas hydrate is formed. This process is common in the Gulf of Mexico.

The accumulations are likely to be small, geometrically complex, and very near the seafloor. Thus, they are likely to be difficult to extract, although, at this point, categorical statements are premature.

In normal, microbially formed, gas hydrate deposits, the common increase in gas hydrate concentration downward through the gas hydrate stability zone toward its base and probable causes were discussed in Section III.C. Direct evidence from drilling and seismic reflection profiling (presence of BSRs) indicates that free gas bubbles typically accumulate just beneath the base of the gas hydrate stability zone, and, presumably, this gas is feeding the gas hydrate accumulation directly above. The concentration of gas beneath the gas hydrate stability zone can vary significantly because the gas can migrate laterally, rising along the sloping sealing surface formed by the base of the gas hydrate stability zone. When it does so, the gas will often reach a site where it becomes trapped at a culmination (shallow spot) and then probably will produce high concentrations of gas hydrate above that spot. Such gas traps can take several forms (Fig. 9). The simplest is formed at a hill on the seafloor, where the base of the gas hydrate stability zone parallels the seafloor and forms a broad arch or dome that acts as a seal to form a gas trap (Fig. 9, top panel). Such a hill can be a sedimentary buildup, such as at the Blake Ridge off South Carolina shown in Fig. 5, where the BSR appears much stronger near the crest, presumably indicating higher concentrations of trapped gas beneath the gas hydrate stability zone there. Alternatively, the seafloor hill could be a fold in an active tectonic setting.

In some cases, a culmination that traps gas at the base of the gas hydrate stability zone can form independent of the morphology of the seafloor. This happens over salt diapirs (plastically rising bodies of lower density salt) as a result of control of the depth of the base of the gas hydrate stability zone by two parameters. First, salt has a higher heat conductivity than sediment, so a warm zone will exist above a salt diapir. Second, the ions that are dissolved out of the salt act as inhibitors (antifreeze) to gas hydrate, just as salt lowers the freezing temperature of water ice. This double effect of chemical inhibition and disturbance of the thermal structure causes the base of the gas hydrate stability zone to be warped upward above a salt diapir, creating a gas trap (middle panel in Fig. 9). A seismic reflection profile across an isolated salt diapir is shown in Fig. 10. Notice that the deep strata appear to be bent up sharply by the rising salt. The base of the gas hydrate stability zone, as indicated by the BSR, also rises over the diapir, and the BSR cuts through reflections that represent sedimentary layers. This does not actually represent a physical bending of the base of the gas hydrate stability zone, but rather a thermal/chemical inhibition that prevents gas hydrate from forming as deeply as it would if the salt diapir
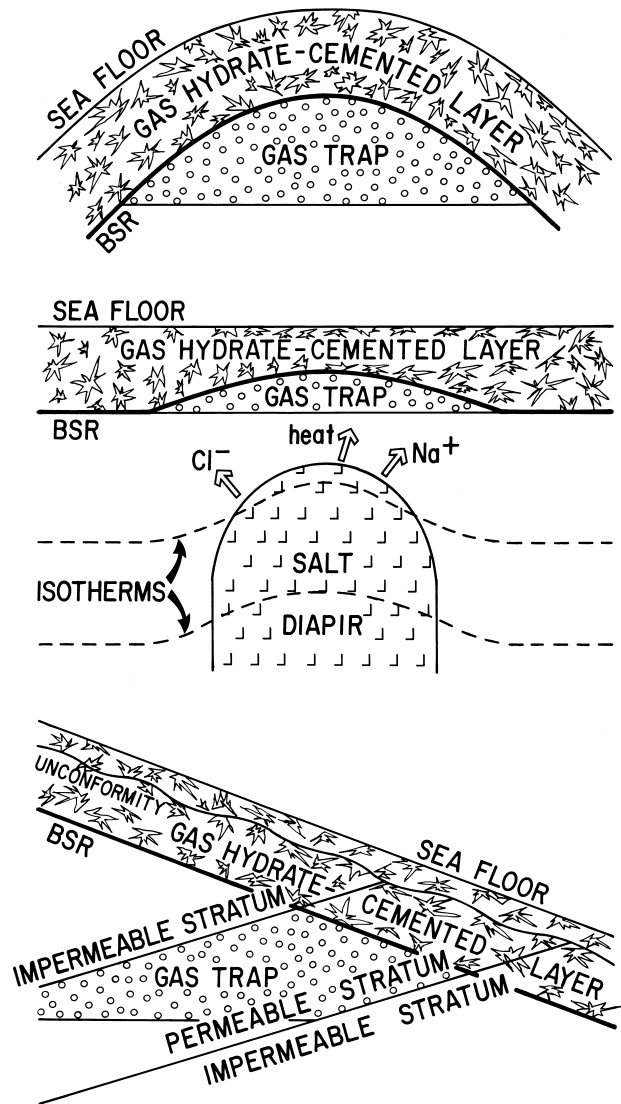


**FIGURE 9**  Conceptual diagrams showing cross sections through three of the many types of traps that confine gas beneath gas hydrate seals.

were not present. In this case, the BSR certainly does not fulfill its name by "simulating" the seafloor because of the strong influence of variable temperature and chemistry, but it does show us where the base of gas hydrate is located. The extremely strong reflections just below the BSR to the left of the diapir indicate that considerable gas is trapped there and is likely to cause concentration of gas hydrate above the trap. Of course, there are innumerable ways in which gas can be trapped beneath the gas hydrate-bearing zone. A common, simple trap (Fig. 9, bottom panel) is formed where dipping strata of alternating permeability are sealed at their updip ends by the gas hydrate-bearing layer, forming traps in the more permeable layers.
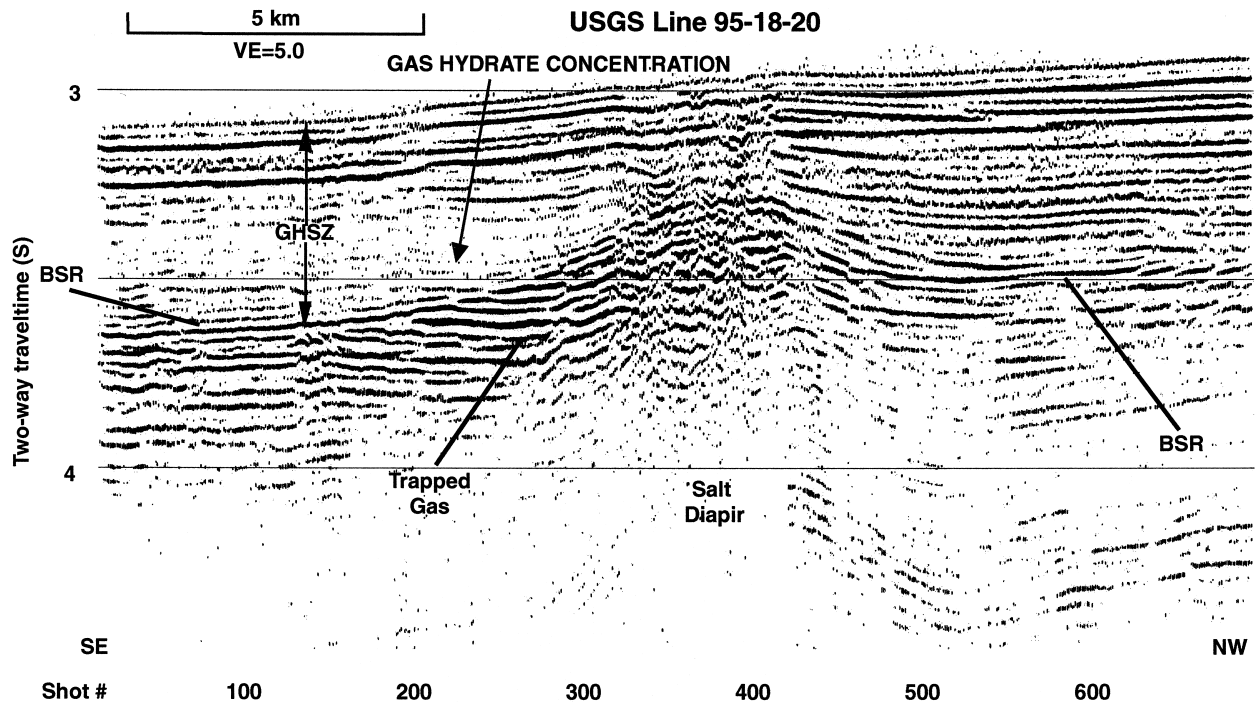
**FIGURE 10** Seismic reflection profile across a salt diapir beneath the continental slope off North Carolina in the southeastern United States. GHSZ indicates gas hydrate stability zone within the sediments. The BSR, denoting the base of the gas hydrate stability zone, rises markedly over the diapir because of the thermal and chemical effects created by the diapir. This doming of the base of gas hydrate stability forms a trap for gas. Compare this to the middle diagram of Figure 9.

## B. Extraction

The best places to look for gas hydrate deposits for methane extraction are the continental slopes, which, as discussed in Section III.C, seem to have the greatest concentrations of gas hydrate because they contain higher concentrations of organic matter for methane generation by microbes and have greater rates of sediment accumulation, which buries the organic matter and allows microbial decomposition to form gas. Furthermore, the shallower depths of the upper slopes allow the greatest concentrations of methane compared to conventional gas traps. From the minimum depth of hydrate stability down to total depths of about 1600 m, gas is held in hydrate in concentrations greater than in conventional traps. This concentration effect happens because gas molecules are held closer together in the crystal lattice than they would be by simple pressurization of free methane. However, the maximum concentration of methane in hydrate is essentially fixed, in contrast to free gas in reservoirs, where pressure increase will cause increased concentration, so at greater depths, free gas reservoirs will be more concentrated as a result of the higher pressure. The fixed concentration of gas in gas hydrate also means that shallower gas hydrate deposits will release gas that will expand more and fill a greater proportion of the pore volume than com-

parable deposits at greater depths (at greater pressures). For example, a 5% pore concentration of gas hydrate at 1000 m will provide the same gas expansion as a 15% concentration at 3000 m. Therefore, there is greater likelihood at shallow depths of exceeding the minimum limit of gas volume needed to generate spontaneous gas flow. Both the concentration factor at shallow depths and the expansion considerations suggest that the shallower part of the gas hydrate range may be more favorable for gas extraction—perhaps in the 1000–2000 m range of water depths. Methane probably will be extracted from hydrate by using a combination of depressurization (perhaps primarily), warming (perhaps using warm fluids from deeper in the sediments), and selective use of chemical inhibitors to initiate dissociation at critical locations.

## C. Environmental Concerns

As previously discussed, methane is a powerful greenhouse gas, so why would we wish to extract it, perhaps risking its escape to the atmosphere, for purposes of using it as a fuel? The primary environmental reason is that methane is considerably less polluting than other fossil fuels. Methane has the highest hydrogen-to-carbon ratio of all fossil fuels (4:1), and therefore, its combustion

produces a minimum amount of carbon dioxide per energy unit—about 34% less than fuel oil and 43% less than coal. Furthermore, methane does not produce any other pollutants such as particulate matter or sulfur compounds, so conversion to methane fuel would significantly clean up our present polluting emissions. The extraction of methane from gas hydrate probably will entail reducing pressure at the base of the gas hydrate stability zone (see Section VI.B). Therefore, overpressures would be reduced, and the likelihood of slides and slumps with release of gas to the ocean/atmosphere system might actually be diminished.

## D. The Future

The future of gas hydrate as an energy resource depends on the evolution of (1) geology (to identify concentration sites and settings where methane can be effectively extracted from gas hydrate); (2) engineering (to determine the most efficient means of dissociating gas hydrate in place and extracting the gas safely); (3) economics (the costs are likely to be higher than for extraction of gas from conventional reservoirs); and (4) politics (increased energy security and independence for energy-poor nations are significant driving forces). We are using other energy resources, which are nonrenewable, at significant rates, and the use of methane as fuel is preferable from a greenhouse-warming point of view because of its favorable hydrogen-to-carbon ratio and other pollution concerns. Therefore, methane hydrate holds promise as a future energy resource.

## SEE ALSO THE FOLLOWING ARTICLES

Energy Resources and Reserves • Fuels • Geochemistry, Low-Temperature • Geochemistry, Organic • Greenhouse Effect and Climate Data • Oceanic Crust • Oceanography, Chemical • Ocean Thermal Energy Conversion • Petroleum Geology

## BIBLIOGRAPHY

Booth, J. S., Winters, W. J., and Dillon, W. P. (1999). "Apparatus investigates geological aspects of gas hydrates," *Oil Gas J.* **Oct. 4,** 63–69.

Buffett, B. A. (2000). "Clathrate hydrates," *Annu. Rev. Earth Planet. Sci.* **28,** 477–507.

Henriet, J. P., and Mienert, J., eds. (1998). "Gas Hydrates: Relevance to World Margin Stability and Climate Change," Spec. Publication 137, 338 p., Geological Society, London.

Holder, G. D., and Bishnoi, P. R. eds. (2000). "Gas hydrates—Challenges for the future," *Ann. NY Acad. Sci.* **912,** 1044 p.

Kvenvolden, K. A. (1988). "Methane hydrate—A major reservoir of carbon in the shallow geosphere?" *Chem. Geol.* **71,** 41–51.

Kvenvolden, K. A. (1993). "Gas hydrates—Geological perspective and global change," *Rev. Geophy.* **31,** 173–187.

Max, M. D., ed. (2000). "Natural Gas Hydrates in Oceanic and Permafrost Environments," Kluwer Science, New York.

Paull, C. K., and Dillon, W. P., eds. (2001). "Natural Gas Hydrates: Occurrence, Distribution, and Dynamics," Monograph GM 124, 316 pp., Am. Geophys. Union, Washington, DC.

Paull, C. K., Matsumoto, R., Wallace, P. J., and Dillon, W. P. (2000). "Proceedings of the Ocean Drilling Program," Scientific Results Volume 164, 459 p., College Station, TX (Ocean Drilling Program).

Sloan, E. D., Jr. (1998). "Clathrate Hydrates of Natural Gases," 2nd ed., 705 p., Dekker, New York.
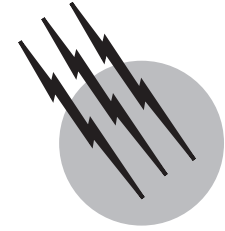
# Ocean Thermal Energy Conversion (OTEC)

## William H. Avery

*Johns Hopkins University (Retired) (Ref.)**

## GLOSSARY

**Heat engine** A system that transforms heat energy into mechanical, electrical, or other types of energy. Steam engines are typical examples.

**Heat exchanger** A device that transfers heat through a membrane from a liquid contained in a vessel(boiler, evaporator) maintained at one temperature, to a heat receiver (condenser) held at a lower temperature.

**Solar energy** Radiation from the sun that impinges on the oceans is strongly absorbed by the seawater, so that only a minuscule amount reaches a depth of 100 m. The absorbed heat warms the surface layers of the tropical

*Present address: 237 N. Main St. South Yarmouth, MA 02664.

oceans until an equilibrium temperature of 22–26°C is reached, when cooling caused by surface evaporation and emission of long-wavelength radiation balances the absorption rate of heat from the sun.

**OTEC IS A TECHNOLOGY** that converts solar heat stored in the upper layers of the oceans into mechanical and electrical energy. OTEC is an important energy option because it offers a practical way to provide clean, renewable fuels and electricity in quantities large enough to replace all fossil fuel and nuclear energy sources. Ships equipped with OTEC powerplants (OTEC plantships) could produce over 200 kW of electric power/km$^2$ of

tropical ocean area. If deployed throughout the ocean regions suitable for OTEC, over 14 million MWe (megawatts) of electric power could be generated continuously. The power would be used, via on-board water electrolysis, to produce hydrogen or to form hydrogen-based fuels that could furnish 300–800 quads per year of inexhaustible energy. The fuels would be stored temporarily on ship-board and then shipped at intervals to world ports. OTEC feasibility has been demonstrated through extensive tests at moderate scale.

## I. INTRODUCTION

OTEC is a heat-engine cycle that uses the difference in temperature between the surface waters of the tropical oceans and the cold water stored at depth in the ocean to convert solar energy to mechanical and electrical energy. Warm seawater is drawn into a heat exchanger (evaporator) containing a fluid with a low boiling point. Transfer of heat from the warm water to the fluid causes it to boil and form a vapor, which is passed through a gas turbine to generate power. The vapor (working fluid) leaving the turbine is condensed in a second heat exchanger, by transferring heat to cold water drawn from the ocean depth. The turbine is connected to an electric generator to produce a continuous source of electric power. There are two basic OTEC systems: *closed cycle*, in which the working fluid, after condensation, is returned to the evaporator; and *open-cycle*, in which the working fluid is discarded after condensation.

OTEC power plants can be sited anywhere in the equatorial regions around the world extending from roughly 10° north latitude to 10° south where warm water at the surface and cold water at depth are always available in the ocean. OTEC systems can be installed on ships or barges (plantships) that will cruise slowly on the tropical oceans. This will enable OTEC plantships to produce power and fuels anywhere on the tropical oceans, and provides a way for OTEC to make a vast potential contribution to world energy needs. OTEC plants can also be installed on land or on the continental shelf at tropical sites where warm and deep cold water are available near shore. This option can add a renewable energy source to many tropical islands and other suitable tropical sites.

## II. OTEC SOLAR ENERGY RESOURCE

### A. Geographical Distribution of Ocean Surface Temperatures

Radiant energy from the sun that falls on the oceans is absorbed in the "mixed-layer" of seawater, 50–100 m thick,

at the surface, forming a reservoir of warm water near the equator that stays at 26–31°C day and night year round. Below the mixed layer the water temperature falls gradually with depth, reaching a temperature of about 4°C. at 800–1000 m. Below this depth the temperature decreases only slightly to the ocean bottom at an average depth of 4000 m. Thus the ocean structure embodies a vast reservoir of warm water at the surface and a vast reservoir of cold water below 1000 m that differ in temperature by 22–27°C. The distributions of temperature differences in the tropical oceans that are suitable for OTEC power production are shown in Fig. 1.

### B. Solar Heat Stored in Tropical Oceans

The area shown in Fig. 1 in which the temperature difference ($\Delta T$) exceeds 22°C is approximately 60 million km². The solar energy absorbed in one square kilometer of the surface layers of the tropical oceans is 235 mJ (megajoules)/sec (235-MWt) (annual 24-hr average). By converting only 0.1% of the solar energy input to electricity (a conservative estimate), OTEC will produce 0.235 MWe of
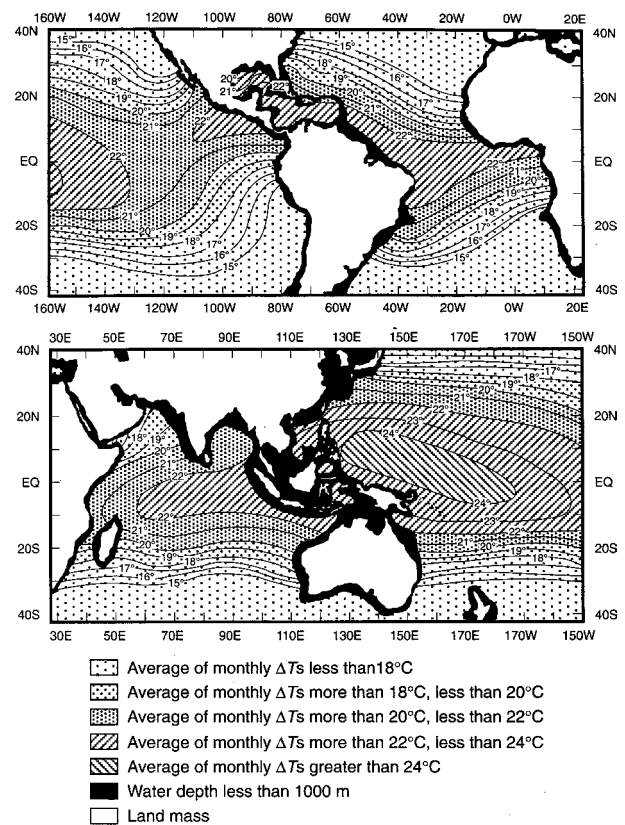


**FIGURE 1** Ocean temperature resource for OTEC (U.S. Department of Energy, Assistant Secretary Energy Technology, Division of Solar Technology).

net power per square kilometer of ocean area. Therefore, if OTEC plantships were sited throughout the suitable ocean area, they could produce more than 14 million MWe of power, at the onboard electric generators. This is 19 times the total installed electric power generation capacity that was available in the United States in 1997.

## III. OTEC SYSTEMS CHARACTERISTICS

An OTEC installation comprises: a power generating subsystem; subsystems to convert the electric power produced to forms useful for power transmission and fuel production; energy storage and transfer equipment; equipment for position control of floating systems; and facilities (ships, platforms, or on-land installations) to house and maintain the total system. Closed-cycle OTEC systems are suitable for installation on ships and, therefore, offer a potential way for OTEC plantships to draw power from the entire area of the tropical oceans and provide a major new source for renewable fuels production.

### A. Closed-Cycle OTEC

An OTEC power-generating subsystem comprises: an evaporator (boiler), a condenser, a turbine-electric generator, pumps to circulate working fluid and warm and cold seawater, and a demister to remove droplets from the working fluid vapor before it enters the turbine. A schematic of the system is shown in Fig. 2a. The operating parameters for a typical OTEC closed cycle are shown in Fig. 2b. In the OTEC closed cycle, a low-boiling liquid in a heat exchanger (boiler) is vaporized by heat from a warm seawater source. The vapor (working fluid) at a pressure between 700 and 1400 kPa (100–200 psi) is used to drive a turbine-generator to deliver electricity to the onboard distribution system. The vapor exhaust from the turbine is condensed in a second heat exchanger (condenser) cooled by cold water from depth. The condensed liquid is then



**FIGURE 2b**  Typical process parameters for the 40-MWe OTEC power system. [From Dugger, G. L. *et al.* (1983). *Solar Energy Handbook*, McGraw-Hill, New York, ch. 19.]

pumped to the boiler to complete the cycle. Closed-cycle OTEC systems have been tested and shown to be feasible at a 1-MWe scale, and engineering designs are ready for 40-MWe pilot plant tests that will provide detailed information on the engineering design and costs of OTEC components and subsystems, as a prelude to development of commercial systems of 100- to 400-MWe power.

### B. Open-Cycle OTEC Systems

In open-cycle OTEC systems, first proposed in 1928 by Georges Claude, a French engineer, seawater is used as the working fluid. A diagram of the Claude-cycle OTEC system is shown in Fig. 3. Warm seawater is drawn into a vessel where the pressure is maintained below the vapor pressure of the seawater, approximately 1/30 atmosphere. Flash evaporation of the water occurs, producing steam, which is used as the working fluid. The vapor passes through a gas-turbine, which drives an electric generator, and is then condensed by a spray of cold seawater drawn from depth. Several other types of open-cycle OTEC systems have been studied. Details of the various designs are discussed in Section X.



**FIGURE 2a**  Diagram of closed-cycle OTEC plantship.
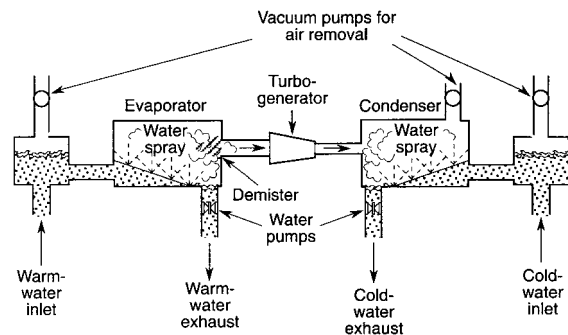


**FIGURE 3**  Diagram of Claude open-cycle OTEC power system.

# IV. OTEC CLOSED-CYCLE SYSTEMS ENGINEERING

Between 1975 and 1980 OTEC development was vigorously supported by the U.S. Department of Energy (DOE) (total funding >$250 million) and by French and Japanese engineers. During this period R & D funding was provided for study of a variety of design concepts, for testing of subsystems and components, and for resolution of problems that might prevent or make impractical commercial development and operation of OTEC systems. By 1980 solutions to all of the basic questions of technical feasibility had been demonstrated on a moderate scale. The next step was to conduct a pilot program to provide quantitative data on costs and engineering details of subsystems and components that would be incorporated in commercial-size systems.

To begin OTEC system development, DOE issued a program opportunity notice (PON) asking industrial teams to propose concepts for construction of OTEC pilot systems of 40-MWe size, for which DOE would share funding. To furnish a basis for evaluation of the various options, the Johns Hopkins University Applied Physics Laboratory (JHU/APL) was given a contract by DOE to develop baseline conceptual engineering designs of floating and moored OTEC powerplants that were to be used for evaluation of the industry proposals. Eight conceptual design plans for OTEC systems construction were submitted by industrial teams in response to the PON. The JHU/APL baselines, and nonproprietary data from industry proposals, provide engineering information that comprises the primary source for the description of OTEC systems in the following sections.

## A. OTEC Closed-Cycle Power Generation

### 1. Cycle Analysis

The efficiency of an OTEC system in converting heat stored in the warm surface water of the tropical oceans into mechanical work has a theoretical limit, called the Carnot efficiency (named after Sadi Carnot, a French engineer who first defined the thermodynamic restrictions on the efficiency of heat engines), where

$$\eta_{(max)} = \text{Carnot efficiency} = (T_w - T_c)/T_w$$

$$T_w = \text{absolute temperature of the warm water}$$

$$T_c = \text{absolute temperature of the cold water.}$$

For the ocean regions most suitable for OTEC operation, the annual average surface temperature is 27–30°C (80.6–86°F). The average temperature of the cold water is 4.4°C (40°F). Thus, the maximum OTEC efficiency will

be 7.5 to 8.1%. However, this ideal efficiency, even without unavoidable reductions caused by friction and heat losses, could be attained only at infinitesimal rates of power production. Maximum power production per unit water flow requires both a high rate of heat transfer in the heat exchangers and a large pressure difference across the turbine. This implies that only about half of the available $\Delta T$ can be used to produce a pressure difference across the turbine. Analysis shows that the maximum gross power production efficiency will be (Wu, 1984)

$$\eta_P = 1 - (T_C/T_W)^{1/2} = 4.0\text{–}4.5\%.$$

Power for pumping warm and cold water and working fluid, and for station keeping and hotel loads will consume 20–30% of the gross power. Therefore, the net efficiency of OTEC conversion of solar heat to electric energy will have a maximum value of approximately 3.5%.

### 2. Heat Exchangers

Commercial heat exchangers (HX) developed for power generation operate at high temperatures, use high operating pressures and require special materials. Costs of materials, and maintenance and replacement costs are high per unit of heat exchanger area for these systems. In contrast, the low-pressure, low-temperature regime of the OTEC cycle can use materials, system components, and operating parameters that are similar to those in commercial refrigeration and air-conditioning equipment, and are of much lower cost. Therefore, OTEC HX costs can be attractive despite the low thermal efficiency of the OTEC cycle. Several aluminum alloys have been shown to be suitable for use in OTEC heat exchangers. OTEC does require large HX area per unit of heat transferred because of the small temperature difference available for heat transfer. It is particularly important, therefore, that OTEC heat exchangers achieve an optimum balance among performance, cost, technical feasibility, use of industrial components and facilities, development risk, and safety.

To address this issue, DOE conducted programs during the period 1977–1981, at total funding over $250 million, to produce design data for OTEC systems, including power system configurations and heat exchanger types and components that would be optimal for the special requirements of OTEC. The programs investigated and defined suitable approaches to aspects of OTEC systems technology that would be critical to successful construction, operation, and durability. Analytical and experimental studies were conducted. Heat exchanger subsystems and components were tested and documented to supply design data for a 40-MWe pilot plant, which would provide quantitative engineering data on component performance and costs prior to commercial development.
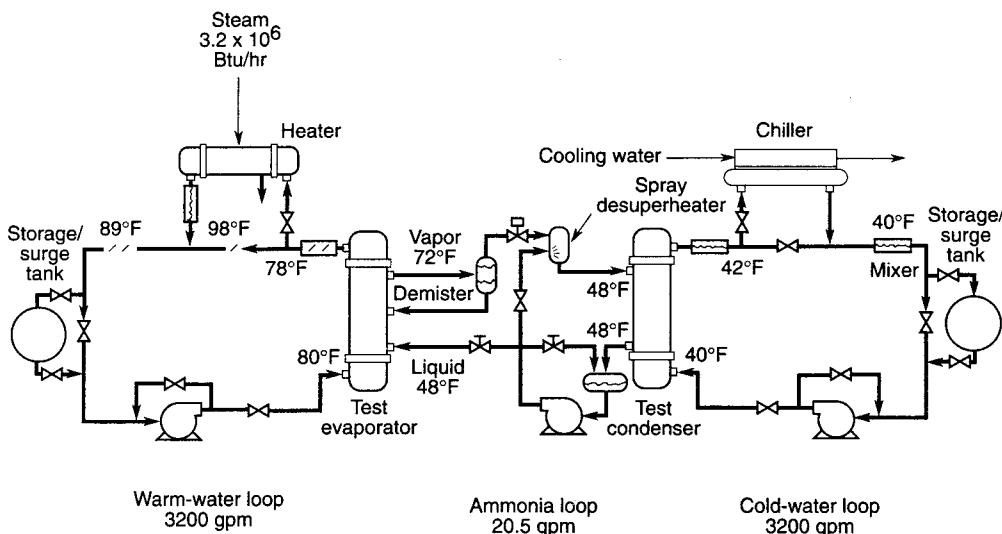
**FIGURE 4** Schematic of ANL test facility for 1-MWt model heat exchangers. [From Lewis, L. G., and Sather, N. F. (1978). *OTEC Performance Tests of the Union Carbide Flooded-Bundle Evaporator*. Argonne National Lab., Dec. ANL/OTEC-PS1.]

Contracts were awarded by DOE for studies and documentation of (a) heat exchanger designs for OTEC conditions, (b) heat exchanger performance for OTEC conditions, (c) working fluid evaluation, (d) materials for heat exchangers, (e) effects of the seawater environment on OTEC design and performance, and (f) subsystem characteristics. To determine the engineering design requirements for heat exchangers, a test facility was constructed at the Argonne National Laboratory (ANL) to test and document the performance, at a 1-MWt (one-megawatt) thermal heat transfer rate for model heat exchangers submitted by potential developers. The test facility characteristics are shown in Fig. 4. To provide a uniform basis for comparison, the heat exchangers submitted by contractors were all evaluated at a 1-MWt rate, with a 2°C temperature change in the water between the inlet and exit. Fresh water was supplied to the evaporators and condensers to avoid surface contamination that would invalidate accurate measurement of heat transfer rates.

## 3. Heat Exchanger Designs

Seven heat exchanger types were submitted in response to the DOE solicitation. (Details of the designs are presented in Avery and Wu (1994). For each type, measurements were made at the ANL facility, of the heat transfer coefficients, the pressure-drop in the fluids flow, the stability of operation, and other relevant behavior.

The rate at which heat is transferred from warm water to the working fluid in a heat exchanger, or from the working fluid to cold water, is:

$$q = UA\Delta T$$

where $U =$ Heat transfer coefficient, $A =$ Area of heat transfer surface. $\Delta T =$ Temperature difference between the water flow and the working fluid. From knowledge of the heat transfer coefficients in the evaporator and condenser, the overall heat transfer coefficient ($U$) can be calculated. The results for the heat exchangers tested at ANL are shown in Table I. Details of the complex calculations are given in Avery and Wu (1994). The ANL tests did not allow optimization of $U$, which involves a trade-off between heat-transfer rates, which increase with the flow velocities of water and working fluid, and power losses from pressure drops that also increase with flow velocity. Thus, two values of $U$ are shown in Table I for each heat exchanger type. The value in the first column is calculated for the ANL test conditions. The number in the second column lists the flow conditions for a maximum value of $U$, as calculated with the Owens formulas.

Since the cold water ducting system is more costly than the warm water system it is desirable to arrange the flow conditions so that there is a large $\Delta T$ (and small water flow) in the cold water flow and a small $\Delta T$ (and large water flow) in the warm water. An example of a choice of flow conditions is shown in the third column listed for the APL heat exchanger design, in which the aim is to achieve maximum performance with a cold water pipe of minimum cost. The data in Table I demonstrate that OTEC heat exchangers can have values of heat transfer coefficients that are equal to or higher than values typical of commercial heat exchangers operating at high temperatures.

**TABLE I   OTEC Power Plant Performance Comparison**

| | APL HX | | | ANL-SP3 | | CMU | | Trane | | Japan | | Alfa-Laval | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (nom) | (opt 1) | (opt 2) | (nom) | (opt) | (nom) | (opt) | (nom) | (opt) | (nom) | (opt) | (nom) | (opt) |
| Heat loading (kWt) | 937 | 1591 | 570 | 937 | 1332 | 938 | 2019 | 938 | 1615 | 4091 | 4664 | 938 | 938 |
| Water flow area (evaporator) ($m^2$) | 0.26 | 0.26 | 0.26 | 0.10 | 0.10 | 0.26 | 0.26 | 0.089 | 0.089 | 0.20 | 0.20 | 0.31 | 0.31 |
| Water flow area (condenser) ($m^2$) | 0.26 | 0.26 | 0.26 | 0.14 | 0.14 | 0.26 | 0.26 | 0.089 | 0.089 | 0.20 | 0.20 | 0.31 | 0.31 |
| Mass flow, warm water (kg/sec) | 202 | 202 | 75 | 202 | 202 | 202 | 202 | 202 | 202 | 409 | 409 | 202 | 202 |
| Mass flow, cold water (kg/sec) | 202 | 202 | 30 | 202 | 202 | 202 | 202 | 202 | 202 | 392 | 392 | 202 | 202 |
| Working fluid (WF) | $NH_3$ | $NH_3$ | $NH_3$ | $NH_3$ | $NH_3$ | $NH_3$ | $NH_3$ | $NH_3$ | $NH_3$ | R-22 | R-22 | $NH_3$ | $NH_3$ |
| Mass flow, WF (kg/sec) | 0.76 | 1.28 | 0.46 | 0.76 | 1.09 | 0.77 | 1.66 | 0.78 | 1.35 | 21.47 | 23.50 | 0.77 | 1.81 |
| $\Delta p$ (sat) (kPa) | 402 | 271 | 268 | 378 | 283 | 446 | 276 | 412 | 279 | 320 | 320 | 440 | 266 |
| $\Delta T$ (sat) (K) | 15.83 | 10.61 | 9.64 | 14.80 | 10.98 | 17.42 | 10.59 | 15.94 | 10.59 | 10.66 | 9.73 | 17.63 | 10.66 |
| $\Delta T$ (lmtd) (evaporator) | 3.27 | 5.55 | 2.75 | 3.52 | 5.01 | 2.26 | 4.86 | 2.51 | 4.32 | 3.82 | 4.35 | 2.48 | 5.02 |
| $\Delta T$ (lmtd) (condenser) | 3.11 | 5.27 | 6.35 | 3.90 | 5.69 | 2.50 | 5.39 | 3.73 | 6.42 | 4.52 | 4.77 | 2.09 | 4.95 |
| $\Delta T$ in WW flow through evaporator (K) | 1.16 | 1.97 | 1.90 | 1.16 | 1.65 | 1.16 | 2.50 | 1.16 | 2.00 | 2.50 | 2.85 | 1.16 | 2.75 |
| $\Delta T$ in CW flow through condenser (K) | 1.13 | 1.91 | 4.61 | 1.13 | 1.60 | 1.13 | 2.43 | 1.13 | 1.94 | 2.43 | 2.76 | 1.13 | 2.67 |
| $U_0$ (evaporator) (W/$m^2$ K) | 1469 | 1469 | 1464 | 2844 | 2844 | 2709 | 2709 | 4030 | 4030 | 2593 | 2593 | 1956 | 2461 |
| $U_0$ (condenser) (W/$m^2$ K) | 1503 | 1509 | 644 | 2678 | 2678 | 3423 | 3423 | 2811 | 2811 | 1886 | 1886 | 2406 | 2406 |
| Net OTEC power (kWe) | 25.69 | 32.31 | 12.86 | 19.65 | 23.24 | 23.92 | 38.97 | 21.49 | 28.33 | 80.27 | 91.77 | 17.52 | 37.68 |
| Cycle net power/max cycle power | 0.825 | 1.000 | 1.000 | 0.891 | 1.000 | 0.683 | 1.000 | 0.817 | 1.000 | 0.994 | 1.000 | 0.663 | 1.000 |
| Net power/total heat exchanger area (kWe/$m^2$) | 0.092 | 0.116 | 0.046 | 0.147 | 0.173 | 0.173 | 0.282 | 0.195 | 0.257 | 0.197 | 0.225 | 0.115 | 0.247 |
| CW flow for 1 MWe (net) power ($m^3$/s) | 7.86 | 6.25 | 2.33 | 10.28 | 8.69 | 8.44 | 5.18 | 9.39 | 7.13 | 4.88 | 4.27 | 11.53 | 5.36 |
| CWP diam. for 40-MWe (net) power (m) | 14.15 | 12.61 | 7.71 | 16.18 | 14.87 | 14.66 | 11.49 | 15.47 | 13.47 | 11.15 | 10.43 | 17.13 | 11.68 |

## 4. Heat Exchanger Materials

Heat exchanger materials used in commercial power generation favor titanium, copper alloys, and stainless steel. These could be used for OTEC but their cost makes them unattractive. Aluminum alloys are used in heat exchangers built for cryogenic systems, which operate at pressures and temperatures in the same ranges as OTEC, are low in cost, and are the preferred materials for OTEC. Before aluminum could be selected for OTEC it was necessary to prove that aluminum heat exchangers would be durable in the seawater environment. This was demonstrated in tests conducted at the Natural Energy Laboratory of Hawaii (NELHA) with warm and cold ocean water, which confirmed the durability of several aluminum alloys (George and Richards, 1980).

## 5. Working Fluid Requirements

Desirable characteristics of working fluids for closed-cycle OTEC include vapor pressure between 700 and 1400 kPa (100–200 psi) at 25°C, low volume flows of vapor per kilowatt of generated power, and high heat-transfer coefficient for the liquid films on the HX surfaces. The thermodynamic properties of four working fluids that are suitable for OTEC HX are listed in Table II. The fluorocarbon, R-22, used in the ANL test program, was banned because of its ozone depleting properties. The hydrofluorocarbon, R-134a, tetrafluoroethane, shown in Table II is considered to be a suitable substitute.

Ammonia was chosen as the preferred working fluid by most of the PON respondents because of its favorable thermodynamic properties. However, ammonia can be corrosive, and does not dissolve and wash away grease which can be deposited on condenser surfaces by vapors emitted by lubricants used in pumps and other HX equipment. This problem can be eliminated by selection of suitable subsystem components. Corrosion problems can be avoided by careful design. Hydroflurocarbons and hydrocarbons may also be used as working fluids. They dissolve grease and are compatible with materials that are suitable for OTEC heat exchangers and other cycle components A working fluid that is selected for OTEC must comply with strict regulations established for safe production, transport, storage, and use of bulk chemicals. New regulations will not be required for OTEC. The safety demonstrated in commercial handling of millions of tons of ammonia,

**TABLE II  Properties of Heat Engine Working Fluids of Interest to OTEC[a] [15.56°C (60°F)]**

| Property | Ammonia | Propane | Butane | Tetrafluoroethane (R 134a) |
|---|---|---|---|---|
| Formula | $NH_3$ | $C_3H_8$ | $C_4H_{10}$ | $C_2F_4H_2$ |
| Molecular weight (M) | 17.03 | 44.06 | 58.08 | 102.0 |
| Density[b] (1) (kg/1) | 616.7 | 506.9 | 583.4 | 1241.0 |
| Density[c] (v) (kg/1) | 5.83 | 16.06 | 4.64 | 23.77 |
| Vapor pressure (sat) (kPa) | 742.1 | 506.9 | 583.4 | 490.5 |
| Heat of vaporization (kJ/kg) | 1204.0 | 351.2 | 370.8 | 186.3 |
| Specific heat (1) (kJ/kg·K) | 4.713 | 2.643 | 2.397 | 1.389 |
| Specific heat (v) (kJ/kg K) | 2.942 | 1.929 | 1.739 | 0.971 |
| Viscosity (1) (Pa s) | 0.1446 | 0.1074 | 0.1741 | 0.2227 |
| Viscosity (v) (Pa s) | 0.00954 | 0.00836 | 0.00734 | 0.0139 |
| Thermal conductivity (1) (W/m K) | 0.5128 | 0.09756 | 0.1106 | 0.0852 |
| Thermal conductivity (v) (W/m K) | 0.0250 | 0.0178 | 0.01536 | 0.0129 |

[a] (Courtesy of NIST).
[b] Saturated liquid (1).
[c] Saturated vapor (v).

hydrocarbons and fluorocarbons, on land and at sea, indicates their acceptability for OTEC design and engineering. The cost of any working fluid selected for OTEC would be a minimal fraction of the power system cost.

## B. Water Ducting

### 1. General Requirements for Floating OTEC Plants

Commercial OTEC plants sited in the tropical oceans will each generate power in the range of 100–400 MWe (net) and will use modular construction. An OTEC plant will require a total water flow of about 3 m$^3$/sec MWe (net). Thus, commercial plants of 100–400 MWe will have total flow rates of 300–1200 m$^3$/sec. Design studies indicate that plants composed of 50- to 60-MWe modules will be optimum and that the flow velocity in the water ducts will be limited to about 2.5 m/sec to avoid excessive drag losses. Therefore, the minimum total duct area will be 60–75 m$^2$/module.

### 2. Cold Water Pipe (CWP) Design and Construction

The CWP is costly because it must extend to a depth of 800–1000 m and be designed to withstand ocean waves and currents that could be generated in a tropical storm of once-in-a-hundred-years intensity. For minimum size and cost, the CWP diameter in each 50- to 60-MWe module will be 7.5–10 m (25–33 ft). Dependant on considerations, other CWP diameters may be chosen. Many designs and materials of construction of CWPs (including steel,

aluminum, and elastomeric materials) have been studied. Scale tests have shown that CWPs made of light-weight concrete or fiber-reinforced plastics (FRP) are suitable for full-scale OTEC floating power plants and plantships. The manufacturing procedures and facilities for construction of pipes made of sections of light-weight concrete, joined end-to-end via a flexible joint, were demonstrated at the 3-m scale. Construction of suspended post-tensioned light-weight concrete pipes with diameters of 10 m is feasible with state-of-the-art facilities. Deployment may be implemented with existing technology available in the off-shore oil industry. Dynamic analyses show that these CWPs will withstand the maximum loads that are anticipated in the tropical ocean environment, in a 100-year operation of floating OTEC plants and plantships.

Analyses and tests done with plastic tubes of 2.5-m diameter show that CWPs made of fiber-reinforced plastic (FRP) are feasible. Cylindrical storage vessels made of FRP are in commercial use in diameters as large as 25 m, and the technology can be adapted to CWP construction. For CWP deployment a double-walled pipe with internal reinforcement will be required to give the plastic sufficient bending strength. The cost was estimated to be about 20% above the estimated costs of concrete CWPs but there could be advantages in deployment of the pipes. The deployment procedure was successfully demonstrated in the Mini-OTEC and OTEC-1 at-sea test programs. The OTEC-1 deployment involved assembly on shore of a 2-m-diameter, 700-m-long CWP, followed by towing to an off-shore ship, attaching the top end to a support, and then allowing the (loaded) other end to sink to orient the pipe in a vertical position. The upper end was then inserted from below into a gimbal connection on the platform.

**TABLE III  Collapsing Loads on CWP [Cold Water Flow 3.4 m³/sec/MWe (net)]**

| | Power MWe (net) | | | | | |
|---|---|---|---|---|---|---|
| | **40** | **40** | **60** | **60** | **80** | **80** |
| $V$ (m/s) | 2.00 | 2.50 | 2.00 | 2.50 | 2.00 | 2.50 |
| $D$ (m) | 9.30 | 8.32 | 11.40 | 10.19 | 13.16 | 11.77 |
| $L$ (m) | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 |
| $P$ (kPa) | | | | | | |
|   Friction | 3.53 | 6.16 | 2.88 | 5.03 | 2.50 | 4.36 |
|   Density effect | 7.65 | 7.65 | 7.65 | 7.65 | 7.65 | 7.65 |
|   Dynamic loss | 1.03 | 1.60 | 1.03 | 1.60 | 1.03 | 1.60 |
|   Miscellaneous | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| $P$(total) (kPa) | 13.20 | 16.40 | 12.60 | 15.30 | 12.20 | 14.60 |
| $P$(total) (psi) | 1.90 | 2.40 | 1.80 | 2.20 | 1.80 | 2.10 |
| $P$(800 m) (kPa) | 12.50 | 15.20 | 12.00 | 14.30 | 11.70 | 13.70 |

### 3. Cold Water Pipe Dynamics

In operation of an OTEC floating plant or plantship, the CWP must withstand collapsing loads created by suction, will be subjected along its length to fluctuating lateral and vertical loads that are caused by waves and currents acting on the ship that supports the pipe, and will be subjected to under-water waves and currents acting along the length of the pipe. These loads determine the design strength requirements of the CWP. The collapsing strength requirements are shown in Table III. Analysis of the dynamics is complex and requires sophisticated computer programs to provide quantitative solutions. Fortunately, the off-shore oil industry has developed analytical techniques to predict the stresses and strains in long suspended pipes in ocean environments. The analytical programs were adapted to predict OTEC CWP dynamics, and were verified in at-sea and laboratory tests. Computer programs were developed by Paulling, TRW, and others that predict strains and stresses as a function of platform movements in the sea state, and motions and moments of segmented and flexible CWPs of various dimensions and materials strengths, as dependent on wave heights, wave frequencies and direction, and underwater forces. The programs include effects of internal and external water flow and are applicable to fixed or flexible connection of the CWP to the platform and include the dynamics of the platform–CWP combination. The predictions of the analysis programs were verified in several programs:

*a. At-sea tests.*  With DOE support and JHU/APL supervision, measurements of strain and stress were made along the length of a segmented 1.52-m-diameter (5-ft) steel pipe ($L/D = 100$) suspended from a semisubmersible, off-shore oil test platform provided by the

Deep Oil Technology Corporation. The tests were conducted near Catalina Island in California. The tests showed that the analysis programs gave good predictions of the observed pipe strains and stresses along the pipe, if the sea state was represented by equations developed by Breitweiser, incorporating a bidirectional wave pattern. With direction and support by the National Atmospheric and Oceanic Administration (NOAA), at-sea tests were also done near Kea-Hole Point in Hawaii, using a 2.44-m-diameter (8-ft) fiber-reinforced FRP CWP, 122-m (400-ft)-long. Results of stress and strain measurements along the length of the pipe were in good agreement with analytical predictions. The tests demonstrate that a satisfactory theoretical base exists for the design and construction of concrete and FRP CWPs of diameter large enough to meet the requirements of 50- to 60-MWe OTEC power plant modules.

*b. Water tunnel tests.*  A series of tests to evaluate the survivability of the 40-MWe OTEC platform–CWP assembly in 100-year tropical storms on the high seas off Brazil and near-shore Puerto Rico were made by JHU/APL using a 1/20th-scale OTEC barge model in a water tunnel operated by the Offshore Techology Corporation in Escondido, California. The facility was able to simulate sea states having waves that approached the barge in two directions, as was observed in the ocean tests near Catalina Island. The test data showed that the 40-MWe OTEC barge, with a flexible attachment to the platform, would survive the open-ocean 100-year worst-storm conditions. At the Puerto Rico site, the moored platform would require mooring lines that would orient the vessel to head into the waves, and would need some fairing of the bow to prevent water from coming over the deck.

### 4. CWP Support Design

Analysis and tests show that a CWP connected to a floating platform by a fixed joint will not survive in an ocean environment. Therefore, a flexible connection to the platform is required at the top of the pipe. The CWP connection with the platform must embody a sea-water seal and allow movement at the joint through an angle of up to 30° at any platform orientation, to survive in the maximum 100-year storm. The connection must also support the weight of the CWP, including loads imposed by heave of the platform.

Several engineering designs for supporting the CWP in a flexible connection were studied in the 40-MWe-demonstration proposals for floating OTEC plants. Features of three designs are shown in Fig. 5. Consideration of the options led to the the ball and socket joint being incorporated in the JHU/APL base-line 40-MWe plant design.

## Options for CWP-platform flexible joint

| Advantages | Disadvantages |
|---|---|
| Ball and socket | |
| No restriction to flow | Possible ball unloading[3] |
| Freedom of rotation in high seas | Large diameter of mounting |
| Feasible sealing | |
| Gimbal | |
| No restriction to flow | Large box beams to transmit loads |
| Universal joint | |
| Minimum structure weight | Some flow restriction |
| Minimum sealing requirement | High bearing loads |
| Possible minimum cost | Submerged bearings |
| Minimum installation diam. | |
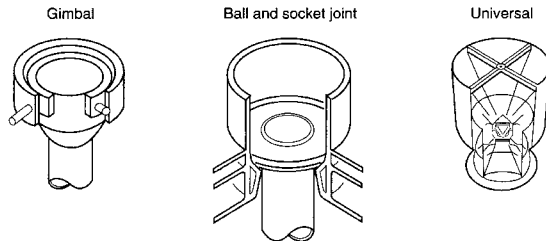


**FIGURE 5** Types of flexible joints studied for OTEC CWP support. [From TRW (1979). *Ocean Thermal Energy Conversion Cold Water Pipe Preliminary Design Final Report.* TRW Energy Systems Group.]

### 5. General Conclusions Regarding CWP Requirements for Floating OTEC Plants and Plantships

Methods of construction for CWPs of 2.5-m (FRP) and 3-m (light-weight concrete) were demonstrated that are directly applicable to CWPs with diameters of 10 m or more. At-sea tests confirm that CWPs will be able to survive storms of maximum severity expected in a 100-year period in the tropical oceans. Lightweight concrete and FRP have been shown to be suitable for CWP construction. Further information is desirable on added-mass effects of external water flow in lateral motions of large-diameter pipes and on effects of multidimensional sea states on CWP dynamics.

### 6. Cold Water Pipes for Shore-Based Plants

The design and construction of shore-based OTEC plants involve significant CWP problems not present in floating plants, namely, the design is site-specific and surveys of the configuration and structure of the sea bottom at proposed sites must be made to define a suitable trajectory for the CWP. Furthermore, data must be available for three ocean regimes: the near-shore regime, which involves high waves and currents, requires heavy structures and/or emplacement below the sea floor to prevent failure of the CWP; a shallow zone on the continental shelf where the water depth increases gradually from near

shore to a depth of 100 m or more; a final zone where the depth increases sharply to reach the cold water source. The total CWP length will be around more than 2 km at the best tropical island sites and 5 km or more at average sites. Decisions must also be made regarding support of the CWP, with possibilities including anchoring the pipe to the sea floor, entrenching it, or using a buoyant pipe held by cables attached to weights on the sea floor. For the cold water supply pipes installed at the NELHA facility in Hawaii, entrenching of the pipe near shore and catenary support for the remaining length have been adopted.

### 7. General Conclusions Regarding CWP Requirements for Shore-Based OTEC Plants

Buoyant polyethylene (HDPE) cold water pipes of ∼1 m in diameter, anchored at intervals with 'cables attached to weights on the sea bottom, have been constructed and deployed at NELHA and are operating successfully. FRP pipes suitable for shore-based plants of several megawatts seem feasible. A preliminary engineering design is available of water ducts for a 50-MWe OTEC plant using concrete ducts for cold deep water, warm surface water, and mixed exhaust water.

## C. Warm and Mixed Effluent Ducting

Warm water from the mixed layer at the ocean surface is pumped into the OTEC vessel through screens at a flow velocity below 1/2 m/sec to prevent fish being trapped against the screens. The warm water flows through the heat exchangers, and is discharged at the bottom of the platform. It may be mixed with the cold water exhaust before leaving the vessel or may be discharged at the bottom where it may be interleaved with the cold water flow. The exhaust flow is denser than the surrounding ocean water and forms a plume which sinks to a depth where the plume density matches that of the ambient ocean water. For floating plants, no external inlet pipes are needed. Dependent on the site, shelf-mounted plants may or may not need external exhaust pipes. OTEC-1 experience showed that the discharge plume descended quickly with insignificant mixing with the ambient ocean layer. This result indicates that external exhaust pipes will not be needed. If they were needed, a new method for their attachment would have to be developed. Attached exhaust pipes failed in some tests. Shore-based plants will require inlet pipes that extend along the bottom to a water depth low enough to minimize wave effects and far enough above the bottom to prevent entrance of sediment dislodged by water flowing to the inlet or by under-sea currents.

## V. ENERGY TRANSFER FROM FLOATING OTEC PLANTS

OTEC power may be supplied to users by direct transmission of electricity to on-shore customers at sites where deep cold water is available near shore, or used on-board to produce a fuel or product that can shipped ashore and stored for later use in motor vehicles or electric power plants. OTEC plantships sited on the tropical oceans will use electric power generated onboard to electrolyze water and form hydrogen based fuels, which will be stored temporarily and then shipped to world ports at monthly intervals.

### A. Water Electrolysis

Water electrolysis provides an efficient mechanism for converting electrical energy to stored chemical energy. In principle, the conversion process can have 100% efficiency. R & D programs indicate that efficiencies in the range of 80–90% will be attainable. Therefore, hydrogen production is a highly efficient way to store electrical energy produced on OTEC plantships.

Electrolysis systems are classed as unipolar or bipolar. Unipolar electrolyzers have positive and negative electrodes immersed in a conducting fluid (electrolyte). In water-based electrolytes an applied voltage causes hydrogen gas to be liberated at the negative terminal (cathode) and oxygen gas at the positive terminal (anode). The electrodes are designed to channel the gases into separate ducts.

Electrolyzer, Inc. has been a principal developer and supplier of unipolar electrolyzers since the 1940s. The electrodes are supported in containers (cells) filled with concentrated KOH. The cells are connected in parallel and operate at voltages in the range of 1.5–2 V. Experimental results indicate that energy conversion efficiencies of 85% or above will be possible in commercial units. A layout for a planned 35-MWe installation is 29 m wide and 95 m long. With these dimensions, a 350-MWe electrolyzer for an OTEC plantship would occupy an area of 2750 $m^2$.

Bipolar electrolysis systems are characterized by the type of electrolyte. The proton exchange membrane (PEM) system, developed by the General Electric Company (GE), uses as the electrolyte a thin membrane of sulfonated fluorocarbon (Nafion$^{TM}$) that conducts electricity when saturated with water. Electrodes are formed by depositing a thin platinum film on opposite sides of the membrane to form a bipolar cell. An electrolyzer is made by stacking 50–200 cells in series, with suitably formed separators to direct the exhaust gases into channels at the sides. Since the membrane is the electrolyte, only pure water needs to be supplied to the cell. When the cell operates, water is protonically pumped through the membrane at about eight times the electrolysis rate. The water provides the source of hydrogen and oxygen, and removes the heat generated by ohmic resistance of the Nafion film. PEM electrolyzers, consisting of 60 cells 15 $cm^2$, were developed by GE and successfully tested at current densities ranging from 2000 to 10,000 amperes/$m^2$ (200–1000 amperes per square foot (ASF)). Tests indicated that cells 1 $m^2$ would be feasible. When funding for the program was terminated some problems with sealing the cells had not been solved. At 1000 ASF, the cell efficiency was 70%.

Alkaline electrolyzers, which use concentrated KOH as the electrolyte, are commercially available in multikilowatt sizes from Teledyne Energy Systems in the United States and several foreign companies—Brown-Boveri (Switzerland), Norsk-Hydro (Norway), and French and Belgian suppliers. The design employs thin membranes of asbestos (or other material) saturated with KOH, with nickel (or proprietary) screens pressed into the membranes, as the electrodes, to form an electroysis cell. Cells are stacked in series, as in the PEM design, to achieve a desired input voltage. The KOH electrolyte provides the source for hydrogen and oxygen production and is circulated for cooling. Cell efficiency improved steadily with R & D support after 1976 leading to a projected value of 85–90% in 1984. As with the GE program, funding was terminated in 1984.

It may be concluded that electrolyzers with efficiencies in the range 60–85% will be available when a significant commercial demand arises.

### B. Storage of OTEC Electrical Energy in Hydrogen-Based Energy Products

OTEC electric energy generated at on-land or near-shore installations may be used to generate hydrogen at off-peak times, and stored as gas or liquid for later power generation. The power requirements for hydrogen liquefaction (~30% of the heat of combustion) make it inefficient to store electric energy in liquid hydrogen unless there is a requirement for direct use of liquid hydrogen. Because of its low density it is impractical to store or ship hydrogen produced on OTEC plantships. However, the hydrogen may be stored efficiently if it is combined with nitrogen on the plantship to form ammonia ($NH_3$), or other hydrogen-rich molecules.

#### 1. Ammonia

Ammonia is of particular interest as a way to store OTEC electric energy because hydrogen generated by OTEC electrolysis on a plantship can be combined with nitrogen

extracted from the air to form ammonia. The reaction is an equilibrium process that is a highly efficient method ($\sim$90%) of storing OTEC energy. The physical properties of ammonia are nearly identical with those of propane, making it practical to synthesize, store, and transport liquid ammonia from plantships to on-land users. Tests of ammonia fuel in internal combustion (IC) engines yield an octane number of 130. It would be a practical, inexhaustible fuel, which does not produce carbon dioxide, to replace gasoline in motor vehicles. It will not produce polluting exhaust products, if used in fuel cells or power plants.

### 2. Methanol

Methanol ($CH_3OH$) synthesis is also an attractive method for storing OTEC energy on plantships. Methanol is a high-octane (110) liquid fuel that may be used directly in motor vehicles, and is the preferred feed-stock for fuel cells. Methanol synthesis is highly efficient because it uses both the oxygen and the hydrogen produced in water electroysis, in the reactions: $C\ (coal) + 1/2O_2 = CO$, $CO + 2H_2 = CH_3OH$. Coal for the process is transported to the plantship in a conventional collier, and converted to CO by injecting oxygen and water into a molten carbonate reactor. If commercialized to use the total tropical ocean resource OTEC methanol synthesis could supply more than 800 quads per year of fuel, roughly three times the total 1998 world fossil fuel consumption. Methanol combustion produces approximately the same amount of $CO_2$ as gasoline. Therefore, OTEC methanol production will not offer a way to reduce global warming.

### C. Electrical Energy Transfer via Electrochemical Cycles

Two electrochemical energy transfer concepts have been investigated. The redox battery system uses electrodes immersed in an electrolyte that can exist in two ionic states such as $Fe^{++}$ and $Fe^{+++}$. Application of a voltage above the ionization potential converts $Fe^{++}$ to $Fe^{+++}$, storing the energy in the electrolyte, which can be shipped to shore where the process is reversed to liberate the stored energy. Costs of shipment of the electrolyte make the system impractical for OTEC energy transfer. The lithium–air battery system uses a slurry of lithium hydroxide, which is prepared on land and shipped to the plantship where OTEC power is used electrolytically to prepare billets of metallic lithium. These are shipped to land where electric energy is produced by reaction of the lithium billets with air in a battery, regenerating lithium hydroxide. Shipping costs make this system impractical for OTEC.

### D. Other Methods of Using Energy Produced on OTEC Plantships

Studies and some testing showed that aluminum refining using power generated on shore-based or floating near-shore OTEC plants at several sites would not be cost effective at 1990 fossil energy costs. Deep-ocean mining of minerals using low-cost OTEC power could make this option economically attractive in the future.

## VI. ENERGY TRANSFER FROM LAND-BASED OTEC PLANTS

At tropical sites where cold, deep water is available near shore OTEC electric power can be transmitted directly to on-land consumers. This can be economically attractive if the power demand is large enough to justify construction of a multi-megawatt OTEC plant. The economic viability of land-based OTEC can be enhanced in favorable locations by using the cold water for air conditioning, mariculture, fresh water production, and coldwater agriculture. Programs conducted at the Natural Energy Laboratory of Hawaii (NELHA) have demonstrated the feasibility of these options.

## VII. OTEC TOTAL SYSTEMS INSTALLATION

An OTEC system requires a structure to house the total OTEC operating complex. It may be designed for direct transmission of power to consumers on land, or designed for at-sea operation to produce a storable product for later shipment. A configuration to be selected will optimize efficiency, ease of construction, cost, ability to withstand environmental stresses, and suitability for operation at selected sites. For floating platforms designed to produce a storable product while "grazing" on the tropical oceans, a barge installation is the best choice. For floating platforms designed for power transmission to shore, the optimum configuration is strongly dependent on the site selection. Designs that have been studied include spar configurations, semisubmersibles, ships, and barges. Problems of construction, deployment, and demonstration of ability to withstand waves and currents generated near shore by "100-year-storms" are involved in the selection, as well as cost.

## VIII. POSITION CONTROL FOR OTEC PLANTSHIPS

OTEC plantships will use thrusters to propel the plantships at roughly 1/4 knot in normal operation, and at 1 knot at maximum speed. Rotatable thrusters mounted below

the platform will maintain direction and orientation in waves and currents. (The low forward speed makes rudder control not feasible.) Analyses of the sea-keeping needs for the 40-MWe demonstration barge provided design information for other configurations and sizes. For the baseline 40-MWe plant the estimated maneuvering power is 5–7 MWe. A total power complement of 8 MWe was found adequate to withstand the maximum winds, waves, and current requirements of the 100-year tropical storm, and to supply the maximum navigation requirements. Thrusters that are commercially available will be practical to install and will provide a force of 15–22 kg/kW of power supplied (1994 data).

## IX. POSITION CONTROL FOR FLOATING OFF-SHORE OTEC POWER PLANTS

Dynamic positioning and mooring have been investigated. A dynamic positioning system must keep the position of the OTEC platform within a circle of about 100-m radius at the surface to allow the power cable to remain attached to the underwater equipment that holds and protects the cable in its trajectory to the onshore facility. Continuous power delivery requires that standby diesel power must be available. An extensive study of the requirements for proposed plant configurations—spar, barge, ship, and semi-submersible—was done to provide quantitative data on thrust requirements and costs; for a moderate environment with surface current of 1.5 m/sec, and for an extreme surface current of 3 m/sec. The data tabulated to show cost as function of plant requirements included vectoring of the water flow in the inlet and exhaust ducts. The cost effectiveness of this option is not known. Mooring of OTEC power plants will require cables reaching to depths greater than 1000 m that can hold the platform within a small surface area without interfering with the power transmission lines. The systems must be designed to operate in storms of 100-year severity or be able to detach the power cables and maneuver out of harm's way when devastating storms are imminent. The mooring requirements for spar designs differ from the other configurations and have received separate design studies. For a platform floating at the surface, a multileg mooring system is required that will hold the vessel within the required area, and will be able to orient the vessel in storms so that waves do not strike it broadside.

Documentation is available for feasible designs using mooring lines made of wire ropes and/or chains. Spar configurations are designed to place the main structure at a depth where forces from waves and currents are relatively small. This makes it possible to hold the plant in position using a single mooring line. For this case the relative merits of multi-anchor-leg (MAL) mooring ver-

sus tension-anchor-leg (TAL) mooring were investigated. In the TAL concept the mooring line is attached between the bottom of the CWP and an anchor emplacement on the ocean floor. The CWP must be designed to withstand the tension loads. Tension-anchor-leg (TAL) mooring offers several attractive features: (i) There is no need for on-board anchor-line handling equipment. (ii) Design simplicity reduces handling and operating needs. (iii) Power cable handling will be simpler. (iv) Life-cycle costs will be lower. There are two serious disadvantages: (i) The tension loads for commercial sized plants far exceed commercial design experience. (ii) Failure of the mooring system will cause catastrophic failure of the entire assembly. Further studies are needed to determine whether the TAL or the MAL system will be the optimum choice.

## X. OPEN-CYCLE OTEC

### A. Claude Cycle

In the Claude open-cycle OTEC system warm ocean water is drawn into a vessel in which the pressure is maintained below the boiling point of the warm water—roughly—1/30 atm (3.8 kPa). This causes the seawater to boil, producing vapor (steam) that becomes the working fluid. The steam passes through a turbine, producing power, and then is drawn into a cold chamber at lower pressure—1/125 atm (0.8 kPa)—where the steam is absorbed in a spray of cold water drawn from the ocean depth. The condensate is then pumped into the ocean. As an alternative, the vapor may be condensed by being directed on to a cold surface, and collected as a source of fresh (desalinated) water. Because the system pressures are so low in the Claude cycle, the turbines, ducts, and containment vessels must be very large. Therefore, open-cycle OTEC systems of the Claude type are impractical for ship or barge installations. Open-cycle 1- to 10-MWe OTEC systems are feasible on shore and could be attractive for installations at tropical sites where cold water near shore can be used by OTEC to furnish fresh water, air conditioning, mariculture and cold water agriculture, in addition to electricity and fuel production. However, closed-cycle OTEC can offer the same options and can be built with less difficulty and cost uncertainties.

A schematic diagram of the Claude cycle system is shown in Fig. 3. Warm seawater passes through a deaerator, then enters a flash evaporator, where it flows into the evacuated vessel through small nozzles, and emerges as a vigorous spray. This induces efficient extraction of heat of vaporization from the surface area of the evaporating droplets, which collect at the outlet with little cooling of the unvaporized water. The vapor passes through the turbine and is then efficiently condensed by the cold

water spray. As a result, the temperature differences between the vapor and water in both vessels can be small. Experiments by Kreith and Bharathan and associates at the National Renewable Energy Laboratory of DOE have provided quantitative data that allow optimization of the heat transfer coefficient as a function of the evaporator design. The results confirm that the cycle efficiency of open-cycle OTEC can be significantly higher than that of closed-cycle OTEC. The open-cycle process eliminates the need for large heat exchangers and also eliminates concerns of environmental risks posed by working fluids used in the closed cycle. The technical feasibility of the process was demonstrated at 210-kW scale in 1998 at the Natural Energy Laboratory of Hawaii (NELHA). The net power was 40-kWe. Further development has been discouraged by problems of construction and operation of the evacuated vessels and process equipment that are foreseen for large-scale systems.

Four other open-cycle systems have received some R & D support:

## B. Panchal–Bell Cycle

In this cycle steam from flash evaporation of warm water is condensed on the surfaces of a heat exchanger containing liquid ammonia, which vaporizes to form the working fluid for the closed cycle discussed in Section II. The condensed steam is discharged to the ocean or may be saved as a source of fresh water. The Panchal–Bell cycle has a number of potential advantages, compared to the Claude cycle: (a) The large low-pressure turbine is eliminated. This allows a major reduction in system volume and eases problems of constructing a vacuum-tight enclosure. (b) The heat transfer coefficient for the steam condensing on the heat exchanger surfaces is high, with benefits in reduced heat exchanger area and improved efficiency. (c) Seawater contact with the heat exchanger surfaces in the evaporator is avoided, eliminating possible contamination and corrosion from the warm seawater. (d) Steam condensation occurs at a pressure of ~1/40 atmosphere rather than ~1/80 atmosphere that is required in the Claude cycle. This significantly reduces the power required for the vacuum pumps, which can be over 10% of the gross power in the Claude system. (e) The cycle may produce fresh water and power in various ratios, depending on needs at a particular site. Further R & D is needed to define a complete system.

## C. Mist-Lift Cycle (Beck)

In this cycle warm seawater is introduced near the bottom of a long vertical tube with a vacuum pump at the top. Flash evaporation of the water produces a mixture of vapor and water droplets which are drawn to the top of the tube (at a theoretical height of 280 m), where they strike a

surface chilled by cold water, coalesce and flow back into the ocean, passing through a hydraulic turbine to generate power. Analysis and R & D have confirmed the validity of the concept. Experiments demonstrated effective coupling of vapor and mist to a height of 50 m. Further design studies and experiments are required to define a complete system.

## D. Foam-Lift Cycle (Zener–Fetkovitch)

In this cycle, cold seawater is drawn through a hydraulic turbine, connected to an electric generator, into a long, vertical, partially evacuated, cylindrical vessel submerged in the tropical ocean. Compressed air containing a foaming agent is introduced at the bottom of the vessel. The foam bubbles reduce the density of the water column, creating an upward flow that causes the mixture to spout from the top of the tube, where a foam-breaker removes the foaming agent. The dense, cold water then flows out into the warm lower-density surface water and sinks. Further R & D is required to determine whether the concept is technically feasible and can be cost effective. (Although the foaming agents may have a concentration as low as 10 ppm their cost will be significant.)

## E. Uehara Cycle

Professor Uehara of Saga University in Japan has invented a new hybrid OTEC cycle that uses a mixture of ammonia and water as the working fluid. Analysis shows that it will have higher efficiency than the Kalina cycle (Uehara *et al.*, 1998), which also uses an ammonia–water mixture as the working fluid. Tests of the concept have been conducted at Saga University. A flow diagram for the test setup is shown in Fig. 6.
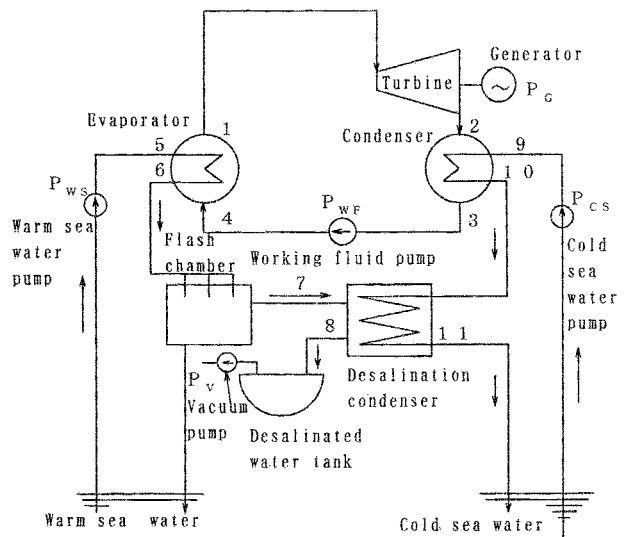


**FIGURE 6** Schematic diagram of the Uehara OTEC cycle.

In operation, a warm mixture of ∼90 wt% ammonia in water is drawn into a vessel maintained at a pressure below the saturation pressure of the mixture, which causes boiling. The vapor passes through the turbines, driving electric generators, and is condensed in the cold water heat exchanger. The test results show that the cycle can result in higher efficiency than closed-cycle systems. Further tests at larger scale will be needed to show what system cost benefits will be possible.

## XI. CLOSED-CYCLE OTEC DEVELOPMENT STATUS

During the period 1974–1985, engineering design studies, R & D on subsystems and components, and tests of OTEC systems, including at-sea and on land programs, were conducted under DOE sponsorship and in French and Japanese OTEC programs that had substantial funding. Total funding for the programs exceeded 250 million dollars (1980 $). In 1980 there was a consensus among OTEC investigators that the technical feasibility of OTEC

development time, and financing options. A program opportunity notice (PON) was issued that invited industrial teams to submit proposals to conduct a program that would construct, deploy, and demonstrate operation of a total OTEC system. The respondents were asked to include a management scheme and proposed cost sharing to accomplish the program objectives. The program was to be done in six phases: *conceptual design, preliminary engineering design, engineering design, construction, deployment and operation, and finally, transfer of ownership to industry*. After the first phase, two or three proposals would be funded, followed by a single choice for the last phases (or perhaps two choices if funds were available). Funds were appropriated by the Congress to begin the program. Six industrial teams submitted conceptual OTEC system designs and development proposals for the first phase of the 40-MWe demonstration programs. A DOE committee was established to review the proposals and select those to be funded. A seventh proposal for a 10-MWe installation in Saipan was not accepted because it did not conform to the PON guidelines. The conceptual design proposals that were submitted in response to the PON were:

| Team leader | System concept | Site | Product | Energy delivery method |
|---|---|---|---|---|
| General Electric Co. (GE) | Shelf-mounted | Hawaii | Electric power | Power cable |
| Ocean Thermal Corp (OTC) | Land-based | Hawaii | Electric power | Power cable |
| Puerto Rico Electric Power Authority (PREPA) | Shelf mounted | Puerto Rico | Electric power | Power cable |
| Solar Ammonia Company (SOLARAMCO) | Plantship | Tropical oceans | Ammonia | Ammonia |
| Virgin Islands Water and Power Authority (VIWPA) | Land-based | Virgin Islands | Electric power | Power cable |
|  | Land-based | Virgin Islands | Fresh water | Pipe line |
| Ocean Solar Energy Association (OSEA) | Shelf-mounted | Puerto Rico | Electric power | Power cable |

had been proved and that the next step in the development of OTEC technology should be construction of a pilot plant that would provide firm data on system performance and costs. A goal for OTEC commercialization was established to demonstrate 100 MWe of OTEC by 1985 and 500 MWe by 1990. DOE was assigned responsibility to plan and carry out a development program to achieve the goals. As the first step, funding was appropriated by the U.S. Congress for DOE to begin a phased program to develop a pilot OTEC system that would demonstrate production of 40 MWe (net) of OTEC power. The objective was to provide engineering data for a broad range of OTEC system concepts, including alternatives in plant installation (mounted on a floating platform, or on a fixed platform on or near shore), geographical location, facility needs, capital and operating cost trade-offs,

To provide a firm technical base for evaluation of the proposed systems, DOE awarded a contract to JHU/APL to conduct preliminary design and testing which would document base-line designs for 40-MWe OTEC floating plants, containing all of the subsystems, that would cruise (graze) near the equator producing an energy storage product, or deliver electric power via underwater cable from OTEC plants moored near shore.

In the first phase of the PON program, funds were appropriated with the intent to support a variety of conceptual designs of OTEC systems, by the six industrial teams. However, with the change in program management to the Reagan administration in 1981, only the proposals by GE and Ocean Thermal Corporation (OTC) were selected for *conceptual design* funding. At completion of this phase, OTC was awarded the only second-phase contract to do

*preliminary engineering design* of an OTEC plant that would be installed on an artificial island near the warm water out-flow from an oil-fired power plant of the Hawaiian Electric Corporation (HECO). The proposal envisioned mixing near-shore ocean surface water with the HECO plant outflow to raise the temperature of the warm water inflow to the OTEC plant from about 25 to 29°C, with an improvement in cycle efficiency.

Volumes of engineering data on OTEC subsystem and components requirements, performance, and costs were developed by the proposal teams. However, when OTEC support was discontinued in 1982 the proposal material was destroyed because it contained proprietary information that DOE could not protect. Part of the material was made available to JHU/APL and is the source of the following discussion of the U.S. OTEC engineering status. Information on French and Japanese programs is taken from Avery and Wu (1994) and Kalina (1987).

Plans were announced in 1997 by the Indian National Institute of Ocean Technology (NIOT) for installation of a 1-MWe floating OTEC plant at a site near Tuticorin, South India. The project is being conducted jointly with a team from Saga University in Japan, headed by Professor Uehara. Operation of the 1-MWe demonstration vessel was scheduled to begin in 2001.

## A. Standard Equipment Available for OTEC

OTEC can directly use technology in commercial use in the shipbuilding industry, in the oil industry, and in power generation. Firm costs are available from these sources for equipment that represents 40–60% of the total OTEC system costs.

### 1. Shipyards

Facilities in the United States suitable for OTEC construction include: Newport News Shipbuilding Co., Avondale Industries, Inc., and Concrete Technology Corporation. Concrete Technology Corporation in 1976 built a 146-m-long concrete barge (the ARCO barge) of roughly the same dimensions as those proposed for the OTEC 40-Mwe pilot plant. The design and construction of that vessel provided detailed information on engineering requirements and costs that were used in the base-line design of the 40-MWe OTEC barge. Shipyards in Germany, Spain, Portugal, Japan, Singapore, and Korea have built supertankers, and their facilities would be readily adaptable for commercial construction of OTEC plantships.

### 2. Construction Procedure for Floating Systems

A complete OTEC system will be constructed in stages. The hull is laid down in a dry dock and then floated to a nearby pier where water depth is 20 m or more. The major equipment items are installed. The vessel is then towed to a tropical site where CWP construction facilities are available and nearby water depth exceeds 1000 m. CWPs are constructed, transported to the floating offshore OTEC vessel, and installed. The complete assembly is then towed, or navigated using onboard power, to the operating site.

## B. OTEC Land- and Shelf-Based Installations

These are site specific. Details of proposed systems are given in Section D.

## C. Pilot-Scale OTEC Systems Tests

### 1. U.S. Mini-OTEC Program

A complete OTEC system, including a 670-m CWP, was constructed, deployed, and tested in the Mini-OTEC program, in 1978–1979. The program was conducted with private funding by an industrial team headed by the Lockheed Missiles and Space Division. Mini-OTEC was constructed in 28 months, using off-the-shelf components. Net power of 18 kW was produced, in close agreement with the predicted performance. The Mini-OTEC program demonstrated OTEC floating system feasibility in 4 months of at-sea operation, including first-time production at sea of net OTEC power.

### 2. OTEC-1 Program and Accomplishments

This program was conducted by DOE to demonstrate OTEC heat exchanger operation at 35-MWt scale in an at-sea environment. A moth-balled Navy T2 tanker was retrofitted to house the OTEC system components. A simulated OTEC power system was installed in which the turbine-generator was replaced by a throttle valve that simulated the pressure drop through the turbine. The equipment included a 35-MWt shell and tube heat exchanger, a 700-m-long, 2-m-diameter plastic CWP and pumps. Provision was made for testing biofouling control methods such as Amertap and chlorine injection. Diesel generators provided electric power to operate pumps and auxiliaries. The CWP was assembled on shore then towed to a site roughly 10 miles off-shore from Kailua-Kona, Hawaii, where it was upended and inserted into a gimbal on the ship. The gimbal allowed movement through an angle of 30°C.

Significant accomplishments demonstrated during the 4-month at-sea operation were (a) The system performed well with a maximum wave height of 4 m, maximum subsurface currents of 3 knots, wind gusts up to 45 knots, and ship roll angle of 18°. (b) The heat exchanger cycle operated in close accord with predicted performance.

(c) Deployment and operation of the CWP were successful. The pipe survived a sea state that caused the pipe to move to a 30° angle at the gimbal. Detachment of the ship from the mooring at this point caused the pipe angle to drop to 10°. This indicated that survival of floating OTEC systems in extreme storm conditions would involve significantly lower risk in the grazing mode than in moored operation. (d) Biofouling control was accomplished with injection of 60 parts per billion of chlorine, 1 hr per day, well below EPA requirements for drinking water. (e) The test provided assurance that the marine operational requirements were well understood. The test program involved no surprises.

### 3. Japanese Shore-Based 100-kW OTEC Pilot Plant at Nauru

In cooperation with the government of the Republic of Nauru, engineers of the Tokyo Electric Power Co. and the Toshiba Corporation in 1980 constructed a 100-kW (gross) OTEC plant at a site on the island of Nauru, which is on the equator 40° east of the Philippine Islands. Power-generating operation was performed from October 1981 to December 1981. Heat-loop operation was continued to July 1982 (Mitsui *et al.*, 1983; Ito and Seya, 1985). The Nauru test was designed to provide accurate experimental data on the performance of the complete power cycle as well as demonstrate total pilot-plant construction and operation. The goals were all successfully accomplished. The plant provided the first demonstration of land-based OTEC net power generation and established a record for total net power production, at 31.5 kW. It also was the first demonstration of OTEC power connection to a utility grid.

### 4. Japanese Mini-OTEC Program

A Mini-OTEC plant was constructed for at-sea tests off the coast of Shimane Prefecture, as an initial phase of the Japanese program directed to demonstration of a 100-MWe moored OTEC system. Deployment took place on October 11, 1979. The work was done by a team headed by Professor Uehara of Saga University. The tests included deployment of a steel CWP, 190 m long, made of 12-m-long tubular sections of 26.7-cm OD and 0.45-cm wall thickness, joined by flanges. The pipe was supported by a float and held in place by a steel cable attached to the pipe and anchored at a depth of 300 m. The OTEC power plant (with plate-fin heat exchangers) and water pumps were mounted on the research vessel, which was attached to the float. A flexible pipe carried the cold water from the float to the vessel. The program demonstrated successful deployment and operation of the system, including heat transfer measurements.

### D. Conceptual Designs of OTEC Systems

Responses to the DOE program opportunity notice (PON) are described here. Only the GE and Ocean Thermal Corp. proposals were funded for the Conceptional Design phase.

### 1. General Electric (GE) 40-MWe Shelf-Mounted Hawaii Plant

GE proposed a tower-mounted OTEC plant which would be installed 2200 m offshore at a depth of 100 m, at Kahe Point, Oahu, Hawaii. Power produced at the OTEC platform is carried by underwater cable to the Hawaiian Electric Company (HECO) plant for distribution to its customers.

### 2. Puerto Rico Electric Power Authority (PREPA)

PREPA proposed a tower-mounted 40-MWe (net) OTEC plant to be sited at the edge of the continental shelf at 75-m depth approximately 200 m offshore from Punta Tuna, Puerto Rico. Electric power is transmitted by underwater cable to the PREPA power distribution system. Cold water is drawn from a depth of 1000 m at a distance of approximately 1500 m from the tower.

### 3. SOLARAMCO Concept of OTEC 40-MWe (net) Grazing Ammonia Plantship

The SOLARAMCO group proposed a floating 40-MWe (net) OTEC ammonia plantship system, designed to graze in a region south of Hawaii where a $\Delta T$ of 22–23°C would be available. The plantship would use the net OTEC electric power for water electrolysis to produce hydrogen for on-board ammonia synthesis. The proposal included provisions for testing four 10-MWe power modules equipped with different heat exchanger types.

### 4. OSEA 100-MWe OTEC Power Plant

A consortium named OSEA, that included SSP (Sea Solar Power, Inc.), GE, and General Dynamics, proposed a 40-MWe floating OTEC plant of the SSP design that would supply power to the Puerto Rico Water and Power Authority. (PREPA). The technical details are proprietary. The following discussion is based on the published reports of the SSP 100-MWe OTEC plant conceptual designs and analyses. Features include (a) design of the power plant components for submerged operation so that the major power plant components can be placed in the ocean outside the hull; (b) turbine-driven pumps for the warm and cold water, which eliminate efficiency losses in conventional designs that result from conversion of turbine power to electric power for operation of electrically driven

pump motors; (c) a new heat exchanger design with projected heat transfer rate four times that of conventional heat exchanger designs; (d) a four-stage vapor cycle to improve the average temperature difference between the water flows and working medium flow. Placing the evaporators 25 m below the condensers allows the condensed liquid, R-22, to flow by gravity to the evaporators, eliminating the need for a working fluid liquid pump and the associated power losses; (f) a "stockade" construction of the CWP designed to permit CWP assembly on the OTEC platform from sections of steel pipe that would be easy to transport. With appropriate valves and seals, the hollow pipes could be used to control the buoyancy of the CWP. Details of the construction, deployment, and operation are not available.

## 5. 200-MWe (nom) OTEC Methanol Plantship

As a follow-up to the preliminary design of the APL 40-MWe OTEC baseline ammonia plantship, DOE supported, a conceptual design study of a 200-MWe (nom) OTEC plantship that would use electrolytic hydrogen, and oxygen reacting with coal, for on-board methanol synthesis. Pulverized coal for the process is transported to the plantship by conventional coal colliers. The methanol is accumulated in on-board tanks for approximately 1 month and then shipped to continental ports in conventional tankers. The process efficiently uses both oxygen and hydrogen produced by water electrolysis in combination with low-cost coal to produce a low-pollution liquid vehicle fuel that can replace gasoline for transportation at competitive costs. The platform configuration is based on scale-up of the JHU/APL 40-MWe plantship to 200 MWe to provide power and space for a 1750-tonne/day methanol plant. Layout diagrams are shown in Figs. 8a,b. The methanol plant is placed at the center of the platform with OTEC power modules at either end. Power for sea-keeping is provided by rotatable thrusters beneath the hull. The power plant includes four power modules at each end of the barge. Each module includes two evaporators and two condensers, supplied with seawater by gravity flow from overhead ponds. CWPs, 7–10 m diameter made of FRP or lightweight concrete, are supported by a universal joint or ball-and-socket mounting. Energy transfer is done by production of methanol on the OTEC plantship with subsequent shipment to world ports. The system will produce 1750 tonnes/day of methanol.

## 6. Conceptual Designs of Japanese 100-MWe Floating OTEC Plants

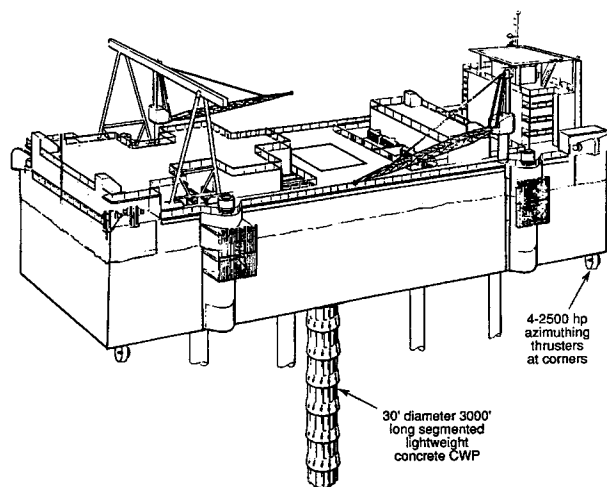During the period 1974–1977, a program, named the "Sunshine Project," was established in Japan to explore



**FIGURE 7a**  Sketch of JHU/APL 40-MWe baseline OTEC ammonia plantship [From George, J. F., and Richards, D. (1980). *Johns Hopkins Appl. Phys. Lab.* SR-80-A&B.]

the feasibility and cost of supplying power- or energy-intensive products to Japan via OTEC. The program began with a first-phase study of a 1-MWe, land-based plant and proceeded to a study of a 100-MWe floating plant. The design included platform structure, station keeping, riser cable design, and heat exchanger improvements. The OTEC system employs moored floating surface ships or submerged discs. The power plant uses ammonia as the working fluid. Plate heat exchangers are incorporated in 25-MWe power modules. An estimate of the total OTEC power that could be produced from the thermal resources of the sea near Japan indicated a potential output of $3.75 \times 10^5$ GWh/year from an area $1°$ wide in longitude and $1°$ in latitude. This output exceeds the annual electric power consumption in Japan.

## 7. Conceptual Designs of French 5-MWe OTEC Plants at Tahiti

A feasibility study of shore-based and floating OTEC plants to be sited in Tahiti was conducted during 1978–1980 under sponsorship of the French Centre pour l'Exploitation des Oceans (CNEXO). The program included detailed site studies, evaluation of closed- and open-cycle options, methods of constructing and deploying the cold water pipe and estimates of costs of favored alternatives. Preliminary surveys led to selection of a site at the port of Papeete for the projected plant. The design does not call for a platform or containment building. Instead the power plant components are mounted separately. Plant layouts for closed-cycle and open-cycle systems were developed. A pond adjoining the power plant installation receives nutrient-rich water leaving the condensers that
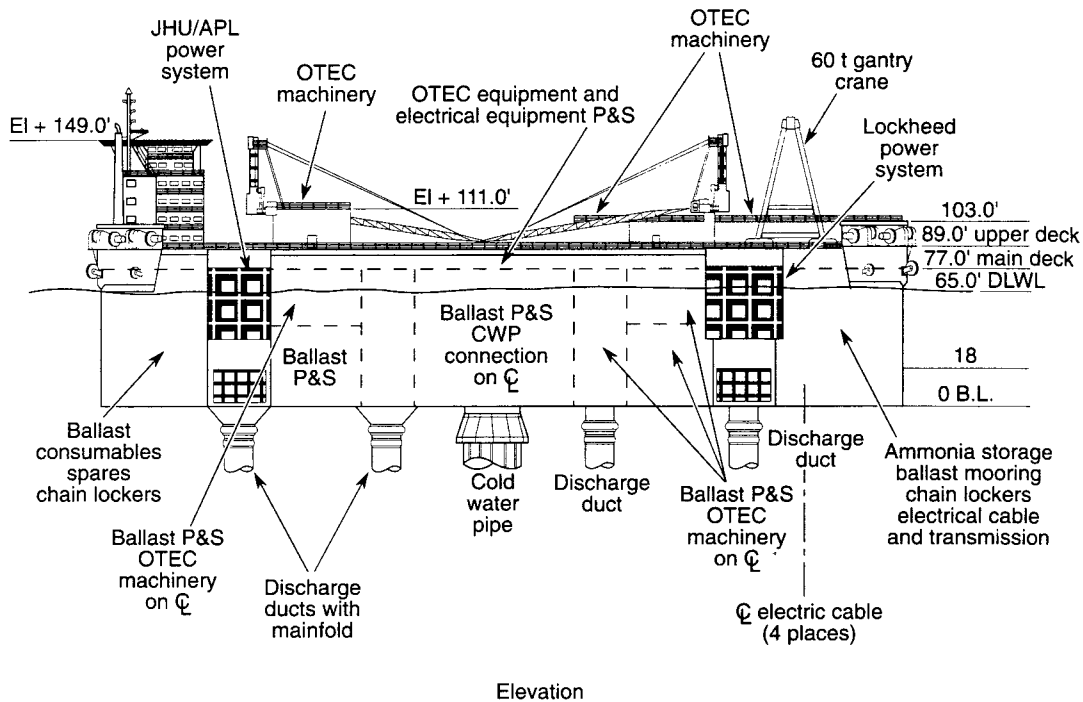
**FIGURE 7b** Subsystems layout of baseline 40-MWe OTEC barge. [From George, J. F., and Richards, D. (1980). *Johns Hopkins Appl. Phys. Lab*. SR-80-A&B.]

are to be used for mariculture programs. The water effluent streams are finally directed to a lagoon from which the water flows to the ocean. Conceptual designs of both closed- and open-cycle OTEC installations were prepared, based on selection of the most attractive of a wide range of alternative designs. Investment comparisons indicated that the costs of the two systems would be within 5% of each other. The closed-cycle plant uses horizontal shell and tube heat exchangers with smooth titanium tubes in both evaporators and condensers. Ammonia was chosen as the working fluid. The open-cycle design uses spout evaporators and direct-contact condensation of the steam with the cold sea water. A variety of pipe designs were evaluated. The preferred overall design that emerged included a double-walled CWP of FRP, enclosed in a 15-m-deep concrete pipe entrenched in the shallow water zone at the shore, extending approximately 100 m to a point where the sea floor drops sharply. The deep-water pipe is supported above the sea floor in a trajectory defined by "hawses" anchored by piles in the sea floor and supported by buoys designed to float at 50 m depth. The neutrally buoyant CWP, 3 m in diameter, assembled on shore as deployment proceeds, is deployed by pulling it through the hawses. Direct transmission of electric power, preparation of fresh water, and mariculture could be combined to make operation of the 5-MWe Tahiti plant of commercial interest.

## 8. Conceptual Design of Taiwan 50-MWe (net) Shore-Based OTEC Plants

In 1984 the Taiwan Power Company initiated a program to determine the feasibility and to prepare conceptual designs of shore-based 50-MWe OTEC plants in Taiwan. Three sites on the east coast of Taiwan were selected for evaluation. Studies led to selection of shore-based plants at Ho-Ping and Chang-Yuan as preferred sites for Conceptual Design. It was decided that the platform (containment structure) should be constructed in a shipyard, where facilities for installing heavy equipment would be available. The platform would then be floated and towed as a unit to the plant site on the beach. Here it would be lowered by flooding the sumps to rest on a prepared bottom foundation behind an appropriate breakwater. There are four 12.5-MWe (net) power modules that use ammonia as the working fluid in horizontal shell-and-tube heat exchangers of conventional overall design. Titanium tubes are employed. The design permits use of either chlorine or Amertap for biofouling control. Warm and cold water enter the platform through pipes that pass through the breakwater and emerge into sumps below the platform floor. Water is forced through the heat exchangers by vertical pumps that draw water from the sumps. The heat exchanger effluents discharge into a common sump from which the water flows out by gravity. A single CWP of 10 m diameter,

made of lightweight concrete, provides cold water for all four power modules. Studies of the variation of net power versus water flow velocity showed that maximum net power of 50 MWe would be possible with a 10-m pipe diameter. Consideration of various methods of pipe construction and deployment led to the conclusion that the pipe should be constructed in sections, 10 m in diameter and 15 m long, with elastomeric bell and spigot configurations at the ends. The pipe sections would be barged to the plant site where they would be joined end to end to form the cold water pipe. Installation of the power plant on the beach allows the generated power to be transferred directly to the onshore utility.

## E. Preliminary Design of OTEC Plants

### 1. Baseline 40-MWe (net) Moored and Grazing OTEC Plants

In 1977, The Applied Physics Laboratory of the Johns Hopkins University (JHU/APL) was awarded a contract by DOE to prepare baseline preliminary engineering designs of moored and grazing 40-MWe (net) OTEC plants that would furnish basic data on systems requirements, performance, and costs. The information would be used to judge the responses to the PON. A barge-type hull made of post-tensioned concrete was selected for the baseline OTEC grazing and moored system designs because of suitability for both missions, ease of installation and replacement of subsystems and components, minimal cost, compatibility with aluminum heat exchangers, and long-term durability. Provisions are made for alternative types of heat exchangers, water ducting including the CWP, pumps, energy delivery systems, station-keeping systems, support facilities, and personnel accommodations. The configuration and susbsystems layout are shown in Figs. 7a,b. The platform length is 135 m. Analysis of the maximum loads that would be imposed by the 100-year storm defined the structural design. The analysis included consideration of the survival requirements for grazing plantships stationed in the equatorial Atlantic Ocean (significant wave height ((Hs) 8 m, period (To) 18.0 sec)) and for moored plants stationed near Puerto Rico ((Hs) 10.9 m, (To) 13.1 sec) or near the island of Oahu in Hawaii ((Hs) 8.4 m., (To) 11.7 sec). The OTEC barge accommodates alternative 20-MWe power systems of designs proposed by JHU/APL and Lockheed Space Sysytems. The dimensions and weight of the two systems are nearly the same. Thus, the hull configuration is suitable for 40-MWe (net) power systems of either design. Other compact power systems such as the Trane or Alfa-Laval designs could also be installed without major changes in the hull dimensions.

The APL power system was designed with two major objectives: (a) to minimize HX cost by using low cost, roll-welded aluminum tubing in a compact folded-tube configuration with internal ammonia flow. This eliminates the bulky containment structure needed in conventional shell and tube HX designs, thereby reducing packaging space and cost; (b) to minimize platform cost by making the water ducting structures compact, integral, load-bearing elements of the hull structure. The power system employs two 5-MWe (net) evaporator and condenser modules combined with a single turbine to form a complete 10-MWe (net) power module. Each 5-MWe (net) unit is housed in a separate bay with vertical sides that form an integral part of the hull structure. Two 10-MWe (net) modules, each having two evaporators and two condensers, are located at the fore and aft ends of the baseline platform. The practicality of fabrication, assembly, installation, and deployment of heat exchangers of the folded-tube type was the subject of a separate design study conducted by the Trane Corporation, which verified the suitability of the JHU/APL HX design for low-cost, mass production. The platform space and HX bays allotted to the APL design are well suited to packaging the Lockheed module, thus no modifications of the baseline hull are indicated. The plate heat exchangers are mounted in two tiers in a heat exchanger bay. Seawater from the overhead pond enters the heat exchangers at the sides, flows horizontally across the heat exchanger plates, and then exits at the bottom of the HX bay. Liquid ammonia enters the evaporator from below and flows vertically upward as it vaporizes. Vapor leaving the turbines enters the condensers at the top and flows vertically downward to the sump from which it is pumped to the evaporators. Analysis by Lockheed showed that this cross-flow arrangement was superior to one using parallel flow.

Cold water is supplied to the OTEC power systems through a 9.1-m-diameter, 930-m-long segmented concrete pipe, supported by a ball-and-socket structure incorporated in the hull at approximately the center of gravity. The joint allows the pipe to pivot freely through an angle of 18° to the vertical. It also allows the pipe to rotate about the vertical axis. The dynamics of the coupled system, analyzed as discussed in Section IV, show that the coupled pipe-hull system will survive the most extreme conditions of the 100-year-storm in grazing operation at the tropical Atlantic site, with a good safety margin, and in moored operation at the Hawaii site. The analysis predicts that the design is marginal for the extreme hurricane condition at the Puerto Rico site, where the combination of extreme values of surface currents and waves leads to a predicted maximum pipe angle of 19.7° (sum of extreme values due to hull deflections and surface currents). The calculation does not include mitigating effects of mooring forces on the platform motions or potential benefits of bow-fairing of the rectangular platform. Further study is required to indicate detail changes in design that might be required

for moored operation at this site. Warm water enters the evaporators through screens mounted at the sides of the hull at a mean depth of 16.6 m. The screen area is about 125 m$^2$ for each 10-MWe evaporator module. In the grazing mode of operation, the warm and cold water effluents from the heat exchangers are discharged at the bottom of the barge, where the higher density of the effluents than those of the surrounding water causes them to sink. Since the grazing barge is always moving into fresh water, there is no recirculation of the discharged water into the warm water inlets. The moored baseline plant is designed with discharge pipes to carry the effluent to a depth below the thermocline to prevent reingestion and possible impacts on the near-shore environment. Eight pipes 4.6-m in diameter and 76-m long are used to carry the mixed effluent below the thermocline, where it is discharged and allowed to sink. A detailed structural analysis of these pipes was not funded but will be necessary if they are to be retained in the design. The OTEC-1 experiment indicated that these pipes may be unnecessary because the mixed plume descends rapidly and may not pose any reingestion problem. However, termination of the program did not permit a conclusion to be reached.

The OTEC on-board electric power generation and distribution system is based on the use of standard commercial equipment for primary and secondary applications and is designed to comply with the requirements of IEEE standard No. 45 and relevant regulations of the U.S. Coast Guard, ABS, NEMA, NFPA, and UL. There are four 16-MWe (gross) ammonia turbine generators, each having its own set of auxiliaries and switch gear. Ammonia storage, liquid nitrogen, circulating cooling water, and generator loading systems are used in common. All of the generators may be used in parallel, and in normal operation all power will come from them. During start up or when the OTEC plant is shut down during storms, reserve power is provided by diesel-generator sets. Tie feeders between the OTEC generators and the reserve power units allow this interchange. The on-board power utilization systems differ for the grazing and moored OTEC options; however, the systems are the same up to the final phase. In the grazing plant, step-down/DC rectifiers are installed to furnish DC current to the electrolyzer cells. In the moored plant, transformers step up the voltage to 115 kV for cable transmission to shore. In OTEC plantships, electric power generated on board is used for water electrolysis, which provides efficient conversion of electrical energy to chemical energy stored in hydrogen-based fuels that can be transported to land sites.

The baseline 40-MWe (net) OTEC plantship uses the power produced by the turbo-generator to produce ammonia (NH3). The on-board ammonia plant includes water electrolyzers to generate gaseous hydrogen (along with gaseous oxygen that is vented to the air), an air separation plant to produce pure nitrogen, a catalytic converter to combine the hydrogen and nitrogen to form ammonia, ammonia liquefaction and storage equipment, and facilities for transferring the liquid ammonia to tankers for shipment to world ports. Details of the ammonia synthesis process are described in Section IV.A.3. Commercial ammonia plants typically produce 1000 tons/day (909 tonnes/day) of liquid ammonia. The output of the 46-MWe (net) baseline plantship would be 125 tons/day.

The feasibility and practicality of installing an ammonia plant on an OTEC platform were investigated in detail in the baseline study, with favorable results, including possible effects of ship motions on the process operations. Smaller and less efficient plant components would be used in the baseline vessel in comparison with commercial plants. The OTEC ammonia plant would be similar to the commercial ammonia plant designs but would be simpler because hydrogen would be produced by water electrolysis rather than by the expensive steam-reforming of methane, which also generates carbon dioxide. A flow diagram of the process is shown in Fig. 8. An installation diagram of a 113-tonnes/day ammonia plant on the OTEC platform is shown in Fig. 9.

It was recognized early in the OTEC program that at favorable sites moored OTEC plants could deliver OTEC power directly to onshore facilities via underwater power cables (Winer, 1977). The requirements for this mode of OTEC operation were also determined in the 40-MWe (net) baseline design study. The site selected for detailed study was near Punta Tuna, Puerto Rico. Power transfer requires on-board equipment to convert the OTEC bus-bar output to suitable voltage and frequency for transmission through the selected power cable, and equipment to support the power cable that must connect to the OTEC platform through a flexible junction. A proposed arrangement for the aft cable installation on the platform is shown in Fig. 10. Experiments with power cables under simulated OTEC conditions indicated satisfactory durability could be achieved, however, a complete design of the power cable system had not been completed when the program was terminated. A design of the electrical system, which satisfies the total requirements of both moored and and grazing OTEC plants, was produced. The principal features of the JHU/APL 40-MWe (nom) OTEC baseline plant designs are listed in Table IV.

### 2. Preliminary Design of Shelf-Mounted OTC OTEC Power Plant for Hawaii

At the completion of the DOE Phase 1 conceptual design competition, a contract was awarded to the Ocean Thermal Corporation (OTC) for preparation of a preliminary
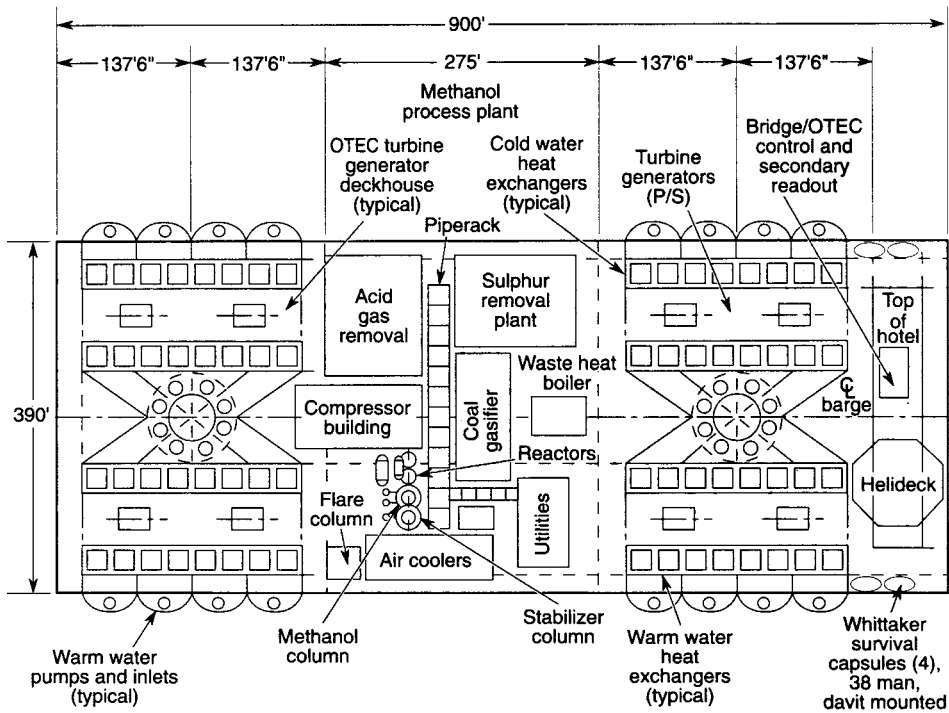
**FIGURE 8a**  Plan view of 1750 mt/d methanol plantship [BARDI 1982. *Coal-to-methanol study using OTEC technology*. For Johns Hopkins University Applied Physics Laboratory. Brown and Root Development, Inc., Houston, Texas.]

design of their proposed shelf-mounted plant. The plant would be sited on the continental shelf 600 m offshore from the Hawaiian Electric Company's (HECO) oil-fired power plant at Kahe Point on the island of Oahu. In the OTC concept, shown in Fig. 11, the warm water outflow

from the HECO plant is mixed with the warm seawater input to the OTEC evaporator, thereby raising the temperature of the OTEC inlet water from 25.6 to 27.9°C, with a significant increase in OTEC performance. The goal of the OTC contract was to develop, with government cost
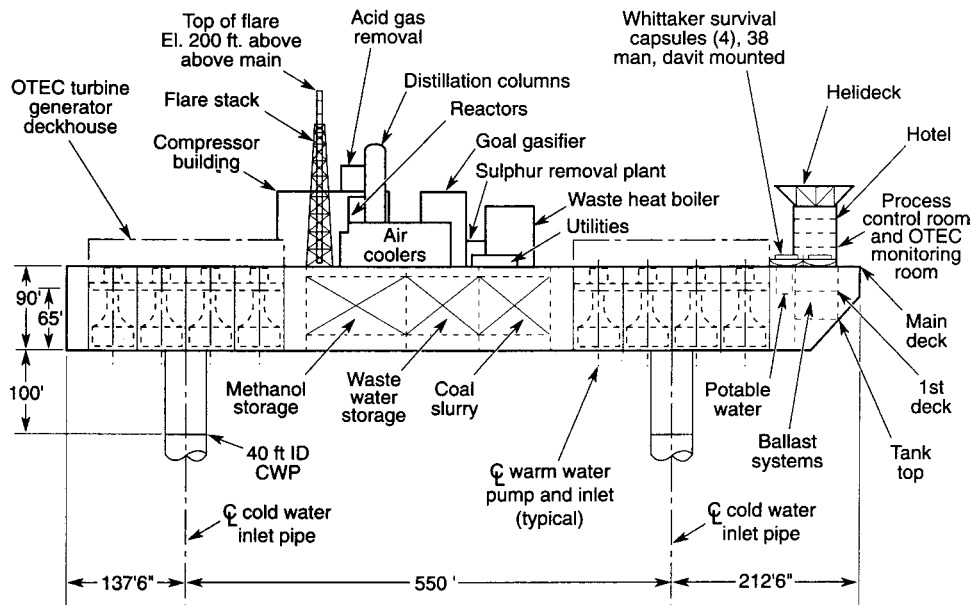


**FIGURE 8b**  Side view of OTEC methanol plantship [BARDI 1982. *Coal-to-methanol study using OTEC technology*. For Johns Hopkins University Applied Physics Laboratory. Brown and Root Development, Inc., Houston, Texas.]

sharing, an engineering and economic evaluation of an OTEC power plant design that would provide a firm basis for detailed engineering and construction of a shore-based 40-MWe demonstration OTEC electric plant. In the conduct of the contract, the following topics were evaluated in some depth:

| | |
|---|---|
| OTEC system performance and operational requirements | Preliminary design system engineering |
| Codes, standards, and recommended practices | OTEC plant system characteristics |
| HECO requirements | OTEC plant system overall layout |
| Possible impact on military operations near OAHU | Land-based containment system (LCBS) design |
| Aesthetic considerations | Power system design |
| Site-specific design conditions | Water ducting design |
| Site description | Asset acquisition |
| Wind, wave design criteria | Project achedule |
| Tsunamis | OTEC plant system deployment |
| Tides | LBCS deployment |
| Currents | Inspection, maintenance and repair |
| Seawater temperatures | Safety analysis |
| HECO condenser outflow conditions | Environmental monitoring and assessment |
| Seismicity | Cost and commercialization prospects |
| Climatology | Risk assessment |
| Biofouling and corrosion | Pipeline deployment |

The LBCS comprises a concrete building 71.8 m wide (236.5 ft), 93.3 m long (306 ft), and 33.5 m high (110 ft), 10 m above sea level mounted on the sea bottom at a

depth of 22.2 m (73 ft), at 550 m (1800 ft) offshore. The LBCS houses the OTEC power plant, water and ammonia pumps, biofouling control equipment, and power conversion equipment. A breakwater structure extending from the sea floor to just above sea level is attached on the seaward end. Design of the CWP system required an extensive study of the physical and hydrological features of the Kahe point site. The studies included detailed bathymetric surveys with a resolution of 50 m covering approximately 5.5 km$^2$ bordering Kahe Point. Bottom core samples and subsurface data representative of the area were also acquired. With this information it was possible to define the pipe structural requirements and to lay out a pipe trajectory that would minimize risks in deployment and long-term operation, and would be the most cost effective. In the final preliminary design of the CWP, water enters the LBCS through a bottom mounted pipe 3660 m long that draws water from an offshore depth of 670 m (2200 ft). The CWP is divided into two major segments. The seaward section of the pipe is of FRP. It is 2560 m long (8400 ft), of 6 m diameter (19.8 ft), and 0.072-m wall thickness (2.83 in.). It is composed of segments 75 m to 90 m long, joined by flexible bellows, and supported on a steel support structure that rests on the bottom. This structure protects the pipe from bottom erosion, relieves tension loads on the FRP, provides weight to ensure stability against hydrodynamic loads, and allows flexibility to accommodate seismic loads, bottom scour, and differential settling. At the top of the escarpment, the FRP pipe connects to a triple-pipe concrete structure that combines a pipe to conduct the
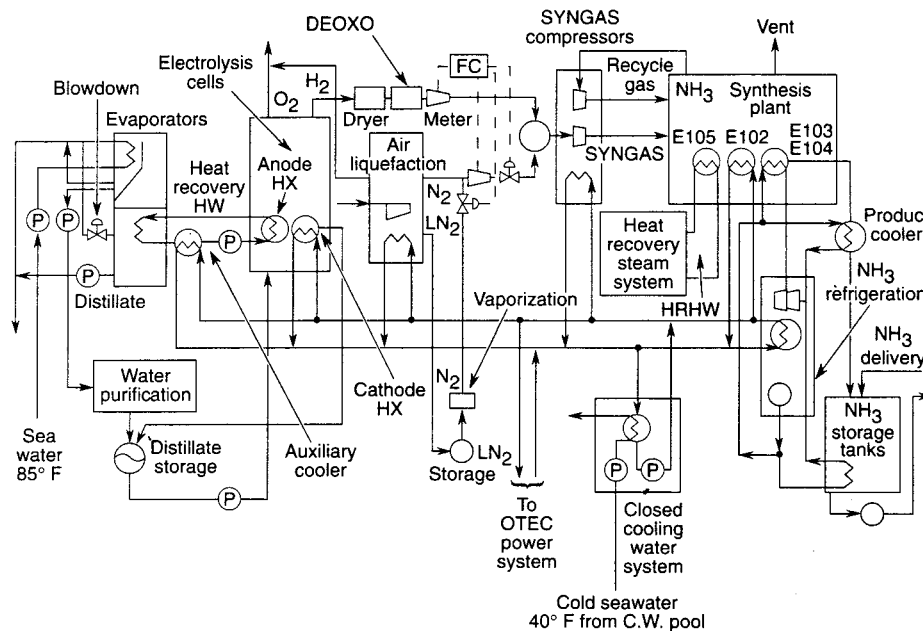


**FIGURE 9** Flow diagram for OTEC ammonia plant. [From George, J. F., and Richards, D. (1980). *Johns Hopkins Appl. Phys. Lab.* SR-80-A&B.]
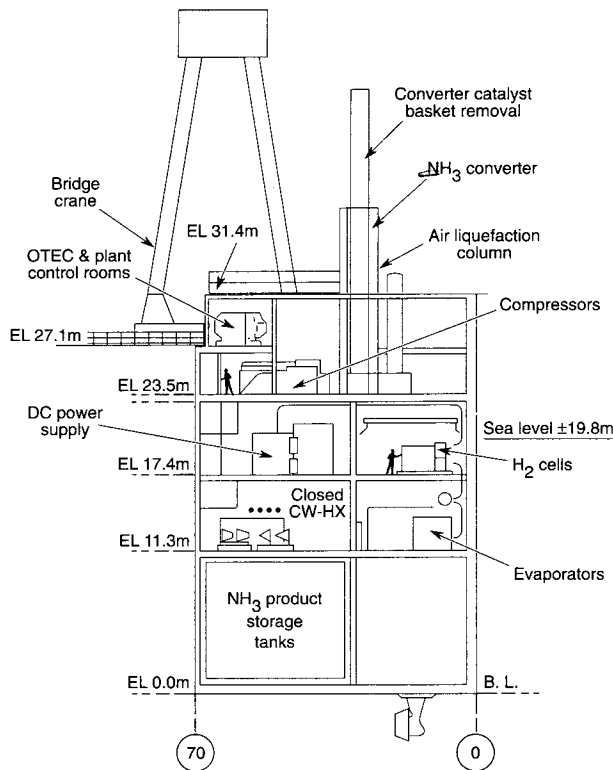
**FIGURE 10a** Schematic diagram of 40-MWe OTEC ammonia plant installation. [From George, J. F., and Richards, D. (1980). *Johns Hopkins Appl. Phys. Lab.* SR-80-A&B.]



**FIGURE 10b** Schematic diagram of power cable equipment installation on 40-MWe moored OTEC plant. [From George J. F., and Richards, D. (1980). *Johns Hopkins Appl. Phys. Lab.* SR-80-A&B.]

cold water to the LBCS, with two parallel pipes that carry the mixed effluent (flowing in the opposite direction) from the OTEC plant to the discharge point. A profile of this section and the concrete triple-pipe structure is shown in Fig. 11. Near-surface seawater enters the LBCS through a structure on the landward side where it is mixed with warm water conducted from the HECO condenser discharge. After passing through the OTEC power system, the exhaust streams from the evaporators and condensers are mixed and then conducted through the bottom-mounted concrete ducting structure to the edge of the escarpment where the combined flow is discharged at a depth of approximately 115 m (380 ft). All support facilities not directly involved in plant power production are sited onshore. These include offices, laboratories, maintenance facilities, ammonia and equipment storage, and related equipment. The principal features of the OTC design are listed in Table V. Cost information is listed in Section XII.

## F. Conclusions about Closed-Cycle OTEC Development Status

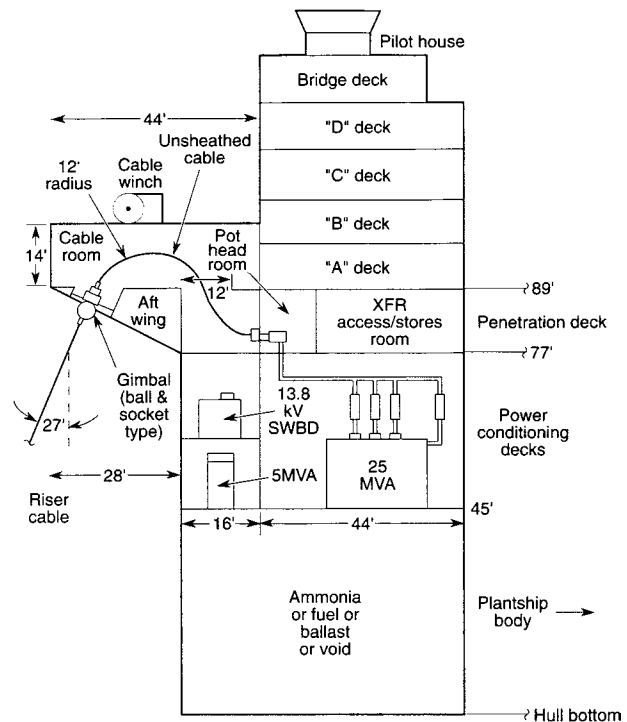The review in this chapter of the OTEC programs conducted in the United States and abroad shows that the technology required for high-confidence pilot plant demonstration of sea-based and land-based OTEC systems was available in 1984. Satisfactory solutions to feasibility questions were demonstrated. It should be noted that a complete OTEC plantship assembly will incorporate commercially available turbines, generators, pumps, power distribution equipment, control equipment for power systems and sea-keeping, hotel and shop facilities, and miscellaneous items. Only the construction of cold water pipes and advanced types of heat exchangers will involve new technology. These items will contribute about 20% of the total estimated cost.

Commercial applications of OTEC have been proposed in three principal categories. The first includes OTEC plantships that would generate a 50- to 400-MWe (net) of onboard electric power. The need to minimize plant size makes it mandatory to use closed-cycle OTEC for these applications. The second category includes land-based or shelf-mounted plants designed to supply power in the 50- to 400-MWe range to municipal utilities. Closed-cycle systems would be suitable. The third category comprises small (5- to 20-MWe) land-based or shelf-mounted OTEC plants designed for island (and some other) applications where electric power generation, mariculture, fresh water production, supply of cold water for air-conditioning systems, cold water agriculture and fuel production could be combined to offer an economically attractive OTEC

**TABLE IV   JHU/APL Preliminary Design of Baseline 40-MWe (nom) OTEC Plants**

Hull

  Rectangular barge

  Dimensions in (ft): $L$ 135 (443.5), $W$ 42.7 (140), $D$ 31.4 (103)

  Hull material: post-tensioned concrete

Power Plant

  Closed Rankine cycle; ammonia working fluid

  Folded tube HXs of JHU/APL design with nested 0.072-m (3-in.) tubes, or plate HXs of Lockheed design using titanium plates

  Internal ammonia flow; 4 10-MWe (net) modules

  Nominal operating characteristics

    Nominal $\Delta T$ (°C)

      Annual average 23.65

      Seasonal $\Delta T$ (°C) (ATL-1): Feb. 23.6, May 24.6, Aug. 23.4, Nov. 23.0

| | JHU/APL power modules | Lockheed power modules |
|---|---|---|
| **Power [MWE (nom)] (optimized for ATL-1)** | | |
| Gross turbo-gen output | 52.0 | 52.9 |
| Cold water pump | 5.76 | 9.66 (both pumps) |
| Warm water pump | 4.08 | |
| Ammonia pumps | 1.51 | 1.18 |
| Noncondensible purge | 0.16 | |
| Ship services | 0.17 | 2.07 |
| Net busbar power | 40.3 | 40.0 |

| | JHU/APL Evap. | JHU/APL Cond. | Lockheed Evap. | Lockheed Cond. |
|---|---|---|---|---|
| Heat load (MWt) | 1,747 | 1,689 | 1,680 | 1,629 |
| Water flow (kg/sec) | 132,300 | 136,720 | 105,600 | 102,900 |
| $NH_3$ turbine flow (kg/sec) | 2,824 | 2,824 | 1,722 | 1,722 |
| $U_0$ (W/m² K) | 2,537 | 2,463 | 3,259 | 3,118 |
| Foul. coeff. (W/m² K) | $3.5 \times 10^5$ | $3.5 \times 10^5$ | $5 \times 10^5$ | $5 \times 10^5$ |
| $H_2O$ in/out temp. (°C) | 27.89/25.27 | 4.56/7.46 | 27.78/23.81 | 4.44/8.36 |
| $NH_3$ in/outlet temp. (°C) | 23.94/21.94 | 9.50/10.00 | 13.37/21.71 | 10.37/10.47 |
| Seawater $\Delta p$ (m) | 2.58 | 2.95 | 3.21 | 3.44 |
|   Inlet duct | 0.51 | 0.88 | 0.51 | 0.88 |
|   HX | 2.07 | 2.07 | 2.70 | 2.56 |
| Seawater velocity (m/sec) | | | | |
|   Inlet duct | 0.3 | 2.0 | n.a. | n.a. |
|   HX | 0.76 | 0.76 | n.a. | n.a. |
| Seawater flow [m³/s/MWe (net)] | 3.24 | 3.33 | 2.59 | 2.51 |

system despite the relatively high cost of power from small OTEC installations. Open-cycle OTEC plants could be the preferred choice at some sites for the third category. Since the importance of OTEC as a renewable energy source will depend on its ability to furnish significant quantities of fuels that will be economically and environmentally attractive, the major emphasis in the following discussion of OTEC economics is placed on the OTEC plantship options.

The estimated investment costs of installed complete OTEC systems, measured in dollars per kilowatt of net OTEC electric power generated, differ significantly among the three categories. Large cost benefits are gained by siting OTEC systems where $\Delta T$ is a maximum, by scaling to optimum size, by large volume production of subsystems and components, and by integrated system design. Small land-based systems are relatively expensive in dollars per kilowatt of power generated but the cost can be offset in part by using the cold water for air-conditioning, mariculture, cold water agriculture, and fresh water production. Experiments at reasonable scale demonstrated practical solutions for all of the technical
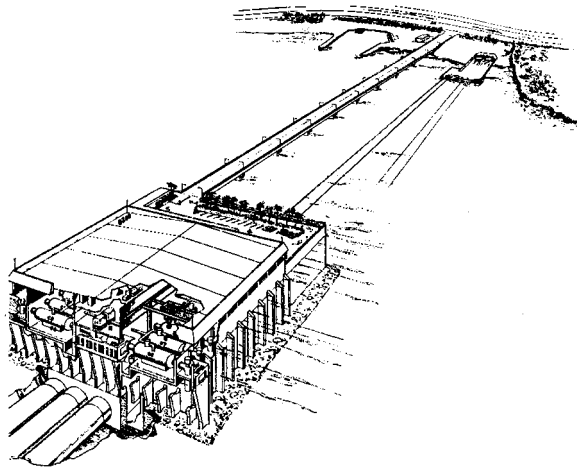
**FIGURE 11** Artist's sketch of the OTC 50-MWe OTEC power plant at Kea-hole Point, Hawaii. [From Ocean Thermal Corporation (1984). *Development of the Land-Based Containment System for the 40 Megawatt Ocean Thermal Corporation OTEC Plant*. Makai Ocean Engineering, Inc., Honolulu, Hawaii.]

problems that were identified as major barriers to OTEC commercial operation in early reviews of OTEC. These include heat exchanger efficiency, heat exchanger biofouling, aluminum heat exchanger durability, OTEC heat engine performance, cold water pipe fabrication, cold water pipe deployment, OTEC plant and plantship ocean durability and survival in tropical 100-year-storm conditions, and OTEC environmental impacts.

## XII. OTEC CLOSED-CYCLE SYSTEMS COST EVALUATION

### A. Cost Uncertainties and Perceived Risk versus Engineering Status

Reliable estimates of the costs of complete systems can be made by dividing the total cost into categories for which the cost uncertainty of subsystems development is low, moderate, or high. A weighted average cost uncertainty can then be calculated for the total cost of the complete system. For OTEC cost estimation, a low uncertainty of $-/+10$–20% was assigned to cost estimates for system components or construction technologies that are in standard industrial use and for which firm quotes were obtained. The fraction of the total cost in this category ranges from 45% of the total for the ammonia plantship to 70% for the methanol plantship. Moderate uncertainty of $-/+20$–35% was assigned to components or construction technologies that are commercially available at a smaller scale or under different working conditions but must be modified for OTEC use. The corresponding fractions of the total cost in this category were 41 and 23% for the two cases. High-cost uncertainty of $-35$ to $+100\%$ was

**TABLE V   OTC Design of a Shore-Mounted 40-MWe OTEC Power Plant**

| | Evaporator | Condenser |
|---|---|---|
| LBCS dimensions m (ft) | | |
| Above water: $L$ 93 (306), $W$ 71 (234), $H$ 11.3 (37) | | |
| Under water: $L$ 116 (380), $W$ 71 (234), $H$ 22.2 (73) | | |
| Power plant | | |
| Closed Rankine cycle; ammonia working fluid; shell-and-tube HXs with titanium tubes 0.01905 m OD, 0.01765 m ID; 4 10-MWe (net) modules | | |
| Nominal operating characteristics | | |
| Nominal $\Delta T$ (K) (mixed seawater and HECO flow)  26.83 | | |
| Power (MWe) | | |
| Gross turbine-generator power  53.8 | | |
| Warm water pump power  4.1 | | |
| Cold water pump power  8.4 | | |
| Ammonia pumps power  0.9 | | |
| Other  0.4 | | |
| Net power output to busbar  40.0 | | |
| Heat load (MWt) | 1,765 | 1,711 |
| Water flow (kg/s) | 13,747 | 11,510 |
| Ammonia flow (kg/s) | 1,434 | 1,434 |
| Number of HX tubes | 40,128 | 37,671 |
| Tube length (m) | 13.65 | 15.03 |
| HX area (m$^2$) | 321,024 | 301,368 |
| Water side | 243,197 | 251,239 |
| Ammonia side | 262,442 | 271,121 |
| Overall heat trans. coeff. (W/m$^2$ K) | 2,516 | 2,601 |
| Fouling coeff. (m$^2$ K/W) $\times 10^5$ | 1.76 | 1.76 |
| Water inlet (outlet) temp. (°C) | 26.8 (22.8) | 5.0 (9.7) |
| Ammonia inlet (outlet) temp. (°C) | (21.44) | 10.66 |
| Log mean temp. diff. (LMTD) ($+°C$) | 2.95 | 2.69 |
| Seawater $\Delta p$ (m) | 2.39 | 2.25 |
| Sum of inlet & outlet $\Delta p$ | 0.60 | 2.91 |
| Heat exchanger $\Delta p$ | 2.25 | 2.39 |
| Seawater velocity (m/s) | 1.37 | 1.22 |
| CWP flow (m$^3$/[s MWe (net)]) | | 2.25 |
| Net power/total HX area (kWe/m$^2$) | 0.081 | |

assigned to fabrication and deployment of the cold water pipe. The uncertainties in this category contribute 13 and 7% to the total estimate. The weighted average risks of cost underestimates or overruns in constructing and deploying OTEC systems were thus judged to be between 20 and 30%. Scaling factors and *experience factors* [Industrial cost surveys show that unit costs of items in commercial production tend to decrease by a constant factor with each doubling of the number produced, i.e., the second unit costs "x" times the cost of the first, the fourth "x" times the cost of the second, and so on.] The factor "x," called the "experience factor," is used in deriving the final estimates shown in Tables VI and VII.

**TABLE VI   Cost Estimates for Floating 40-MWe OTEC Subsystems, Components, Facilities, and Procedures ($M 1990)[a]**

| Subsystem | 46 MWe (net) | | | | 40 MWe (net) | | | |
| | Grazing plantship | | Est. cost uncertainty | | Moored plant | | Est. cost uncertainty | |
| | Mid-1980$ | Mid-1990$ | −% | +% | Mid-1980$ | Mid-1990$ | −% | +% |
|---|---|---|---|---|---|---|---|---|
| Platform | 47.3 | 64.8 | 13 | 13 | 66.6 | 91.2 | 14 | 28 |
|   Hull | 20.4 | 27.9 | 15 | 20 | 20.4 | 27.9 | 15 | 20 |
|   Water pumps | 9.7 | 13.3 | 15 | 15 | 9.7 | 13.3 | 15 | 15 |
|   Position control | 7.6 | 10.4 | 10 | 10 | 24.2 | 33.1 | 15 | 50 |
|   Other | 9.6 | 13.1 | 10 | 10 | 12.3 | 16.8 | 10 | 10 |
| Power systems | 42.0 | 57.5 | 24 | 17 | 42.0 | 57.5 | 24 | 18 |
|   Heat exchangers | 28.3 | 38.7 | 30 | 20 | 28.3 | 38.7 | 30 | 20 |
|   $NH_3$ systems | 3.5 | 4.8 | 15 | 15 | 3.5 | 4.8 | 15 | 15 |
|   Power generation | 7.6 | 10.4 | 15 | 15 | 7.6 | 10.4 | 15 | 15 |
|   Other | 2.6 | 3.6 | 10 | 10 | 2.6 | 3.6 | 10 | 10 |
| Water ducts | 9.4 | 12.9 | 25 | 99 | 13.0 | 17.8 | 22 | 76 |
|   WW & disch. pipes | 0 | 0.0 | 15 | 15 | 3.5 | 4.8 | 15 | 15 |
|   CWP | 6.6 | 9.0 | 25 | 100 | 6.7 | 9.2 | 25 | 100 |
|   CWP/hull joint | 2.7 | 3.7 | 25 | 100 | 2.7 | 3.7 | 25 | 100 |
|   Other | 0.1 | 0.1 | 15 | 15 | 0.1 | 0.1 | 15 | 15 |
| Energy transfer | | | | | | | | |
|   $NH_3$ production | 27.7 | 37.9 | 23 | 25 | | | | |
|   Liquid $N_2$ plant | 3.2 | 4.4 | 15 | 15 | | | | |
|   Electrolysis | 12.0 | 16.4 | 30 | 35 | | | | |
|   Plant $NH_3$ synthesis | 8.7 | 11.9 | 20 | 20 | | | | |
|   Other | 3.8 | 5.2 | 15 | 15 | | | | |
|   Elec. power to shore | | | | | 23.1 | 31.6 | 25 | 76 |
|   Control and bus | | | | | 2.2 | 3.0 | 15 | 15 |
|   Conditioning | | | | | 1.4 | 1.9 | 15 | 15 |
|   Riser cable | | | | | 9.0 | 12.3 | 30 | 100 |
|   Bottom cable | | | | | 4.1 | 5.6 | 30 | 100 |
|   Buoys | | | | | 2.4 | 3.3 | 25 | 25 |
|   Misc. hardware | | | | | 2.4 | 3.3 | 15 | 15 |
|   Potheads, gimbals, disconnects | | | | | 1.6 | 2.2 | 15 | 15 |
| Deployment | 11.6 | 15.9 | 31 | 84 | 20.7 | 28.3 | 34 | 90 |
|   Platform | 1.3 | 1.8 | 15 | 15 | 0.9 | 1.2 | 15 | 15 |
|   Power system | 1.1 | 1.6 | 20 | 40 | 0.7 | 1.0 | 20 | 40 |
|   CWP | 9.2 | 12.5 | 35 | 100 | 9.4 | 12.9 | 35 | 100 |
|   Discharge pipe | | | | | 1.7 | 2.3 | 35 | 50 |
|   Electric cable | | | | | 8.0 | 11.0 | 35 | 100 |
|   Acceptance, ind. fac., E& A | 6.4 | 8.8 | 20 | 20 | 8.4 | 11.5 | 20 | 20 |
|   Weighted cost uncertainty (%) | | | 22 | 30 | | | 26 | 54 |
| Direct cost ($ M) | 144.4 | 197.7 | | | 170.2 | 233.0 | | |
| Interest dur. constr. (15% of direct cost) | 22.1 | 30.2 | | | 26.0 | 35.6 | | |
| Tot. investment ($M) | | | | | | | | |
|   Nominal | | 228 | | | | 269 | | |
|   Minimum | | 182 | | | | 199 | | |
|   Maximum | | 296 | | | | 414 | | |

[a] 1990$ = 1980$ × 1.369 = 1983$ × 1.128. [From Avery, W. H., and Wu, C. (1994). *Renewable Energy from the Ocean: A Guide to OTEC*. Oxford University Press, New York, p. 361.]

**TABLE VII   Cost Estimates for Grazing OTEC Methanol and Ammonia Plantships ($M 1990)[a]**

| Subsystem | First 200-MWE methanol plantship | Subsystem cost uncertainty | | First 368-MWe ammonia plantship | Subsystem cost uncertainty | |
|---|---|---|---|---|---|---|
| | | $-\%$ | $+\%$ | | $-\%$ | $+\%$ |
| Platform | 190.7 | 14.0 | 16.9 | 247 | 13.8 | 16.3 |
| Hull | 109.7 | 15 | 20 | 124 | 15 | 20 |
| Water pumps | 42.7 | 15 | 15 | 65 | 15 | 15 |
| Position control | 20.7 | 10 | 10 | 31 | 10 | 10 |
| Other | 17.6 | 10 | 10 | 27 | 10 | 10 |
| Power systems | 180.4 | 23.4 | 17.6 | 273 | 24.5 | 17.6 |
| Heat exchangers | 124.3 | 30 | 20 | 188 | 30 | 20 |
| $NH_3$ system | 12.2 | 15 | 15 | 18 | 15 | 15 |
| Power generation | 30.5 | 10 | 10 | 46 | 10 | 10 |
| Other | 13.4 | 15 | 15 | 20 | 15 | 15 |
| Water ducts | 30.8 | 24.9 | 99.0 | 47 | 24.9 | 99.0 |
| WW & disch. pipes | 0.0 | 0 | | | | |
| CWP | 21.9 | 25 | 100 | 33 | 25 | 100 |
| CWP/hull joint | 8.5 | 25 | 100 | 13 | 25 | 100 |
| Other | 0.4 | 15 | 15 | 1 | 15 | 15 |
| Energy transfer | | | | | | |
| $NH_3$ production | | | | 175 | 18.7 | 18.7 |
| Liquid $N_2$ plant | | | | 23 | 15 | 15 |
| Electrolysis plant | | | | 86 | 15 | 15 |
| $NH_3$ synthesis | | | | 62 | 20 | 20 |
| Other | | | | 27 | 15 | 15 |
| $CH_3$ OH production | 311.2 | 16.5 | 16.5 | | | |
| Coal prep. molten $Na_2CO_3$ gasifier | 47.0 | 25 | 25.0 | | | |
| Gas clean-up, CO shift, sulfur recov. | 104.6 | 15 | 15.0 | | | |
| $CH_3OH$ synthesis | 68.0 | 15 | 15.0 | | | |
| Steam syst. | 25.3 | 15 | 15.0 | | | |
| $H_2O$, waste treatment $Na_2CO_3$ regen. | 27.1 | 15 | 15.0 | | | |
| Serv. & maint. | 39.2 | 15 | 15.0 | | | |
| Electrolysis | 58.4 | 30 | 35.0 | | | |
| Deployment | 39.0 | 30.6 | 81.4 | 59 | 30.6 | 81.4 |
| Platform | 4.9 | 15 | 15 | 7 | 15 | 15 |
| Power system | 3.7 | 15 | 15 | 6 | 15 | 15 |
| CWP | 30.5 | 35 | 100 | 46 | 35 | 100 |
| Discharge pipe | 0.0 | | | | | |
| Electric cable | 0.0 | | | | | |
| Acceptance, ind. fac., E& A | 45.1 | 20 | 20 | 68 | 20 | 20 |
| Weighted cost uncertainty (%) | | 13 | 24 | | 20 | 26 |
| Direct cost ($M) | 855.6 | | | 869 | | |
| Interest dur. constr. (15% of direct cost) | 131 | | | 133 | | |
| Tot. investment ($M) | | | | | | |
| Nominal | 986 | | | 1002 | | |
| Minimum | 789 | | | 802 | | |
| Maximum | 1292 | | | 1313 | | |

[a] 1990$ $= 1980\$ \times 1.369 = 1983\$ \times 1.128$. [From Avery, W. H., and Wu, C. (1994). *Renewable Energy from the Ocean: A Guide to OTEC*. Oxford University Press, New York, p. 361.]

All complete OTEC installations will incorporate a platform system or land-based containment system to support and house the OTEC subsystems, including water pumping and distribution, platform station-keeping, and auxiliary platform services; a power plant(s) including heat exchangers, turbines, electric generators, water and working-fluid pumps, and control equipment; water ducting equipment; an energy transfer system which may include a process plant and related equipment for fuel production, storage and transfer; equipment for routing electric power to consumers, mariculture installations, fresh water production, and use of the cold water effluent for air conditioning and/or cold water agriculture; support facilities and equipment including shops, personnel accommodations, etc.

## B. Estimated Costs of OTEC Systems

In the OTEC R & D programs sponsored by DOE cost estimates were made for many alternative designs of subsystems and components. In the United States, the most detailed OTEC development cost studies were conducted in the JHU/APL and OTC preliminary design programs, which drew heavily on industrial experience and costs, and work by other DOE contractors. These studies have been used as the primary basis for the estimates listed in the tables presented in later sections of this chapter. Cost studies in less detail have also been reported in other American studies, and by Japanese, French, and Taiwan investigators. Some of these are also listed. Because of unknown differences in estimating procedures the results may not be strictly comparable but are deemed useful to include.

### 1. Baseline 40-MWe (nom) OTEC Grazing Plant-ship and Moored Plan

A cost assessment of floating OTEC systems at the **Preliminary Design** level has been completed for the baseline 40-MWe (nom) grazing ammonia plantship and for moored electric power plant designs. Optimization studies show that the grazing plantship stationed near the equator where the water average annual temperature difference ($\Delta T$) is 23.3°C will deliver 46–MWe (net) at the onboard busbar. The moored plant at Punta Tuna, Puerto Rico, with the same power system but with inlet water $\Delta T$ of 22.3°C, will deliver 40 MWe (net). These values of the generated net power are used in Table VI. The estimated cost of net power at the onboard bus-bar is $3200/kW for the plantship and $3700/kW for the Puerto Rico plant.

### 2. SOLARAMCO 40-MWe (nom) OTEC Demonstration Plantship

Solar Ammonia Company (SOLARAMCO) prepared a **Conceptual Design** of a 40-MWe (nom) ammonia

plantship, to be sited on the ocean south of Hawaii, that would produce 125 mt/d of liquid ammonia. The estimated cost of this system was $250 M (1990$). ($6250/kW, including the ammonia plant)

### 3. Commercial Sized OTEC Grazing Plantships

**Conceptual Design** studies have been completed for a 200-MWe methanol plantship. The data have been used also for an estimate of the cost of a 320-MWe (nom), (368-MWe des. est.) ammonia plantship. The data provide a basis for estimating the costs and cost uncertainties of OTEC systems that would enter commercial production. Cost information available from conceptual design studies of other OTEC systems and preliminary designs of subsystems, as well as prices of standard items and components, have been used in these estimates. The cost data are based on industrial estimates made in the period 1978–1982. Termination of OTEC development funding prevented the acquisition of comparable estimates of current costs. Therefore, the quoted estimates of total system costs in 1990$ represent an updating of earlier data to account for inflation in the costs of industrial equipment and construction. The inflation factors are taken from the periodic surveys of construction costs published by Chemical Engineering magazine. (McGraw-Hill) Estimated costs for the **Conceptual Design** of commercial-size OTEC methanol and ammonia plantships are listed in Table VII.

### 4. Sea Solar Power, Inc. (SSP) 100-MWe Floating OTEC Power Plant

The cost of the 100-MWe plant was estimated as $250 M (1985$) Anderson, J. H. (1985). "Ocean thermal power-the coming energy revolution," *Solar and Wind Power Technology* **2,** 25.

### 5. Japanese Floating OTEC Plants

**Conceptual Designs** of moored OTEC plants sited in the Sea of Japan were discussed in Section VI. The costs estimated by the Japanese engineers are listed in Table VIII.

**TABLE VIII  Estimated Costs of 100-MWe (net) Japanese Moored OTEC Power Plants (1980$)**

|  | At Osumi | At Toyama |
|---|---|---|
| Type of platform | Submerged disc | Surface ship |
| Net power output (MWe) | 100 | 100 |
| Unit construction cost (yen/kW) | $78 \times 10^5$ | $59 \times 10^5$ |
|   ($/kW) | 3550 | 2690 |
| Unit power cost at busbar (yen/kW h) | 12.21 | 12.36 |
|   ($/kW h) | 0.0555 | 0.0562 |
| Power transmission cost (yen/kW h) | 0.66 (AC) | 0.66 (AC) |
|   ($/kW h) | 0.003 | 0.003 |

## C. Estimated Costs of Shore-Based OTEC Systems

### 1. Ocean Thermal Corporation (OTC) Shore-Based Power Plant

During 1981–1983 a **Preliminary Design** of a 40-MWe (nom) shelf-mounted OTEC power plant to be sited at Kahe Point on the island of Oahu, HI, was completed by Ocean Thermal Corporation (OTC) under DOE auspices. The program is described in Section VII.F. Details of the system design costs are proprietary but the estimated costs of the major subsystems are available and are presented in Table IX. Performance analysis by OTC in 1983 showed that the nominal 40-MWe plant would produce a design power output of 50 MWe (net), with state-of-the-art components. The cost uncertainties listed in the table are those of the authors of this article, based on a review of the OTC preliminary design report and comparison with other relevant information. The much larger system costs, compared with those of the floating baseline moored power plant, are caused by the use of titanium heat exchangers instead of aluminum in the OTC design, and by the expensive construction and deployment of the 3-km-long, bottom mounted, CWP. Estimated costs of the OTC OTEC plant are shown in Table IX.

### 2. General Electric **Conceptual Design** of a Tower-Mounted 40-MWe OTEC Power Plant for Hawaii

Costs estimated in the design study, discussed in Section XI , are listed in Table X.

### 3. Puerto Rico Electric Power Co. (PREPA) **Conceptual Design** of a 40-MWe (nom) Tower-Mounted OTEC Plant

The costs quoted for the PREPA installation are presented in Table XI.

**TABLE IX   Estimated Costs of OTC 40-MWe (nom) Land-Based OTEC Power Plant [50-MWe (net) Final Design] ($M 1985)**

| | | | |
|---|---|---|---|
| Land-based containment system | 60.3 | Energy transfer system | 5.0 |
| Structure | 38.1 | Power control & bus | 1.4 |
| Pumps and motors | 12.1 | Power conditioning | 0.5 |
| Facility support | 7.7 | Cable system | 1.0 |
| Outfit and furnishings | 2.4 | Shore-grid station | 1.8 |
| Cold water pipe system | 73.1 | Biofouling corrosion control | 0.3 |
| Pipe | 67.7 | Deployment services | 54.2 |
| Intake | 0.1 | LBCS deployment | 6.8 |
| CWP/LBCS transition | 0.7 | CWP deployment | 27.4 |
| CWP embedment/anchoring | 1.7 | WWP deployment | 6.8 |
| Biofouling corrosion control | 2.9 | Effluent pipe deployment | 11.2 |
| Warm water pipe system | 3.7 | Power system deployment | 2.0 |
| Pipe | 3.4 | Energy transfer syst. deployment | 0.1 |
| WWP/HECO basin transition | 0.2 | Industrial facilities | 12.0 |
| WWP/LBCS transition | 0.2 | Construction facilities | 11.5 |
| Mixed effluent pipe system | 5.0 | Logistic support facilities | 0.5 |
| Pipe | 2.3 | Acquisition, acceptance, testing | 2.0 |
| Effluent pipe/LBCS transition | 1.1 | Engineering & detail design | 5.2 |
| Outfall & discharge channel | 0.5 | Total direct cost | 327.8 |
| Embedment/anchoring | 1.1 | Construction interest | 50.1 |
| Power system | 107.2 | Plant investment (M 1984$) | 378 |
| Evaporators | 39.7 | Plant investment (M1990$) | 419 |
| Condensors | 37.1 | | |
| Turbine/generators | 15.4 | | |
| Working fluid loop | 4.8 | | |
| Nitrogen purge | 0.3 | | |
| Instrument & control | 5.4 | | |
| Biofouling corrosion control | 3.0 | | |
| Auxiliary systems | 1.6 | | |

**TABLE X   Estimated Costs of GE 40-MWe (nom) Tower-Mounted OTEC Plant**

| System cost estimates | $M (1982$) |
|---|---|
| GE tower-mounted design | |
| Tower system | 90 |
| Fabrication | 63 |
| Transportation | 3 |
| Installation | 17 |
| Engineering | 4 |
| Insurance | 4 |
| Cold water pipe system | 64 |
| Pipe material | 49 |
| Templates | 1 |
| Deployment | 12 |
| Weather allowance | 2 |
| Power plant | 166 |
| Heat exchangers | 34 |
| Ammonia charge | 1 |
| Water pumps | 6 |
| Turbine-generators | 20 |
| Deploy, install, misc. | 105 |
| Engineering, misc. | 8 |
| Total system cost | 320 |
| $/kWe (net) | 8000 |

**TABLE XI   Estimated Costs of PREPA 40-MWe (net) Pilot Plant ($M 1982)**

| | |
|---|---|
| Tower system | 21 |
| Fabrication | 13.9 |
| Transportation | 0.6 |
| Installation | 6.5 |
| Power plant system | 117.4 |
| Heat exchangers | 62.8 |
| Seawater systems | 12.9 |
| Turbine-generator | 10.0 |
| Biocontrol | 7.4 |
| Ammonia systems | 5.0 |
| Startup/standby power | 7.0 |
| Control and elect. | 10.6 |
| Miscellaneous | 1.5 |
| Water ducting | 22.2 |
| CWP incl. deployment | 18.6 |
| Warm and mixed systems | 3.6 |
| Energy transfer system | 66.0 |
| Acceptance testing | 2.0 |
| Deployment services | 20.0 |
| Industrial facilities | 2.0 |
| Eng. and detail design | 2.0 |
| Total | 259.4 |
| $/kW | 6485 |

### 4. French (CNEXCO) (ONERA) Designs of 5-MWe OTEC Land-Based Plant for Tahiti

Conceptual designs of alternative 5-MWe closed-cycle and open-cycle plants, including equipment for the production of fresh water, were completed. The cost comparison for the two options, is shown in Table XII. (IFREMER 1985. *Ocean thermal energy conversion, the French version, a pilot electric power plant for French Indonesia* March.

### 5. Taiwan 100-MWe (gross) Land-Based OTEC Power Plant

Estimated costs are listed in Table XIII. Liao, T., Giannoti, J. G., van Mater, P. R., and Lindman, R. A. (1986). "Feasibility and design studies for OTEC plants along the coast of Taiwan, Republic of China," *Proc. International Offshore Mechanics and Arctic Engineering. Symp.* **2,** 618.

### 6. Design of OTEC Land-Based OTE Plant for Jamaica

Plans for a 1-MWe OTEC pilot plant to be sited near Kingston, Jamaica, yielded a cost estimate of $6000/kW to $9000/kW, for a 40-MWe demonstration plant. The $\Delta T$ is 21.3°C at the site. Details of the cost estimate are not

available. [For a $\Delta T$ equal to that in Puerto Rico (22.3°C), the estimated cost range would be $5300–8000/kW.]

## XIII. CAPITAL INVESTMENTS AND SALES PRICES FOR OTEC PRODUCTS

The cost of supplying energy from a facility is determined by the operating costs plus the cost to amortize the capital investment. Thus, a knowledge of the plant investment and operating costs allows an estimate to be made of the sales price of power or products that will return an attractive profit to a private investor, or an allowable

**TABLE XII   Estimated Costs of French Land-Based 5-MWe (net) OTEC Electric Power and Fresh-Water Plants for Tahiti (FF 1985)**

| | Closed cycle | Open cycle |
|---|---|---|
| General | 23 | 23 |
| Power system | 144 | 167 |
| Land structures (incl. land part of CWP and civilian work) | 221 | 221 |
| CWP | 135 | 136 |
| Total | 523 | 547 |
| FF/kW | 105,200 | 106,900 |

**TABLE XIII  Power and Costs of Taiwan 50-MWe (nom) Land-Based OTEC Plants (1980$)**

| Site | Cold-water pump | Warm-water pump | Module pump power | Total pump power | Gross turb/gen power | Optimized plant Net power output (nom) | Optimized plant Power output (ann.avg) | (Investment M$) |
|---|---|---|---|---|---|---|---|---|
| Ho-Ping (long CWP) | 1.54 | 3.07 | 4.62 | 18.46 | 68.46 | 50.0 | 53 | 301.5 ($5690/kW) |
| Ho-ping (short CWP) | 1.50 | 3.07 | 4.57 | 18.27 | 68.27 | 50.0 | | |
| Chang-Yuan (long CWP) | 1.71 | 3.13 | 4.84 | 19.37 | 69.37 | 50.0 | 74 | 364.5 ($4920/kW) |
| Chang-Yuan (short CWP) | 1.60 | 3.13 | 4.73 | 18.93 | 68.93 | 50.0 | | |

return to a publicly owned enterprise. OTEC systems can be financially competitive if funding can be supplied to bring OTEC to a commercial status, and if sales prices are determined by rules that establish a "level playing field." This requires that prices quoted for products from competitive systems be based on replacement costs, and include the costs of environmental impacts, waste disposal costs, and expenses of maintaining the security of nuclear power sources and imported oil supplies. With these requirements, OTEC product prices can be compared in a valid way with prices of products from existing commercial and proposed energy delivery systems.

Project financing in general requires two types of investors: (a) the project sponsors who build and operate the system, sell the product, and receive the profits; and (b) lenders who provide the funds for construction and implementation, and receive interest on the money lent. The first types of investors are "venture capitalists," the second, bankers and other financial institutions. The total investment consists of equity and debt.

## A. Financial Analysis

A financial analysis procedure specifically devised by Mossman (1989) for comparison of OTEC price estimates with other energy sources allows a valid comparison to be made of the relative commercial attractiveness of energy production methods. It establishes uniform guidelines to estimate the sales price that would return an acceptable profit for investors in alternative energy production systems. The Mossman procedure eliminates distortions and misunderstandings caused by failure to include the effects of government subsidies or hidden costs that differ widely among the various options.

The Mossman financial analysis of alternative fuel and power production options assumes 2 plus years of financed construction, followed by an operating life of 25 years, and assumes that the project will be undertaken by a firm or consortium that will use all possible tax shields provided by depreciation and interest payments. The analysis calculates a quantity termed "The Adjusted Present Value"

(APV) that measures the profitability of a potential venture. If the APV has a positive value, the venture will return a profit sufficiently high to attract entrepreneurs to invest in it rather than to select alternative profit-making opportunities. If the APV has a negative value the project will not return a satisfactory profit or will be too risky to be considered. The APV methodology calculates the price at which at which the product must be sold if it is to achieve a positive APV, by following procedures described by Mossman Full details are presented in Avery and Wu (1994). With the above guidelines, a financial analysis computer program was prepared that enables one to calculate the profitability of alternative energy production ventures, as they depend on the values of the capital investments, plant life, interest rates for construction and amortization, operating costs and other parameters. Product sales prices are varied in the program to find the price that will give a positive number for the APV. The financial analysis program has been used to estimate sales prices of methanol and ammonia produced on OTEC plantships and delivered to U.S. ports, and of electric power generated on moored OTEC plants and transmitted via underwater cable to land-based utilities. The same financial analysis is used for estimating sales prices of fuels and electric power from alternative energy sources. A valid comparison can then be made of the profitability of OTEC products versus the alternative sources. The results are discussed briefly in the next section and are displayed in the accompanying tables and graphs.

## XIV. OTEC ECONOMICS

The economic contribution that OTEC will make to the solution of the nation's energy problems depends on whether its virtues relative to existing and alternative sources of energy and fuels are recognized, and whether it can elicit funding for commercialization. By using the Mossman financial analysis method to estimate the prices that must be charged for profitable operation, a decision to fund OTEC commercialization can be based on an unbiased comparison of prices of OTEC fuel or electricity, with prices

for profitable construction and operation of the proposed alternatives.

## A. Commercial Viability of Alternative Energy Options

The information needed to compare the profitability of alternative energy production systems in the Mossman analysis includes:

1. Capital investment. Costs of related facilities as well as the plant investment
2. Annual payments to retire plant investment.
3. Fuel costs for electric plants and refineries. These costs vary with time in an unpredictable manner. Past forecasts have been in error by large factors.
4. Operating and Maintenance Costs (O&M). For the options considered, these costs vary annually from a few percent to about 10% of plant investment.
5. Federal, state, and local taxes, licensing fees, insurance, and incidental costs. These costs are a few percent of plant investment. State and local property taxes will be zero for OTEC plantships.
6. Demand for the energy option, and potential revenues. Many options may be profitable for special applications but the resource that could be developed is too small to have a major role in satisfying national needs.
7. Environmental charges. Environmental impacts of energy production are typically ignored or charged to the public through state or local taxes. These costs should be paid by the energy producer. The environmental costs can be large but the financial implications are difficult to quantify, particularly if the effects are cumulative. Compliance by existing nuclear power installations with environment and safety regulations are estimated to require expenditures in excess of $200B, and are federally funded. Environmental costs will be large also for oil shale and tar conversion systems and for coal conversion plants. Solar-based options are generally nonpolluting and are expected to involve minimal environmental costs. Because OTEC plantship operation will enhance fish production, the overall environmental effects are expected to be beneficial.
8. Vested interests of public, industry, and government organizations. Commercialization of new energy sources is dependent on the support that can be developed from the existing energy producing and distributing organizations and their powerful political support groups. The absence of a strong political base for solar energy options is a major impediment to their commercialization.

## XV. ECONOMIC, ENVIRONMENTAL, AND OTHER ASPECTS OF OTEC COMMERCIALIZATION

As noted earlier, the plant investments and market prices of energy sources involve subsidies, tax breaks, and other factors, including R & D expenditures and environmental impacts, that are paid by the government or the general public. These factors disguise the true costs. An objective evaluation will determine the desirability of new energy options based on the price of the fuel or power that can be offered profitably to consumers. The price comparison must also include the replacement costs of conventional power sources. The Mossman financial analysis allows comparison of all of the alternatives on a uniform basis, and is used in evaluating the options discussed in the following and expressed quantitatively in Tables XIV and XV.

## A. Production of Liquid Fuels from Existing and Proposed Fuel Sources

The material in this section is based mainly on price quotations from industrial suppliers of components and equipment that were obtained in the DOE-funded programs that ended in 1985. The costs were updated to 1990 in preparation of Section XV.A.1. Information on fuel and electricity prices was available from reports issued by the Energy Information Administration of the U.S. Department of Energy (EIA/DOE). Other sources are listed in the references.

### 1. Petroleum-Based Fuels

Petroleum must be refined to make products suitable for use in transportation, in industrial or residential boilers, and in miscellaneous applications. The refining processes add to the cost. In November 1999 a barrel of crude oil costing $25 ($0.595/gal) was associated with a wholesale price of $0.745/gal ($31.3/bbl) for gasoline and $0.686/gal ($28.8/bbl) for diesel fuel. Thus, the wholesale gasoline price exceeded the quoted oil price ($/gal) by the factor $74.5/59.5 = 1.25$.

### 2. Synthetic Fuels

*a. Fuels from oil shale.* Two programs initiated with support by the now defunct Synfuels Corporation received major funding in the early 1980s to demonstrate technologies for extracting motor fuel from oil shale: above-ground retorting, and underground gasification. The Occidental Oil Shale Company proposed a demonstration program, involving a plant requiring an investment of $225M, that would produce 1200 bbl/day of motor fuel. The process

**TABLE XIV   Alternative Motor Fuels Price Comparison ($ 1990)[a] (ITC = 10%; APV = 0)**

| Plant type | PI ($M) | Product | Output (gal/y) | Annual cost ($M/y) | Sales price (APV = 0) ($/gal) | Gasoline mpg equivalent ($/gal) | Number of modules for $10^5$ bbl/d gasoline equivalent | Average plant investment ($B) | Sales price (APV = 0) ($/gal) | Gasoline equivalent price ($/gal) | Investment for $10^5$ bbl/d gasoline equivalent ($B) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| OTEC $CH_3OH$ (coal $50/t) | 960 | $CH_3OH$ | $1.99 \times 10^8$ | 51.1 | 0.70 | 1.27 | 13 | 0.72 | 0.620 | 1.12 | 9.3 |
| OTEC $NH_3$ | 975 | $NH_3$ | $1.70 \times 10^8$ | 25.6 | 0.74 | 1.98 | 24 | 0.70 | 0.540 | 1.44 | 16.9 |
| Shale oil | 225 | HC fuel | $1.47 \times 10^7$ | 7.5 | 1.51 | 1.51 | 83 | 0.14 | 0.662 | 0.66 | 12.0 |
| Shell-FT coal liquefaction (coal $22.7/t) | 4400 | HC fuel | $1.28 \times 10^9$ | 450 | 0.76 | 0.76 | 1 | 4.32 | 0.762 | 0.76 | 5.2 |
| NG to $CH_3OH$ (NG $3/mBtu) | 237 | $CH_3OH$ | $2.44 \times 10^8$ | 117 | 0.58 | 1.04 | 6.3 | 0.19 | 0.461 | 0.83 | 1.7 |
| NG to $NH_3$ (NG $3/mBtu) | 100 | $NH_3$ | $1.53 \times 10^8$ | 44 | 0.37 | 1.02 | 26 | 0.074 | 0.337 | 0.93 | 1.9 |

[a] Does not include carbon tax or automobile modification costs.

would also generate 35 MW of electric power that would add to the profitability.

*b. Liquid fuels from tar sands.*   This approach seems to be worth developing but cost information on U.S. tar sands fuel production is not available.

*c. Liquid hydrocarbons from coal.*   Approaches that have been supported include (i) coal pyrolysis, which produces a liquid fuel (+ gas) and coke; (ii) coal gasification and liquefaction, involving reaction of coal with steam followed by catalytic combination of the products to form liquid hydrocarbon fuels; (iii) hydrogen addition to coal, heavy oils or tars to form a liquid fuel. Research and development by industry with government participation showed that the product cost for the hydrogen addition processes and the coal gasification processes were nearly the same. The pyrolysis approach is not considered to be competitive with the other processes if used alone but could be promising in combination with electric power generation using the residual coke.

*d. Liquid fuels formed by natural gas reforming.* Methanol and ammonia are both produced at low cost in large quantities ($\sim$20 million tons/year in the United States) by natural gas reforming, followed by

**TABLE XV   Alternative Electricity Sources Price Comparison (Does Not Include Carbon tax) (ITC = 10%; APV = 0)**

| Plant type | Fuel | Fuel cost ($/GJ) | Plant investment ($M) | Output (kWh/y) | Annual cost ($M/y) | Break-even sales price ($/kWh) | Number of modules for $10^{10}$ kWh/y | Sales price ($/kWh) | Investment for $10^{10}$ kWh/y ($B) |
|---|---|---|---|---|---|---|---|---|---|
| Coal steam with FGD | Coal | 2.2 | 141 | $6.1 \times 10^8$ | 20.1 | 0.055 | 16 | 0.047 | 1.7 |
| Nuclear | Uranium | 7.0 | 2350 | $5.3 \times 10^9$ | 84.0 | 0.069 | 2 | 0.063 | 4.5 |
| Photovoltaics | Solar actinic energy | 0.0 | 125 | $5.9 \times 10^7$ | 1.25 | 0.229 | 169 | 0.137 | 12.4 |
| Solar–thermal–electric | | | | | | | | | |
|   Central receiver | Solar heat | 0.0 | 70 | $3.1 \times 10^7$ | 2.30 | 0.288 | 326 | 0.164 | 12.5 |
|   Focused dish | Solar heat | 0.0 | 200 | $8.8 \times 10^7$ | 12.0 | 0.258 | 114 | 0.160 | 13.9 |
|   Cylindrical mirror | Solar heat | 0.0 | 225 | $2.5 \times 10^8$ | 6.8 | 0.117 | 46 | 0.071 | 6.3 |
| Wind | Wind | 0.0 | 18 | $1.6 \times 10^7$ | 0.16 | 0.118 | 633 | 0.064 | 5.8 |
| Methanol fuel cell | $CH_3OH$ | 10.5 | 50 | $3.9 \times 10^8$ | 28.5 | 0.098 | 25 | 0.073 | 0.89[a] |
| | | 6.7 | 50 | $3.9 \times 10^8$ | 16.4 | 0.056 | 25 | 0.051 | 0.89[a] |
| Ammonia fuel cell | $NH_3$ | 12.7 | 50 | $3.9 \times 10^8$ | 29.9 | 0.090 | 25 | 0.085 | 0.89[a] |
| | | 8.4 | 50 | $3.9 \times 10^8$ | 20.1 | 0.065 | 25 | 0.060 | 0.89[a] |

[a] Includes only the cost of the fuel-cell fuel plant.

catalysis. Ammonia plant costs are similar to those for methanol. Coal gasification to produce fuel-grade methanol could be competitive with natural gas reforming if large use of methanol or ammonia in transportation caused the price of natural gas to increase.

*e. Synthesis of methanol or ammonia by natural gas reforming followed by purification and catalytic steps to produce a liquid fuel.*  Coal gasification can also provide the feed stock,

*f. Production of hydrogen by water electrolysis followed by further reactions to produce a liquid fuel.* In addition to ammonia and methanol, hydrogen can be combined with heavy oils or tar to produce a liquid fuel.

## B. Production of Electricity from Existing and Proposed Power Plants

Electric power may be supplied to consumers from a variety of existing and proposed sources.

1. Coal-powered steam electric plants with flue gas desulfurization (FGD). The estimated plant investment for a 1000-MWe electric power plant based on coal-powered steam generation with flue-gas desulfurization (FGD is quoted as $1.41 B. O & M costs are estimated as $166 M/year.
2. Nuclear steam–electric power plants. Costs of 1000-MWe nuclear power plants, estimated by Starr (1987), are plant investment (PI) $1700–$3000 M, fuel cost/year $32 M, O & M cost/year $42 M, delivery cost/year $37 M.
3. Hydroelectric power. Hydroelectric power is the cheapest source of electricity and provides approximately 10% of U.S. electric power. No significant increase in output is foreseen.
4. Natural-gas-fueled steam electric power. Natural gas can substitute for coal in conventional coal-fired electric plants. The capital cost is estimated at $780/KW (design capacity) and O&M at 2.6% of plant investment.
5. Oil-fueled steam electric plants. Fuel oil is a convenient source for steam-powered electric plants but has been more expensive than coal and therefore used only in cases where the convenience outweighed the added cost.
6. Photovoltaic (PV) electric power generation. Photovoltaic cells convert solar energy directly into electricity and could supply a significant fraction of electric power needs in the southwestern regions of the United States if system costs can be reduced. Energy conversion efficiencies of 10% or more, of

insolation are now being achieved. If PV installations, in, e.g., Arizona, were designed that intercepted one fourth of the incoming radiation the peak electric power output would be 7 MW/km$^2$ or a daily average power of 1.25 MWe for clear sky conditions. This implies that the total electric power needs of Arizona (15,000 MW in 1996) could be supplied by PV installations occupying 12000 km$^2$ (4630 square miles), or 4% of the state area. Installed costs for grid-connected PV systems have been predicted to fall to $2500 per peak kilowatt after the year 2000.
7. Solar–thermal–electric systems. Three types of solar–thermal–electric concepts have received substantial development support: (a) Central receiver systems. These systems employ an array of mirrors (heliostats) that focus solar energy on a central heat sink connected to a heat-engine electric-generator combination for electric power production. For a 10-MWe installation, the plant investment was estimated to be $70 M. (b) Focused-dish array. This system employs an array of parabolic mirrors with a small heat-engine electric generator at the focus of each mirror. It has demonstrated 15% efficiency. Predicted peak power efficiencies are in the range 16–28%. Plant investment was estimated to be $200 M for a 100 MWp plant. (c) Cylindrical mirror assembly. This system has cylindrical one-axis parabolic mirrors that concentrate solar energy on an axial pipe at the focus that contains a heat absorbing liquid. This is pumped to a central station where the hot liquid is used to generate power. The estimated power cost was $225 M for 80 MWe output.
8. Electric power generation by wind turbines. Electric power generation by wind turbines has achieved commercial status. The estimated installation cost (PI) is $(900–1000)/kWp.
9. Fuel-cell electric power generation. Ammonia, methanol, natural gas, or hydrogen could become significant, environmentally attractive, sources of electric power if used in fuel cells as local sources of electricity. Whether methanol or ammonia produced by OTEC will be used as fuel sources will depend on the cost relative to the costs of the fuels derived from natural gas.

## C. Comments

The financial analysis program can estimate fuel or power sale prices for a range of values of plant investment and operating cost. Comparisons can also be made of the effects on profitability of improvements or deficiencies in performance or cost, which might result from changes in interest rates, capital investments, or from further R & D

or manufacturing experience. In commercial production of mechanical systems, construction costs are reduced as component sizes and numbers increase. Therefore, estimates of projected product prices should be based on the same total output of fuel or electric energy for each system compared. Typically, the product cost is reduced by a constant factor, called the experience factor, when the number produced is doubled. When the unit size is increased, the ratio of costs depends on the ratio of sizes, with an exponent, called the "scaling factor." Conservative values of the exponents have been selected for all of the options to provide a means for judging the importance of the scaling and experience factors in comparing the options. Values of 0.93 for the experience factor and 0.6 for the scaling factor have been chosen, drawn from industrial experience in the shipbuilding, petroleum, and utility industries.

## XVI. ECONOMIC IMPACT OF OTEC COMMERCIALIZATION

As a means of judging the economic impact of a national plan that would ensure a sustainable energy future beyond the year 2050, the plant investments and fuel or electricity prices have been estimated for profitable operation of new alternative systems that would each install enough plants to yield a total energy equal to the fuel or electric power delivered to U.S. consumers in 1999. Examples of the economic features of proposed future energy systems are compared in Tables XIV and XV. The pertinent data, including plant investment (PI), plant output, and estimated delivered energy prices, are shown for each proposed system. The capital costs estimated in 1990 for the systems listed, based on DOE supported programs, are used in the estimates in lieu of unavailable up-to-date values. The number of plants (modules) needed to supply the total energy demands is shown for each system.

The tables show that huge capital investments will be necessary to install any new system designed to supply gasoline or electric power in the amounts consumed in the United States in 1999. Annual capital investments of billions dollars per year will be required to satisfy the national energy demands by 2050, whether existing systems are replaced or new types are installed.

The commercial prices of fuels and electricity are determined by the market and by the effects of government actions on capital investments, interest rates, and other factors. The data presented in Tables XIV and XV are estimates of sales prices for fuels and power, based on data developed in the DOE OTEC program that are useful in comparing the relative merits of alternative energy

options. OTEC methanol and ammonia are shown to be marginally profitable in comparison with other options. When plans are developed for particular systems, the analysis program can be used to provide more accurate estimates, which will use up-to-date information for all of the relevant variables.

Tables XVI and XVII show examples of the dependence of sales prices, and profits to investors, of OTEC methanol and ammonia production, as the number of installed plantships increases, and when the effects on sale prices of the environmental costs associated with fossil fuel combustion and costs to the national economy of importing crude oil are included. [The rationale for the estimates is explained in Avery and Wu (1994).] For the assumed values of the variables, production and sale of OTEC methanol would be profitable beginning with the first plantship. Production of ~128 plantships would be necessary before OTEC ammonia sales would be profitable. If each system were developed in sufficient numbers to replace all U.S. gasoline consumption, and gasoline taxes were imposed to pay for economic impacts of importing oil and of $CO_2$ additions to the atmosphere, substitution of OTEC methanol for gasoline at the 1990 average wholesale price per gallon would save \$152B/year; and OTEC ammonia, \$174B/y.

OTEC commercialization has not occurred because industry and government have concentrated funding on projects that have strong political support or a short-term payoff. The large capital investment required for OTEC commercialization has been regarded as an impregnable barrier despite the fact that equivalent funds to support other less attractive developments have been spent by government and industry. Wise consideration of the long-term energy needs of the United States is needed to make it evident that OTEC development should be supported. OTEC has unique features that merit giving it high priority in future plans: (i) It uses stored solar energy that is inexhaustible. (ii) OTEC energy is available 24 hr a day. (iii) OTEC ammonia fuel generates no greenhouse gases. (iv) OTEC produces no hazardous wastes in construction or operation. (v) OTEC is environmentally friendly, and would eliminate the "not-in-my-backyard" roadblock to on-land energy installations. (vi) With commercial development, OTEC will be financially competitive if the environmental and national security costs of the of alternative options are included in the financial comparisons.

The commercialization of OTEC to its full potential will have a significant impact on the U.S. economy through elimination of oil imports and of the huge sums required to protect the imported oil sources, through elimination of air pollution caused by combustion of fossil fuels, and by abating hazards associated with power production and waste disposal from nuclear sources. Inexhaustible OTEC

**TABLE XVI** Profit to Investors Selling Methanol Produced on OTEC Plantships To Replace Gasoline Used in U.S. Automobiles (Debt Interest 11%) (Construction Interest 9%) (ITC 10%) (APV 0) ($1990)

| Plant no. | Relative cost | Sum | nth CH3OH plant PI ($M) | CH3OH total PI ($B) | Avg PI ($M) | Total CH3OH prod/n (Bgal/y) | CH3OH Sales price ($/gal) | Equiv. gasoline price ($/gal) | Gasoline market price ($/gal) | Gasoline replaced (Bgal/y) | Gasoline "real price" with import impact ($/gal) | CH3OH mileage equivalent price incl automobile costs (8) ($/gal) | Savings using CH3OH to replace gasoline ($B/y) | Gasoline price incl import and carbon tax ($/gal) | CH3OH mileage equiv price incl auto cost and carbon tax ($/gal) | Net profit / Net profit to investors replacing gasoline with CH3OH ($B/y) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.000 | 1.00 | 934 | 0.93 | 934 | 0.20 | 0.684 | 1.23 | 0.79 | 0.11 | 1.36 | 1.34 | 0.00 | 1.73 | 1.50 | 0.025 |
| 2 | 0.800 | 1.80 | 747 | 1.68 | 841 | 0.40 | 0.642 | 1.16 | 0.79 | 0.22 | 1.36 | 1.26 | 0.02 | 1.73 | 1.42 | 0.068 |
| 3 | 0.800 | 2.60 | 747 | 2.43 | 809 | 0.60 | 0.627 | 1.13 | 0.79 | 0.33 | 1.36 | 1.24 | 0.04 | 1.73 | 1.39 | 0.110 |
| 4 | 0.744 | 3.34 | 695 | 3.12 | 781 | 0.80 | 0.614 | 1.11 | 0.79 | 0.44 | 1.36 | 1.21 | 0.06 | 1.73 | 1.37 | 0.157 |
| 8 | 0.692 | 6.32 | 646 | 5.90 | 738 | 1.59 | 0.594 | 1.07 | 0.79 | 0.88 | 1.36 | 1.18 | 0.16 | 1.73 | 1.33 | 0.347 |
| 16 | 0.643 | 11.9 | 601 | 11.1 | 692 | 3.18 | 0.573 | 1.03 | 0.79 | 1.77 | 1.36 | 1.14 | 0.38 | 1.73 | 1.30 | 0.76 |
| 32 | 0.598 | 22.2 | 559 | 20.7 | 647 | 6.37 | 0.553 | 1.00 | 0.79 | 3.54 | 1.36 | 1.10 | 0.90 | 1.73 | 1.26 | 1.65 |
| 64 | 0.557 | 41.3 | 520 | 38.6 | 603 | 12.7 | 0.532 | 0.96 | 0.79 | 7.08 | 1.36 | 1.06 | 2.06 | 1.73 | 1.22 | 3.56 |
| 128 | 0.518 | 76.9 | 483 | 71.8 | 561 | 25.5 | 0.513 | 0.92 | 0.79 | 14.2 | 1.36 | 1.03 | 4.61 | 1.73 | 1.19 | 7.61 |
| 256 | 0.481 | 143 | 450 | 134 | 522 | 50.9 | 0.495 | 0.89 | 0.79 | 28.3 | 1.36 | 1.00 | 10.1 | 1.73 | 1.46 | 7.4 |
| 423 | 0.457 | 224 | 427 | 209 | 494 | 84.2 | 0.482 | 0.87 | 0.79 | 46.8 | 1.36 | 0.97 | 18.0 | 1.73 | 1.47 | 12.3 |
| 512 | 0.448 | 266 | 418 | 249 | 486 | 102 | 0.479 | 0.86 | 0.79 | 56.6 | 1.36 | 0.97 | 21.9 | 1.73 | 1.46 | 14.9 |
| 965 | 0.419 | 469 | 391 | 438 | 454 | 192 | 0.464 | 0.84 | 0.79 | 107 | 1.36 | 0.94 | 44.1 | 1.73 | 1.46 | 44.3 |
| 1024 | 0.416 | 496 | 389 | 463 | 452 | 204 | 0.463 | 0.83 | 0.79 | 113 | 1.36 | 0.94 | 47.0 | 1.73 | 1.46 | 29.8 |

| | |
|---|---|
| Total U.S. crude oil consumed (BBL/D) (1990) | 1.30E+07 |
| Total imported crude oil consumed (BBL/D) | 5.69E+06 |
| U.S. gasoline consumed (1990) (BBL/D) | 6.96E+06 |
| Ratio gasoline consumed to crude oil consumed | 0.533 |
| U.S. gasoline consumed (1990) (GAL/Y) | 1.08E+11 |
| Equivalent CH3OH amount to match gasoline road mileage (gal/year) | 1.92E+11 |
| OTEC CH3OH plant output (gal/year) (345 D) | 1.99E+08 |
| Number of plants to replace all U.S. gasoline consumed | 965 |
| Ratio imported oil to total oil consumed | 0.438 |
| Total gasoline made from imported oil | 4.67E+10 |
| Number of plants to replace gasoline equivalent of imported oil | 423 |

**TABLE XVII  Profit to Investors Selling Ammonia Produced on OTEC Plantships To Replace Gasoline in U.S. Automobiles (Construction Interest 9%) (Debt Interest 11%) (ITC 10%) (APV 0) (1990 Average Wholesale Gasoline Price without Taxes)**

| Plant no. | nth NH$_3$ plant PI ($M) | Total NH$_3$ PI ($B) | Avg PI ($M) | Total NH$_3$ prod\n (Bgal/y) | NH$_3$ Sales price ($/gal) | Equiv. gasoline price ($/gal) | Gasoline market price w/o tax ($/gal) | Gasoline replaced (Bgal/y) | Gasoline "real price" with carbon impact ($/gal) | Savings using NH$_3$ to replace gasoline ($B/y) | Gasoline price incl import and carbon taxes ($/gal) | NH$_3$ equiv price incl auto cost ($/gal) | Savings using NH$_3$ to replace gasoline ($/gal) | Net profit to investors replacing gasoline with NH$_3$ ($B/y) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 948 | 0.95 | 948 | 0.17 | 0.66 | 1.77 | 0.79 | 0.064 | 1.36 | −0.03 | 1.65 | 2.07 | −0.42 | −0.027 |
| 2 | 758 | 1.71 | 853 | 0.34 | 0.61 | 1.62 | 0.79 | 0.13 | 1.36 | −0.03 | 1.65 | 1.92 | −0.27 | −0.034 |
| 3 | 758 | 2.46 | 822 | 0.51 | 0.585 | 1.56 | 0.79 | 0.19 | 1.36 | −0.04 | 1.65 | 1.86 | −0.21 | −0.041 |
| 4 | 705 | 3.17 | 806 | 0.68 | 0.57 | 1.53 | 0.79 | 0.25 | 1.36 | −0.04 | 1.65 | 1.83 | −0.18 | −0.047 |
| 8 | 656 | 5.99 | 749 | 1.36 | 0.54 | 1.43 | 0.79 | 0.51 | 1.36 | −0.04 | 1.65 | 1.73 | −0.08 | −0.043 |
| 16 | 610 | 11.2 | 702 | 2.72 | 0.51 | 1.35 | 0.79 | 1.02 | 1.36 | 0.01 | 1.65 | 1.65 | 0.00 | 0.00 |
| 32 | 567 | 21.0 | 656 | 5.44 | 0.48 | 1.27 | 0.79 | 2.0 | 1.36 | 0.18 | 1.65 | 1.57 | 0.08 | 0.16 |
| 64 | 528 | 39.2 | 612 | 10.9 | 0.45 | 1.19 | 0.79 | 4.1 | 1.36 | 0.67 | 1.65 | 1.49 | 0.16 | 0.63 |
| 128 | 491 | 72.9 | 570 | 21.8 | 0.42 | 1.12 | 0.79 | 8.1 | 1.36 | 1.98 | 1.65 | 1.42 | 0.23 | 1.90 |
| 256 | 456 | 135.7 | 530 | 43.5 | 0.39 | 1.05 | 0.79 | 16.3 | 1.36 | 5.09 | 1.65 | 1.35 | 0.30 | 4.9 |
| 512 | 424 | 253 | 493 | 87.0 | 0.37 | 0.98 | 0.79 | 33 | 1.36 | 12.4 | 1.65 | 1.28 | 0.37 | 12.0 |
| 894 | 400 | 415 | 464 | 152 | 0.35 | 0.93 | 0.79 | 57 | 1.36 | 24.5 | 1.65 | 1.23 | 0.42 | 23.9 |
| 1024 | 395 | 470 | 459 | 174 | 0.34 | 0.91 | 0.79 | 65 | 1.36 | 29.1 | 1.65 | 1.21 | 0.44 | 28.5 |
| 1676 | 375 | 727 | 434 | 285 | 0.33 | 0.88 | 0.79 | 107 | 1.36 | 51.6 | 1.65 | 1.18 | 0.47 | 50.6 |
| 2048 | 367 | 874 | 427 | 348 | 0.32 | 0.87 | 0.79 | 130 | 1.36 | 64.4 | 1.65 | 1.17 | 0.48 | 63.2 |

| | |
|---|---|
| Total U.S. crude oil consumed (BBL/D) (1990) | 1.30E+07 |
| Total imported crude oil consumed (BBL/D) | 5.69E+06 |
| U.S. gasoline consumed (1990) (BBL/D) | 6.96E+06 |
| Ratio gasoline consumed to crude oil consumed | 0.533 |
| U.S. gasoline consumed (1990) (gal/year) | 1.07E+11 |
| Equivalent NH$_3$ amount to match gasoline road mileage (gal/year) | 2.85E+11 |
| OTEC NH$_3$ plant output (gal/year) (345 D) | 1.70E+08 |
| Number of plants to replace All U.S. gasoline consumed | 1676 |
| Ratio imported oil to total oil consumed | 0.438 |
| Total gasoline made from imported oil | 4.67E+10 |
| Number of plants to replace gasoline equivalent of imported oil | 734 |

energy could be produced in amounts exceeding world demands. This would remove the need to develop fusion power, which receives substantial funding from government and industry, because it is regarded as a future source of inexhaustible energy that could replace power now produced from fossil fuels and from nuclear fission, both of which can become depleted in the future.

A detailed discussion and analysis of the overall economic and social impacts of full commercialization of OTEC is given in Avery and Wu (1994).

## SEE ALSO THE FOLLOWING ARTICLES

ENERGY FLOWS IN ECOLOGY AND IN THE ECONOMY • ENERGY RESOURCES AND RESERVES • OCEAN-ATMOSPHERIC EXCHANGE • OCEAN SURFACE PROCESSES • OFFSHORE STRUCTURES

## BIBLIOGRAPHY

Abraham, R., Jayashankar, V., Ikegami, Y., Mitsumori, M., and Uehari, H. (1999). Analysis of Power Cycle for 1 MW Floating OTEC Plant. Proceedings of the International OTEC/DOWA Conference '99 Imari, Japan, p. 123.

Avery, W. H., and Wu., C. (1994). "Renewable Energy from the Ocean: A Guide to OTEC," Oxford Univ. Press, New York.

George, J. F., and Richards, D. (1980). "Baseline designs of moored and grazing OTEC pilot plants," *Johns Hopkins Appl. Phys. Lab.* SR-80-A&B.

Ito, F., and Seya, Y. (1985). "Operation experience of the 100-kW OTEC and its subsequent research," *Japanese–French Symp. on Ocean Development*, Tokyo, Sept.

Kalina, A. I. (1987). Regeneration of the working fluid and generation of energy. Japanese Patent Sho62-39660.

Lin, G., Lee, T. T., and Hwang, K. K. (1983). "Ocean thermal energy conversion technology development program in Taiwan, Republic of China," *Proc. Oceans* '83 (Aug. 2), 723.

Marchand, P. (1985). "French thermal energy conversion program," *Proc. French-Japanese Symp. Ocean Devel.* Tokyo, Sept. 1.

Mitsui, T. F., Soya, Y., and Nakamoto, Y. (1983). "Outline of the 100-kW OTEC pilot plant in the island of Nauru," IEEE/PES 1983 Winter meeting, New York, Jan. 212.

Owens, W. L. (1983). "Optimization of closed-cycle OTEC plants," *Proc. ASME-JSME Joint Conf.* Honolulu **2,** 227.

Starr, C. (1987). "Technology innovations and energy futures," *Proc. Energy Technology Conf. XIV* Govt. Institutes Rockville, MD, 21.

Uehara, H. (1982). "Research and development on ocean thermal energy conversion in Japan," *Proc.17th IECEC Conf.* **3,** 1454.

Uehara, H., Ikegami, Y., Sakaki, K., and Nogamim, R. (1998). "The experimental research on ocean thermal energy conversion using the Uehara cycle," *Proc. International OTEC /DOWA Conf.*, Imari, Japan.

Wu, C. (1987). "A performance bound for real OTEC heat engines," *Ocean Eng.* **14,** 349.
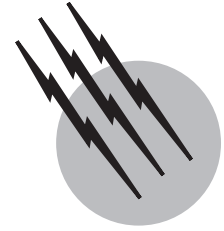
# Petroleum Geology

**Richard C. Selley**

*University of London*

## GLOSSARY

**Anticline**  Upfold of strata, a common form of petroleum trap.

**Cap rock**  Layer of impermeable strata sealing petroleum within a reservoir.

**Evaporites**  Group of rock-forming minerals resulting from the evaporation of seawater.

**Kerogen**  Solid hydrocarbon, insoluble in petroleum solvents, that generates petroleum when heated.

**Petroleum**  Fluid hydrocarbons, both liquid (crude oil) and gaseous.

**Reservoir**  Porous permeable rock, generally a sediment, capable of storing petroleum.

**Seal**  Layer of impermeable strata that caps a petroleum reservoir.

**Sedimentary basin**  Part of the earth's crust containing a thick sequence of sedimentary rocks.

**Source rock**  Fine-grained organic-rich sediment capable of generating petroleum.

**Trap**  Configuration of strata, such as an anticline, capable of trapping petroleum.

**Unconformity**  Break in deposition, often marked by an angular discordance between two sets of strata.

**PETROLEUM GEOLOGY** is concerned with the way in which petroleum forms, migrates, and may become entrapped in the crust of the earth. It studies the generation of organic matter on the surface of the earth, the way in which this may be buried and preserved in sediments, and the physical and chemical processes that lead to the formation of petroleum and its expulsion from sedimentary source beds. Petroleum geology also embraces the dynamics of fluid flow in porous permeable rocks. It is concerned with the depositional environments and post-depositional history of the sandstones and limestones that commonly serve as petroleum reservoirs. Petroleum geology also studies the ways in which tectonic forces may deform strata into configurations that may trap migrating petroleum.

## I. FIELD OF PETROLEUM GEOLOGY

Petroleum is the name given to fluid hydrocarbons, both the gases and the liquid crude oil. It is commonly noted that petroleum occurs almost exclusively within sedimentary rocks (sandstones, limestones, and claystones). Petroleum is seldom found in igneous or metamorphic rocks. Thus petroleum geology is very largely concerned with the study of sedimentary rocks. It deals with processes operating on the surface of the earth today to discover how organic detritus may be buried and preserved in fine-grained sediments. It considers where and how porous sands become deposited, because these may subsequently serve as reservoirs to store petroleum. It is also important to know how porosity and permeability are distributed in sedimentary rocks. This is essential to understand the way in which petroleum emigrates from the source beds, how it may subsequently be distributed in reservoir beds, and how it may flow once that reservoir is tapped.

Petroleum geology is also concerned with the structural configuration of the crust of the earth. This interest includes both large- and small-scale crustal features. On the broadest scale, petroleum geology is concerned with the formation of sedimentary basins. This is because some basins are highly favorable for petroleum generation, while others are less suitable. The deposition of petroleum source beds varies with basin type, as does the heat flow that is responsible for generating oil and gas from the source rock.

On a smaller scale, petroleum geology is concerned with the way in which strata may be deformed into folds, and how they may be ruptured and displaced by faults. Such structural deformation may form features capable of trapping migrating petroleum.

## II. PROPERTIES OF PETROLEUM

Petroleum, as previously defined, is the name given to the fluid hydrocarbons. A closer examination of the physical and chemical properties of petroleum is perhaps necessary as a prerequisite to an analysis of the geology of petroleum.

The term kerogen is applied to solid hydrocarbons that are insoluble in the normal petroleum solvents, such as carbon tetrachloride, yet which have the ability to generate fluid petroleum when heated. One of the best-known examples of kerogen is coal and its associates peat, lignite, and anthracite. Kerogen is also found disseminated in many fine-grained sediments and sedimentary rocks (the clays, claystones, and shales). This is seldom of direct economic value itself, but is of great interest to petroleum geologists because it is the source material from which commercial quantities of petroleum are derived.

Liquid petroleum is generally termed crude oil, oil, or simply "crude." In appearance, oils vary from colorless to yellow-brown or black. The former are generally of low viscosity and low density, while the black oils tend to be heavier and more viscous. The gravity of petroleum is commonly expressed by reference to a scale proposed by the American Petroleum Institute (API). The API gravity is calculated as follows:

API gravity = (141.5/specific gravity at $60°F$) − 131.5

Chemically, petroleum consists principally of hydrogen and carbon, but also contains small percentages of oxygen, nitrogen, sulfur, and traces of metals, such as vanadium, cobalt, and nickel. The common organic compounds include alkanes (paraffins), naphthenes, aromatics, and heterocompounds. The specific gravity of oil is variable. Heavy oils are defined as those with a specific gravity greater than 1.0 ($10°API$). These are rich in heterocompounds and aromatics. The heavy oils grade with increasing density and viscosity into the plastic hydrocarbons (generally termed pitch, tar, or asphalt). The light oils are rich in alkanes and naphthenes. They grade into "condensate." This is petroleum that is gaseous in the earth's crust, but which condenses to a liquid at the surface.

Petroleum gas is generally referred to as "natural gas," though this ignores the nonhydrocarbon natural gases of the atmosphere. Natural gas is generally composed of methane and varying amounts of ethane, propane, and butane. Petroleum gas composed almost exclusively of methane is termed "dry gas." Petroleum gas with substantial amounts of the other gaseous alkanes is termed "wet gas." Petroleum gases are also associated with traces of other gases, notably hydrogen sulfide, carbon dioxide, and nitrogen.

Gas hydrates are compounds of frozen water that contain gas molecules. Naturally occurring gas hydrates contain substantial amounts of petroleum gas, principally methane. These are only stable under certain critical pressure/temperature conditions. They occur today in the surface sediments of deep ocean basins and in permafrost. Vast reserves of petroleum are trapped in gas hydrates but they cannot yet be extracted safely and commercially.

## III. ORIGIN OF PETROLEUM

The origin of petroleum has long been a matter for debate. From the days of Mendeleev, some chemists and astronomers have argued that petroleum is of inorganic abiogenic origin. Petroleum geologists have seldom doubted that it is of organic origin. It is known that hydrocarbons can form in space. Jupiter and Saturn are believed to be composed of substantial amounts of methane. A particular

type of meteorite (the stony meteorites, or chondrites) contain complex hydrocarbons up to the level of amino acids and isoprenoids. Petroleum is found in some igneous rocks (rocks formed from the cooling of molten magma). In most of these occurrences the petroleum infills fracture and have obviously migrated into the rock long after it cooled. Rarely, however, it occurs in gas bubbles within the rock itself. In some cases it can be demonstrated that the magma has intruded organic-rich sedimentary rocks from which the petroleum may have been acquired, but this is not universally true. Methane occurs in many rift valley floors, both on land (as in the East African rifts) and in the mid-ocean ridge rift systems and in permafrost and deep ocean gas hydrates. Carbon isotope analysis is able to show that this is not always of shallow biogenic origin, but that it is often of deep thermal origin.

These observations have caused some people to argue that petroleum is of inorganic abiogenic origin; that methane occurs in the mantle of the earth from which it seeps to the surface, being intermittently expelled up fault systems aided by earthquakes. The fundamental reaction is believed to be between iron carbide and water (analogous to the reaction whereby acetylene is produced by calcium carbide and water):

$$CaC_2 + 2H_2O = C_2H_2 + Ca(OH)_2$$
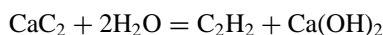
Once it was generally believed that all life forms began with photosynthesizing algae, and that these formed the base of ecosystems. Hence petroleum must be of organic origin. The discovery of hot vents on the deep ocean floor termed "smokers," shows that ecosystems can be based on methanogenic bacteria.

An abiogenic origin for petroleum is popular with some astronomers, but, except in the Commonwealth of Independent States (CIS), these views are seldom held by geologists in general, and petroleum explorationists in particular. Most geologists, and certainly those engaged in the commercial quest for petroleum, believe that it is of organic origin.

There are many lines of evidence to support the thesis that commercial petroleum accumulations are of organic origin. Some of the evidence is chemical and some geological. Oil exhibits levorotation, the ability to rotate polarized light; this property is peculiar to biogenically synthesized organic compounds. Oil contains traces of complex organic molecules that are common in living plant and animal tissues. Examples include porphyrins, steroids, and derivatives of chlorophyll. In many instances oil is found in porous permeable reservoirs totally enclosed by impermeable shale. It is hard to see how petroleum could have migrated into the reservoir, either from the earth's mantle or from outer space. In such instances the enclosing shales generally contain kerogen. Gas chromatography demonstrates the similarity of petroleum generated from the kerogen in the shale with that found in the reservoir.

Finally there is the ubiquitous observation that petroleum is commonly found in sedimentary rocks and not in igneous or metamorphic ones. These observations lead to the conclusion that although petroleum may form inorganically in space and in the earth's mantle, commercial accumulations of petroleum are formed from the thermal maturation of biologically synthesized organic compounds in the earth's crust. The foregoing analysis of the origin of petroleum points us toward the study of sedimentary rocks and their contained fluids. It is now generally recognized that the four basic requirements for a commercial petroleum accumulation are (1) a thermally mature source rock, (2) a porous permeable reservoir to retain the petroleum, (3) an impermeable cap rock, or seal, to retain the petroleum within the reservoir, and (4) a configuration of rock, such as an anticline, to trap the petroleum within the reservoir. These items will now be considered in turn.

## IV. GENERATION AND MIGRATION OF PETROLEUM

### A. The Production and Preservation of Organic Matter

Photosynthesis is the process on which all organic matter is based. In this reaction, carbon dioxide and water are turned into sugar and oxygen in the presence of sunlight, according to the following reaction:

$$6CO_2 + 12H_2O = C_6H_2O_6 + 6H_2O + 6O_2$$
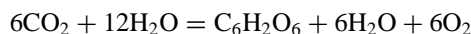
The sugar that is produced by photosynthesis is the starting point for the construction of the more complex organic molecules found in plants and animals. In the ordinary way, when life forms die, they are eaten or broken down by the normal processes of bacterial decay. Thus organic matter is seldom preserved in sedimentary rocks.

There are, however, certain depositional environments that favor the preservation of organic matter (Fig. 1). In lakes, a thermal stratification of water sometimes develops. Sunlight warms the shallow water and encourages blooms of algae. As these give off oxygen they provide the base for animal food chains. In the deeper waters, sunlight does not penetrate. The water is thus cooler and denser. Lacking photosynthesizing algae this zone may become anoxic. The density stratification inhibits the mixing of oxic and anoxic waters. Organic detritus falling from above may be preserved in the stagnant sediments of the lake floor.

Thus ancient lacustrine deposits are often petroleum source beds. A similar stratification, with an upper
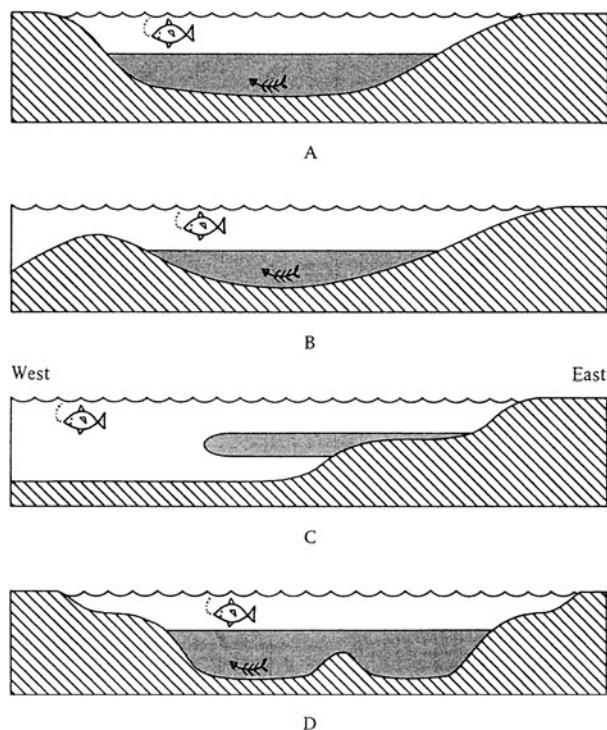
**FIGURE 1** Geophantasmograms to illustrate four common depositional models for organic-rich petroleum source bed deposition. (A) Thermally stratified lake. (B) Hypersaline lagoon, with briny deep water recharged by normal-salinity open sea water. (C) Anoxic conditions occur on the edge of the continental shelf on the eastern sides of oceans. Shallow water is oxygenated by phytoplankton. Cold polar oxygenated waters sink to ocean floors. (D) Pan-oceanic anoxic event. This occurs when polar glaciation is absent during episodes of uniform warm global temperature, such as may have occurred during the Cretaceous period. [From Selley, R. C. (1998). "Elements of Petroleum Geology," 2nd ed., Fig. 5.8, Academic Press, San Diego.]

oxygenated and a lower anoxic layer of water, sometimes develops in arid climate lagoons. In this environment the stratification is due not to thermal layering, but to evaporation. Dense anoxic brines concentrate on the lagoon floor and favor the pickling of organic detritus in the underlying muds. Organic sediments may also be deposited in ocean basins. In modern oceans, the shallow waters are oxygenated by plankton photosynthesis. The deep ocean basins are seldom anoxic. But an oxygen-deficient zone commonly occurs between 200 and 1500 m on the eastern sides of the major ocean basins. The upper limit of this zone commonly coincides with the edge of the continental shelf. Thus organic matter may be preserved in sediments deposited on the continental slope and rise.

Global rises of sea level cause marine transgressions and favor the deposition of blanket petroleum source beds across continental shelves. Today the flow of cold dense oxygenated polar water into the ocean basins of lower lat-

itudes prevents their floors from becoming stagnant. It has been argued that when the earth has had a uniform equable climate, this mixing would not occur. Global anoxic events may then favor the ubiquitous preservation of organic matter on the ocean basin floors. These are the four types of sedimentary environments that favor the preservation of organic detritus in fine-grained sediments.

## B. Generation of Petroleum from Kerogen

As muds are buried they compact, lose porosity, and are lithified into shales or mudstones. The organic matter that they contain undergoes many changes. These can be conveniently grouped into three stages: diagenesis, catagenesis, and metagenesis. Diagenesis occurs in the shallow subsurface at near normal temperatures and pressures. Methane, carbon dioxide, and water are driven off from the organic matter to leave a complex hydrocarbon termed kerogen (already defined as hydrocarbon that is insoluble in normal petroleum solvents, but expels petroleum when heated).

There are four main types of kerogen. Type I, sapropelic kerogen, is formed from algal matter. It has a high potential for generating oil. Type II, liptinic kerogen, can generate both oil and gas. Type III, humic kerogen, is predominantly gas prone. This type of kerogen forms as peat on the surface of the earth. As this is buried and matured it evolves through lignite, bituminous coal, and anthracite to graphite (pure carbon). With increasing burial the source bed is subjected to rising temperature and pressure. Diagenesis merges into catagenesis. This is the principal phase of petroleum generation. Initially, oil is given off once temperatures are above about 60°C. The peak of oil generation is at about 120°C. Above this temperature, oil generation declines, to be replaced by gas generation and ceases at about 200°C. This temperature marks the transition between catagenesis and metamorphism. At this stage, the kerogen has yielded up all its petroleum. The residue is pure carbon, graphite.

The fourth type of kerogen is reworked from older sedimentary formations. It is essentially inert, and has no petroleum-generating potential.

## C. Petroleum Migration

The migration of petroleum has long been a matter for debate. It is normal to differentiate primary from secondary migration. Primary migration refers to the emigration of petroleum from the source rock into a permeable carrier bed. Secondary migration refers to the migration of petroleum from the carrier bed into the reservoir, and to subsequent flow during petroleum production. Secondary migration is relatively simple to understand. Oil and gas

move in response to buoyancy and pressure differentials within large pore systems.

Primary migration presents a different problem. Shales are impermeable and normally provide excellent seals that inhibit upward petroleum movement (see later). How then can petroleum emigrate from an impermeable source bed into a permeable carrier formation? It is not enough to invoke simple compaction of clay during burial. When clays compact, most of the porosity is lost in the first kilometer or so of burial. Temperatures are still too low to permit kerogen maturation. At the depths at which petroleum generally forms (some 3 to 4 km) little porosity loss takes place.

Many different mechanisms have been proposed to explain primary migration. These invoke solution of oil and gas in water, improbably high temperatures, abnormally high pressures, and the use of soapy micelles as petroleum solvents which then vanish once they have done their task.

It is often noted that petroleum is commonly produced from source rocks that are, or once were, overpressured, that is to say, whose fluid pressure is/was above the hydrostatic pressure. There is evidence that the pressure in an overpressured shale decreases over time to hydrostatic pressure, not gradually, but episodically. A trigger, such as an earthquake, causes a sudden expulsion of fluids through faults and fractures. These immediately close up once the hot fluid flush has passed through. This mechanism could well explain how petroleum emigrates from impermeable organic-rich shales.

Though no single theory has received universal acceptance, the last mentioned is one of the most popular. Perhaps an understanding of the mechanism is not too important. The significant fact is that gas chromatography can correlate kerogen in source beds with the trapped petroleum that once emigrated from them.

The *first requirement for a commercial accumulation of petroleum* is the existence of a thick sequence of organic-rich shale, with an excess of 1.5% organic carbon. Whether the kerogen yields oil or gas, or both, is dependent on the type of kerogen and its level of thermal maturation. These parameters are commonly displayed on graphs that plot the ratio of hydrogen:carbon against oxygen:carbon (Fig. 2).

## V. PETROLEUM RESERVOIRS AND CAP ROCKS

### A. Porosity and Permeability

The *second requirement for a commercial accumulation of petroleum* is the existence of a reservoir capable of storing the oil or gas. There are two basic requirements
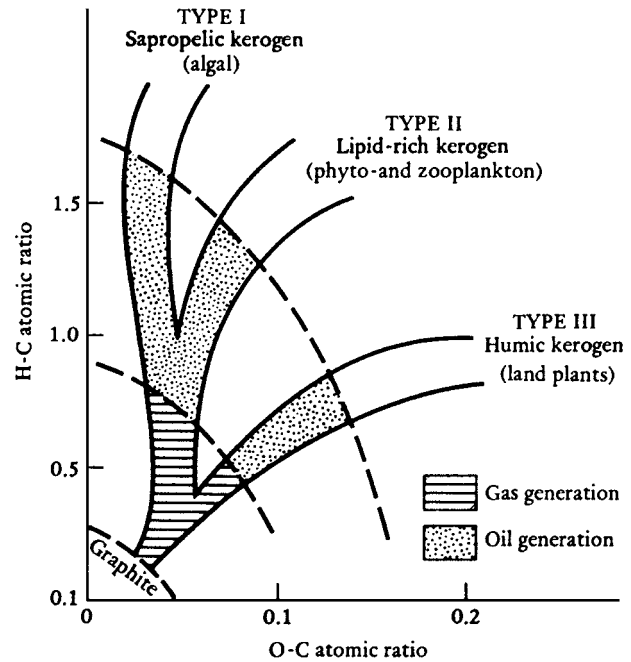
**FIGURE 2** Graph of hydrogen/carbon ratio plotted against oxygen/carbon ratio. This shows the composition of the different types of kerogen. Type I kerogen generates principally oil with minor gas. Type II kerogen generates both oil and gas. Type III kerogen generates only gas. Type IV kerogen, not shown, is inert. [From Selley, R. C. (1998). "Elements of Petroleum Geology," 2nd ed., Fig. 5.13, Academic Press, San Diego.]

for a reservoir: porosity and permeability. Porosity is the storage capacity of the rock. This is sometimes expressed as the void ratio, or more commonly as a percentage:

$$\text{Porosity} = (\text{void volume/rock volume}) \times 100\%$$

Permeability is the rate of flow of fluid through the rock. It is expressed by Darcy's Law. This states that the rate of flow of a homogeneous fluid in a porous medium is proportional to the pressure gradient and inversely proportional to the fluid viscosity. This is generally expressed mathematically thus:

$$Q = \frac{[K(P_1 - P_2)A]}{mL}$$

Where $Q$ is the rate of flow, $K$ a constant, $(P_1 - P_2)$ the pressure drop, $A$ the cross-sectional area of the sample, $L$ the length of the sample, and $m$ the viscosity of the fluid. Permeability is measured in Darcies (D). One Darcy is the permeability that allows a fluid of 1 cP viscosity to flow at a velocity of 1 cm/sec for a pressure drop of 1 atm/cm. Since most petroleum reservoirs commonly have permeabilities of less than 1 D the millidarcy (1/1000 D) is usually used. Permeability can be measured in the laboratory from a core or a plug cut from a core. It can also be calculated

from a production well test of a petroleum reservoir. Some geophysical borehole logs can also measure permeability in a roundabout way.

Theoretically, any rock can act as a petroleum reservoir so long as it has these two essential properties of porosity and permeability. Igneous and metamorphic rocks are generally composed of a tight interlocking mosaic of mineral crystals. They thus seldom have porosity and permeability unless they have been fractured or weathered beneath an unconformity (an old land surface). Only about 10% of the world's oil reserves have been found in igneous and metamorphic basement. Most petroleum reservoirs are found in the sedimentary rocks (sandstones, limestones, and shales) because, by their very nature, they form from the deposition of granular particles with microscopic pores between the grains. Shales (compacted mud) are commonly porous. But, because of the small diameter of their pores, shales are generally impermeable and seldom act as reservoirs. About 45% of the world's petroleum reserves have been found in sandstones, and about 45% in carbonates (limestones and dolomites).

Table I lists the various types of pores systems encountered in petroleum reservoirs. There are two main types of pores: primary (or syndepositional) and secondary (or postdepositional). Primary pores are those that form when a sediment is deposited. They may be either intergranular, occurring between grains, or intragranular, occurring within grains. The former is common in terrigenous sediments and is therefore the major type of pore found in sandstone reservoirs (Fig. 3). Intraparticle pores occur within shells in skeletal lime sands. During burial, however, these pores are often destroyed by compaction or by cementation.

There are several different types of secondary porosity. Solution due to migrating fluids may generate moldic or vuggy porosity. Moldic porosity is fabric selective. That is to say, only one element of the rock is leached out, such as fossil shell fragments. Vuggy pores cross-cut the fabric of the rock, grains, matrix, and cement. Vugs are thus

commonly larger than moldic pores. They grade up with increasing size into cavernous porosity. Solution porosity is characteristic of limestone reservoirs. This is because calcium carbonate is less stable in the subsurface than the quartz (silica) of which sandstones are largely composed. Unlike primary intergranular porosity, however, solution pores are often isolated from one another. Thus, such rocks may have considerable porosity but negligible permeability.

The other main type of porosity is that which is caused by fracturing. Fractures can develop in any brittle rock, including igneous or metamorphic basement, limestone, or cemented sandstone. Fracturing is generally caused by tectonic processes, such as folding and faulting. Once initiated, fractures may be enlarged by solution or infilled by cementation. Fracturing may not increase the porosity of a rock very much, but it may cause a dramatic increase in permeability. Thus if a well has drilled into a petroleum reservoir of low permeability, it is common to fracture the rock adjacent to the borehole by explosive charges.

The last main type of porosity to consider is the intercrystalline variety. This is the type of pore system that is typical of dolomite reservoirs. Dolomite is a mixed calcium magnesium carbonate ($CaMg[CO_3]_2$). Some dolomite forms at the surface of the earth. This is generally fine grained, and, though porous, generally tight. In the subsurface, however, dolomite may form by the replacement of limestone. The reaction is a reversible one:

$$2CaCO_3 + Mg^2 = CaMg(CO_3)_2 + Ca^2$$
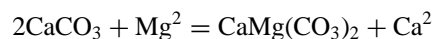
When calcite is replaced by dolomite, there is an overall volume shrinkage of as much as 13%. This results therefore in a considerable increase in porosity if rock bulk volume remains unchanged. The resultant dolomite rock commonly has a friable saccharoidal appearance. These secondary dolomites may thus serve as very effective petroleum reservoirs because they exhibit both porosity and permeability.

Figure 4 shows the relationship between porosity and permeability for some of the different pore systems that have just been described. This shows the very complex relationship that exists between the texture of a rock and its reservoir potential. A major part of petroleum geology is concerned with understanding the depositional environments of sediments, so as to predict their distribution in the subsurface. Similar interest is shown in the postdepositional geochemical reactions that take place between sediments and migrating pore fluids. These cause sediments to be cemented and turned into solid rock, thus destroying primary porosity. But selective leaching can subsequently give rise to secondary solution pore systems of erratic distribution.

**TABLE I   A Classification of Reservoir Porosity**

Primary or syndepositional porosity
    Intergranular or interparticle (characteristic of sandstones)
    Intragranular or intraparticle (occur within fossils)
Secondary or postdepositional porosity
    Solution porosity (common in limestones)
    Moldic (fabric selective)
    Vuggy–cavernous (cross-cuts rock fabric)
    Intercrystalline (characteristic of dolomites)
    Fractures (develop in any brittle rock, sedimentary, igneous, or metamorphic)

**FIGURE 3**  Photomicrograph of a thin section through a sandstone petroleum reservoir showing primary intergranular porosity between the quartz framework grains. (Simpson Group (Ordovician) Arbuckle Mountains, Oklahoma.) Width of photograph is 3 mm.

## B. Cap Rocks or Seals

The *third essential requirement for a commercial accumulation of petroleum* is a cap rock or seal. This is a sedimentary stratum that immediately overlies the reservoir and inhibits further upward movement. A cap rock need have only one property: It must be impermeable. It can have porosity, and may indeed even contain petroleum, but it must not permit fluid to move through it. Theoretically any impermeable rock may serve as a seal. In practice it is the shales and evaporites that provide most examples. Shales are probably the commonest, but evaporites are the more effective. We saw earlier how mud is compacted during burial into mudstone, or shale. These rocks are commonly porous, but because of the narrow diameter of the pore throats, they have negligible permeability. Thus shales generally make excellent seals to stop petroleum migration. When strata are folded or faulted, however, brittle shales may fracture. As described earlier, fractures enhance permeability most dramatically. In such instances, petroleum may leak from an underlying reservoir and ultimately escape to the surface of the earth.

The evaporites are a group of sedimentary rocks that were originally thought to have formed from the evaporation of sea water. They are composed of crystalline layers of halite, or rock salt (NaCl), anhydrite ($CaSO_4$), carnal-lite (KCl), and other minerals. The evaporites exhibit a number of anomalous physical properties. Unlike most sedimentary rocks, when they are subjected to stress they do not respond by fracturing, but by plastic flow. Thus, a petroleum saturated reservoir may undergo all sorts of structural deformation and may even be fractured, but an overlying evaporite may still provide an effective unfractured impermeable seal.

## VI. PETROLEUM TRAPS

### A. Terminology and Classification

The *last major requirement for a commercial accumulation of petroleum* is the existence of a trap. Source rock, reservoir, and seal must be arranged in such a geometry that migrating petroleum is trapped in some configuration of strata and cannot escape to the surface of the earth. The simplest type of trap is an upfold of strata termed an anticline. Figure 5 illustrates such a structure and some of the terms applied to traps. A trap may contain oil, gas, or both. This will be a function of the pressure/temperature conditions in the reservoir, and will also be determined by the type of kerogen that exists in the source bed, and in its level of thermal maturation.
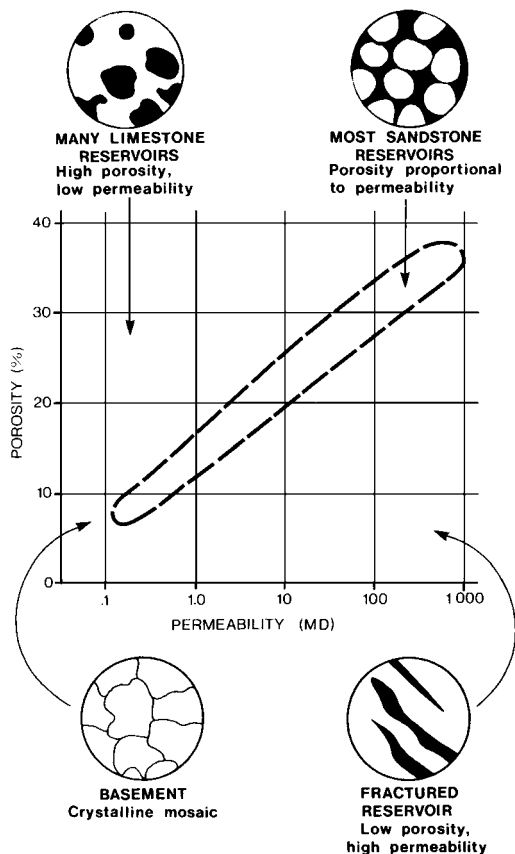
**FIGURE 4** Graph of porosity plotted against permeability showing the reservoir characteristics of the different types of pore systems. [From Selley, R. C. (2000). "Applied Sedimentology," 2nd ed., Academic Press, San Diego.]

**TABLE II   A Classification of Petroleum Traps**

Structural traps
    Anticlines
    Fault traps
Diapiric traps
    Salt diapirs
    Mud diapirs
Stratigraphic traps
    Unconformity related (truncation and onlap)
    Unassociated with unconformities (channels, bars, and reefs)
Diagenetic traps (due to solution or cementation)
Hydrodynamic traps (due to water flow)
Combination traps (due to a combination of any 2 or more
   of the above)

Gas is lighter than oil and oil lighter than water. There is thus a gravity segregation of the reservoir fluids. The top may consist of a gas zone, or cap, separated by the gas:oil contact from the oil zone. The oil zone is separated from the water zone by the oil–water contact. The reservoir beneath the oil zone is termed the bottom water, and that adjacent to the field area is termed the edge water.

The vertical interval from the crest of the reservoir to the petroleum:water contact is termed the pay zone. Not all of this interval may be productive. It may also contain impermeable strata. Thus it is usual to differentiate between the gross pay and the net effective pay. The vertical interval from the crest of a reservoir to the lowest closing contour on a trap is termed the "closure." The lowest closing contour is termed the spill plane. The nadir of the spill plane is termed the "spill point." Depending on the amount of petroleum available a trap may or may not be full to the spill point. The term "field" is applied to a petroleum-productive area. An oil field may contain several separate "pools." A pool is a petroleum accumulation with a single petroleum–water contact.

There are many different types of petroleum traps. They are commonly classified into five main categories and associated subcategories as shown in Table II. These will now be defined and then described in turn. Structural traps are those caused by tectonic forces in the earth's crust. They thus include anticlines and fault traps. Diapiric traps are due to density contrasts in sedimentary rocks, normally evaporites and overpressured clays. Stratigraphic traps are due to depositional, erosional, or diagenetic processes.



**FIGURE 5** Cross section through a simple anticlinal petroleum trap to illustrate terminology. [From Selley, R. C. (1998). "Elements of Petroleum Geology," 2nd ed., Fig. 7.1, Academic Press, San Diego.]

Hydrodynamic traps are caused by water flow. Combination traps are due to a combination of any two or more of the above processes.

Statistical studies have shown that about 75% of the world's petroleum comes from structural traps, principally anticlines, about 5% from diapirs, and about 6% from stratigraphic traps. The rest are of combined origin. It must be remembered, however, that these figures pertain to the petroleum that has been discovered to date. No one knows what the actual proportions are. These figures thus reflect our ability to find petroleum, and probably indicate our obsession with drilling anticlinal prospects.

## B. Structural Traps

Structural traps are those that are caused by earth movements such as folding or faulting. Folds may be caused by compression or tension of the earth's crust. Compression causes a shortening of the crust, deforming it into a series of upfolds, or anticlines, and downfolds or synclines. Compressive anticlines are the main types of trap for the oil that comes from the foothills of the Zagros Mountains of Iran.

When the earth's crust is subjected to tension it begins to subside, developing into a basin several hundred kilometers or more in diameter. This basin may be subsequently infilled by sediments. The floor of such a basin commonly splits into a mosaic of fault blocks. The uplifted blocks are termed "horsts" and the downdropped blocks "grabens." As the basin floor is buried by sediments, the strata are draped over the horsts. Subsequent compaction enhances the vertical closure. As the topography is mantled by successive layers of sediment, the closure of the anticlines diminishes upward. These anticlines are therefore different in origin from those due to compression. They are commonly termed drape or compactional anticlines. The Forties field of the North Sea is a classic example of a compactional anticlinal trap.

*Fault traps* are of several varieties, but in all cases it is essential for the fault to be impervious and sealing. This is by no means always the case, and many faults are open conduits to fluid movement. It is seldom possible to establish whether a fault is open or closed without drilling to test. Tension of the crust generates "*normal*" *faults* across which part of the sedimentary section is absent (Fig. 6B). The Fahud field of Oman is a good example of a normal fault petroleum trap. When a normal fault moves, the strata on the downthrown side often flop down into the space created by the fault movement. These are called "rollover" anticlines. It is often noted that strata thicken when traced from the upthrown to the downthrown side of the fault, and that the throw of the fault increases when measured downward from sediment increment to increment (Fig. 6C).
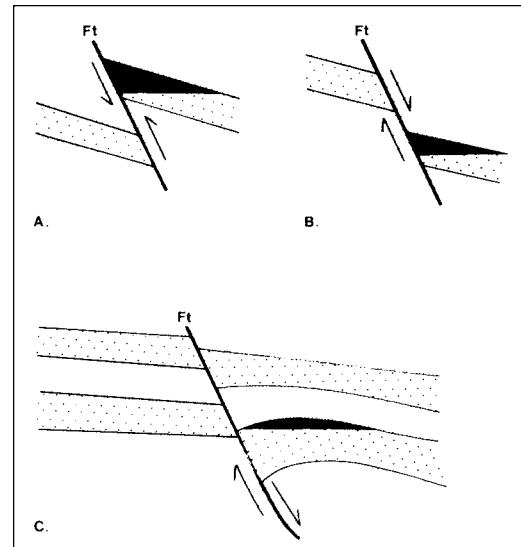


**FIGURE 6** Cross sections to illustrate the different types of fault traps. (A) Reverse fault caused by compression. (B) Normal fault caused by tension. (C) Growth fault with petroleum trapped in adjacent rollover anticline.

This demonstrates that the fault is removed repeatedly through time. Such faults are thus termed *growth faults*. Growth faults host petroleum traps, both where reservoir beds are truncated and sealed by the fault, and in rollover anticlines on the downthrown side. Examples are common in the Tertiary deltas of Nigeria and the Gulf Coast of the United States.

Compression causes *reverse faults* that give rise to the vertical repetition of strata (Fig. 6A). *Thrust faults*, as these are often called, are commonly associated with compressive anticlines. Such traps are found along the edges of mountain fronts. The Turner Valley and Winterton fields of the Rocky Mountains are examples of such traps.

Transverse movements of the crust give rise to *wrench faults*. These commonly consist of a single fault at depth that bifurcates upward into several branches. These often generate an anticline coincident with the wrench fault. These are commonly termed "flower" structures. Oil fields in such traps are found, for example, in the Los Angeles basin of California.

## C. Diapiric Traps

Diapiric traps are caused by differences in the density of sediment layers. In the ordinary way as sediments are buried they become compacted due to the overburden pressure. They lose porosity and their density increases. There are two exceptions to this general rule in which low-density sediments may be overlain by denser sediment. This situation is inherently unstable. The deeper less dense material is displaced upward as the overburden bears

down. The upward movement is localized into domes or diapirs only a few kilometers across. These structures can be foci for petroleum entrapment.

The two rocks that can generate diapirs are the evaporites and overpressured clay. The evaporites commonly have a density of about 2.03 g/cm³. This is greater than that of freshly deposited sand and clay. But as normal sediments compact, their density exceeds that of evaporites at about 800 m of burial. Below this depth, therefore, diapiric deformation may be expected to commence. Movement may be initiated by structural forces, but can also apparently occur spontaneously.

The second type of diapir is produced by overpressured clays. In some environments, especially deltas, sedimentation is so rapid that some sediments are unable to dewater before they are buried and sealed by younger detritus. With increasing burial, the situation can occur in which dense compacted sediment overlies less dense undercompacted clay. This may be squeezed upward in cylindrical diapirs analogous to those produced by salt movement. There are many ways in which petroleum may become trapped over and adjacent to a diapir (Fig. 7). This can range from simple domal traps over the crest to cylindrical fault traps in which the dome has pierced through the surrounding strata. Salt dome oil fields include the Ekofisk and associated fields of the North Sea. Clay diapirs trap petroleum in the Beaufort Sea of Arctic Canada. Both salt and mud diapirs generate petroleum traps in the Gulf of Mexico area of Texas and Louisiana.

## D. Stratigraphic Traps

As previously defined, stratigraphic traps are due to depositional, erosional, or diagenetic processes (Table II). Though the beds involved may be tilted from the horizontal, folding and faulting are absent in a pure stratigraphic
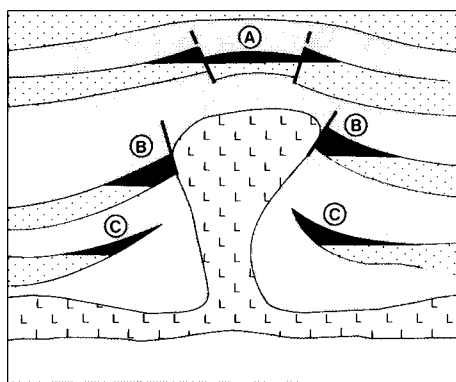


**FIGURE 7** Cross section through a salt dome to show the different types of associated petroleum traps. (A) Crestal dome. (B) Flank fault traps. (C) Flank pinchout traps.

trap (Fig. 8). Depositional, erosional, and diagenetic stratigraphic traps will now be described in turn.

The three main types of depositional stratigraphic traps are reefs, bars, and channels. Reefs are a major type of depositional stratigraphic trap. Modern coral reefs are highly porous and permeable. They may, in the fullness of time, be buried beneath organic-rich marine muds, thus becoming potential petroleum traps (Fig. 8A). There are many ancient petroliferous reefs. Today reefs thrive only in warm clear shallow seawater. The evidence of ancient reefs suggests a similar ecology, though corals have only become important reef-building organisms in recent geological time. Calcareous algae have always been important reef builders, as have bryozoa, bivalves, and several extinct groups of lime-secreting colonial organisms. Recent reefs are highly porous and permeable. But they are composed of calcium carbonate which is highly unstable in the subsurface. Acid rain water may leach lime from the upper part of a reef and reprecipitate it as a cement in the primary pore spaces deeper down. Subsequent uplift and weathering may generate secondary solution and fracture porosity. Thus ancient reefs are not always porous and permeable. Even where they possess these properties their distribution may be unrelated to the initial arrangement when the reef first formed. Despite these problems, there are many ancient reefs that serve as stratigraphic traps for petroleum. Notable examples occur in the Devonian rocks of Alberta, and the Cretaceous rocks of Mexico and the Arabian Gulf.

Beaches, barrier islands, and offshore bars are all sedimentary environments that deposit clean well-sorted sands of high porosity and permeability. Where these are enveloped in organic-rich marine sales, they may become stratigraphic traps (Fig. 8B). Notable examples occur in the Cretaceous basins of the Rocky Mountain foothills from Alberta in the north to New Mexico in the south.

There are several sedimentary environments in which channels become infilled with sand that may serve as a petroleum reservoir (Fig. 8C). These range from alluvial flood plains, through deltas, to deep-water submarine channels at the foot of the continental slope. In all of these situations, sand channels may have cut into and been overlain by impermeable muds. These muds may act as source beds and seals, generating petroleum that is then trapped within the channel sands. There are many examples of alluvial channel sand traps in the Cretaceous basins of Colorado and Wyoming and western Canada. Petroliferous deltaic distributary channel sand traps are common in the Pennsylvanian sediments of Illinois and Oklahoma. Channel and bar stratigraphic traps both have linear geometries. They are thus colloquially referred to as "shoestring" sands. It is important to note, however, that channels will tend to trend down the old depositional slope into the center of a sedimentary basin. Barrier bar
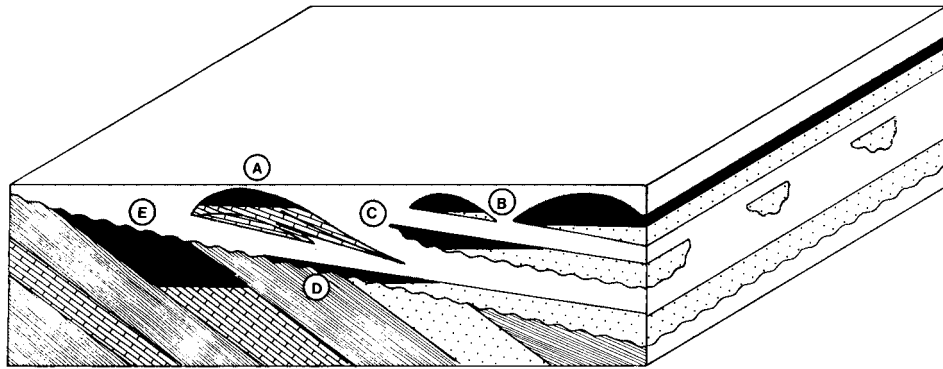
**FIGURE 8** Block diagram to illustrate the different types of stratigraphic traps. (A) Fossil reef limestone. (B) Coastal barrier-bar sand. (C) Sandstone infilled channel. (D) Onlap of sand above unconformity. (E) Truncation of limestone beneath unconformity.

and beach sands, however, will tend to be elongated at right angles to channels, being aligned parallel to the paleoshoreline.

The second main group of stratigraphic traps to consider comprises those that are associated with unconformities. An unconformity is a surface that marks a major break in the depositional history of an area. The underlying rocks may be of any type, igneous, metamorphic, or sedimentary. Where the latter is the case, the strata may have been folded or tilted before the deposition of the overlying sediments. There is often evidence of weathering and fracturing in the rocks beneath the unconformity, though subaerial exposure has not always occurred. Unconformities play an important role in petroleum entrapment for two main reasons. Weathering beneath the unconformity may generate secondary porosity and permeability in all sorts of rocks, both sedimentary and basement. These weathered zones may serve both as reservoirs and as conduits to permit petroleum migration. Second, unconformities permit the juxtaposition of source beds and reservoirs. Sometimes source overlies reservoir and sometimes the reverse occurs. Thus there are two types of stratigraphic traps associated with unconformities: onlap or pinchout traps and subcrop or truncation traps.

At its simplest, an onlap trap may be a blanket sand that pinches out up-dip (Fig. 8D). It is sealed by impermeable rocks beneath and by an onlapping shale (generally the source rock, as well as the cap). Many unconformities are old land surfaces, however, and sands may be deposited in old topographic lows. Alternating hard and soft sediments may have been weathered and eroded to form scarps, dip slopes, and strike valleys. Fluvial or shallow marine sands may have been deposited along the old valleys and sealed by marine muds. Stratigraphic traps of this type are known as the Mississippian: Pennsylvanian unconformity of Oklahoma. Alternatively, the unconformity may have been a planar land surface that was locally incised by alluvial valleys. These may have been sand filled and drowned by

an advance of the sea that deposited impermeable source beds above them. Such alluvial valley sands may thus act as stratigraphic petroleum traps.

Subcrop truncation traps can be developed in sandstones or limestones (Fig. 8E). Some sort of closure, either depositional or structural, is required to seal the reservoir off along the strike of the truncated reservoir. Some of the world's largest fields occur in truncation traps. The East Texas field and the Messla field of Libya are two such examples.

The third type of stratigraphic trap is due to diagenesis. This term encompasses the physical and chemical changes that take place in a sediment after it has been deposited. Diagenesis generally leads to a gradual destruction of porosity as the sediment is compacted due to the weight of the overlying detritus, and as primary porosity is destroyed by the precipitation of minerals from percolating groundwater. Sometimes, however, diagenesis leads to the generation of secondary porosity. This is either by leaching, giving rise to molds, vugs, and caverns, or by mineral replacement, such as dolomitization, giving rise to intercrystalline porosity. There are few petroleum traps that can be attributed solely to diagenesis. There are, however, many traps in which diagenesis has played a major role in the distribution of porosity and permeability. This is more true of carbonate than sandstone reservoirs, because calcium carbonate is chemically less stable than silica in the subsurface environment. We have already observed the effects of diagenesis on the porosity evolution of reefs and subunconformity traps.

## E. Hydrodynamic and Combination Traps

The last two types of traps to consider are those due to hydrodynamic flow, and those caused by a combination of factors. In some sedimentary basins, groundwater flows down toward the basin center. In so doing it may encounter oil migrating up toward the surface. Oil is lighter than
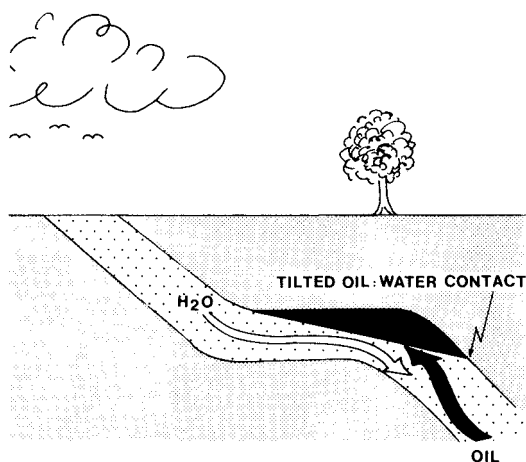
**FIGURE 9** Cross section through a hydrodynamic trap. Petroleum moving up through permeable strata is trapped in a minor flexure by downward flowing water. [From Selley, R. C. (1998). "Elements of Petroleum Geology," 2nd ed., Academic Press, San Diego.]

water, so within a given permeable bed, oil will tend to occupy the upper part and water the lower part. Strata sometime contain flexures which lack vertical closure. That is to say, a regional dip may have local horizontal wrinkles to it. In the ordinary way, oil could not be trapped beneath such benches; but where there is hydrodynamic flow, entrapment can take place in such situations. The oil:water contact will, of course, be tilted, and the petroleum will remain trapped in the flexure only as long as the downward flow of water continues (Fig. 9). Very few pure hydrodynamic traps have been discovered to date. There are, however, many oil fields around the world in which the oil:water contact is not a horizontal plane, but is in fact tilted. The lateral flow of water beneath such traps is one of several causes of tilted oil:water contacts.

This leads logically to a discussion of combination petroleum traps. These were previously defined as those due to a combination of two or more trap-forming processes: structural, diapiric, stratigraphic, or hydrodynamic. There is a multiplicity of ways in which these several processes may combine to form traps. The interplay of hydrodynamic and diagenetic processes on the more common trap types has been already noted. Probably the commonest type of combination trap is where a structural high has been growing during sedimentation. The uplift itself may have been initiated by a fold, fault, or diapir. Such highs are likely to be subject to erosion that can cause unconformity-related traps. These will include crestal subunconformity truncations as well as depositional pinchouts around the edges. Carbonate reefs may also be located on syndepositional structural highs. Large fields that are formed by a combination of structural and stratigraphic processes include the Brent and associated fields

of the North Sea, and the Prudhoe Bay field on the North Slope of Alaska. The permutations of ways in which combination traps may form are infinite.

## VII. CONCLUSIONS: EXPLORATION METHODS

The preceding account has described the fundamental facts and concepts concerning the generation, migration, and entrapment of petroleum in the earth's crust. It does not tell you how to find and oil field. In the old days, petroleum was found by wandering about with a naked light, a spirit of optimism, and a sense of adventure.

The first exploration method involved no more than looking for an oil seep and collecting the gunck in a bucket, as in the days of Noah (Genesis 6: 13–14). Then wells were sunk adjacent to petroleum seeps, as in China and Japan 2 millenia ago. The first application of science to petroleum exploration came toward the end of the nineteenth century with the realization that petroleum was commonly trapped underground in anticlinal folds. Then the main exploration tool was to send geologists out and about mapping anticlines. Wells were then drilled on the anticlinal crests.

Nowadays, however, geophysics provides the main exploration tool. Gravity and magnetic surveys define the overall architecture of sedimentary basins. Seismic surveys can map traps, not only anticlines, but also channels, barrier bars, and reefs. Seismic surveys can now be used to image petroleum:water contacts within a reservoir. If seismic surveys are shot at regular intervals during the productive life of a field, it is even possible to image the movement of gas:oil and oil:water contacts as the field depletes.

## SEE ALSO THE FOLLOWING ARTICLES

Coal Geology • Evaporites • Geology, Earthquake • Petroleum Refining • Sedimentary Petrology • Stratigraphy

## BIBLIOGRAPHY

Hunt, J. M. (1996). "Petroleum Geochemistry and Geology," 2nd ed., Freeman, New York.
Selley, R. C. (1998). "Elements of Petroleum Geology," 2nd ed., Academic Press, San Diego.
Selley, R. C. (2000). "Applied Sedimentology," 2nd ed., Academic Press, San Diego.
Stoneley, R. (1995). "An Introduction to Petroleum Exploration for Nongeologists," Blackwell, Oxford.
Tissot, B. P., and Welte, D. H. (1984). "Petroleum Formation and Occurrence," 2nd ed., Springer-Verlag, Berlin.
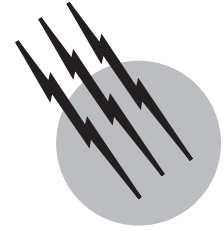
# Renewable Energy from Biomass

**Martin Kaltschmitt**
**Daniela Thrän**

*Institute for Energy and Environment, Leipzig*

**Kirk R. Smith**

*Center for Occupational and Environmental Health,*
*University of California, Berkeley*

## GLOSSARY

**Bio-chemical conversion** This involves biological processes. The most important to date are alcohol production from biomass containing sugar, starch and/or celluloses and biogas production from organic waste material (like animal or human waste). Both technologies are state of the art and widely used for fuel making in developing as well as in industrialized countries. Additionally heat can be produced during composting of biomass waste, although this heat is difficult to tap economically.

**Biofuel** Biomass used as fuel. Sometimes used to refer to processed liquid, solid, or gaseous fuels derived orginally from biomass.

**Biomass** This term is used somewhat more narrowly than its common use in ecology, where it refers to all living matter. Biomass refers here to organic material produced initially by photosynthesis with or without human support. In practice, this includes:

• Recently living plant material*

*Although dried fish and other animal remains have been used occasionally as fuels, they do not make up a significant portion of energy use and are not discussed here.

- Agricultural residues including animal manure and crop wastes
- All material made directly or as a by-product during biomass processing (e.g., residues from food, feed, fiber, lumber, and organic fertilizer production, slaughterhouse wastes, black liqueur, paper residues, organic fraction of municipal solid waste, sewage sludge, plant oil, and alcohol).

We do not regard peat as biomass, although it is so defined in some Scandinavian countries.

**Herbaceous biomass** refers to all types of biomass with a grassy nature, such as cereal and hay. Such biomass is provided mainly by agriculture. Compared to wood herbaceous biomass it is more difficult to burn such types of biomass in an environmentally sound way due to a higher amount of trace elements like chlorine and a lower ash melting point.

**Lower heating value (LHV)** The LHV (net calorific value) characterizes the energy released during complete oxidation of dry biomass without counting the latent heat of the water produced. If the water vapor in the flue gas produced from fuel hydrogen during oxidization is condensed within the conversion device, additional energy is released. The higher heating value (HHV—gross calorific value) includes this energy as well.

**Physico-chemical conversion** This may be used to produce liquid fuel by physical (e.g., pressing) or chemical (e.g., esterification) means. The only important process with a certain market potential so far is vegetable oil production from oil seed and esterification of this vegetable oil to fatty acid methyl ester as a substitute for diesel fuel. This technology is used to some extent in Europe.

**Thermo-chemical conversion** This is used to convert biomass into a solid, liquid, or gaseous fuel by heating under various conditions with some access to air. Gasification, pyrolysis, and charcoal production are the main examples. Of these, only charcoal production is widely used to date, but gasification in connection with electricity production seems to be promising in both developed and developing countries. Biomass gasification also shows advantages in developing countries for production of process heat in small-scale industrial applications such as food drying, silk manufacture, and cremation. Pyrolysis for production of liquid fuel useable in engines is a future option.

**Woody biomass** This can be produced by forestry (i.e., firewood from natural or intensively managed forests), agriculture (i.e., short-rotation plantations of fast-growing trees such as poplar or willow), or as residues from wood production for nonfuel purposes such as lumber. Compared to herbaceous biomass, it is a more promising source of energy. Besides the fact that the content of trace elements in wood is lower compared with herbaceous biomass, the ash melting point is also usually higher. Therefore from an environmental point of view wood is more promising as a solid biofuel compared to straw or hay.

**ENERGY** from biomass currently contributes 10–12% of gross worldwide energy. Due to geographical, economic, and climatic differences, the share of biomass energy in relation to total consumption differs widely among different countries, ranging from less than 1% in some industrialized countries like the United Kingdom and The Netherlands to significantly more than 50% in some developing countries in Africa and Asia. It is by far the most important renewable energy source, being significantly larger in energetic terms than the second largest, hydropower. The energy from these, the oldest fuels utilized by humanity, is much larger in absolute terms than the energy from the newest, nuclear fuels (International Energy Agency, 2000). In terms of numbers of people, biofuels dominate the world picture for it is probably true to say that most of the people of the world still rely on biomass fuels for most of their energy, a situation that has not changed since the mastery of fire a hundred thousand years ago. This is due to the relatively low costs and easy access of biofuels, such as wood, dung, and crop residues, in the poor rural areas of developing countries where about half of humanity still lives (Smith, 1987). In addition, however, biofuels such as wood chips and biogas commonly play an important role in many developed regions where their cost and access compare favorably with alternatives. The potential size of total biofuel resources is considerable throughout the world. Depending on how utilized, biofuels also have the advantage of being relatively environmentally sound compared to many other sources of energy, a characteristic that has led to a resurgence in interest throughout the world.

Cooking with biofuel is extremely rare today in rich countries although there is significant use of biofuel for space heating in regions with easy access to forests. Specific space heat demand is generally decreasing because of improved insulation of old and new buildings due to regulations and other incentives. Average living area per citizen in most industrialized countries is still increasing, however, while population itself is not growing significantly. The net result is that space heat demand is likely to remain more or less stable in industrialized countries for the next decade. The household market in developing countries is still increasing due to increasing population and economic growth. The rapid urbanization common in

many developing countries acts as a countervailing trend, however, because urban households tend to rely less on biofuels than do rural households. The net result is probably similar to that in developed countries, i.e., overall consumption of household biofuels is relatively stable, although in this case with large regional variations. China may well be an important exception because there seems to be an ongoing trend to substitute coal for biomass in rural household applications.

The demand for electric energy is increasing in industrialized as well as in developing countries. Therefore, electricity production from biomass is often seen as one of the most important future markets for biomass worldwide. The same is also true, in principle, for the production of transportation fuels from biomass. The need for mobility is still increasing worldwide, particularly in the developing countries of Asia. Therefore finding ways to provide cheap and environmentally more benign fuels based on renewable energy in general and biomass in particular is an extremely active endeavor. To date, however, only Brazil and the United States have processed significant amounts of biomass into liquid fuel (ethanol) for transport.

More broadly, expanded use of biomass in clean applications offers a way toward more sustainable energy systems in all countries, an issue of growing international concern (Kaltschmitt and Reinhardt, 1997 and Kaltschmitt et al., 1997). Because many renewable energy sources offer the potential to provide useful energy with reduced emissions of greenhouse gases and be additionally environmental advantageous compared to fossil fuel energy, their attractiveness has been increasing during the last decade. In particular, biomass is considered environmentally and climatically sound because

- If operated with high combustion efficiency and with renewable harvesting, biomass fuel cycles are greenhouse neutral, i.e., the carbon is completely recycled and there is no net increase of greenhouse gases in the atmosphere.
- Biomass combustion normally has low sulfur and nitrogen emissions with consequent small contributions to secondary particle formation and acid precipitation downwind.
- Biomass has few other intrinsic contaminants that can contaminate the environment such as the toxic elements, mercury, lead, arsenic, fluorine, etc., found in some fossil fuels.
- In many circumstances, the ash produced during combustion can be usefully recycled back onto the land from which the biomass has been grown.

Among the various possibilities for exploiting solar energy, biomass shares with a hydropower reservoir the char-

acteristic of being a form of "stored" solar energy. This offers significant advantages over wind or direct solar energy, which need to be linked with expensive storage systems to be available reliably throughout the day, month, and year. In addition, being a chemically based fuel, the energy density of biomass is fairly high, although not as high as that of most fossil fuels.

Biofuel is rarely directly affected by energy crises, being an indigenous energy carrier (i.e., biomass is, in most cases, produced close to the place where it is used) characterized by short supply chains with low risks of failure. Also the production and utilization of biomass is widely accepted (unlike nuclear and coal power in some countries) and offers benefits for rural areas related to employment, rural infrastructure, the conservation of cultivated areas, and hence the attractiveness of rural regions.

Most biomass used for energetic purposes is directly combusted to produce heat and/or power, but a huge variety of additional possibilities are available to provide environmentally sound heat and/or electricity as well as transportation fuels from organic material. The most important conversion routes available now or in the near future will be discussed according the framework shown in Fig. 1.

It is useful to consider two major forms of biomass fuel: unprocessed and processed. For unprocessed biofuel in which the material is used essentially in its natural form (as harvested) direct combustion usually supplies cooking, space heating, or electricity production needs, although there are also small- and large-scale industrial applications for steam raising and other processes requiring low-to-medium temperature process heat. Processed biofuels in the form of solids (mainly charcoal), liquids (mainly alcohols), or gases (mainly mixtures with methane or carbon monoxide), can be used for a wide range of applications, including transport and high-temperature industrial processes.

The purpose of biomass conversion is to provide fuels with clearly defined characteristics that meet given fuel quality standards. Such defined solid, liquid, or gaseous fuels can then be used to meet a specific supply task efficiently. To ensure that these fuel quality standards are met and these biomass-based fuels can be used with a high efficiency in conversion devices (like engines, turbines) upgrading is needed. Here, a distinction is made between thermo-chemical, physico-chemical, and bio-chemical conversion processes to accomplish this aim (see Glossary).

These upgraded biomass fuels can be used in specially adapted engines, turbines, boilers, or ovens to provide thermal and/or mechanical energy (i.e., heat and power), which in turn can be converted into electrical energy. Additionally, liquid and (potentially) gaseous fuels can
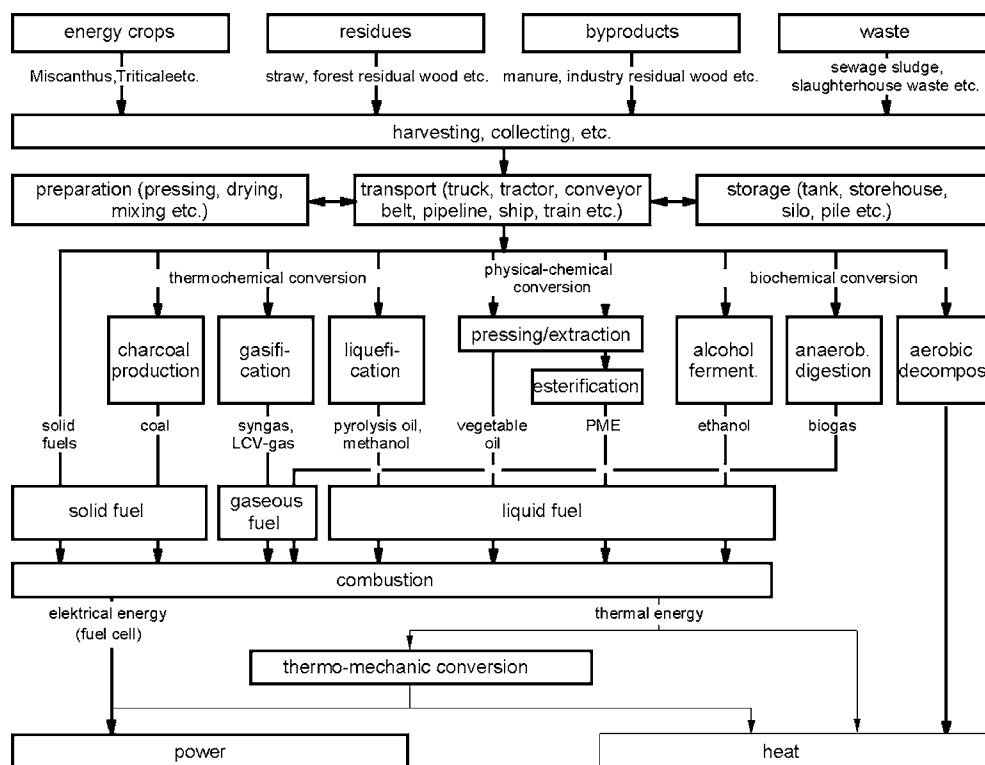
**FIGURE 1** Possibilities to provide heat and/or power as well as fuels from biomass. [From Kaltschmitt, M. *et al.* (2000). International Workshop on Integrating Biomass Energy with Agriculture, Forestry and Climate Change Policies in Europe, Impact and Opportunities from Agenda and Agriculture, Forestry and Climate Change Policies in Europe, Impact and Opportunities from Agenda 2000 and the Kyoto Protocol, Imperial College London; Kaltschmitt, M., and Wiese, A., eds. (1997). "Renewable Energies, Systemtechnology, Economics, Environmental Aspects," Springer, Berlin.]

be used directly or after treatment as transportation fuels.

Heat production and electricity generation are the most important uses worldwide for biomass fuel so far. Direct combustion devices are widely distributed with thermal capacities ranging from a few kW in household stoves up to heating plants with several tens of MW. The conversion efficiencies vary from 8 to 18% for simple stoves up to approximately 90% and above for modern heating units with high-end technology. Electricity production has been based mainly on the conventional steam cycle with efficiencies around 30% and capacities of several 100 kW and above.

The production of transportation fuels is of less importance on a worldwide scale but has been important in some countries. In Brazil, for example, production of ethanol from sugar provided nearly 14 billion liters per year for automobiles during the mid-1990s. In the United States, ethanol from corn has provided as much as about 4 billion liters per year, although production depends on financial subsidies. There is potential for increased demand as an agent to reduce air pollution emissions from vehicle fuel, however (Goldenberg, J. *et al.*, 2000).

# I. BIOMASS SOURCES

## A. Terms, Definition, and Description

Biomass consists primarily of the elements carbon (C), hydrogen (H), and oxygen (O) (Table I). Additionally, significant amounts of trace elements can be found in some types of biomass. For example, straw can contain fairly high amounts of chlorine (Cl) and/or silicon (Si) and rapeseed shows a relatively high amount of nitrogen (N). These trace elements can sometimes cause problems. For example, during combustion Cl is responsible for high temperature corrosion in boilers, Si can lead to boiler slagging, and nitrogen can cause high $NO_x$ emissions.

In most practical applications the energy content of biofuel is best described by the LHV (see Glossary). The LHV is greatly influenced by the water content of the biomass as well as fuel hydrogen content. The actual LHV of biomass containing a known percentage of water can be calculated from the LHV of the absolute dry biomass, which is available from the literature. In Eq. (1) $H_{u(w)}$ describes the LHV (in MJ/kg) of the biomass at a specific water content, $H_{u(wf)}$ the LHV of the fully dry biomass, and $w$ the water

**TABLE I** Energy Content and Concentrations of Some Elements in Untreated Biomass Compared with Coal

| Type of biomass | Heating value[a] in MJ/kg | Ash content in % | Volatile compounds in % | C | H | O | N | K | Ca | Mg | P | S | Cl |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | in % of dry substance | | | | | |
| Sprucewood (with bark) | 18.8 | 0.6 | 82.9 | 49.8 | 6.3 | 43.2 | 0.13 | 0.13 | 0.70 | 0.08 | 0.03 | 0.015 | 0.005 |
| Beech-wood (with bark) | 18.4 | 0.5 | 84.0 | 47.9 | 6.2 | 45.2 | 0.22 | 0.15 | 0.29 | 0.04 | 0.04 | 0.015 | 0.006 |
| Poplar wood (short rotation) | 18.5 | 1.8 | 81.2 | 47.5 | 6.2 | 44.1 | 0.42 | 0.35 | 0.51 | 0.05 | 0.10 | 0.031 | 0.004 |
| Willow wood (short rotation) | 18.4 | 2.0 | 80.3 | 47.1 | 6.1 | 44.3 | 0.54 | 0.26 | 0.68 | 0.05 | 0.09 | 0.045 | 0.004 |
| Bark (softwood) | 19.2 | 3.8 | 77.2 | 51.4 | 5.7 | 38.7 | 0.48 | 0.24 | 1.27 | 0.14 | 0.05 | 0.085 | 0.019 |
| Rye straw | 17.4 | 4.8 | 76.4 | 46.6 | 6.0 | 42.1 | 0.55 | 1.68 | 0.36 | 0.06 | 0.15 | 0.085 | 0.40 |
| Wheat straw | 17.2 | 5.7 | 77.0 | 45.6 | 5.8 | 42.4 | 0.48 | 1.01 | 0.31 | 0.10 | 0.10 | 0.082 | 0.19 |
| Triticale straw | 17.1 | 5.9 | 75.2 | 43.9 | 5.9 | 43.8 | 0.42 | 1.05 | 0.31 | 0.05 | 0.08 | 0.056 | 0.27 |
| Barley straw | 17.5 | 4.8 | 77.3 | 47.5 | 5.8 | 41.4 | 0.46 | 1.38 | 0.49 | 0.07 | 0.21 | 0.089 | 0.40 |
| Rape straw | 17.1 | 6.2 | 75.8 | 47.1 | 5.9 | 40.0 | 0.84 | 0.79 | 1.70 | 0.22 | 0.13 | 0.27 | 0.47 |
| Corn straw | 17.7 | 6.7 | 76.8 | 45.7 | 5.3 | 41.7 | 0.65 | | | | | 0.12 | 0.35 |
| Sunflower straw | 15.8 | 12.2 | 72.7 | 42.5 | 5.1 | 39.1 | 1.11 | 5.00 | 1.90 | 0.21 | 0.20 | 0.15 | 0.81 |
| Hemp straw | 17.0 | 4.8 | 81.4 | 46.1 | 5.9 | 42.5 | 0.74 | 1.54 | 1.34 | 0.20 | 0.25 | 0.10 | 0.20 |
| Rice straw | 12.0 | 4.4 | | | | | | | | | | | |
|   Husk | 14.0 | 19.0 | | | | | | | | | | | |
| Rye whole crop | 17.7 | 4.2 | 79.1 | 48.0 | 5.8 | 40.9 | 1.14 | 1.11 | | 0.07 | 0.28 | 0.11 | 0.34 |
| Wheat whole crop | 17.1 | 4.1 | 77.6 | 45.2 | 6.4 | 42.9 | 1.41 | 0.71 | 0.21 | 0.12 | 0.24 | 0.12 | 0.09 |
| Triticale whole crop | 17.0 | 4.4 | 78.2 | 44.0 | 6.0 | 44.6 | 1.08 | 0.90 | 0.19 | 0.09 | 0.22 | 0.18 | 0.14 |
| Rye grain | 17.1 | 2.0 | 80.9 | 45.7 | 6.4 | 44.0 | 1.91 | 0.66 | | 0.17 | 0.49 | 0.11 | 0.16 |
| Wheat grain | 17.0 | 2.7 | 80.0 | 43.6 | 6.5 | 44.9 | 2.28 | 0.46 | 0.05 | 0.13 | 0.39 | 0.12 | 0.04 |
| Triticale grain | 16.9 | 2.1 | 81.0 | 43.5 | 6.4 | 46.4 | 1.68 | 0.62 | 0.06 | 0.10 | 0.35 | 0.11 | 0.07 |
| Rape grain | 26.5 | 4.6 | 85.2 | 60.5 | 7.2 | 23.8 | 3.94 | | | | | 0.10 | |
| Miscanthus | 17.6 | 3.9 | 77.6 | 47.5 | 6.2 | 41.7 | 0.73 | 0.72 | 0.16 | 0.06 | 0.07 | 0.15 | 0.22 |
| Switch grass | | | | | | | | | | | | | |
| Sugar cane stalk (bagasse) | 8.0 | 4.0 | 80 | 45 | 6.0 | 35 | 0.0 | | | | | 0.0 | |
| Hay from various sources | 17.4 | 5.7 | 75.4 | 45.5 | 6.1 | 41.5 | 1.14 | 1.49 | 0.50 | 0.16 | 0.19 | 0.16 | 0.31 |
| Road side green | 14.1 | 23.1 | 61.7 | 37.1 | 5.1 | 33.2 | 1.49 | 1.30 | 2.38 | 0.63 | 0.19 | 0.19 | 0.88 |
| Hard coal | 29.7 | 8.3 | 34.7 | 72.5 | 5.6 | 11.1 | 1.3 | | | | | 0.94 | <0.13 |
| Lignite | 20.6 | 5.1 | 52.1 | 65.9 | 4.9 | 23.0 | 0.7 | | | | | 0.39 | <0.1 |

[a] LVH.

From Kaltschmitt, M., and Hartmann, H., eds. (2001). "Energy from Biomass," Springer, Berlin, Heidelberg. (In German.)

content (in%). The constant 2.44 results from the heat of evaporation of water.

$$H_{u(w)} = H_{u(wf)}(100 - w) - 2.44w/100 \qquad (1)$$

Figure 2 shows that the LHV of wood decreases from approximately 18.5 MJ/kg with increasing water content. The LHV is zero at water content of approximately 88%. Normally the water content of air-dried wood is between 12 and 20% yielding a heating value of 13–16 MJ/kg. Freshly harvested wood is characterized by a water content of about 50% or more. A low LHV is the result.

For a discussion of different conversion routes, it is useful to subdivide the biomass into three categories: woody biomass, herbaceous biomass, and other biomass (see Glossary).
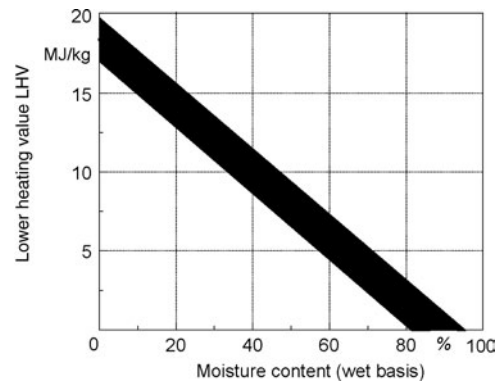


**FIGURE 2** LHV of wood depending on the water content. [From Kaltschmitt, M., and Hartmann, H., eds. (2001). "Energy from Biomass," Springer, Heidelberg, Germany. (In German.)]
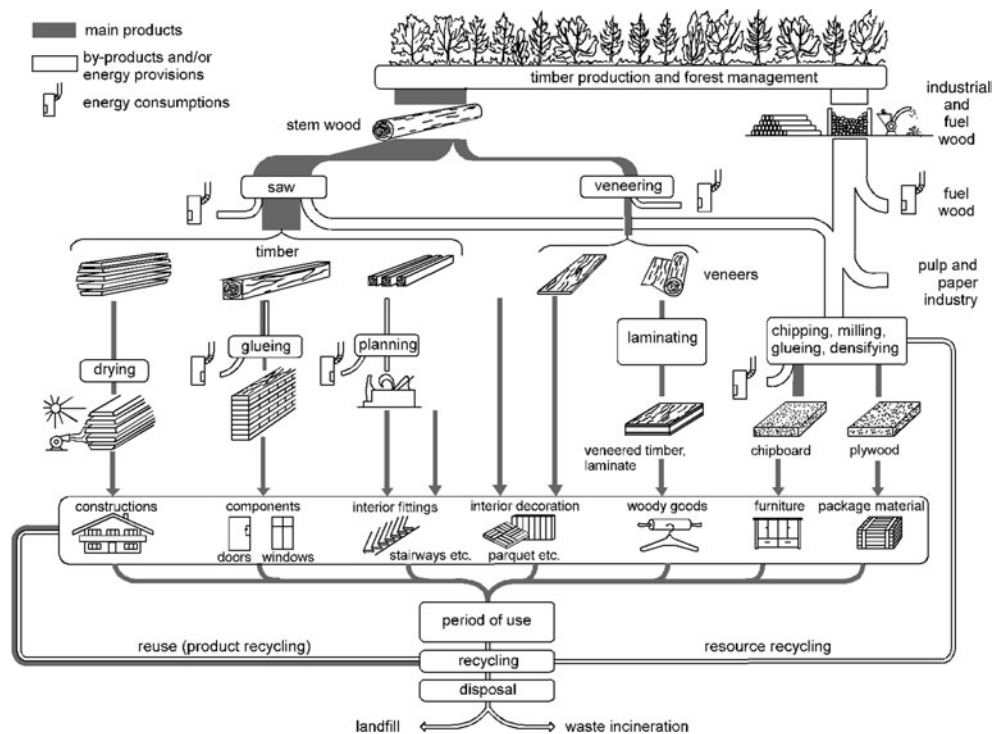
**FIGURE 3**  Material flow of wood within the overall economy. [From Kaltschmitt, M., and Hartmann, H., eds. (2001). "Energy from Biomass," Springer, Heidelberg, Germany. (In German.)]

### 1. Woody Biomass

In addition to being grown directly for fuel, wood fuel can derive from production and consumption of wood and wood products as a residue, a by-product and/or a waste material all along the overall supply chain of wood, and pulp and paper products (Rösch, C., *et al.*, 2000) (Fig. 3). For example, thinning wood and forest residues are produced as a by-product during the production of stem wood as a raw material for paper and pulp production or for cupboard and furniture production. Industrial residual wood, bark, and wood dust are by-products and/or waste materials resulting from the production of timber and the manufacturing of wood structures or products based on wood as a raw material. At the end of the product lifetime and after recycling of some wood fractions (such as old furniture or roof timbering, and chipboards), partly contaminated woody material and waste wood remain, which can be used as fuel. If the contamination with heavy metal due to coloring is too high, such materials have to be treated first.

### 2. Herbaceous Biomass

Here, a differentiation is made between energy crops grown specifically as fuels (like whole crop cereal, rape or sunflower seed, miscanthus, switch grass, or sugar cane) and by-products of other production (like straw). The lat-

ter arise primarily from the production processes for the provision of food, fiber, and fodder (e.g., the production of grain is necessarily connected with the production of straw). Additionally herbaceous biomass is also produced, for example, in landscape conservation areas, parks, cemeteries, gardens, and on roadside areas.

### 3. Other Biomass

Various other organic materials usable as energy sources are available within the overall economy, including by-products and waste materials from agriculture (e.g., animal manure), industrial food and fodder processing processes (e.g., slaughterhouse residues, vegetable waste, olive and fruit pits), food and fodder trading (e.g., spoiled food and fodder), households (e.g., the organic fraction of solid municipal waste), as well as from waste water processing (i.e., sewage sludge).

### B. Use and Potential

These different biomass sources can be used to meet energy demand in industrialized as well as in developing countries. A substantial fraction of biomass fuel is not traded in markets, particularly in developing countries, but is gathered and used within individual households or enterprises. Thus, biomass fuel use is poorly represented in official energy statistics and the data available are often

based on sample surveys of uncertain validity and timeliness. Only in the late 1990s, for example, has the International Energy Agency (IEA) included biomass fuel in its official statistics and then only in fairly aggregated forms (International Energy Agency, 2000). More and more developing countries, however, are starting to include fuel use questions on census forms and other national surveys that will help reduce these uncertainties.

Recognizing these data constraints, it seems that biomass contributes about one-third of the primary energy consumption in developing countries but varies from over 90% in less developed countries such as Uganda, Rwanda, Tanzania, and Nepal to about 45% in India, 30% in China and Brazil, and 10–15% in Mexico and South Africa. By comparison, the share of primary energy provided by biomass within industrialized countries is estimated to be only about 3%. Importantly, however, the absolute consumption per capita varies by a much smaller amount worldwide. Indeed, cross-sectional studies seem to indicate that economic development does not usually result in less overall absolute use of biomass fuel, although its fraction of total energy declines and use shifts from households to other sectors. Overall, current commercial and noncommercial biomass fuel supplies about 20–60 EJ/y worldwide. Recent IEA estimations, for example, indicate approximately 40 EJ/y (Table II).

Use of biofuel differs significantly throughout the world (Table II ). Residues concentrated at industrial processing sites (like bark and sawdust in saw mills) are currently the largest commercially used biomass source. For example, bagasse, the fiber remaining after juice extraction in sugar cane processing, often provides energy for extracting the juice and processing to pure cane sugar or alcohol at the sugar mill and, in addition, surplus bagasse is used to supply electricity for the local grid.

In the United States, some 8000 MWe of biomass-fueled electricity generating capacity are installed primarily as combined heat and power production systems. In Europe biomass is mainly used for the provision of heat and, to a minor extent, for electricity production. In rural areas biomass is used traditionally for heating purposes. Electricity production has become more important in recent years due to environmentally motivated policy measures (e.g., the Feed-in Law for electricity from renewables in Germany, the $CO_2$-tax in Sweden). In average, however, only the use of firewood in households as well as industrial application in the wood processing industry are of major importance in Europe (for example, firewood covers almost 75% of biofuel use in Europe). The use of wood and other organic material for district heating is important in only a few countries (e.g., Denmark, Finland, Austria).

In developing countries, biomass is used mainly in open combustion devices for cooking and, to a lesser extent, space heating. The poor combustion conditions in these devices results in high emissions of health-damaging substances as products of incomplete combustion and low fuel-use efficiencies. Additionally, biomass is sometimes used in a nonsustainable way in that more is removed from the natural biomass resources (e.g., forests, grasslands) than is regrown. Therefore, the available biomass

**TABLE II   Biomass Energy Potentials and Current Use in Different Regions**

| | North America | Latin America | Asia[b] | Africa | Europe | Middle East | Former Russia | World |
|---|---|---|---|---|---|---|---|---|
| Biomass potentials in EJ/a | | | | | | | | |
| Wood | 12.8 | 5.9 | 7.7 | 5.4 | 4.0 | 0.4 | 5.4 | 41.6 |
| Straw | 2.2 | 1.7 | 9.9 | 0.9 | 1.6 | 0.2 | 0.7 | 17.2 |
| Dung | 0.8 | 1.8 | 2.7 | 1.2 | 0.7 | 0.1 | 0.3 | 7.6 |
| (Biogas)[c] | (0.3) | (0.6) | (0.9) | (0.4) | (0.3) | (0.0) | (0.1) | (2.6) |
| Energy crops | 4.1 | 12.1 | 1.1 | 13.9 | 2.6 | 0.0 | 3.6 | 37.4 |
| Sum | 19.9 | 21.5 | 21.4 | 21.4 | 8.9 | 0.7 | 10.0 | 103.8 |
| Use in EJ/a | 3.1 | 2.6 | 23.2 | 8.3 | 2.0 | 0.0 | 0.5 | 39.7 |
| FPEC and HE[a] in EJ/a | 104.3 | 15.1 | 96.8 | 11.0 | 74.8 | 15.4 | 37.5 | 354.9 |
| Shares in % | | | | | | | | |
| Use/potential | 16 | 12 | 108 | 39 | 22 | 7 | 5 | 38 |
| Use/PFEC and HE[a] | 3 | 17 | 24 | 76 | 3 | <1 | 1 | 11 |
| Potential/PFEC and HE[a] | 19 | 143 | 22 | 195 | 12 | 5 | 27 | 29 |

[a] Primary fossil energy consumption (PFEC, including nuclear power) and hydroelectricity (HE).

[b] In Asia, current use exceeds available potential (i.e., biomass is used nonsustainably).

[c] Potential of biogas production from the available animal dung.

From Kaltschmitt, M., and Hartmann, H., eds. (2001). "Energy from Biomass," Springer, Berlin, Heidelberg, Germany. (In German.)

resources are degraded, with negative environmental impacts on land, soil, and water resources. Research has shown, however, that fuel demand is rarely the sole or even primary driver of deforestation in developing countries, but it can be a contributing factor in some areas, particularly where semi-legal commercial forms of forest exploitation such as charcoal production are undertaken.

The total sustainable worldwide biomass energy potential seems to exceed 100 EJ/y (Table II), which is just below 30% of total global energy consumption today (Kaltschmitt, 1999 and Goldenberg *et al.*, 2000).

## II. BIOMASS PRODUCTION AND SUPPLY

Biomass feed to the conversion facility is accomplished through a supply chain. Because biomass grows in a distributed manner, supply chains are usually characterized by a significant transport step to gather sufficient biomass at one site to take advantage of economies of scale in conversion. Additionally, depending on the biomass source, processing or treatment steps, as well as storage and/or drying activities, may be connected with this transport chain.

For household and other small-scale use in developing countries, supply chains are often operated by household members using solely human and/or animal labor. This is especially so for low-quality biomass fuels, such as dung, crop residues, and brush that are typically gathered near the point of use. Wood fuel can be transported economically further and is often provided commercially to cities and towns from rural areas by trains, trucks, or buses. The charcoal fuel cycle that is important for many African cities involves charcoal made in simple kilns in remote areas that is then transported by truck, sometimes many hundreds of kilometers.

### A. Industrialized Countries

Here a differentiation can be made between woody biomass and herbaceous biomass (see Glossary).

#### 1. Residual Wood

Wood log production in conventional forests requires periodic thinning to ensure the availability of high-quality stem wood. The wood produced during such thinning can be a source of energy. Trees removed from the forest with a breast height diameter (BHD) from 8 to 20 cm are highly suited for the production of wood chips. To operate a chain like this, a variety of different technologies is applied, including chain saws, mechanical harvesters, or forwarder (large machines that move whole trees) to

move the wood to the main forest road where mobile chippers are operated.

After felling and moving the trees or wood pieces to the forest road, the wood may be stored in the forest until the end of summer to dry out by respiration and evaporation. During the following season with higher heating demand, the wood pieces are chipped at the forest road and transported to the conversion plant using trucks or tractors.

#### 2. Woody Energy Crops

Wood-chip production from relatively fast-growing poplar or willow trees is especially promising for many industrialized countries, although eucalyptus can also be attractive in some areas. For a plantation, 12,000–20,000 plants per hectare are needed. An average yield of around 10 t/(ha/y) of dry matter is typical in countries like the United Kingdom or Germany. During the first year, weed control is necessary. In the following years, only fertilizer is needed in most cases. Harvesting is typically done during winter to early spring every third year. After the biological/economic lifetime of such plantations, removing roots from the soil is necessary to make the land available again for planting annual crops.

For such short-rotation plantations, the so-called "green chain" is used in most cases in which trees are harvested according to demand at the conversion plant. Thus, the harvested wood chips are transported directly to the plant without any additional storage.

#### 3. Straw

During grain harvesting, straw is often left on the field. For transport to the conversion facility, straw can be collected from the field and pressed into bales by a tractor-driven straw press. With such locally available machinery, the loose straw material can be compressed into small or large rectangular bales, cylindrical bales, straw pellets, or compact rolls. Depending on the requirements of the conversion technology used to transform the straw into heat and/or electricity, storage of bales within a barn or on the field wrapped with plastic film is possible.

Handling of bales in and out of storage is done with equipment already available on the farm in most cases (e.g., front loader for a tractor). On demand, the straw is transported from the storage to the conversion plant with a truck or a tractor-driven trailer depending on distance.

#### 4. Herbaceous Energy Crops

A distinction is made between annual and perennial crops. The former are sown and harvested within one year (e.g., triticale) and the latter are used for a couple of years.

Annual crops are mainly cereal crops produced with available technology for the production of grain for feed

and fodder. Biomass is then harvested and pressed into bales like straw. After a storage period, it can be used in suitable combustion devices. Compared to straw the losses are higher because cereal grains can not be used fully with available technology. Also pest losses during storage are often fairly high.

Perennial crops like switch grass, miscanthus, or sugar cane are grown on plantations with a technical/economic lifetime of approximately a decade and even more. Here these crops are produced like grass on pasture for feeding cattle. The harvest for some crops is like that for conventional hay production. Alternatively, some crops are harvested with a specially developed harvester in early spring when the biomass from the last growing season is fairly dry and the soil is frozen. In most cases bales are produced and the further provision chain is comparable to the provision of straw. Sugar-rich biomass (sugar cane, sugar beet) is difficult to store because of the high intensity of biodegradiation, so that such biomass is treated seasonally without a long time storage. Like for annual crops, provision chains for perennial crops are not very much developed. Only very few chains are realized so far on a commercial basis (i.e., cultivation sugar cane for ethanol production).

## B. Developing Countries

Logs, brush, crop residues, and large-animal dung are the principal fuels used by the majority of rural households living in developing countries. They are generally harvested locally, although in some areas people must spend an average of 1–3 hours per day or more to walk to and from the areas where it is available. Being gathered directly by the users, a large fraction of this fuel does not pass through markets and is thus "noncommercial." Thus, it tends to be poorly accounted by the authorities.

By definition, agricultural wastes (dung and crop residues) are harvested renewably; i.e., the carbon released through burning is recaptured from the atmosphere in the next crop cycle. In contrast, in some areas brush and wood are sometimes harvested nonrenewably putting pressure on local forests and other biomass resources and directly contributing to greenhouse gas emissions. In general, however, it has been difficult to find examples of large-scale deforestation that is driven mainly by fuel demand. In most cases, deforestation is driven by demand for new agricultural land, clear-cutting for lumber, road-building, and other land-use changes.

Examples of deforestation significantly driven by fuel demand mostly stem from urban demand for wood or charcoal. In such cases, trucks, buses, and trains are used to supply these markets from some distance. In the case of charcoal in East Africa, for example, such "wood sheds" can extend hundreds of kilometers from large cities. Although not as prevalent as in rural areas, there is, nevertheless, significant biomass use as household fuels in developing country cities, particularly among the one-third of the urban population thought, on average, to live in informal settlements (slums or shanty towns).

## III. THERMO-CHEMICAL CONVERSION

The most widely use of biomass is the provision of heat released during combustion of solid biofuels. But combustion is only one conversion process among others based on a heat-induced chemical conversion of the organic material. Therefore within the following explanations we first discuss the basics of such thermo-chemical conversion processes. Then we focus on the technical implementation of these principles in conversion plants.

## A. Basics

During combustion, biofuels are oxidized primarily to carbon dioxide, water, and heat. Equation (2) shows the approximate process for wood, which can described as $C_nH_mO_p$.

$$C_nH_mO_p + \left(n + \frac{m}{4} - \frac{p}{2}\right)O_2 \rightarrow nCO_2 + \frac{m}{2}H_2O + \text{heat}$$

(2)
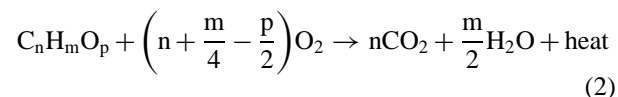
### 1. Basic Processes

Depending on conditions, biomass combustion consists of the steps: Heating and drying, pyrolytic decomposition, gasification, and oxidation (Fig. 4).

*a. Heating and drying.* Before any chemical reaction of the organic material can take place, physically bound water is evaporated at temperatures up to 200°C. The water may leave the reaction region with the flue gas or become involved in reduction to be converted to $H_2$. This process requires energy (i.e., is endothermic).

*b. Pyrolytic decomposition.* Within this step the biomass macromolecules are decomposed by heat in absence of oxygen and the volatile compounds are driven out of the biomass material due to thermal effects. At about 200°C and more, such pyrolysis processes start to take place and the hemicellulose and lignine which are parts of solid biomass-like wood are decomposed. Starting at about 300°C is decomposition of cellulose, which is another component of wood. The volatile components of the biofuel are vaporized at temperatures below 600°C
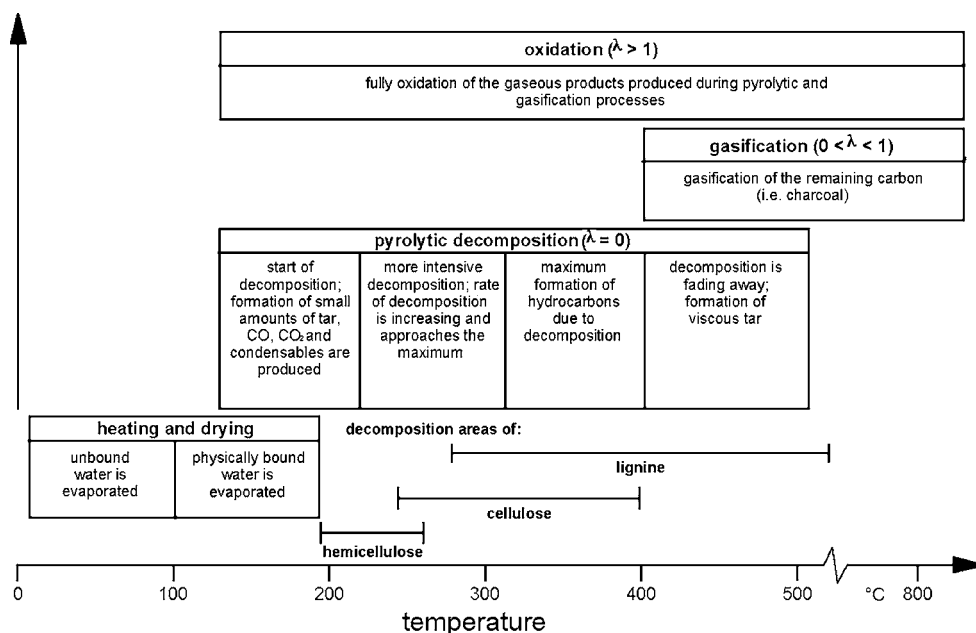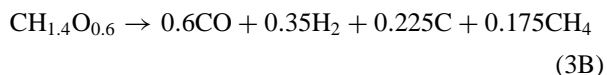
**FIGURE 4** Example of thermo-chemical conversion of wood. [From Kaltschmitt, M., and Hartmann, H. (2001). "Energy from Biomass," Springer, Berlin, Heidleberg, Germany. (In German.)]

by a set of complex decomposition reactions not fully known and understood. As a result of such decomposition processes, the following components are formed:

- Volatile compounds, such as hydrogen ($H_2$), carbon monoxide (CO), methane ($CH_4$), carbon dioxide ($CO_2$), nitrogen ($N_2$) and steam ($H_2O$)
- A carbon-rich solid fraction (char)
- Low molecular weight organic compounds and high molecular weight (condensable) compounds (i.e., liquid products)

Fixed carbon (char) and ash are the by-products that are not vaporized.

The pyrolytic decomposition of biomass (represented for wood by the formula $CH_{1.4}O_{0.6}$) can be characterized by the following equations:

$$CH_{1.4}O_{0.6} \rightarrow 0.6CO + 0.7H_2 + 0.4C \tag{3A}$$

$$CH_{1.4}O_{0.6} \rightarrow 0.6CO + 0.35H_2 + 0.225C + 0.175CH_4 \tag{3B}$$

The reactions taking place during a specific conversion process are dependent on process conditions. Therefore, a broad variety of different processes take place, often concurrently in different parts of the reaction region.

*c. Gasification.* It is obvious, that solid carbon necessarily remains as a product of the pyrolytic decomposition [Eqs. (3A) and (3B)]. In order to convert this solid

carbon into a gas, an oxygen-containing agent such as air or pure oxygen is required. At a temperature range between 700 to 1500°C the solid carbon as well as gaseous products (CO, $H_2$, $CH_4$) are oxidized as follows:

| Partial oxidation of solid carbon | $C + O_2 \rightarrow CO$ | $\Delta H = -111$ kJ/mol | (A3) |
| Complete oxidation of solid carbon | $C + O_2 \rightarrow CO_2$ | $\Delta H = -197$ kJ/mol | (A4) |
| Oxidation of carbon monoxide | $CO + O_2 \rightarrow CO_2$ | $\Delta H = -283$ kJ/mol | (A5) |
| Oxidation of hydrogen | $H_2 + O_2 \rightarrow H_2O$ | $\Delta H = -242$ kJ/mol | (A6) |
| Oxidation of methane | $CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$ | $\Delta H = -802$ kJ/mol | (A7) |
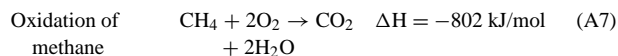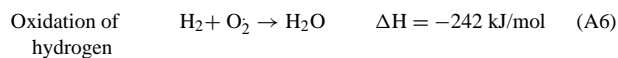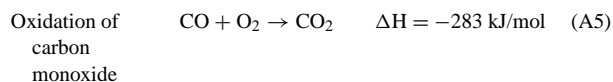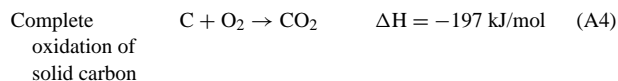
Additionally, balance reactions take place. The most important are the reduction of $CO_2$ to CO (Boudouard reaction) and of $H_2O$ to $H_2$ (heterogeneous water gas reaction). Simultaneously, carbon can be gasified to $CH_4$.

| Boudouard reaction | $C + CO_2 \rightarrow 2CO$ | $\Delta H = 173$ kJ/mol | (A8) |
| Heterogeneous water gas reaction | $C + H_2O \rightarrow CO + H_2$ | $\Delta H = 131$ kJ/mol | (A9) |
| Heterogeneous methane production | $C + 2H_2 \rightarrow CH_4$ | $\Delta H = -87$ kJ/mol | (A10) |
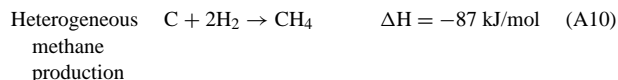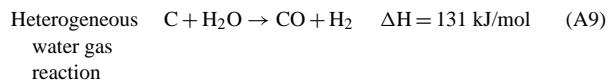
During the complicated mix of these and other reactions making up gasification, sometimes energy is needed and sometimes energy is released, depending on conditions. Additionally these reactions can also take place during the pyrolytic decomposition at higher temperatures because the biomass itself contains oxygen. Therefore the separation between these two conversion steps is not always clearly defined.

*d. Oxidation.* Within this last step of the thermo-chemical conversion, the gaseous products produced during the steps already performed are fully oxidized to carbon dioxide and water releasing energy (i.e., exothermic).

### 2. Technical Implementation

If Eq. (2) occurs in one step, we speak of full oxidation. Under these conditions the excess air ratio is 1.0 or above [the excess air ratio is defined as the ratio between the amount of oxidizing agent fed to the conversion process and the amount of oxidizing agent needed to fully oxidize all reaction products; per definition the excess ratio is 1.0 if the conversion process is realized exactly as shown in Eq. (2)].

Oxidation can also be realized in two steps in which the excess air ratio of the first step is below 1.0. Under these conditions the reaction products can be further oxidized at a second step releasing the rest of the available energy. Carbon monoxide and/or hydrocarbons are typically produced at the first step and transported to another device for full oxidation. Within such processes the procedure described above is paused, for example, after the gasification step (e.g., within a gasifier) and the oxidizing step is realized at another time and at another place (e.g., within the engine).

If in such a two-stage process the intermediate product is a liquid we call the process pyrolysis (the procedure described above is than suspended after the pyrolytic decomposition). Under these conditions the excess air ratio of the pyrolysis process is zero. Solid, liquid, and gaseous products are formed in varying amounts depending on the process conditions (e.g., temperature, heating rate, pressure, water content). If a gas is to be produced at the first step, the excess air ratio is between zero and one (Table III). The gas, which often mainly contains carbon dioxide, is called producer gas.

## B. Direct Combustion

The thermo-chemical conversion of biomass or biomass products into heat is called combustion. This heat released during the oxidation of organic material mainly into carbon dioxide and water can be used directly at the conversion plant (e.g., for cooking or space heating) or can be transported by means of a heat carrier (i.e., hot water or steam) to the place of consumption (e.g., district heating systems). The thermal energy can also be converted via a steam turbine or other combined heat and power (CHP) process into electricity and low temperature heat. Because of technical and economic limits, the temperature and pressure of steam processes based on biomass cannot approach those using fossil fuels. Therefore, only fairly low conversion rates of biomass fuel energy into electrical energy (maximum of 25–30%) are possible at present. The technology required to optimize combustion depends on the capacity, fuel consistency, water content, ash melting behavior, trace contaminants, and other factors. Where these conditions vary, for example, with changeable mixtures of biomass species in the feed, performance will inevitably suffer.

As noted above, due to the relatively high volatile content and other characteristics of biomass, spatial separation is usually provided in modern combustion devices between the fuel gasification process and the full oxidation of the producer gas into $CO_2$ and $H_2O$. The former occurs with primary air fed into the glowing fire and the latter through secondary air fed into the burning gas, preferably in an after-burning chamber. To achieve low pollutant emissions, good mixing of air and combustible gas

**TABLE III  Excess Air Ratio and Reaction Product Composition for Thermo-Chemical Conversion Processes**

| | Excess air ratio | Temperature in $^\circ$C | Pressure in bar | Products Gas | Liquid |
|---|---|---|---|---|---|
| Oxidation | $\lambda \geq 1$ | 800–1,300 | 1–30 | $H_u = 0$ | |
| Gasifcation | $0 < \lambda < 1^a$ | 700–900 | 1–30 | $H_u > 0$ | |
| Pyrolysis, coalification | $\lambda = 0$ | 350–550 | 1–30 | $H_u \geq 0$ | H |

$^a$ In most cases $0.2 < \lambda < 0.5$.

From Kaltschmitt, M., and Hartmann, H., eds. (2001). "Energy from Biomass," Springer, Berlin, Heidelberg, Germany. (In German.)

**TABLE IV   Typical Combustion Technologies and Their Characteristics**

|  | Combustion technology | Typical thermal capacity | Biofuels |
|---|---|---|---|
| Manually fed systems | Open/closed chimney | 2–15 kW | Wood logs, briquette |
|  | Single stove | 3–12 kW | Wood logs, briquette |
|  | Tiled stove | 2–15 kW | Wood logs, briquette |
|  | Pellet stove | 3–10 kW | Pellets |
|  | Wood log stove | 10–500 kW | Wood logs, briquette |
| Automatic | Gasification firing system | 20 kW–2 MW | Wood chips |
|  | Under feed system | 20 kW–2 MW | Wood chips, wood shavings and filings |
|  | Grate firing system for wood | 150 kW–15 MW | Wood, bark |
|  | Fluidized bed system | From 10 MW | Wood, bark, sewage sludge, black liqueur |
|  | Grate firing system for herbaceous biomass | 50 kW–20 MW | Bales, chipped herbaceous biomass |
|  | Blow in firing system | 200 kW–50 MW | Dust, shavings and filings |

is necessary. Additional design features assist in reducing particle emissions, including those for keeping flue gas velocities low so that ash particles are not entrained. Table IV gives an overview of combustion technologies primarily used in industrialized countries.

### 1. Household Cookstove

A typical household stove used in developing countries for cooking utilizes a semi-enclosed combustion chamber made from mud, bricks, or rocks with no flue or chimney (Baldwin, 1986). It basically operates as a single stage combustor, i.e., there is no secondary air or other system to separate gasification from the oxidization. As a result, combustion efficiency can be low because gasification products escape from the combustion zone before being fully oxidized. Typical biomass stoves in India, for example, have nominal combustion efficiencies of 80–95%, which refers to the percentage of the fuel carbon converted to carbon dioxide. The remainder is released as products of incomplete combustion (PIC), which consist mainly of carbon monoxide, hydrocarbons, and airborne soot. A large range of other products have also been identified, however, including various carbonyls and polyaromatic hydrocarbons.

A cookstove, like other combustion devices designed to produce useful heat, has two internal efficiencies that determine total efficiency ($\eta_t$). The combustion efficiency ($\eta_c$) indicates how much chemical energy in the fuel is converted to heat. The heat-transfer efficiency ($\eta_h$) indicates what fraction of the heat produced goes into the heating vessels. Thus

$$\eta_t = \eta_c \times \eta_h$$

Figure 5 shows the carbon balance for a well-functioning open wood cookstove with 18% total efficiency. Many stove/fuel combinations in common use have total efficiencies of less than 10%, however. Note that a significant amount of the energy is lost to PIC, which, in turn, consists of many health-damaging species including respirable soot particles, carbon monoxide, and hazardous organic compounds such as formaldeyhde, benzo-$\alpha$-pyrene, and benzene. In poorly ventilated conditions such as a typical village kitchen, the resulting concentrations of pollutants often greatly exceed levels required to protect health. Because of the widespread use of such simple stoves in developing countries, it has been estimated that more than a million premature deaths due to associated respiratory and other diseases are caused annually worldwide from exposures to smoke from such stoves, mostly in women and young children who spend most time near the fire (Smith, 2000). In addition, methane and other greenhouse gases are also a component of PIC and in inefficient stoves may actually comprise a significant warming potential. Because so much carbon is diverted to PIC, most of which have higher warming potentials than carbon dioxide, stoves with low combustion efficiencies can actually produce greenhouse gas warming even if the biomass itself is harvested renewably (Smith *et al*., 2000).

Improved biomass stoves have been promulgated in many countries, with mixed results. China has introduced some 180 million since 1980, commonly with flues and ceramic combustion chambers. Total efficiencies have apparently risen and smoke exposure fallen, although no systematic review seems to have been done since 1990. The second largest stove program, that of India, has introduced about 30 million over the same period, but has had difficulty with quality control, lifetime, and maintenance,
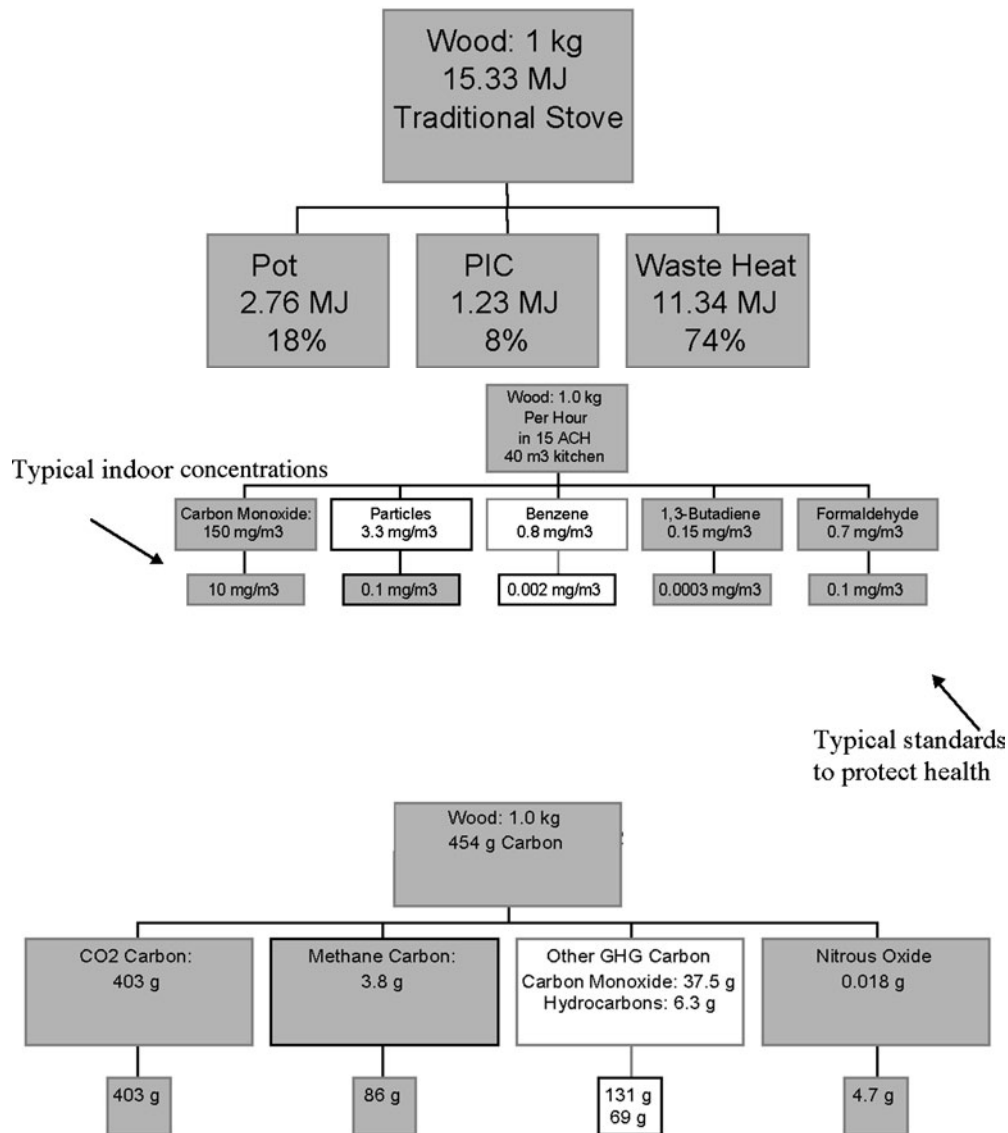
**FIGURE 5** (a) Energy flows in well-operating traditional wood-fired cookstove. PIC, traditional productions of incomplete combustion. (b) Indoor health-damaging pollutants (HDP) concentrations from typical wood-fired cookstove. Many dozen other HDPs are also known to be in woodsmoke. (c) Greenhouse warming commitment for typical wood-fired cookstove. Figures in lowest line indicate global warming commitments of each of the non-CC equivalents.

partly because of very strict cost limits. As a result only a few million are thought to be still in use. Dozens of other countries have local programs.

The basic approach to improving simple biomass stoves is to enclose the combustion zone and vent to the outside through a flue/chimney. Unfortunately, such designs can sometimes actually lower total efficiency because of the high airflow induced by natural draft in the flue reduce heat transfer to the pot. Shutting back the airflow with dampers can lead to a rise in total efficiency, but at the same time may reduce combustion efficiency sufficiently to actually increase PIC (pollutant) emissions per meal. If vented out-

doors, of course, the emissions are less damaging to health, but in crowded villages or urban slums can lead to a significant degree of "neighborhood" pollution. Longer term solutions would seem to entail improving combustion efficiency through secondary air or other means.

A promising approach, for example, is to run the stove in a downdraft mode in which all gasification products pass through the flame zone before exiting and may be clean enough to vent directly into a living space. It is difficult to maintain stable combustion conditions in these devices, however, without use of a small electric fan, which would not be practical in many parts of poor countries. With

such a fan, however, reliably high efficiencies and low emissions can be achieved (Reed and Larson, 1996) and, as electrification proceeds, such devices may come to have widespread potential.

### 2. Pellet Combuster

Automatically fed combustion devices have been developed for standardized wood pellets. On the back of such a device (Fig. 6), a container is located to store fuel for automatic operation over a few days. The fuel pellets are transported with a screw to a pipe from where the pellets fall into the combustion chamber. Primary air is fed by nozzles through the bottom of the combustion bowl. Secondary air is fed above the burning fuel via ring-shaped nozzles to ensure full conversion of the flue gas. Additional air is fed into the system via the fall pipe to prevent fire from backing into the fuel container. This well-developed air-feeding system produces low emissions, even to the level of burners fed by natural gas. The ash produced during the combustion of the wood pellets is normally removed manually. This combustion residue normally contains less than 1% carbon and can be used as a fertilizer or taken to a landfill.

The thermal efficiency of such systems can be 90% or more and the emissions comparably low when used with a standardized and well-defined biofuel. Such heaters can be

operated over a wide range of power settings, i.e., between approximately 30 and 100% of the rated capacity.

### 3. Tiled Stove

Tiled ceramic stoves are used traditionally in Europe. Due to the large mass of ceramic, a significant amount of heat can be stored. This allows the fire to be set at high power for a relatively short period (2–3 hr) and still provide steady heat into the room overnight or even up to 24 hr. This is convenient and thermally efficient, as well as producing relatively low emissions because combustion efficiency is better during high-power burning.

A variant is the tiled stove with a heating insert made from cast iron or metal sheets. Cold room air enters the gap between the metal insert and ceramic stove body and is heated before returning to the room. Unlike the high-mass ceramic stoves, such insert stoves store heat for relatively short periods, providing more flexibility for varying heat demand but less convenience.

### 4. Oven for Wood Chips

A throw-charging system is used where wood chips are transported from fuel storage with a stoker scroll and thrown with the help of a centrifugal wheel into a combustion chamber equipped with a stiff grate. Such a fully automatic feeding system allows the smaller fuel particles
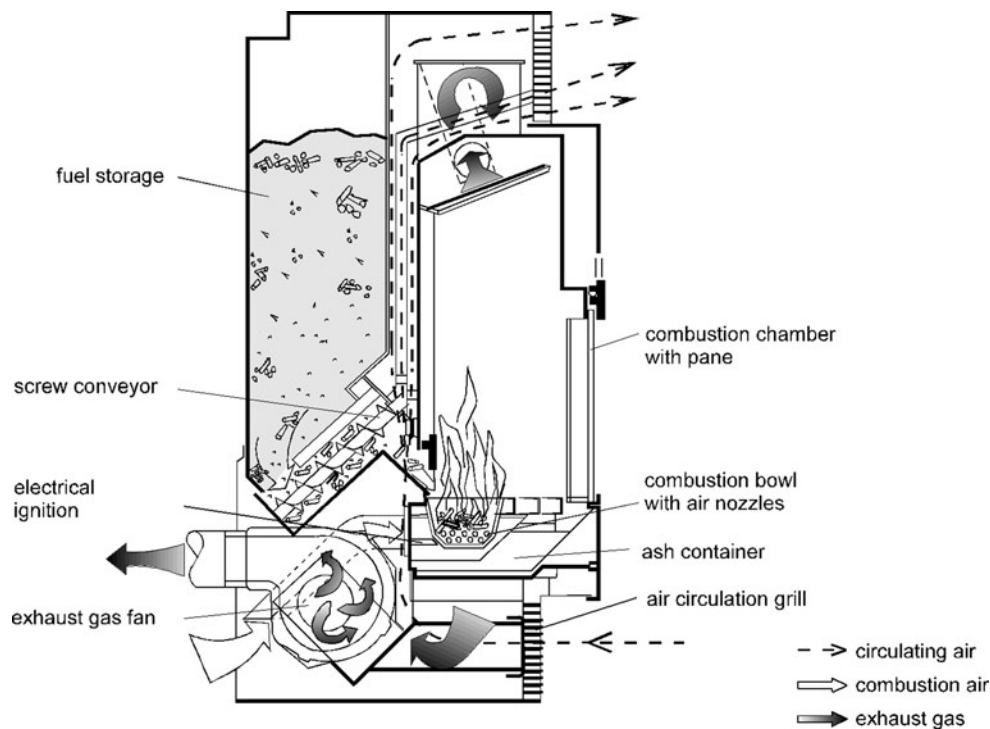


FIGURE 6   Example of a combustion device fired with wood pellets. [From Kaltschmitt, M., and Hartmann, H., eds. (2001). "Energy from Biomass," Springer, Berlin, Heidelberg, Germany. (In German.)]

to be combusted during flight while more coarse fuel particles are burned on the grate. Such a feeding system also has the advantage that the fuel is fed gently on the fire, thus reducing airborne particle emissions.

The primary air is blown with an automatic ventilator through holes in the grate into the glowing fire. The secondary air is blown at the top of the burning fuel. The air feeding system adjusts itself automatically according to information from a sensor fixed in the flue gas outlet, minimizing pollutant emissions. No additional flue gas treatment is needed and ash can be used as a fertilizer or put on a landfill site.

The heat is extracted from the flue gas via a heat exchanger located on top of the combustion device. Within the heat exchanger, the flue gas is cooled down and water is heated up—depending on the heat utilization system—either close to the boiling point (for hot water systems) or to steam.

Such combustion systems are characterized by a fully automatic operation and available on the market in the range of thermal capacities of 300 up to approximately 3500 kW. They can be used either for the heat provision for a single- or multifamily house or for industrial process heat.

## 5. Blow-in Combustion Units

Blow-in combustion units have been developed for using dusty biomass residues (e.g., sawdust, shavings, filings) from the wood processing industry (Fig. 7). The fuel particles are transported pneumatically from the fuel silo into the combustion chamber itself or—as shown in Fig. 7—into a precombustion chamber. Additionally, primary air is blown into the precombustion or combustion chamber to ensure a full conversion of the fuel into gas. Secondary air is blown into the combustion chamber to ensure a complete combustion. To guarantee that the fuel particles ignite by themselves during their flight through the combustion chamber, fuel water content has to be below 15–20%. To first ignite such a unit, a burner based on light oil or natural gas is used to heat the precombustion or combustion chamber to 450–500°C.

Such a combustion system is characterized by relatively low emissions of gaseous pollutants. Because fuel is fed in very small pieces, however, the particle content in the flue gas is relatively high. To meet particle emission regulations in most countries, the use of a flue-gas cleaning is necessary. Therefore such systems are often equipped at least with a cyclone, but often this is not sufficient. In some cases, therefore, a fabric or bag house filter or even an electrostatic dust removal system is also used. Such combustion units are characterized by fully automatic operation and low emissions at thermal capacities of several MW up to some 10 and more MW and are mainly located at places where wood dust is produced as a residue, such as the furniture industry.

## 6. Cigar Burner

Compared to wood it is much more demanding to combust herbaceous biomass in an environmentally sound way due to its poorer fuel characteristics. One approach is
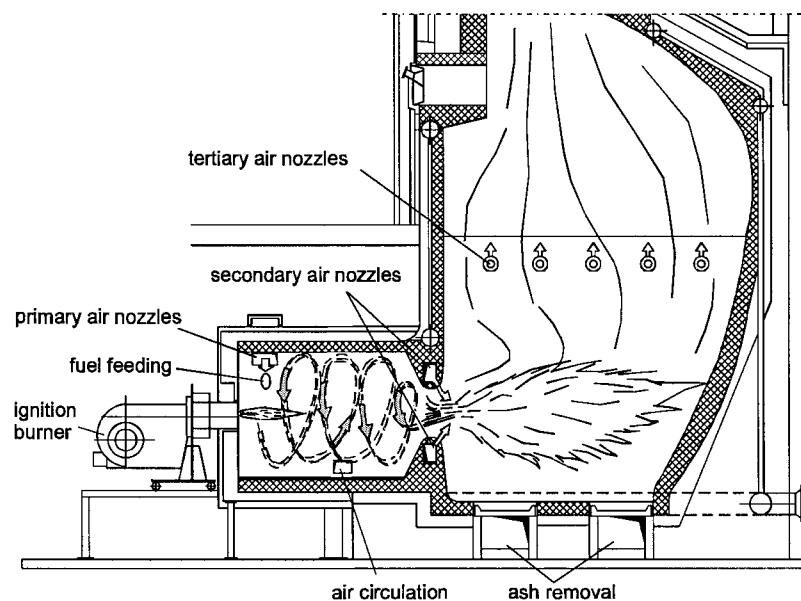


**FIGURE 7** Example of a blow-in combustion unit for wood dust. [From Kaltschmitt, M., and Hartmann, H., eds. (2001). "Energy from Biomass," Springer, Berlin, Heidelberg, Germany. (In German.)]
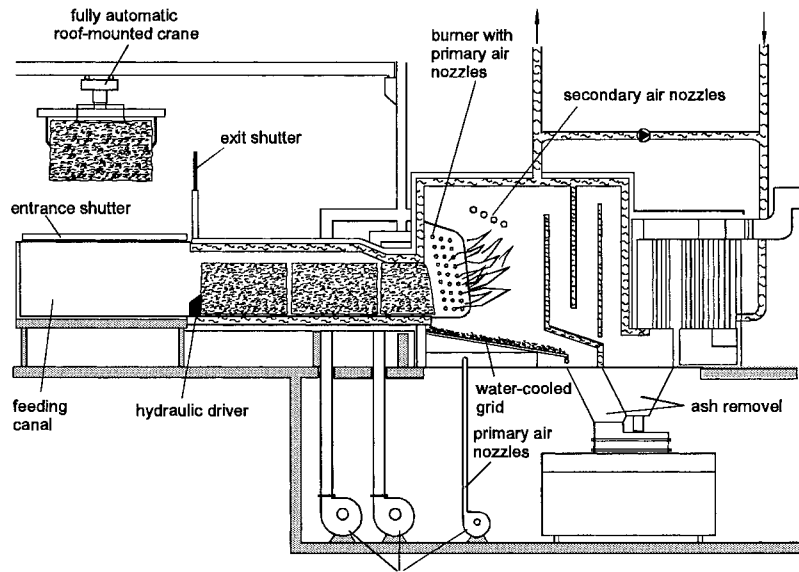
**FIGURE 8**  Example of a cigar burner for straw bales. [From Kaltschmitt, M., and Hartmann, H., eds. (2001). "Energy from Biomass," Springer, Berlin, Heidelberg, Germany. (In German.)]

illustrated by a device developed in Denmark for heat provision from straw bales (Fig. 8). The bales are fed continuously without any preparation into the combustion chamber within a horizontal feeding canal via a hydraulic cylinder. They are transported from a storage facility (barn) close to the plant via a fully automatic roof-mounted crane system to the end of this feeding canal. Such feeding arrangements allow a fully automatic operation of the overall combustion plant. To ensure an easy storage without rotting of the fuel (or self-ignition) as well as an environmentally sound combustion, the water content of the straw has to be below 20%.

Within the combustion chamber the primary air is blown directly on the forehead of the bale where the combustion takes place. Secondary air to ensure complete oxidation is blown into the after-burning zone located close after the flame front (Fig. 8). Incompletely burned parts of the fuel sometimes fall down at the forehead of the burning bale. Such fuel particles are fully burned afterwards on a moving grate system located right below the flame front. A backburning of the flames through the feeding system is prevented by a water-cooled system to lower the temperature, one-way valves, and maintenance of a minimum velocity of the bales being fed into the combustion chamber. The coarse ash deposits within the combustion chamber and is removed by an automatic ash removal system.

The flue gas flows through a heat exchanger and is then cleaned via a flue gas treatment system if necessary. Depending on regulations, a cyclone with or without fabric filter is used. The removed fine ash is in most cases taken to landfill and the coarse ash can be used as a fertilizer. Such

combustion devices are adjustable between 30 and 100% of installed capacity. Operation at the lower end of this adjustment range, however, often leads to relatively high pollutant emissions. Due to the fixed dimensions of the feeding canal, only bales with clearly defined dimensions can be used. Additionally with the necessary minimum feeding velocity of such a plant to avoid back-burning, the thermal capacity of one unit is fixed to 2–3 MW. Smaller thermal capacities are not possible. Larger thermal capacities can only be realized by adding additional feeding canals (i.e., building up a modular system). Such systems are mainly used for the provision of heat for district heating systems in some European countries.

## C. Coalification

In this process, woody biomass is heated in a nearly oxygen-free environment. Up to about 200°C, drying occurs, followed at higher temperatures with pyrolytic decomposition of the organic compounds. Substantial liquid and gaseous residues such as tar and carbon monoxide are released in the course of creating the final product, charcoal.

Such processes can be realized with quite different technologies. In developing countries for example, charcoal kilns are usually made from earth or brick, although simple metal kilns are sometimes found. The efficiency of such devices is low (less than 25%) and the airborne emissions of volatile organic compounds are high, and there can be substantial production of toxic liquid residues. In industrialized countries, charcoal is produced in large fully automatic industrial devices. Here a differentiation is made

between retort and flush gas processes. Within the former the charcoal is produced in a batch mode in a closed container (i.e., retort) where the thermo-chemical conversion from wood to charcoal takes place. The heat necessary to allow this process to take place is obtained from the combustion of the gaseous and liquid products produced as a by-product. There are also continuous charcoal production systems where the wood is fed constantly through a big reaction container. Within this container are different zones with various settings of reaction conditions to ensure that the charcoal is produced during the migration of the material through this container. Also here the gas and liquids produced as a by-product are used as a source of energy to keep the coalification process going.

In general the importance of such processes is relatively low in industrialized countries because charcoal plays only a minor role within the energy systems. But charcoal is used in such countries to a certain extend as a raw material (e.g., to produce activated charcoal for filtering) and as fuel for leisure. In developing countries charcoal is used as a clean fuel for heating and cooking. Additionally in some countries charcoal is also used for industrial purposes as in Brazil where charcoal is used for steel-making. The use of charcoal in these markets is dependent on political support and economic conditions.

## D. Pyrolysis

The decomposing processes realized during pyrolysis are similar to those during charcoal production, but here the process conditions are set to ensure that the main product is liquid rather than solid. There have been a broad variety of technologies developed during the last decades attempting to make pyrolysis oil for use in engines without any additional processing.

Most promising is flash pyrolysis in which fuel particles are heated very rapidly (more than $1000°C/sec$), but remain in the hot zone for a short time (in general less than 1 sec). After this short time, the liquid compounds produced from the solid biomass by decomposing the organic compounds of the biomass have to be removed and cooled rapidly to avoid further decomposition into gases. To date, flash pyrolysis reactors have been developed mainly to the laboratory stage.

Reactors with ablative impact designs decompose the biological raw material primarily into liquid components on the surface of a hot rotation wheel. On the surface of this wheel the solid biomass is "melted." To avoid further decomposition, the produced liquid components have to be removed quickly from the hot zone close to the rotating wheel. Necessarily, gaseous and solid components are also produced, which are used to heat the wheel.

Pyrolysis oil is a mixture of different hydrocarbons, many partly oxygenated, along with charcoal, ash, and water. The actual composition is strongly dependent on the pyrolysis process as well as the specific process conditions. This is especially true for the average composition and structure of the hydrocarbons (e.g., chain length, degree of double bonding). This oily liquid can be toxic and is in most cases not stable in air. Additionally the conversion efficiency of the available fast pyrolysis processes is relatively low. The average heating value of pyrolysis oil is approximately 40% that of petroleum-based fuels.

The goal of pyrolysis is the provision of liquid that can be used directly in engines, but the current state of technology does not achieve it. Upgrading of the produced oil is necessary to ensure easy use in existing engines. For example, charcoal and ash particles need to be removed to ensure a sediment-free fuel and, to increase stability, double bonds need to be broken by adding hydrogen. Additionally other measures have to be taken to ensure the required viscosity and combustion behavior. Like the fast pyrolysis process itself, these upgrading technologies are not fully developed yet. Because of these technical and economic constraints, there are no applications known for the provision of pyrolysis oil as a source of energy on a fully commercial basis.

## E. Gasification

Gasification describes the complete conversion of solid biomass at high temperatures to a gaseous fuel by adding a small amount of oxidizing reactants to the gasification process. Unlike coalification and pyrolysis, gasification of biomass is realized in the presence of some oxygen. The main objective of gasification is to transfer the maximum possible share of the chemical energy within the feedstock to the gaseous fraction remaining by producing a high yield of low molecular weight products. This so-called "producer gas" can be used as a fuel for the provision of heat through direct combustion or used in engines or turbines. Even use in fuel cells might be possible. Electricity production seems to be a promising option because electricity production via gasification allows much higher efficiency in principle compared to processes driven only by direct combustion.

Biomass gasification consists of the following, more or less spatially distributed, steps: heating and drying of the biomass, pyrolytic decomposition of the biomass (i.e., extracting the volatile components by heating), and gasification (i.e., partial oxidation of the biomass, partial reduction of the oxidation products ($CO_2$ and $H_2O$ to CO and $H_2$), and simultaneous transformation of solid carbon to CO).

The physical and chemical processes of biomass gasification are carried out in a variety of different forms of

equipment and technical concepts. Each of them offers certain advantages and disadvantages concerning feedstock possibilities, plant size, and gas quality.

The gasification techniques can be distinguished related to different criteria such as reactor type (fixed bed or fluidized bed), gasifying agent (air, oxygen, or steam), heat supply (directly or indirectly heated), and reactor pressure (atmospheric or pressurized). Over the years, a considerable variety of biomass gasifiers have been developed.

- In fixed-bed reactors, the feedstock is exposed to the gasifying agent in a packed bed that slowly moves from the top of the gasifier to the bottom, where the ash is discharged. Fixed-bed gasifiers are dense-phase gasifiers characterized by a relatively large amount of fuel exposed to a limited amount of reactive gas. In fixed-bed reactors the feedstock occupies most of the reactor volume.
- Fluidized-bed reactors are lean-phase gasifiers having a low ratio of solid to reactor volume. Typically, the feedstock is occupying only a small fraction of the total reactor volume. Fluidized-bed gasifiers are classified, depending on the intensity of fluidization, as bubbling fluidized-bed gasifiers, circulating fluidized-bed gasifiers, or entrained flow gasifiers.
- Gasifiers have also been developed that cannot easily be classified as belonging to one of the former groups but have features of both.

Commonly, the goal of gasification is not to provide gaseous fuel itself but to provide an easy to handle and environmentally sound intermediate energy carrier with clearly defined characteristics that can be converted easily into another, more valuable, energy carrier, for example, electricity or methanol. To reach that goal, gas cleaning is usually necessary to ensure a long lifetime of the conversion device (e.g., engine or turbine) because the gas produced within the gasifier does not usually match the fuel requirements in terms of condensable organic compounds and/or particles (cf. Kaltschmitt and Bridgwater, 1997 and Kaltschmitt, Roesch, and Dinkelbach, 1998).

### 1. Gasifier

*a. Fixed-bed gasifier.* In general, fixed-bed gasifiers rely upon gravity feed for the fuel and the extraction of the coarse ash at the bottom. Due to its continuous decomposition, the biomass slowly moves downward passing the zones of drying, degasification, oxidation, and reduction in the reactor. Fairly distinct zones are established within the gasifier for each of these gasification steps. The heat necessary to keep the gasification process going is supplied by partial combustion of the feedstock (Fig. 9).

Although feedstock moves from the top to bottom, the gas stream can be co-current, countercurrent, or cross-flow leading to several types in all:

- In downdraft gasifiers, both fuel and the gas move downward.
- In updraft gasifiers, fuel moves downward and gas upward.
- In co-current gasifiers, the fuel and gas move in the same direction, usually downward (downdraft gasifiers), but the movement can also be upward.
- In countercurrent gasifiers, fuel and gas move in opposite directions; updraft if the fuel moves up but
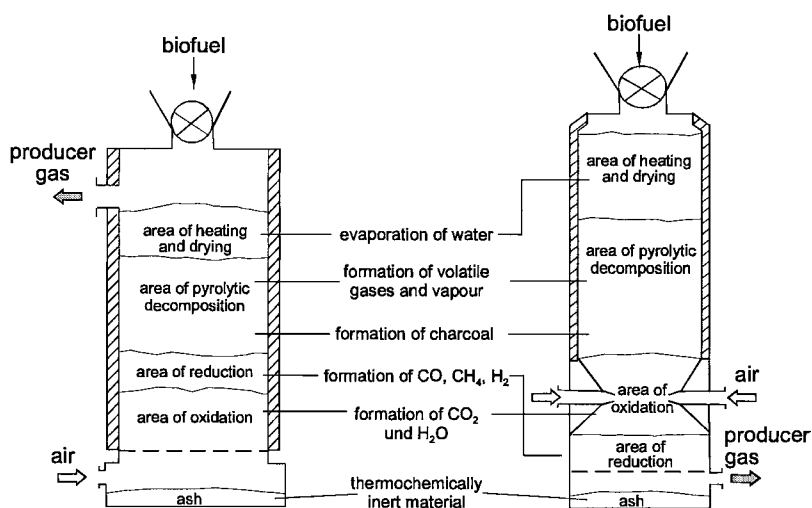


**FIGURE 9** Schematic of an updraft gasifier (left) and downdraft gasifier (right). [From Kaltschmitt, M., and Hartmann, H., eds. (2001) "Energy from Biomass," Springer, Berlin, Heidelberg, Germany. (In Germany.)]

flows can be reversed. In cross-draft or cross-current reactors, fuel moves downward and gas moves at right angles, i.e., horizontally.

Fixed-bed gasifiers were developed first in the 19th century and have been sized from a few 10–100 kW thermal capacity. Updraft gasifiers have been developed for higher thermal capacities, but only for the provision of heat. Fixed-bed gasifier technology is relatively well-developed, but sometimes losses (i.e., heat, unburned material) are fairly high. Unfortunately, gas quality often does not match the requirements of gas conversion units (e.g., engine and turbine) due to high particle and tar content in the producer gas. Also the environmental impact is sometimes fairly high.

*b. Fluidized-bed gasifier.* Fluidized-bed gasifiers were originally developed for coal gasification and have been adapted for biomass conversion. The gasifying agent is fed from the bottom into the gasifier at a speed high enough to "fluidize" or swirl the feedstock within the conversion reactor. Unlike fixed-bed gasifiers, no distinct reaction zones exist within this fluidized bed. All different ongoing thermo-chemical conversion processes are taking place in the same volume. Fluidized beds are versatile and insensitive to fuel characteristics, except for fuel size, which may require feedstock to pass through a size-reduction stage.

Fluidized beds have high rates of heat and mass transfer and provide good mixing, meaning that reaction rates are high, the residence time of the fuel particles are short, and the temperature is more or less constant in the bed. Fluidized beds can be operated at low capacity, but this is not economic in most cases. Therefore, such gasifiers are usually commercially viable with thermal capacities of several megawatts. They can generally be scaled up easily to several tens of megawatts or even higher.

Fluidized-bed gasifiers can be operated either under atmospheric pressure or as pressurized systems. Pressurized gasification has the advantage that downstream compressors are not needed to reach the pressure requirements of the turbine where the producer gas might be used. Furthermore, high-pressure gasification plants can be constructed more compactly and have, in principle, higher efficiencies than atmospheric-pressure plants, but they are also more costly in most cases. There is no clear judgment yet as to which is best for power production from an overall technical and economic point of view.

There have been several fluidized-bed gasifiers built recently. Their disadvantage is that the producer gas is quite dirty, requiring demanding gas-cleaning systems. Nevertheless, this gasification approach seems to be the most promising for future large-scale applications.

## 2. Gas Cleaning

Depending on feedstock, process conditions, and gasifying agent, the producer gas consists mainly of different fractions of carbon monoxide (CO), carbon dioxide ($CO_2$), methane ($CH_4$), hydrogen ($H_2$), and nitrogen ($N_2$). In addition, it could contain steam ($H_2O$), nonmethane hydrocarbons (e.g., ethylene, propane), as well as unwanted components, especially condensable organic compounds (i.e., tars) plus unburnt carbon, ash, and other particles. Biomass gasification with air results in a high $N_2$ (60%) content in the producer gas, and thus in a gas with a low heating value, i.e., 4–6 $MJ/m_n^3$. The heating value can be higher if pure oxygen or water is used as a gasifying agent.

Each application of the producer gas requires a different level of gas cleaning to prevent erosion, corrosion, deposition, and other problems downstream of the gasifier. Gas cleaning usually starts with the separation of particulates and condensable organic compounds from the raw producer gas. Depending on needs, additional removal of minor gas impurities, e.g., alkali metals, can be required. For particle removal, several filter techniques are available, such as cyclones, fabric filters, electrostatic precipitators, granular filter beds, and ceramic candle filters as well as washing towers. Technologies are often combined to take advantage of the individual strengths and avoid the weaknesses of each.

In gasification processes, it seems to be impossible to totally prevent the formation of condensable organic compounds by the design of the reactor and modification of gasification conditions. Reducing the amount of high- and low-molecular-weight condensable organic compounds by adsorption techniques, wet scrubbing, or catalytic decomposing into noncondensable organic compounds allows the gas to be cooled for entry into downstream cleaning equipment preventing condensation of organics that may cause severe plugging and fouling.

Cheap, reliable gas cleaning is still the greatest need in biomass gasification for electricity production. No available single or combination technique is able to fulfill the requirements of gas engines or turbines in a commercial way, although there have been promising demonstration projects (e.g., Vermont Project, United States; Värnamo Plant, Sweden). This arena has a high priority for additional research and development.

## 3. Gas Utilization

The gas generated from biomass gasification may be fired directly, e.g., for process heat within a retrofitted furnace or boiler, or fed to an engine, turbine or fuel cell for the provision of mechanical power or electricity (along with heat in a CHP system).

If producer gas is cleaned to the level of purity found in natural gas, there is a wide range of possible applications, as noted above. Since this is currently not economically possible, there are efforts to develop more robust engines and turbines that might tolerate less clean gas.

The conversion of gas to liquid secondary energy carrier is also technically possible. The most widely known process converts the carbon monoxide in the gas to methanol by adding hydrogen. This hydrocarbon can be used as an additive to conventional petrol as well as a raw material for the chemical industry. To date, however, no commercially viable facility has been developed. In large plants it is also feasible to use the producer gas to produce chemicals such as methanol, hydrogen, or ammonia.

## IV. PHYSICO-CHEMICAL CONVERSION

The technology for producing fuel oils from biomass is similar to that for the production of vegetable oil as food or fodder. Some biomass contains fatty and/or oily components, for example, rape and sun flower seeds, coconuts, peanuts, and corn. In some cases, the potential yield is high. For example, rapeseeds contain 40–45% vegetable oil. These crops can only be cultivated in certain regions and require relatively high agricultural efforts to produce and thus the potential for their use is localized.

In most cases, vegetable oil from these sources is more valuable as food or fodder, but use as a fuel is possible. Vegetable oil is easy to store, has a relatively high energy content, and can be used without major problems in transport. Additionally, use in small-scale CHP plants for rural electrification in developing and in industrialized countries as well as in heating plants is possible. Because the fuel characteristics of conventional diesel fuel and raw vegetable oil are quite different, treatment of the vegetable oil is often needed. Therefore the vegetable oil is converted into fatty acid methyl ester (FAME).
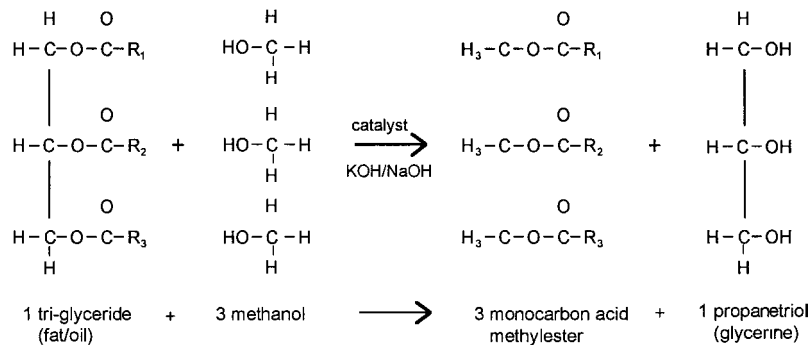
## A. Basics

Vegetable oil combusts differently from conventional diesel fuel because it is a triglycerid (Fig. 10). Therefore an adaptation of existing diesel engines to the specific fuel characteristics of vegetable oil is necessary. This adaptation is expensive and reduces the lifetime of the engine. Therefore, vegetable oil is sometimes chemically treated to better match the combustion behavior of diesel fuel by converting it into FAME. This is done by removing the glycerine from the triglycerid (Fig. 10). To saturate the ester bindings instead of the glycerin, methanol is added to the remnant. This reaction is realized with the help of a catalyst.

## B. Oil Production

Three approaches are available for producing oil: pressing only, extraction only, and a combination of both. During the pressing, the oil is squeezed out of the oil seeds based on a one- or two-stage pressing process leaving vegetable oil and oil cake. The latter still contains 4–10% oil. Sometimes this oil cake is used as a cattle feed where the remaining oil contributes to the nutrition value of this product. In developing countries and within traditional oil production processes in industrialized countries (e.g., traditional olive oil production), this simple technology is still of considerable importance.

Extraction occurs by means of a solvent applied in a countermovement in large-scale units. This technology has been widely employed for many decades. Such devices produce two different material streams: the solvent saturated with oil and the oil-free extraction residue saturated with solvent. The solvent must then be removed from both streams by heating and reused. Compared to simple pressing, a much higher share of the oil originally within the oil seed can be removed. After extraction the meal contains significantly less than 1% oil and can be used as cattle feed.



**FIGURE 10**  Esterification of vegetable oil from rape seed to fatty acid methyl ester. [From Kaltschmitt, M., and Hartmann, H., eds. (2001). "Energy from Biomass," Springer, Berlin, Heidelberg, Germany. (In German.)]

For biomass with high oil contents, like rapeseed, pressing followed by extraction will maximize production and be economic. The produced vegetable oil contains about 0.5–6% oil-free solid residues originating from the seed. The amount of such residues mainly depend on the condition of the pressing aggregate, the flow-rate, the kind of pressed seed, and the water content in the oil seeds. The solid residues have to be separated from the oil by filtration and sedimentation to prevent possible problems during the further downstream processing of the vegetable oil. Even after this cleaning step, however, the vegetable oil is not qualified for direct utilization because it still contains unwanted substances (fatty acids, ketones, wax, pigments, heavy metals, pesticides) reducing shelf-life and complicating subsequent processing and use. Further refining is needed and typically leads to a loss of 4–8% of the usable oil mass through de-acidification, de-coloring, and steaming.

## C. Esterification

Figure 11 shows a continuous two-stage esterification in which the mixture of cleaned vegetable oil together with the catalyst and the methanol is pumped with a low velocity through a vertical pipe. The low velocity of the liquid ensures that the glycerin produced during the esterification process can settle in the reactor and removed for use as a raw material. Additionally glycerin is removed within a separation tank as well as by a separation unit. After the removal of the remaining methanol, the liquid is cleaned by a multistage washing process. The produced FAME is now ready to be used as a fuel.

There are FAME plants operating in Europe. In Germany, for example, plants with a capacity of 300,000 t/y FAME are in operation and more are planned. Manufacturing FAME is therefore state of technology, although the energy balance is fairly poor compared to other biomass conversion routes. Accordingly, the costs are relatively high.

## D. Use

Pure vegetable oil can be used in some existing engines, but it lowers engine lifetime and increases maintenance requirements. But under specific conditions this could be a promising option to provide power especially in rural areas also for rural decentralized electrification. This is true for developing as well as for industrialized countries.

Fatty acid methylester can be used directly as a diesel substitute in conventional diesel engines (Table V). Large-scale demonstration projects with FAME based on rape oil (RME) have shown that there are no significant problems, although the more aggressive chemistry of RME compared to conventional diesel fuel requires fuel lines to be made from a more resistant material like Teflon® (Heinz et al., 2000).

FAME has promising environmental advantages compared to conventional diesel fuel in certain niche applications. It is more easily degraded in nature and shows lower
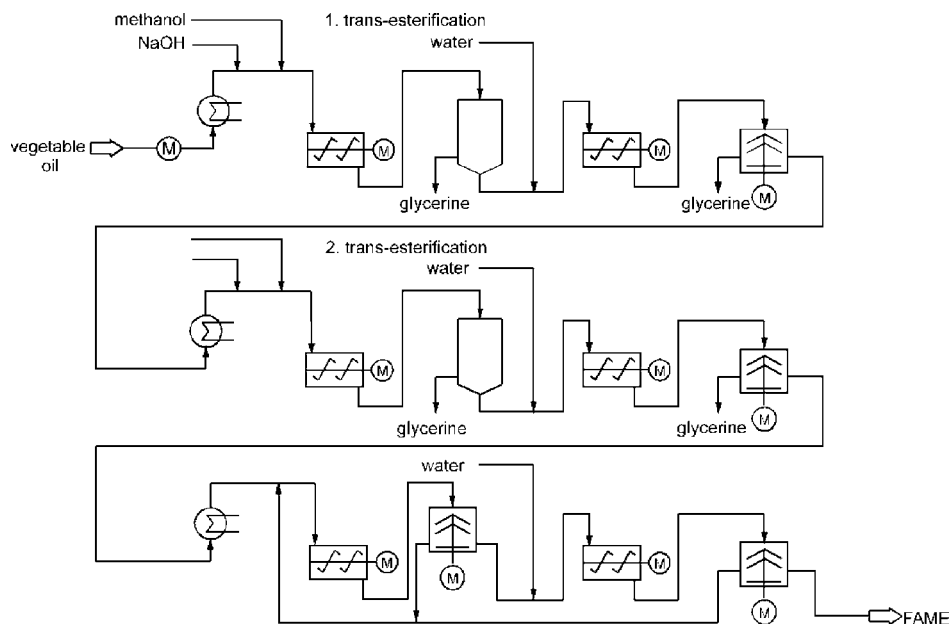


**FIGURE 11** Example for a continuously working esterification process. [From Kaltschmitt, M., and Hartmann, H., eds. (2001). "Energy from Biomass," Springer, Berlin, Heidelberg, Germany. (In German.)]

**TABLE V  Selected Properties of Diesel Fuel, FAME, and Rape Oil**

|  | Diesel fuel | FAME | Rape Oil |
|---|---|---|---|
| Density (15°C) in kg/m$^3$ | 820–845 | 875–900 | 900–930 |
| Viscosity (40°C) in mm$^2$/s | 2.0–8.0 | 3.08 | 78.7 |
| Flashing point in °C | >55 | Min. 110 | Min. 220 |
| CFPP-Value in °C (summer/winter) | Max. 0 | Max. 0 |  |
|  | Max. −20 | Max. −20 |  |
| Sulfur content in mg/kg | Max. 350 | Max. 100 | Max. 20 |
| Cetan-Value | Min. 51 | Min. 49 |  |
| Calorific value in MJ/kg | 42.7 | 37.9 | Min. 35 |

particle emissions, for example, thus showing promise for use in water conservation areas and congested cites.

## V. BIO-CHEMICAL CONVERSION

Nearly all biomass is eventually decomposed naturally through biological processes. Some of these processes can be harnessed to produce fuels.

- Composting occurs if oxygen is available. During composting the biomass is degraded by bacteria mainly to carbon dioxide ($CO_2$) and water ($H_2O$) while releasing low temperature heat that can, in principle be used via a heat pump. But for the time being this option is only of theoretical importance.
- Under anaerobic (oxygen-free) conditions, a variety of degrading processes are employed by microorganisms. Most relevant to energy are ethanol production via alcohol fermentation and biogas (methane) production via anaerobic digestion.

### A. Alcoholic Fermentation

Sugar ($C_6H_{12}O_6$) is converted to ethanol ($C_2H_5OH$), carbon dioxide ($CO_2$), and low temperature heat by yeast under anaerobic conditions. Because starch and even celluloses can be converted more or less easily into sugar, such biomass streams are also a potential resource for the production of ethanol in addition to naturally sugar-containing crops like sugar cane and sugar beet. For fuel, ethanol is needed in a pure form and therefore further steps are needed after fermentation.

#### 1. Basic Processes

Sugar ($C_6H_{12}O_6$) is converted to ethanol ($C_2H_5OH$), carbon dioxide ($CO_2$), and low temperature heat by yeast under anaerobic conditions:

$$C_6H_{12}O_6 \rightarrow 2C_2H_5OH + 2CO_2 + 88\,kJ/mol$$

According to the above equation 100 kg of sugar is converted into 51.14 kg of ethanol and 48.86 kg of carbon dioxide and 400 kJ of heat. This heat is needed to allow the microorganisms to live and grow. To reach a quick transformation of the sugar into ethanol, additional growth supporter and mineral nutrients, as well as optimal temperature and pH are needed. Most processes are realized at a temperature level between 25 and 40°C. This process is well known because of humanity's long and extensive production of alcoholic beverages.

Because starch and even celluloses can be converted more or less easily into sugar, such biomass fractions are also a potential resource for the production of ethanol in addition to naturally sugar-containing crops like sugar cane and sugar beet. For fuel, ethanol is needed in a near pure form and therefore further steps are needed after fermentation.

#### 2. Substrate Production

For fermentation, sugar is needed in a water solution. Therefore sugar has to be extracted from biomass containing sugar (like sugar cane, sugar beet) and/or has to be produced from biomass containing starch (like grain, potatoes) or celluloses (like wood).

- *Biomass containing sugar.* Sugar can be directly removed from sugar cane and sugar beet and thus these two species are the most widely used for this purpose. Sugar cane is a grass growing mainly under tropical climatic conditions containing a sweet liquid that is easily removed by pressing and accounts for about 30% of the weight of fresh cane. The fiber remaining after pressing is called bagasse. It is used often as a solid biofuel for the provision of the energy needed to run a sugar mill and/or an alcohol production factory. After several preparation and processing steps this liquid can be used as a feedstock for alcoholic fermentation. Sugar beet, which grows in temperate climates, is first chipped into small pieces from which the sugar is extracted by means of water applied countercurrent to the flow direction of the chips. The remaining products are the nearly sugarfree extracted beet which can be used as a cattle feed.
- *Biomass containing starch.* Cereal, corn, potatoes, topinambur, maniok, and other similar crops contain starch, which is a polysaccharide consisting of long chains of glucose modules. To produce sugar from starch, this polysaccharide has to be decomposed in single saccharid rings by means of biological processes based on enzymes. Due to the variety of starch containing biomass crops, there is no typical chemical process. For each raw material, a process has been developed and optimized in the past because these

materials are often used also for the production of alcoholic beverages. In most cases, however, water is added to the chipped or milled material. Due to swelling, the starch becomes a paste to which is added the respective enzyme adapted to each type of biomass. Under the right process conditions, the starch is then converted into sugar, which can be used as a feedstock for the alcoholic fermentation.

- *Biomass containing celluloses.* All biomass contains celluloses as this is one of the most important elements of plants. Like starch, celluloses is built of sugar components. Therefore celluloses can be split again into sugar, in this case both through biological and nonbiological processes. The former are based on special enzymes, but are relatively slow and inefficient. The latter are realized with an acid catalyst-based hydrolysis. Such a process is technologically very demanding and thus alcoholic fermentation based on sugar from celluloses has no importance for the time being.

### 3. Alcoholic Fermentation

The sugar-containing feedstock is next inoculated with yeast. Continuous operation is usually done for small-scale production of high-quality spirits such as beverages and discontinuous (batch) operations are used in large-scale systems for production of ethanol as a raw material for the chemical industry. After the fermentation, the yeast is removed from the slurry and recycled.

### 4. Production of Pure Alcohol

After fermentation the material contains 8–10% alcohol, the rest being primarily water and the residues from the sugar- or starch-containing organic material. Distillation or rectification is done to purify the product using a crude alcohol column. The result is an alcohol-water-mixture with a high share of alcohol (more than 80%) and a slurry without significant alcohol content. The latter is used as an animal fodder and/or as a fertilizer. It is also suitable for biogas production. This technology is in large-scale operation worldwide.

At best, distillation and rectification will only achieve 95–96% pure ethanol. Engine fuel must be 99.9% pure, which can be achieved via a further step called absolutation. Here, a third chemical (expident) is added to the alcohol-water-mixture producing a mixture of alcohol-chemical plus water-chemical. Afterward the chemical is removed from both mixtures and reused.

### 5. Use

In engines, ethanol can substitute fully or partly for gasoline although the LHV of ethanol is lower than that for

**TABLE VI   Selected Properties of Fuels**

|  | Ethanol | Gas |
|---|---|---|
| Composition in % |  |  |
| Carbon | 52 | 86 |
| Hydrogen | 13 | 14 |
| Oxygen | 35 | 0 |
| LHV in MJ/kg | 26.8 | 42.7 |
| in MJ/1 | 21.3 | ca. 32.0 |
| Density (15°C) in kg/l | 0.794 | 0.72–0.78 |
| Viscosity (20°C) in mm$^2$/s | 1.5 | 0.6 |
| Boiling point in °C | 78 | 25–215 |
| Flame point in °C | 12.8 | −42.8 |
| Ignition temperature in °C | 420 | ca. 300 |
| Evaporation heat in kJ/kg | 904 | 380–500 |
| Minimum air volume in kg/kg | 9 | 14.8 |
| Octane-value | 107 | 93 |

gasoline (Table VI). The combustion of ethanol, however, required a lower air volume. Therefore the heating value of the mixture pressed into the cylinder is more or less the same for ethanol and gas. This is the reason why an engine powered by ethanol produces the same power as an engine driven by gas.

Internal combustion engines must be specially adapted to ethanol, because it shows different combustion behavior compared to gasoline. Adaptation kits have been used successfully in several countries, for example, Brazil. Under current petroleum prices, pure ethanol is a high-priced alternative as well as requiring modified distribution networks and such problems as water retention during storage. In a number of countries, however, ethanol is mixed with gas to a maximum of 10% in which form it can be used without any known problems in existing engines and distribution networks. Additionally, ethanol can be converted into ethyl tertiary butyl ether (ETBE) for use as a pollution-reducing gasoline additive.

## B. Anaerobic Digestion

During anaerobic digestion, organic material is decomposed in an oxygen-free atmosphere by bacteria that produce a gas containing approximately two-thirds methane ($CH_4$) and one-third carbon dioxide ($CO_2$) plus some impurities. Such decomposition occurs in nature, for example, at the bottom of lakes and moors within the sediments containing organic material, and takes place in landfills and dumps where the organic fraction of the waste material is converted into landfill gas.

### 1. Basics

In addition to degradable organic input and freedom from air, additional nutrient for the bacteria and absence of
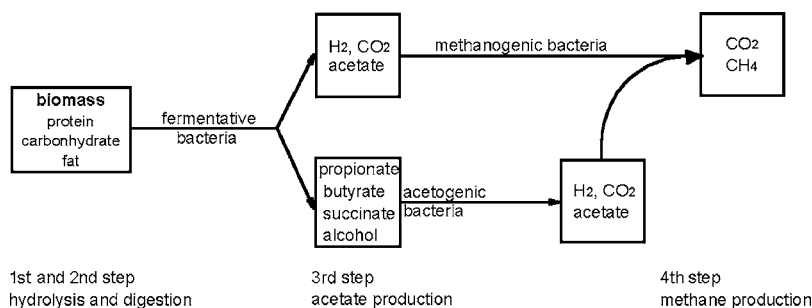
**FIGURE 12** Anaerobic process scheme. [From Neubarth, J., and Kaltschmitt, M., eds. (2000). "Renewables in Austria—Systemtechnology, Potentials, Economics, Environmental Aspects," Springer, Vienna, Austria. (In German.)]

harmful, pathologic, and inhibiting substances are also needed for success. Also the organic material has to be

bic digestion are agricultural residues (e.g., liquid manure, leaves from sugar beet), residues from the food processing industry (e.g., slurry of fruit and potato processing, some types of slaughterhouse residues), and sewage sludge. Only lignin-containing biomass like wood cannot be degraded by anaerobic bacteria.

During anaerobic digestion, organic matter is degraded by three different kinds of bacteria: fermentative, acetogenic, and methanogenic. The first two bacteria families degrade the complex organic compounds of biomass into simpler intermediates (Fig. 12). These intermediate products are then converted to methane and carbon dioxide by the methanogenic bacteria.

Anaerobic digestion relies on a dynamic equilibrium among the three bacterial groups, which is strongly affected by temperature. Most digesters operate in the mesophilic temperature regime (with a peak in microbe activity at around 35°C). Some operate in the thermophilic regime (with a peak in microbe activity around 55°C). Up to the temperature of peak microbial activity, higher operating temperatures produce greater metabolic activity within either regime. The pH value, the composition of the biomass in relation to easily degradable biomass compounds, and rate of loading of the feedstock affect the bacterial balance also. For optimal fermentation, pH-values between 6.8 and 7.2 are often maintained by use of buffers.

The composition of the produced biogas is typically about two-thirds methane and one-third carbon dioxide. Additionally the biogas may contain trace substances such as $H_2S$, depending on the composition of the biomass, the process conditions, and other parameters. The LHV of biogas, which depends mainly on methane content, ranges from 14 to 29 $MJ/m^3$.

Expected biogas yields for different substrates are given in Table VII. Already digested substrates like liquid manure and droppings show, in general, smaller yields than fresh material, e.g., roadside green or waste fat. Economic viability, however, also depends on the rate.

### 2. Biogas Production

The bacteria use relatively little of the biomass energy for their own survival and only a small amount of heat is produced, making the conversion process of the organic material into biogas fairly efficient. To ensure optimal process conditions with a good access of the bacteria to the organic material so that gas yields are maximized, the organic material needs to be in a slurry with a water content of more than two-thirds. To optimize production rates, the slurry should be kept at temperatures of 28–35°C or 50–70°C, which correspond to the ranges appropriate to different bacterial species. Production will occur at lower temperatures, but essentially stops at 10°C.

**TABLE VII   Yield Targets of Biogas for Different Substrate**

| Material | Yield of biogas in $m^3$/t organic dry matter | Material | Yield of biogas in $m^3$/t organic dry matter |
|---|---|---|---|
| Liquid manure from beef | 250 | Paunch content | 420–520 |
| Liquid manure from pork | 480 | Rey straw | 300–350 |
| Droppings from chicken | 450 | Potato herbs | 560 |
| Sewage sludge | 400 | Sugar beet leaves | 550 |
| Organic waste from households | 170–220 | Food residues | 80–120 |
| Waste fat | 1040 | Waste water from brewing industries | 500 |
| Roadside green | 550 | Waste water from sugar industries | 650 |

In Europe, the main task of this conversion route is to treat organic waste or residues to reduce waste disposal costs; the production of energy is only a by-product in most cases (like the treatment of sewage sludge).

Typical feedstock for a biogas production is organic material available at low (or zero) costs. A big advantage occurs if it is already available with high water content (i.e., in most cases significantly more than 90% water), as with sewage sludge, animal manure, and organic residues from some types of agricultural production and the food processing industry. Within biogas plants, preparation of the feedstock is needed that might include short-time storage, sedimentation of mineral contaminants (like sand), reducing the material into small pieces, mixing of different types of feedstock to maximize the gas yield, and heating to the needed temperature level. Then the feedstock is pumped into the biogas container or reactor where anaerobic fermentation takes place. For successful operation, the bacteria must always be well-mixed with the organic material. It is also important to realize a good temperature distribution within the reactor. The biogas is removed from the top of the plant and, after removal of impurities like water, stored before use as a source of energy is realized.

The digested material is removed from the reactor and stored in a tank where a small amount of additional biogas is produced. This digested slurry is used as a fertilizer, because it contains the nitrogen originally part of the feedstock. Since biogas technology helps to close the nutrient cycle, it is of increasing importance in environmentally sound waste management. In China, human as well as animal waste is often added to household biogas digesters as well, giving them a sanitation as well as an energy function.

Figure 13 shows a biogas plant commonly used in countries like Denmark, Germany, and Austria. The animal manure from the stable is mixed, stored for a short time, heated, and pumped into the biogas reactor. This mixing facility is driven by a slow-moving electric motor.

## 3. Use

Biogas can be used in a boiler, stove, or engine for the provision of heat, mechanical power, and/or electricity. A part of the heat and electricity produced is often partly applied to warming up the feedstock and running of the plant. This is a system layout most widely used in industrialized countries. Alternatively or additionally, biogas can also be fed into the local natural gas system, to do so requires a demanding upgrading of the biogas to fulfill the local natural gas standards. This is done in Switzerland and Sweden, for example. Use of the biogas as a transportation fuel is possible and has been shown in demonstration projects, but it is not clear at present if such options will become more important in the future.

In developing countries the use of the biogas for household cooking and lighting is becoming common in areas with appropriate climates, sufficient animals to provide dung as feedstock, and access to low-cost capital funds to build the digesters. The technology is simpler than used in large-scale facilities, for example, not involving a heating system and relying on manual feedstock preparation. This results in lower biogas yields, higher operating cost per unit energy, and low or zero production during the cold season. Community-scale facilities in which one plant supplies gas to many households are used in several Asian countries, although experience has been mixed.
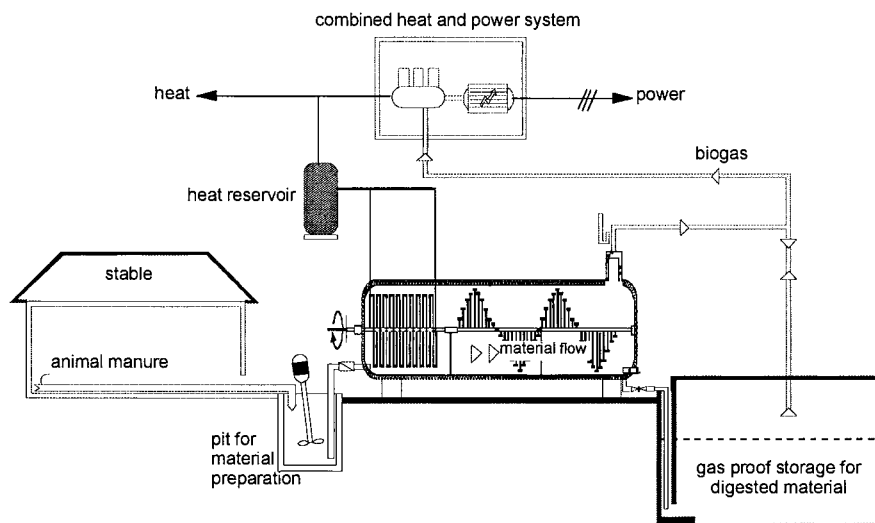


FIGURE 13 Example of a biogas plant using animal manure as a feedstock. [Form Neubarth, J., and Kaltschmitt, M., eds. (2000). "Renewables in Austria—Systemtechnology, Potentials, Economics, Environmental Aspects," Springer, Vienna, Austria. (In German.)]

Large-scale plants, similar to those used in industrialized countries, are also in use in developing countries at pig farms and other facilities with concentrated feedstock availability.

## VI. OUTLOOK

Biomass fuel will continue to be important locally around the world. As oil and gas prices move upward over time, more efficient modern technologies will be employed to convert biomass energy to these easier to use forms along with expanded use for power and co-generation (Kartha and Larson, 2000 and Hall *et al.,* 2000). In developing countries, although economic growth will lead to a shift up the household energy ladder to liquid and gaseous fuels for hundreds of millions of people, for many decades there will still be a billion or more of the poorest people who will be unable to switch from using unprocessed traditional fuels. Moving these people to cleaner and more efficient alternatives would take concerted efforts involving new technologies, subsidies, and education at the scale now employed to provide clean water and sanitation. There are also many attractive uses for biomass in commercial and small-scale industrial applications in developing countries (Larson, 2000). There are prospects for advanced biomass conversion technologies to liquids and gases, such as application of the Fischer-Tropsch process to create synthetic LPG from crop residues, that show potential, but are at present too speculative for accurate evaluation. Should oil and gas prices remain high for long periods, however, there may be sufficient incentive to develop such biomass alternatives. Finally, if serious long-term efforts are made to reduce carbon emissions through taxes, subsidies, and other incentives, biomass fuels will likely become much more attractive. As much as half of world energy could be provided by 2050 with a major effort according to some estimates (Goldenberg *et al.,* 2000). At the scale required to significantly reduce global emissions from fossil fuel, however, competition with agriculture and other sectors for land, water, and other resources would become issues.

## SEE ALSO THE FOLLOWING ARTICLES

BIOMASS, BIOENGINEERING OF • BIOMASS UTILIZATION, LIMITS OF • ENERGY EFFICIENCY COMPARISONS AMONG COUNTRIES • ENERGY FLOWS IN ECOLOGY AND IN THE ECONOMY • ENERGY RESOURCES AND RESERVES • POLLUTION, AIR • WASTE-TO-ENERGY SYSTEMS

## BIBLIOGRAPHY

Baldwin, S. (1986). "Biomass Stove Technologies: Engineering Design, Development, and Dissemination," Volunteers in Technical Assistance and Princeton University Center for Energy and Environmental Studies, Princeton, NJ.

Goldenberg, J. *et al*. (2000). "World Energy Assessment," United Nations Development Programme, New York.

Hall, D. O., House, J. I., and Scrase. I. (2000). An overview of biomass energy. *In* "Industrial Uses of Biomass Energy: the Example of Brazil," (F. Rosillo-Calle, H. Rothman, and S. V. Bajay, eds.), Taylor & Francis, London.

Heinz, A., Kaltschmitt, M., Hartmann, R., and Hitzler, G. (2000). "The First Year of Operation of the Biodiesel-Fuelled Energy Supply System of the New German Parliament Building in Berlin—Analysis of the Technical Feasibility, Environmental Aspects and the Costs," 11th European Biomass Conference, Sevillia, Spain.

International Energy Agency, World Energy Outlook, Paris, 2000.

Kaltschmitt, M. (1999). Utilization of Biomass in the German Energy Sector. *In* "Strategies and Technologies for Greenhouse Gas Mitigation" (J.-F. Hake, N. Bansal, and M. Kleeman, eds.), Ashgate, Aldershot, U.K.

Kaltschmitt, M., and Bridgwater, A. V., eds. (1997). "Biomass Gasification and Pyrolysis—State of the Art and Future Prospects," CPL Scientific, Newbury, U.K.

Kaltschmitt, M., and Hartmann, H., eds. (2001). "Energy from Biomass," Springer, Berlin, Heidelberg. (In German.)

Kaltschmitt, M., and Reinhardt, G. A., eds. (1997). "Regrowing Energy Carrier; Basics, Processes, Eco Balances," Vieweg, Braunschweig/Wiesbaden. (In German.)

Kaltschmitt, M., Reinhardt, G. A., and Stelzer, T. (1997). Life cycle analysis of biofuels under different environmental Aspects. *Biomass Bioenergy* **12,** 121–134.

Kaltschmitt, M., Rösch, C., and Dinkelbach, L., eds. (1998). "Biomass Gasification in Europe," European Commission, DG XII, Brussels, Belgium.

Kaltschmitt, M., and Wiese, A., eds. (1997). "Renewable Energies, Systemtechnology, Economics, Environmental Aspects," Springer, Berlin. (In German.)

Kartha, S., and Larson, E. D. (2000). "Bioenergy Primer, Modernized Biomass Energy for Sustainable Development," United Nations Development Programme, New York.

Larson, E. D., ed. (2000). Modernized biomass energy, special issue of *Energy Sustainable Deve.* **4**(3).

Neubarth, J., and Kaltschmitt, M., eds. (2000). "Renewables in Austria—Systemtechnology, Potentials, Economics, Environmental Aspects," Springer, Vienna, Austria. (In German.)

Reed, T. B. and Larson, R. (1996). A Wood-Gas Stove for Developing Countries. *In* "Developments in Thermochemical Biomass Conversion," (A. V. Bridgwater, ed.), Blackie Academic Press, London.

Rösch, C., Kaltschmitt, M., and Limbrick, A. (2000). "Standardisation of Solid Biofuels in Europe," 11th European Biomass Conference, Sevillia, Spain.

Smith, K. R. (1987). "Biofuels, Air Pollution, and Health," Plenum, New York.

Smith, K. R. (2000). National burden of disease in India from indoor air pollution. *Proce. Nati. Acad. Sci.* **97**(24), 13286–13293.

Smith, K. R., Zhang, J., Uma, R., Kishore, V. V. N., Joshi, V., and Khalil, M. A. K. (2000). Greenhouse implications of household fuels: An analysis for India. *Annu. Rev. Energy Environ.* **25,** 741–763.
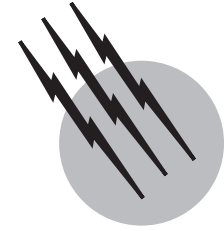
# Solar Energy in Buildings

**S. V. Szokolay**

*University of Queensland*

## GLOSSARY

**Active systems** Solar–thermal systems in which heat transfer takes place by mechanically driven fluid (water, air) flow, usually in closed channels (pipes or ducts) where pumps or fans (blowers) use a significant amount of conventional form of energy, if CoP < 20 (see definition of coefficient of performance CoP).

**Auxiliary energy $E_a$** Energy from conventional sources (boilers, furnaces, etc.) used to supplement the energy from the solar source should the solar source fall short of the demand.

**Coefficient of performance CoP** Ratio of energy of solar origin $E_s$ delivered by the system as useful heat (excluding $E_a$) to the parasitic energy used; $CoP = E_s/E_p$ (this definition differs somewhat from that used for cooling machines or heat pumps).

**Hybrid systems** Essentially passive systems with some small amount of conventional energy used to assist the otherwise spontaneous heat transfer processes, if $50 > CoP > 20$.

**Low temperature systems** Solar–thermal systems for delivering heat at a temperature below $100°C$, most often between 40 and $80°C$.

**Parasitic energy $E_p$** Energy from conventional sources (e.g., electricity) used to drive the heat transfer mechanisms of the system, such as pumps, fans (blowers), and controls.

**Passive systems** Solar–thermal systems in which heat transfer takes place predominantly by spontaneous (natural) processes, if $CoP > 50$.

**Selective surfaces** Generally, surfaces having absorption and emission characteristics varying as a function of radiation wavelength; specifically, surfaces with high

absorptance for solar wavelengths (∼6000°K), but low emittance at normal operating temperature (∼350°K).

**Solar fraction** $F_s$ Fraction of the total load $L$ (e.g., heating requirement) supplied from the solar source; the remainder is the auxiliary energy. $F_s = E_s/L = 1 - (E_a/L)$.

**Solar–thermal systems** Systems for converting radiant solar energy into thermal energy (heat) and delivering this heat where and when it is required for use.

**SOLAR ENERGY** can be utilized in many ways (e.g., photovoltaics, solar–thermal–electrical systems, solar ponds, and biomass conversion), but only its application to provide (or contribute to) the heat requirements of buildings will be discussed here (with a brief mention of building-integrated photovoltaics). These are all low-temperature heat requirements and may include the heating of water (for domestic hot water systems or swimming pool heating), the heating of the building space itself, and possibly also space cooling.

The three essential functions of such systems are

1. Collection. The conversion of radiant solar energy into heat
2. Storage. The system used to keep the collected heat and make it available *when* required (i.e., distribution in time)
3. Distribution (in space). The way to supply this heat *where* it is required.

The heat transfer process must be driven and controlled, and an auxiliary heat source must be included should the supply from the solar source be insufficient at times.

These systems may be mechanical installations, essentially independent of the building (although perhaps attached to or integrated with some elements of the building), which are generally referred to as *active systems*, or they may be constituted by the building itself (where elements of the building and their configuration may be designed to perform the three functions), which are known as *passive systems*. For the discussion of the latter, it is essential to consider the thermal behavior of buildings because all buildings are to some degree passive solar systems.

## I. SOLAR COLLECTORS

If a surface is exposed to solar radiation, it will be heated. In this sense, any such exposed surface is a solar collector. The degree of this heating will depend on (1) the amount of radiation incident on the surface and (2) the absorp-

tance of the surface. Radiation incident on a surface can be measured as

1. Irradiance, or the instantaneous density of radiant flux (energy flow rate or power density), in watts per square meter (W/m²). The term intensity is often (but erroneously) used to mean this quantity.
2. Irradiation, or energy density, i.e., the amount of energy incident on the surface over a specified time period (an hour, a day, etc.) in joules per square meter (J/m²) or watt-hours per square meter (Wh/m²).

Radiation from the sun can reach a surface three ways; thus the total irradiance (denoted G for global) of a surface may consist of three components: $G_b$, which is the direct beam irradiance arriving along an unobstructed straight path from sun to surface; $G_d$, which is the diffuse irradiance (i.e., radiation scattered by the atmosphere) apparently arriving from all directions within the sky hemisphere; and $G_r$, which is reflected irradiance from the ground surface or from other objects in the environment. Thus, $G = G_b + G_d + G_r$.

Radiation falling on an opaque surface is either reflected or absorbed. The proportions reflected and absorbed are measured by the surface qualities reflectance $\rho$ and absorptance $\alpha$. The sum of the two must be unity: $\rho + \alpha = 1$. For a surface exposed to radiation, the rate of energy input is $G\alpha$. As the surface is heated and becomes warmer than its surroundings, it will lose heat at a rate depending on the surface conductance ($h_o$) and the temperature difference between the surface ($T_s$) and the air ($T_a$): $h_o(T_s - T_a)$. The surface will reach an equilibrium temperature when the heat input rate equals the heat loss rate, that is, when $G\alpha = h_o(T_s - T_a)$, from which the surface temperature can be expressed as $T_s = T_a + (G\alpha/h_o)$. This is the maximum temperature the surface could reach in the absence of any heat flow into the body of material behind that surface, and it is often referred to as the sol-air temperature. (More precise definitions would include a radiant emission term, so that the numerator would become the net irradiance absorbed by the surface.)

If we have an absorber plate that incorporates fluid channels through which some liquid or air can be circulated, then the heat collected by the absorber plate can be removed and transported to storage or to a location where it is needed. Figure 1 shows some possible arrangements of such an absorber plate.

The flat plate collector consists of an absorber plate in a casing, incorporating back and edge insulation and one or two sheets of transparent cover (most often glass). The thermal processes in such a collector are illustrated in Fig. 2.
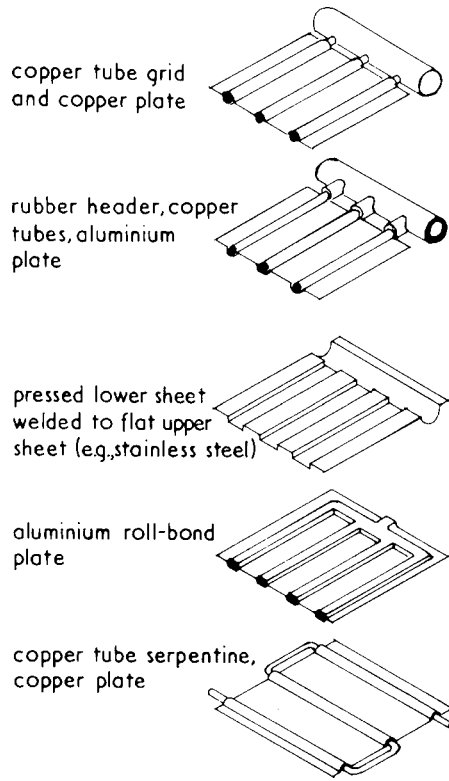
**FIGURE 1** Some absorber plate constructions.

The instantaneous collector efficiency is

$$\eta = \frac{\text{useful heat delivered}}{\text{irradiance on collector surface}} = \frac{Q_{\text{use}}}{GA_c}, \quad (1)$$

where $A_c$ is the collector surface (aperture) area (in square meters) and $G$ the global irradiance, because flat plate collectors can utilize the diffuse and reflected components, as well as the beam component (as opposed to optical
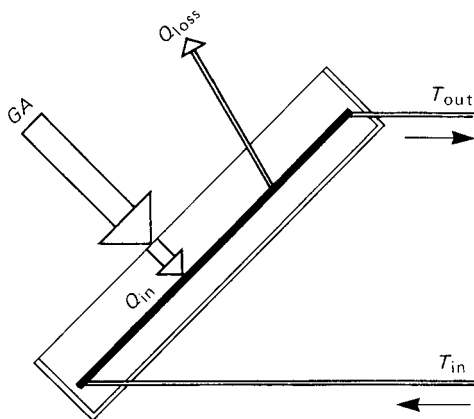


**FIGURE 2** Thermal processes in a flat plate collector.

concentrators, that is, lens or mirror-type collectors, which can only respond to the beam component).

$$Q_{\text{use}} = mC_\rho(T_{\text{out}} - T_{\text{in}}),$$

where $m$ is the fluid mass flow rate through the collector (in kilograms per second, $C_\rho$ the specific heat of fluid (joules per kilogram Kelvin), $T_{\text{out}}$ the output fluid temperature (in degrees Celsius), and $T_{\text{in}}$ the inlet fluid temperature (in degrees Celsius).

An analytical expression for useful heat delivered is

$$Q_{\text{use}} = Q_{\text{in}} - Q_{\text{loss}}, \quad (2)$$

where the heat input is

$$Q_{\text{in}} = GA_c(\tau\alpha)F \quad (3)$$

where $\tau$ is the transmittance of the glass cover, $\alpha$ is the absorptance of the surface, $F$ the absorber heat removal factor; the heat loss is

$$Q_{\text{loss}} = A_c U F \Delta T, \quad (4)$$

where $U$ is the global heat loss coefficient of collector per unit aperture area (in watts per square meter Kelvin), and the difference in temperature is $\Delta T = [(T_{\text{out}} + T_{\text{in}})/2] - T_a$.

Substituting Eqs. (3) and (4) into (2), and Eq. (2) into (1), we get

$$\eta = (\tau\alpha)F - UF\Delta T/G.$$

This is the collector efficiency function (Fig. 3) of type $y = abx$, where the two collector constants $a = (\tau\alpha)F$ and $b = UF$ are found by testing.
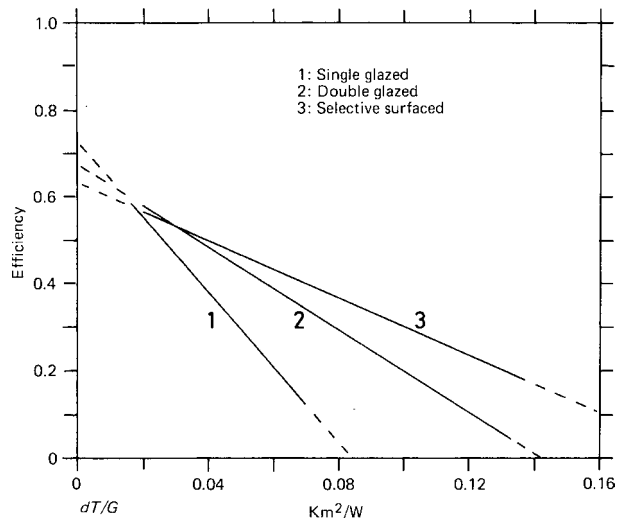


**FIGURE 3** Efficiency functions of three collectors: (1) single glazed, $\eta = 0.73 - 8.7\ (\Delta T/G)$; (2) double glazed, $\eta = 0.68 - 4.9\ (\Delta T/G)$; and (3) selective surfaced, $\eta = 0.63 - 3.3\ (\Delta T/G)$.

If $\Delta T = 0$ (i.e., there is no loss), then $\eta_o = (\tau\alpha)F$, which is the no-loss efficiency, the $y$-axis intercept. The collector stagnation temperature is reached when the efficiency is reduced to zero; thus

$$(\tau\alpha)F = UF\Delta T/G,$$

which is the $x$-axis intercept in Fig. 3.

Figure 3 shows three typical efficiency functions: (1) a single-glazed, (2) a double-glazed, and (3) a double-glazed selective surfaced collector.

For any surface, absorptance equals emittance for the same wavelength of radiation. For ordinary black surfaces, the absorptance for high-temperature ($\sim 6000°K$), short-wave solar radiation ($\alpha_{6000}$) is about the same as the emittance for long-wave radiation at normal operating temperatures around $350°K$ ($e_{350}$).

Special selective surfaces can, however, be produced that have a much greater $\alpha_{6000}$ value than $e_{350}$; therefore, the losses can be substantially reduced. In some cases the $a_{6000}/e_{350}$ ratio can be as high as 14, and reliable and durable selective surfaces with ratio of about 8 are commercially available. Table I summarizes the properties of some selective surfaces.

If selective surfaces reduce the radiant losses, the convective losses can be reduced by a convection suppression device, which is usually a transparent honeycomb structure (or lamellae) inserted between the glazing and the absorber surface. The outer surface of the glass may receive an antireflective treatment (to increase transmittance), while the inside surface can have an infrared reflective coating (to reduce outward radiation). Still further improvements can be achieved by enclosing the absorber in a vacuum, which eliminates all convective losses.

The structurally strongest shape is the tube; this led to the development of a variety of evacuated tubular collectors (ETCs), as shown in Fig. 4. These can produce temperatures in excess of $200°C$ and can provide good efficiencies up to about $150°C$.

Because improved collectors cost more, the choice must depend on a cost versus benefit evaluation. High efficien-
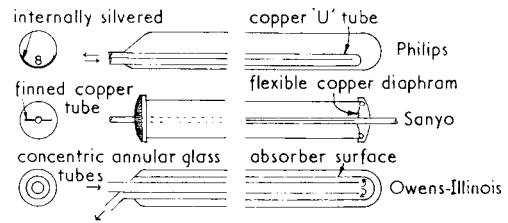


**FIGURE 4**  Three evacuated tubular collectors (ETCs).

cies can be achieved by inexpensive collectors at low temperature operation. If higher temperature ($>70°C$) heat is needed, then improved insulation (small $U$ value), double glazing, and a selective surface are warranted. Such improvements would be more beneficial in a cold situation, where the operational $\Delta T$ is large, but less so in a mild climate.

For best results the collector should generally face due south (or north in the southern hemisphere), but a deviation of about $20°$ from this either way would not make much difference. To get the maximum annual total irradiation, the collector should face the annual mean position of the sun, which means that the angle of tilt (from the horizontal) should be the same as the geographical latitude. In most places, however, less sun is available in winter; thus it is usual to give the tilt a winter bias, that is, to set the tilt as the latitude $+10°$ to reduce the difference between winter and summer performance.

## II. HOT WATER SYSTEMS

The simplest thermosiphon domestic hot water (DHW) system is shown in Fig. 5. As the water in the collector is heated, it expands, becomes lighter, moves upward, discharges into the top part of the storage tank, and draws in cooler, heavier water from the bottom of the tank. A natural thermosiphon circulation is thus developed, which gradually heats up the volume of water contained in the tank. It has been observed that this circulation exhibits

**TABLE I**  Radiation Properties of Some Selective Surfaces[a]

| Surface treatment | $\alpha_{6000}$ | $e_{350}$ | $\alpha_{6000}/e_{350}$ |
|---|---|---|---|
| Nickel black on galvanized steel | 0.81 | 0.17 | 4.8 |
| Copper black (NaOH + NaClO$_2$ treatment of copper) | 0.89 | 0.17 | 5.2 |
| Copper oxide (CuO surface conversion) | 0.93 | 0.11 | 8.5 |
| Copper oxide on aluminium | 0.93 | 0.11 | 8.5 |
| Chromium black electrodeposited on nickel plated steel | 0.95 | 0.09 | 10.5 |
| Nickel black on nickel plated steel | 0.94 | 0.07 | 13.4 |

[a] After Duffie, J. A., and Beckman, W. A. (1980). "Solar Engineering of Thermal Processes," Wiley (Interscience), New York.
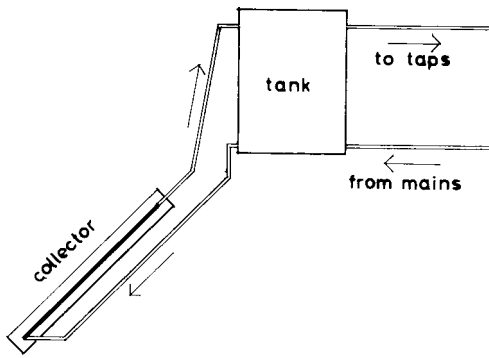
**FIGURE 5**  A simple thermosiphon domestic hot water system. (In terms of Table II: solar, direct, closed, filled, thermosiphon, recirculating system, with separate tank.)



**FIGURE 6**  A forced circulation drain-back system: when pump A stops, the content of the collector flows back into the tank.

self-regulating tendencies: the flow rate will adjust itself so that the temperature increment from collector inlet to outlet ($T_{out}-T_{in}$) is approximately 10 K, at least in the typical thermosiphon systems, and this holds true for widely varying temperature and radiation conditions.

This system requires that the tank be at a level higher than the top of the collector. In all cases, but especially if it is exposed to outdoor conditions, the tank must be very well insulated to prevent undue tank heat losses. If for some reason the tank cannot be placed in such an elevated position, then a small pump must be introduced to drive the fluid circulation. The pump must then be controlled, for example, by a differential thermostat. This will then be a forced circulation system, classified as active, while the thermosiphon system is accepted by most authors as a passive system.

The cold water feed can be provided by direct connection to the mains (in which case all components must be able to withstand mains pressure) or from a feeder tank (which is in an elevated position and filled from the mains through a float valve similar to the water closet-cistern float valve). For mains pressure systems most utilities or water supply authorities require the installation of a nonreturn valve. Whereas low pressure (feeder tank) systems are open to the atmosphere through a vent pipe, mains pressure systems must be protected against overpressure (due to thermal expansion of water as heated) by a pressure relief valve, as well as against collapse (implosion), which could be caused by a suction effect (vacuum formation) in case of an accidental water discharge, by an air intake (or breather) valve.

In areas where freezing can occur, the system must be protected from frost damage. In mild winter climates, where freezing occurs rarely, some heat may be sacrificed by a slow night circulation of water or by a small heating element built into the lower part of the collector. In colder climates drain-back systems can be used (Fig. 6), which
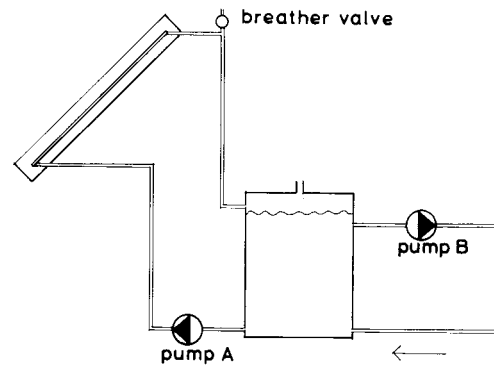
automatically empty the collector when the temperature drops to near freezing. Alternatively, an indirect system can be used (Fig. 7) in which the primary (collector) circuit is filled with an antifreeze solution (e.g., ethylene or propylene glycol) or with a nonaqueous liquid. This is a closed circuit and the heat is transmitted to the consumable water via a heat exchanger. If the antifreeze agent is toxic, then a double-walled heat exchanger may be necessary as a safety measure. Such control and safety devices can make an otherwise simple system quite complicated.

A solar DHW system could be designed to provide 100% of the hot water requirement (i.e., a solar fraction of 1), but it would not be economical except in the most favorable situations. Figure 8 shows the solar fraction as a function of collector area for a typical DHW installation, and it is a clear illustration of the law of diminishing returns. Most designers find the economic limit to be between a solar fraction of 0.6 and 0.8.

The remaining heating requirement must be supplied by an auxiliary heater, based on some form of conventional
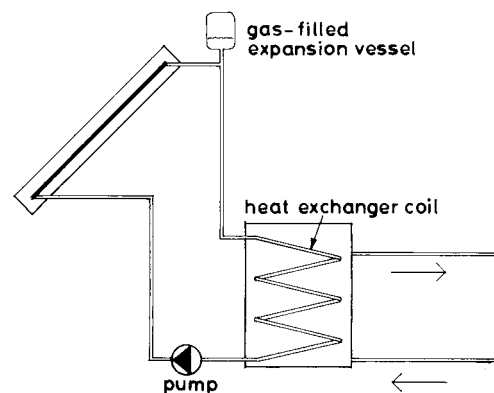


**FIGURE 7**  An indirect domestic hot water system. (In terms of Table II: solar, indirect, closed, filled, pumped, recirculating system, with separate tank.)
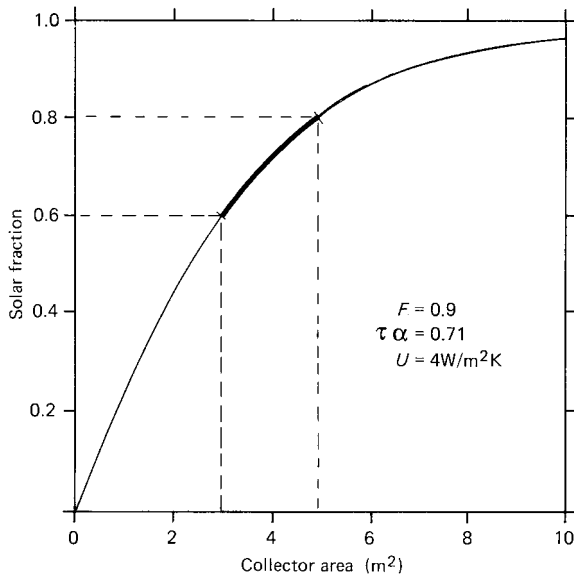
**FIGURE 8** Solar fraction as function of collector area (for 180 liter/day hot water use, in Brisbane, using a double glazed, selective-surfaced collector), an illustration of the law of diminishing returns.

source of energy, unless the users are willing to change their life styles and put up with lukewarm water at times or have a hot shower and do the washing only when solar hot water is available. The technically simplest form of auxiliary heat is the use of an electric immersion heater element built into the hot water tank itself. It is recommended that this should be positioned at one half to two thirds of the way up the tank so that it heats only the top third or half of the water volume. Should sunshine become available, the cooler water at the lower part of the tank remains ready for solar heating, so no heat is wasted.

Gas- or oil-fired heaters can also be used as auxiliary heat sources, but these would be external to the tank, circulating its content through the heater or having a separate heat transfer loop to supply heat to the hot water tank through a heat exchanger. From the point of view of environmental sustainability gas auxiliary heaters are much preferred as they cause much less $CO_2$ emissions.

An auxiliary heat source can also be installed as an after-heater. The solar part of the system would thus become a preheater. Only when draw-off occurs would the solar-heated water flow into the after-heater, at whatever the temperature it is, and be brought up to the required temperature level. This after-heater may be a flow-through-type (instantaneous) or a storage-type device. Under favorable conditions it may not be called on to operate, but it would serve as a standby unit. Table II shows a classification of solar DHW systems according to seven attributes.

The heating of swimming pools is an ideal application for solar energy. Because the temperature requirement is very low ($<30°C$), inexpensive often unglazed, collectors can be used. Several collector systems that use plastic or rubber materials, which can be fitted to an existing roof surface, are on the market. In almost all cases, these are active (pumped) systems, but generally the pool filter pump is sufficient to drive the circulation through the collectors.

**TABLE II  Classification of Solar Domestic Hot Water Systems[a]**

| Attributes | Descriptors | | |
|---|---|---|---|
| | a | b | c |
| 1. Energy source | Solar only | Solar preheat | Solar + auxiliary |
| 2. Coupling | Direct | Indirect | — |
| 3. Closure | Open | Vented | Closed |
| 4. Operation | Filled | Drain-back | Drain-down |
| 5. Circulation | Thermosiphon | Forced (pumped) | — |
| 6. Flow type | Recirculating | Series connected | — |
| 7. Storage | Remote (separate) | Close coupled | Integral |

[a] Any system should have one descriptor in each of the seven attribute categories.

## III. SPACE HEATING

Systems very similar to those just described can be used for the heating of buildings if scaled up by an order of magnitude. Whereas a DHW system for an average family house would have a collector of 3 to 5 m$^2$ and a tank of 220 to 350 liters, a system for space heating would be about ten times larger. With such a collector size, a thermosiphon system is rarely practicable, so pumps are almost always used; therefore, these are invariably classified as active systems. Figure 9 shows the basic system diagram, including the three main functions: collection (C), storage (S), and distribution (D) of heat.

The distribution system can use a range of different heat emitters, such as fan convector units, panel radiators, and embedded pipe-coil floor-warming systems. All these are identical to emitters used in conventional central heating
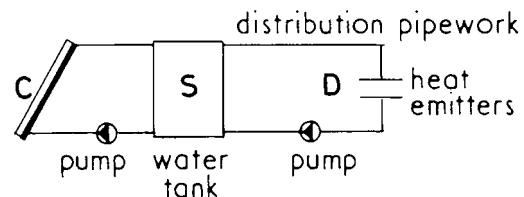


**FIGURE 9** Principles of a water-based (hydronic) solar space heating system. C, collection; S, storage; D, distribution.

systems. The last type is particularly attractive for solar operation because it requires only quite low temperatures (~30°C); therefore, higher collection efficiencies can be achieved.

Active solar space heating systems are rarely designed to give a solar fraction greater than about 0.7. The remaining heating requirement can be provided by an independent conventional heating system or by an auxiliary heater coupled with the solar system. The heat can be produced by electricity, by burning gas, oil, or coal, or even by using firewood. The coupling (as shown in Fig. 10) may be in (1) the storage vessel, (2) the flow pipe to the emitters (in series with the solar), or (3) a storage bypass of the emitter loop (parallel to the solar source). If there is a solar heating system, then the domestic hot water supply can also be coupled with this. The coupling must be an indirect one, that is, through a heat exchanger, because the heating system recirculates the water (or fluid) it is filled with, whereas the content of the DHW system is consumed and replaced by fresh cold water. There is an almost infinite range of possibilities for the combination of space heating, DHW, and auxiliary heating systems.

Air can also be used as the heat transfer fluid in collectors built for this purpose. The heat store in this case will most often be a rock bin or a rock bed, that is, an insulated container filled with crushed rocks or pebbles (called a bin if it is an upright container or a bed if it is a shallow layer under the floor). The solar heated air will warm the rocks, and at a later time the air flow in the distribution circuit will remove this heat and transport it to the spaces where it is needed. Figure 11 shows a basic air heating system diagram. Instead of small pipes, more bulky air ducts must be used, and instead of pumps, fans, or blowers. A much larger quantity of air must be circulated than the volume of
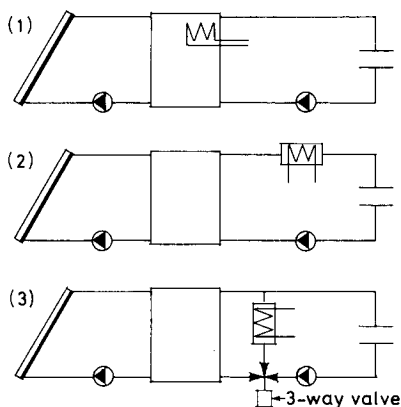
**FIGURE 10** Three ways of coupling an auxiliary heater with the solar space heating system: (1) in the storage tank, (2) in the distribution circuit flow pipe (series connection), and (3) in a storage bypass (parallel connection).
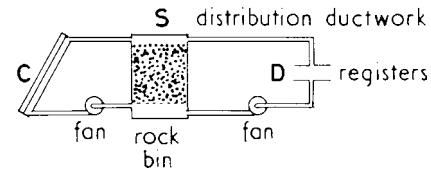
**FIGURE 11** Principles of a solar air heating system. C, collection; S, storage; D, distribution.

water that would be necessary to transport a given quantity of heat because air has a low specific heat capacity and has very low density. Thus, the fan power requirement will be greater than the pump power required for an identical performance. Efficiencies of air heater collectors are lower than of water heater collectors. In spite of these factors, air systems are attractive for many reasons:

1. There is no need for frost protection.
2. There is no need for the many safety devices.
3. Air leaks are not as damaging as are water leaks.
4. There is no need for heat exchangers and emitter devices (warm air can be discharged directly into the rooms).
5. The system is generally simple (and thus lower in cost).

The use of phase-change materials (PCMs) for heat storage in both water and air systems is now on the increase. These PCMs are usually eutectic salt solutions, but paraffin wax and various other mineral oil products are also used, which go through a phase change (liquid to solid and vice versa) at a set temperature. The latent heat of phase change can be used for heat storage without change in temperature, in addition to any sensible heat storage due to temperature changes. For the storage of a given amount of heat, the volume required will be only about one sixth or one eighth the volume of a water tank having a similar heat capacity. Such phase-change materials are marketed encapsulated in plastic containers of various forms and shaped to provide a heat exchanger surface in contact with water or air.

The last decade saw few radically new ideas but numerous refinements and improvements. Much attention has been directed at the development of manufacturing technologies to achieve better engineered and more cost-effective products.

In domestic hot water (DHW) systems the most popular model now is a close-coupled indirect system. The collector fluid is an aqueous glycol solution, with corrosion inhibitor added, which circulates through a heat exchanger jacket, surrounding the horizontal cylinder attached to the top of the collector. This system allows the use of ordinary steel sheet for the collector; thus, it is much less expensive than copper or stainless steel collectors.

In larger, pumped systems the variable volume/variable flow rate method ensures optimum operating conditions: the collector flow rate is increased as the solar radiation input increases; thus the output temperature remains near constant. If there is a hot water draw-off from the tank during a no-sun period (e.g., at night), the tank water level is allowed to drop, that is, no make-up water is admitted (which would reduce the tank temperature), until the next day's solar input period.

There are some doubts about the durability and cost-effectiveness of active solar space heating systems, but passive systems are generally recognized as worth having under almost any circumstances. A perennial concern is the optimization of solar window size. This provides the solar input, but it is also the biggest loser of heat during no-sun periods. Movable (night-) insulation was thought to be the only solution, but the systems so far developed are cumbersome to use or expensive or both. Great improvements have been achieved in recent years in transparent insulation materials and constructions. These can act as quasi-diodes: admitting solar radiation but preventing conductive heat loss. Several types are already commercially available, but the development work continues.

Much work has been done in recent years, especially in cold winter climates, on central (district-) space heating systems using seasonal storage. The potential of natural formations (caves, aquifers) for thermal storage has been proven.

## IV. THERMAL BEHAVIOR OF BUILDINGS

Buildings are exposed to two sets of thermal influences from the environment:

1. Air temperature, acting through heat conduction by the enclosing elements or convection through air exchange, that is, ventilation or air infiltration
2. Solar radiation, entering primarily through windows, but also warming the outside surfaces of solid enclosing elements and two sets of man-made influences
3. Incidental internal gains, that is, the heat produced by people, lights, and appliances
4. Deliberate heating or cooling

The designer's aim is to create a controlled environment in which the temperatures (as well as other environmental factors) are kept within narrow limits, as required by the occupants. Two sets of tools can be used to serve this aim:

1. Active controls, which are mechanical systems of varying complexity that provide heating or cooling and

rely on some form of energy supply (e.g., firewood, oil, or electricity); and
2. Passive controls, which are the controls of heat flow by the building itself through its careful design; these may include its orientation, shape, surface qualities, thermal properties of its fabric, and the control of windows and openings.

The terms active and passive controls differ from the terms active and passive solar systems. Passive controls may include passive solar systems, but also other thermal characteristics of the building. Active solar systems constitute only a small subset of active controls, a term that includes the whole range of conventional heating, ventilating, and air conditioning (HVAC) systems.

Climatic design is the manipulation of passive controls to create the best possible indoor conditions in a given climate and thus to eliminate (or reduce) the need for active controls and consequent energy consumption. In climates having outdoor air temperature higher than that desired indoors, both environmental influences tend to cause discomfort. Solar heat input should be eliminated or minimized and ways found to dissipate the unwanted heat. When the outdoor temperature is below the comfort limit and thus causes an outward (negative) heat flow, solar radiation can be used to provide a positive heat input. In very cold climates, all available means must be employed to minimize heat losses and maximize solar heat inputs. In moderate climates, the designer must carefully balance solar heat inputs against heat losses. Solar overheating can occur even in the winter and even at latitudes greater than $50°$.

## V. PASSIVE SYSTEMS

It is difficult to draw the line between a well-designed house and a passive solar one. A window oriented toward the equator (south in the northern hemisphere and north in the southern hemisphere) will admit a large amount of solar radiation. This will be absorbed by the floor and other internal surfaces, causing a temperature increase. Some of this heat will be transferred to the room air by convection or emitted as long-wave infrared radiation toward other (cooler) room surfaces. Window glass is practically opaque to such long-wave radiation, so it will be trapped in the room. (This is known as the greenhouse effect.) Another part of the absorbed heat may be conducted into the body of material behind the absorbing surface, elevating its temperature. This material will in effect store the energy as sensible heat. At a later time (e.g., at night) as the room cools down, the heat flow is reversed and the stored heat is re-emitted into the room (Fig. 12). This
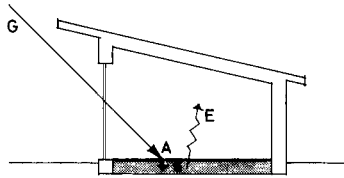
**FIGURE 12** Principles of a direct gain (passive) system. G, global irradiance; A, absorption; S, storage; E, emission.



**FIGURE 14** Principles of a mass wall (passive) system. Abbreviations as in Fig. 12.

mechanism is referred to as a direct-gain. system. A good direct-gain system will have large windows facing the equator and highly absorbent (dark) internal surfaces in areas reached by direct solar radiation. These dark-surfaced elements should have a thermal capacity sufficient to store the radiant heat input without an excessive rise in temperature. Dark colored quarry tiles or slates is a favored solution. [Thermal capacity is the product of the volume $V$, density $\rho$, and specific heat capacity $C_\rho$ of the material: $V \times \rho \times C_\rho$ (m$^3 \times$ kg/m$^3 \times$ J/kgK = J/K).]

Excessive summer heat input can be avoided by simple but carefully designed shading devices. The noon solar altitude angle on equinox dates (March 21 and September 23) is 90° minus the latitude (Fig. 13). If a canopy or awning above the equator-facing window is set to this angle, it will exclude all sun penetration for the summer half of the year (the sun moves up 23.5° from the equinox position) and admit an amount of solar radiation increasing to a maximum at mid-winter, when the sun is 23.5° below its equinox position. This provides an automatic seasonal adjustment, utilizing the sun's apparent movement. In a cooler climate the cut-off date can be changed to allow more heating, or in warmer climates to reduce the heat input. This mechanism works only for a window with an equatorial orientation.

One disadvantage of direct-gain systems (especially if the thermal capacity is inadequate) is the large diurnal (day and night) fluctuation of indoor temperatures. The reason for this is inherent in the system in that the absorber surface, which can reach quite high temperatures, is also the emitter surface. Therefore much heat is added to the room
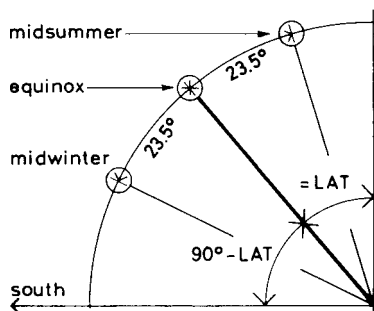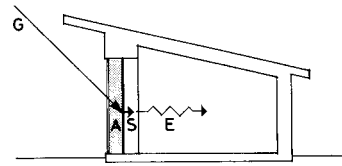
air simultaneously with the solar heat input and only the remainder is conducted into the storage mass. Delayed gain can be provided by so-called mass-wall systems. These are typically heavy masonry or concrete walls (Fig. 14) with a dark colored outer surface covered by glazing. The outer surface is the absorber and the inner surface is the emitter, so the heat flow is one directional and avoids the reversal of input/output heat flow of the direct-gain system. The inside surface will reach its peak temperature with a time lag dependent on the thermal capacity and conductivity of the wall (typically 6–10 hr), so the maximum heat input to the room can be provided at a time when it is most needed.

A variant of this is referred to as the water-wall system. Water in suitable containers is used as the thermal storage mass, but functionally it is the same as the masonry mass. There is a slight difference in behavior: whereas the masonry wall can show a steep temperature gradient through its thickness during the heating-up period, temperatures in the water container will tend to be more homogeneous. It is an efficient form of storage, but it gives a much shorter time lag than does solid masonry.

The Trombe wall (named after the French physicist Felix Trombe, who investigated the behavior of such walls) is a mass wall, with ventilator openings near both the floor and ceiling (Fig. 15). If these vents are open during the solar heating period, an upward air flow is generated in the space between the heated absorber and the glazing, discharging the heated air into the room through the top vent and drawing in cooler air from the room through the bottom vent (a thermosiphon air loop). This provides a heating effect simultaneously with the solar heat input. If the vents are closed, it behaves as does the mass wall. By regulating the opening and closing of the vents, the proportion of simultaneous and delayed gains can be adjusted at will.
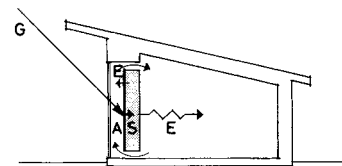


**FIGURE 13** Solar altitude angles at noon.



**FIGURE 15** Principles of a Trombe wall (passive) system. Abbreviations as in Fig. 12.
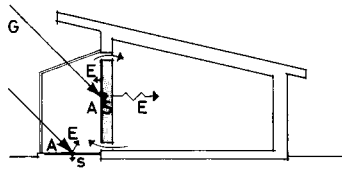
**FIGURE 16** Principles of an attached sun space (passive) system. Abbreviations as in Fig. 12.



**FIGURE 17** Principles of a water roof (Skytherm) system.

The attached sunspace system utilizes a separate space (e.g., a greenhouse or sun porch) for heat collection during the day. It is sometimes referred to as the indirect-gain system. In principle, it is the same as the Trombe wall, except that the approximately 100-mm space between the absorber and the glazing is extended to 2 m or more. There will be a large diurnal temperature fluctuation in this space; thus there should be a facility to isolate it from the room it serves (Fig. 16). During the day, the top and bottom vents in the dividing wall are open, allowing a warm air flow into the room. This can even be assisted by a small through-the-wall fan (this would be a hybrid system). At night the temperature in the sun space may drop, but then the vents are closed. The dividing wall can also act as the mass wall described earlier, giving a delayed heat input into the room.

By placing some mass (e.g., concrete blocks or water-filled vessels) in the sun space, its diurnal temperature fluctuation will be reduced, at the expense of heating the room. If this space is a greenhouse or conservatory, this is highly desirable, but it must be realized that there is a conflict of interest between the greenhouse use and the sun space as a heat source for the habitable room behind it. In many cases, greenhouses can become net losers of energy over the 24-hr cycle, but they are still popular as pleasant habitable spaces in their own right, albeit only during sunshine periods.

In all four systems mentioned, double (or even triple) glazing can be used to reduce heat losses, but it should not be forgotten that multiple glazing also reduces the transmission of incoming solar radiation. Perhaps a more effective solution is the use of movable night insulation, which would substantially reduce the heat loss rate for the larger part (up to 18 hr) of the winter day.

The last, and perhaps most attractive, generic system is the water-roof (Skytherm) system patented by Harold Hay (of Los Angeles). The movable insulation is the essence of this system, which can be used both for heating and cooling (Fig. 17). A metal roof deck becomes the ceiling, exposed internally, with water-filled plastic bags placed on it. Sliding insulating panels can cover the roof or can be removed and "parked," exposing the water bags. The seasonal operation is reversed: in winter, the water bags are exposed during the day and the water is heated by the sun;
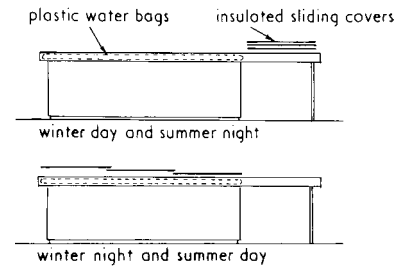
they are covered during the night, and the heat is preserved while radiant ceiling heating is provided; in summer, they are exposed overnight, and the water is cooled by convection and radiation to the sky; they are covered during the day, and solar heating is excluded while the cool ceiling provides a radiant heat sink and cools the room air by convection.

A barrier to greater acceptance of the water-roof system is the complicated sliding cover mechanism and the unusual character of the resulting building. In contrast, the direct-gain system enjoys the greatest popularity because it can be incorporated into any conventional building type. Mass-wall and Trombe-wall systems are relatively rare because they tend to dominate the architecture unless handled with some finesse. Probably the combination of direct gain (for instantaneous heat input) with some (non-ventilated) mass walls (for delayed gain) offers the most balanced performance as well as the greatest potential for architectural integration, for imaginative solutions.

## VI. SOLAR COOLING

It is a contradiction in terms to talk about passive solar cooling. Passive cooling techniques exist, but these are (if anything) anti-solar. Solar and other heat gains should be eliminated.

Night-time heat dissipation combined with some heat storage capacity can provide passive cooling. (If the mass is cooled down overnight, it will act as a heat sink during the day.) In dry climates evaporative cooling can be used to cool the ventilation air. Indirect evaporative coolers (Fig. 18) can extend the usefulness of this technique: the exhaust air may be cooled evaporatively and led through a heat exchanger, which will then cool the intake air without the addition of moisture. This could be considered as a hybrid system: the cooling is passive (evaporation) but it uses two electric fans and a small pump.

The stable (in hot periods, relatively cool) temperature of the ground can be made use of, either by direct coupling of the building with the ground (e.g., earth-covered or
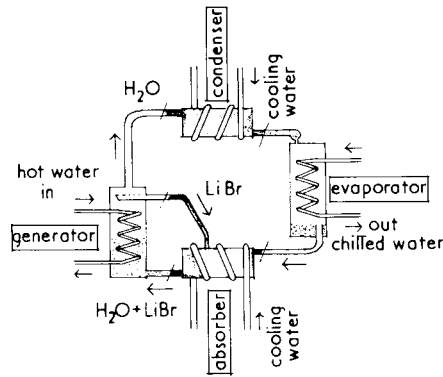
**FIGURE 18** An indirect evaporative cooler: white arrows show the passage of return air from the room which is evaporatively cooled by a spray in the vertical spaces of the plate heat exchanger, then discharged; solid arrows show the path of fresh air through the horizontal gaps of the heat exchanger, where it is cooled, then supplied to the room.



**FIGURE 19** Schematic diagram of an absorption chiller.

earth-integrated buildings) or by using underground pipes to cool the ventilation air. These are passive cooling methods, but not solar ones.

The use of solar energy for the purposes of cooling is an attractive proposition because in the annual cycle, the peak demand coincides with the maximum availability of solar radiation (in the summer). This is not so with solar heating, where the peak heat demand (in winter) occurs when there is little solar radiation available.

Active solar cooling and air conditioning systems can be of two kinds: closed or open cycle. Closed-cycle systems use absorption refrigeration machines (chillers). These are binary systems, using an absorbent and a refrigerant (as opposed to compression systems, which use a single fluid as refrigerant). Two systems are popular:

| Absorbent | Refrigerant |
|---|---|
| Water ($H_2O$) | Ammonia ($NH_3$) |
| Lithium bromide (LiBr) | Water ($H_2O$) |

These types of chillers are commonly used in gas or kerosene operated refrigerators. The difference is that here solar-heated water would be used to provide heat input in the generator (see Fig. 19). In the system shown the refrigerant is expelled from the rich solution and the weakened solution is returned to the absorber. The refrigerant vapor dissipates some of its heat to the environment (may be water cooled) and condenses. This liquid enters the evaporator through a throttle valve (or a high-resistance thin pipe), where, in the low-pressure chamber, it evaporates and takes on some heat from its environment (e.g., from a water jacket), thus it produces chilled water. The vapor then returns to the absorber and is absorbed by the ab-
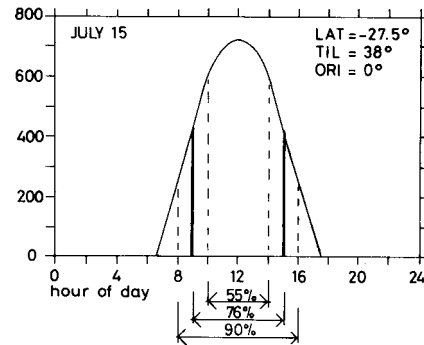
sorbent. This is an exothermic process, so heat is released to the environment.

Typically, the generator temperature needed is 80–90°C and chilled water of 8–10°C can be produced. This chilled water is then circulated through an air-to-water heat exchanger (e.g., a fan-coil unit) and will cool the air supplied to the rooms.

The coefficient of performance (CoP) of chiller systems is defined as

$$\text{CoP} = \frac{\text{rate of heat removal}}{\text{rate of energy input to drive the system}}.$$

While compression chillers can achieve a CoP of 2 to 4, absorption chillers rarely produce more than 0.5 to 0.6. Their use is justified only if inexpensive, low-grade (80–90°C) heat is available in ample quantities.

Solar systems can produce hot water at such temperatures, but either very good quality flat plate or evacuated tubular collectors (ETCs) must be employed; thus the collection system will be expensive. Numerous installations of such systems have been completed, many of which are technically successful, but very few (if any) would be judged as economically viable.

In humid climates the effectiveness of evaporative cooling is limited by the already high moisture content of the air and the supply air would also become unpleasantly humid. Moisture can be removed by either solid or liquid desiccants, the former in a slowly rotating wheel enclosed in a wire mesh, the latter in the form of a fine spray. These are incorporated in "open cycle" cooling systems. Several such systems are in the developmental stage, some components are already in use in conventional air conditioning systems.

Their name comes from the fact that air from the atmosphere is taken through the system and discharged to the rooms (so there is no closed circulation of some refrigerant, as in cooling machines). All variants rely on evaporative cooling and the removal of moisture by some sorbent material (desiccant), which will then be reactivated by
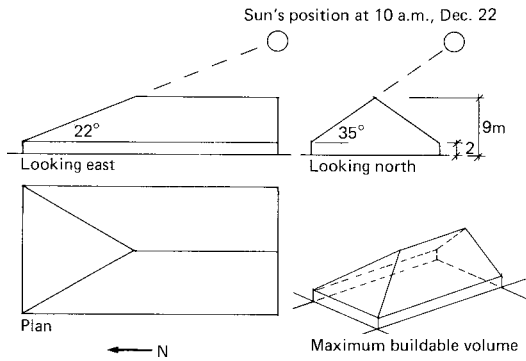
**FIGURE 20**  Two open cycle cooling systems.

solar heating. Figure 20 shows two such systems in diagrammatic form: the first with a moisture transfer wheel containing silica gel, the second using a glycol spray.

## VII.  DESIGN METHODS

The thermal design of any building, as well as that of distinct solar systems, must be based on some predictive model. The first approach to an evaluation of expected heat flows can be based on steady-state assumptions by taking the daily mean indoor-outdoor temperature difference and calculating the 24-hr average solar gain. This is a relatively simple calculation, but it only approximates reality. The outdoor temperatures vary between a maximum, shortly after midday and a minimum, normally just before sunrise. The solar heat input is usually restricted to less than one third of the day, increases to noon, then decreases to sunset; cloud conditions introduce a further variation. Indoor temperature requirements, as well as internal heat gains, may also vary with the time of day. Consequently, the time-dependent thermal response of the building should also be evaluated.

Many sophisticated mathematical models have been constructed and built into powerful computer programs for the simulation of the thermal behavior of buildings. These tend to use very detailed climatic data and analytically follow the heat flows through various building elements of differing exposure and construction. The time step used is often 1 hr, but it can be as short as 10 min, and simulations can be run for a full year to predict indoor temperatures (for a free-running building) or the energy requirements for systems to keep set indoor temperatures. These programs use hourly meteorological data, assembled to form a TMY (typical meteorological year) or TRY (typical reference year). The two major methods in use are based on finite difference and response factor methods, respectively.

In the United States the most sophisticated and most widely used such program is DOE2, which is now available also for PCs, in a Windows environment. In Europe the most refined one is ESPr (of the University of Strathclyde, Glasgow), but this is rather a research tool than a design tool. In Australia a comparable program is CheeNat of the CSIRO (Commonwealth Scientific and Research Organisation), used also as the simulation engine in the National House Energy Rating Scheme (NatHERS).

The previously cited are building thermal response, heating-, cooling load calculation, and energy use simulation programs; but as such they can also cope with most passive systems. For active solar systems the best simulation package is TRNSYS, a modular program developed by the University of Wisconsin.

For the practicing designer, the simplified methods are more suitable. These are based on correlations of major system parameters with the resulting performance, as predicted by multiple runs of some sophisticated program. The most widely used of these is the solar-load ratio (SLR) method, developed by the Los Alamos Scientific Laboratories and the unutilizability method of the University of Wisconsin. In Europe the Method 5000 is the most generally accepted, adopted as the minimum requirement for Commission of European Communities (CEC) projects.

For active solar systems, the f-chart method is generally used. All these can be used manually or with modest sized microcomputers. These methods predict the solar fraction, that is, the fraction of the total load (e.g., heating requirement) that is likely to be provided by the solar system. The simplicity of such methods allows the designer to examine many possible alternatives and select the most appropriate ones.

Multiple simulation runs are also used for parametric studies, in which one design variable at a time is systematically changed and the performance of the system is expressed as a function of that parameter. Sets of sensitivity curves are thus produced, providing useful advice to the designer.

With the dramatic increase in computing power of PCs and laptops a number of packages have been developed incorporating some sophisticated simulation system, but with front ends and outputs to suit the designers' or architects' way of thinking. The program ENERGY10 of the NREL (National Renewable Energy Laboratory, in Golden, CO.) or the IEA (International Energy Agency) package ISOLDE are perhaps the best known examples.

## VIII.  USER INFLUENCE

It is relatively easy (and reliable) to predict the thermal behavior of a building and its installations as physical

systems. Performance measurements in unoccupied buildings have been used for the validation of many computer-based simulation programs. However, as soon as people occupy these buildings, the predicted and measured results start to diverge. This divergence is particularly pronounced in the case of passive systems.

With active systems, the controls are mostly automatic, heat flows are well contained, and heat transfer processes are accurately predictable. With passive systems, heat flows are more diffuse and complex interchanges occur, which can be grossly influenced by user interference. Often the appropriate adjustment of controls (e.g., of shading devices or of Trombe-wall vents) by the user at the right time is required to achieve optimum performance of the system. Both the competence and the attitude of the user can drastically influence the results.

A passive solar house, given to an unsympathetic (possibly even hostile) user or a user lacking an understanding of the system, can use more energy than a conventional house with an ordinary fuel-based heating system. The most successful solar houses are the ones designed and built by owner-occupiers. These owners are intimately involved with the system, are capable of adjustments, maintenance, and even modification of the system, and are often emotionally committed to its success.

For tenanted houses, an instruction manual should be provided, but preferably also an explanation and demonstration should be given. Perhaps the financial incentive (saving on fuel bills) may be enough to obtain the tenant's cooperation.

## IX. ECONOMICS

If an average conventional house is taken as the base case and its annual energy use taken as 100%, it is suggested that about 30% of this can be saved by no-cost improvements and by the investment of a little thought and knowledge at the design stage. An increased expenditure at the design–construction stage, carrying the energy conservation measures to their full potential, would increase this saving to about 50%. If explicit passive solar systems are also built into the design, the average saving could be 80% and, in favorable situations, even higher.

For the evaluation of such additional investments, the concept of pay-back period is often used. The simple pay-back period ($P_{PB}$) is

$$P_{PB} = \frac{\text{annual benefit}}{\text{extra cost}} = \frac{B}{C}$$

in terms of years. The annual benefit is taken as the solar contribution times the cost of conventional (auxiliary) energy, which would otherwise be used. A more accu-

rate definition would compare the present worth of future savings with the extra cost.

The length of an acceptable pay-back period depends on financing arrangements. With a passive system, it may be the period of the mortgage loan (20–25 years). With an active system, it should not be more than the life expectancy of the collector array (18–20 years), but the average householder would not consider a period greater than 10 to 15 years. In a commercial or industrial situation, pay-back periods of 3 to 5 years are often demanded.

## X. URBAN PROBLEMS

In most countries it is customary to build the front of houses facing the street. The front of a house generally means the location of the largest windows of the primary habitable rooms. This may give good results for plots on the north side of east-west-running streets. Because a southern or near-southern orientation is best in the northern hemisphere, this should be considered at the stage of subdivision and street layout. So far, most solar buildings have been erected on selected (therefore favorable) sites in low density areas. If passive solar design is to become the norm (rather than the exception, as at present), then all new subdivisions should be designed to permit this.

Some changes in attitude may also be necessary, such as the acceptance of buildings aligned other than parallel to the street, giving a stepped (sawtooth) streetscape, or the acceptance of major windows facing the back. If higher densities are aimed at, the streets must run east–west or nearly so. Wider plots are necessary for north–south streets to allow sideways orientation of houses. With diagonally running streets (north–west–southeast or northeast–southwest) wider plots are necessary to permit a "twisted" positioning of the house on the site.

One of the major problems is that of avoiding overshadowing of solar collectors or solar apertures by other buildings, that is, undisturbed solar access must be ensured. The greater the building density, the more difficult this problem will be. Much work has been done on this subject from the legal and urban planning points of view, and attempts have been made in some areas to create legislation for the protection of solar access. Perhaps more significant (at least so far) are the few existing solar village projects produced by individual initiative and effort. Davis, California, is a notable example. The whole area is run on a cooperative basis and it seems to work well both physically and socially.

Several authors have suggested that the best way of ensuring solar access is by use of a system of covenants rather than legislative measures. Whichever method is adopted,
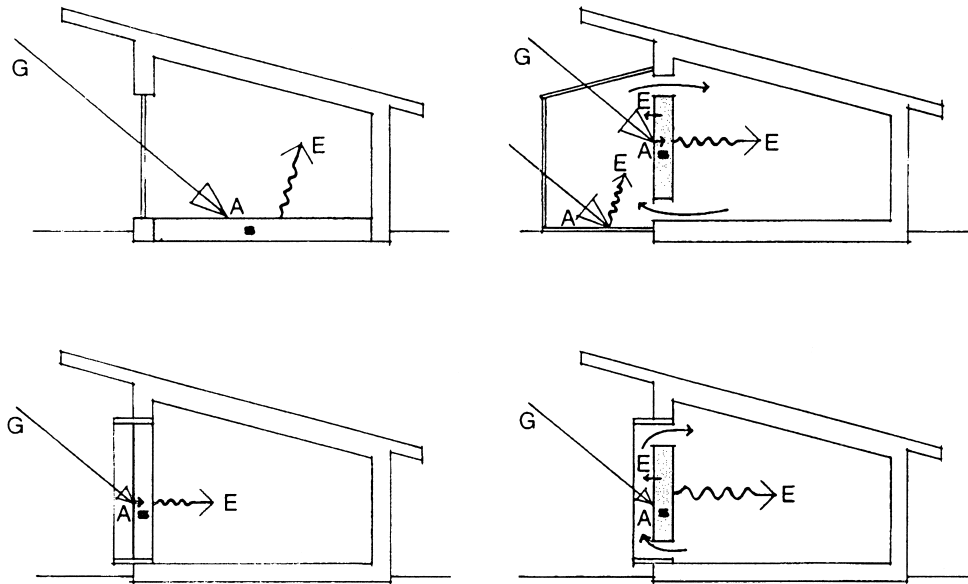
**FIGURE 21** An example of daily irradiation of a collector (day's total $= 4853$ Wh/m$^2$): between 10 a.m. and 2 p.m., 2681 Wh/m$^2 = 55\%$. between 9 a.m. and 3 p.m., 3689 Wh/m$^2 = 76\%$; and between 8 a.m. and 4 p.m., 4375 Wh/m$^2 = 90\%$ of total.

the technical questions are as follows: what are the criteria, what is the physical basis of measurement, and what degree of solar access is taken as reasonable without unduly restricting development?

A roof-mounted solar collector is not as vul-nerable to overshadowing as a solar window of a passive system, which is near ground level, but ensuring total solar access at all times, even for this, would be unreasonable. With low sun angles (just after sunrise and just before sunset), very long shadows are cast. However, the irradiance at such times is very low, therefore an obstruction would not cause a great loss.

Some authors suggest the criterion of no overshadowing between 9 a.m. and 3 p.m. (i.e., 3 hr on either side of solar noon). Others are more demanding, suggesting 4 hr before and after solar noon (i.e., 8 a.m. to 4 p.m.); yet others are more permissive, with limits of 10 a.m. and 2 p.m. (i.e., $\pm 2$ hr). All of these limits would apply to the mid-winter day. Figure 21 shows a typical irradiance curve for a north-facing collector (located in the Southern Hemisphere) with a 9 a.m. and 3 p.m. cut-off (other periods are shown by dashed lines). It is obvious that most of the radiation is received around solar noon, and the fraction lost (outside these time limits) is not excessive.

Two methods may be used for measurement as a basis for enforcement: the solar envelope and the solar window methods. The former would define the maximum build-able volume possible on any site without casting undue shadows outside the site boundaries (Fig. 22). The latter would define in three-dimensional terms the solid angle subtended at an existing (established) solar device that should not be penetrated by any obstruction (Fig. 23). The former is more generally applicable, giving protection not only to existing solar users but also to any future buildings.

## XI. RECENT DEVELOPMENTS

The earlier enthusiasm for solar cooling and air condition-ing systems has subsided: in spite of the improvement in CoP to 0.9 or even 1, they are often not only expensive but also use so much parasitic energy that there would be no pay-back period even in energy terms. With the im-provement and cost reduction of solar photovoltaic cells, it may now be more effective to use such PV cells to gen-erate electricity, which then drives a compression chiller. With direct coupling of electrical output to the compressor
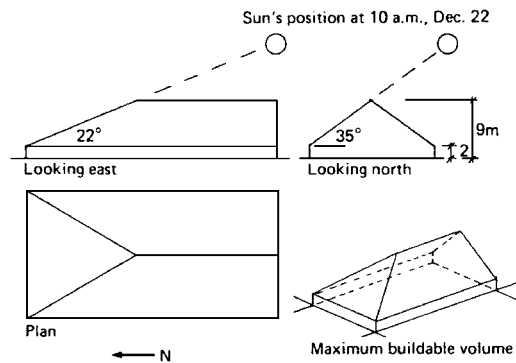


**FIGURE 22** Solar envelope for a $20 \times 40$ m building site to ensure solar access of adjoining properties between 10 a.m. and 2 p.m. on a mid-winter day.
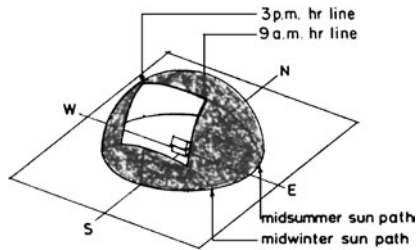
**FIGURE 23** Solar window (or unobstructed sky space) to ensure solar access for an established device between 9 a.m. and 3 p.m. throughout the year.

motor, most cooling would be provided when the solar radiation is strongest, A system with refrigerant storage may altogether eliminate the need for battery storage of electricity, which would otherwise be a large cost component of solar–electric systems.

Passive systems, on the other hand, are gaining grounds. More and more architects, designers and builders understand and accept that every house is a solar house, every house can make use of solar energy to a greater or lesser extent. Correct orientation is recognised as the most important decision, which attracts no extra cost.

The design of the building envelope also shows an improving trend. Much more attention is paid to insulation and the prevention of thermal bridge effects. Perhaps the greatest development occurred in windows. It used to be accepted that the window is the weakest point of the building envelope. In many climates the solar gain through a window was less than the heat loss.

While a moderately good wall would have a $U$-value (thermal transmittance of 1 W/m$^2$K, a single glazed aluminium window would be over 6 W/m$^2$K. With conventional double glazing units this may be reduced to 3 W/m$^2$K. Low emittance coating added gives a further reduction to 1.7 and if the cavity of the glazing unit is filled with argon (instead of air) the result is about 1.6 W/m$^2$K. Triple glazing, with two surfaces low-e coated and the unit filled with krypton: the $U$-value can be as low as 0.85 W/m$^2$K, if the frame is also improved with thermal breaks or if the metal is replaced by vinyl. These are commercially available products. Experimentally, with a multilayer "super"-window 0.63 W/m$^2$K has been achieved (two outer glass layers and two internal low-e coated plastic films). A double glazing unit, with low-e coating and vacuum between the glass sheets (which necessitates the use spacer studs) reached so far the best results: 0.56 W/m$^2$K.

As the solar radiation transmittance of such multi-layer glazing is still 0.5–0.6, the designer can successfully use large (well oriented) windows without unduly increasing heat loss.

Another area of rapid development is building-integrated photovoltaics (BIPV). With stand-alone PV systems the "balance of system" costs (mainly the support structure and the storage batteries) is practically as much as the PV cell arrays. The support structure cost is saved if the arrays are mounted on a building and there is no need for batteries if the system is grid-connected. Building elements are produced to incorporate PV cells and are used as for example roofing tiles or spandrel panels in curtain walls. They are replacing nonsolar elements, which gives a cost saving.

Germany started the "thousand solar roofs" project in 1990 and by 1995 some 2250 roof installations were completed, each between 1 and 5 kW, giving a total installed capacity of over 6 MW. Switzerland was not much behind. The European Commission installed over 11 MW PV capacity since 1979 and the Netherlands alone (Shell) has an annual production output of 20 MW. A new German plant has an annual output of 25 MW.

The United States launched a "million solar roofs" initiative in June 1997, which is to be completed by 2010 (this includes both PV and solar heater panels).

An increasing number of projects in the UK, Germany, Switzerland, the Netherlands, Denmark, Sweden, Japan, the United States and several other countries use PV panels as wall cladding on large scale, often multistorey buildings, referred to as "PV facades."

Two directions of development represent two entirely different philosophical attitudes to sustainable development. One is self-sufficiency, aiming for the "autonomous house," which produces its own energy requirements (as well as rainwater collection, waste water reuse, vegetable gardening, etc.). The other is acknowledging interdependence, the benefits of connecting to the electrical grid, of water supply and other public services and attempts to contribute to sustainability at the level of society as a whole. The utilization of solar energy in buildings is an important element in both.

## SEE ALSO THE FOLLOWING ARTICLE

Energy Efficiency Comparisons Among Countries • Energy Resources and Reserves • Photovoltaic System Design • Radiation Physics • Simulation and Modeling • Ventilation, Industrial

## BIBLIOGRAPHY

Anderson, B. (1977). "Solar Energy Fundamentals in Building Design," McGraw-Hill, New York.

Department of Energy (1980–1983). "Passive Solar Design Handbook," Vols. I–III, U.S. Department of Energy, Washington, DC.

Duffie, J. A., and Beckman, W. A. (1980). "Solar Engineering of Thermal Processes," Wiley (Interscience), New York.

Givoni, B. (1994). "Passive and Low Energy Cooling of Buildings," van Nostrand Reinhold, New York.

Hastings, S. R. *et al.* (1994). "Passive Solar Commercial and Institutional Buildings," I. E. A.—John Wiley & Sons, Chichester.

Mazria, E. (1979). "The Passive Solar Energy Book," Rodale Press, Emmaus, PA.

Szokolay, S. V. (1975/1977). "Solar Energy and Building," Architectural Press, London.

Szokolay, S. V. (1980). "World Solar Architecture," Architectural Press, London, and Wiley (Halsted), New York.

Tabb, P. (1984). "Solar Energy Planning," McGraw-Hill, New York.

Watson, D., and Labs, K. (1983). "Climatic Design," McGraw-Hill, New York.
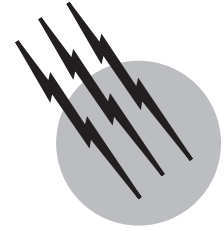
# Thermionic Energy Conversion

**Fred Huffman**

*Thermo Electron Corporation*

## GLOSSARY

**Arc drop**  Minimum electrostatic potential difference that is required to maintain a plasma in an ignited discharge.

**Barrier index**  Figure-of-merit parameter for characterizing the performance of a thermionic converter that is given by the sum of the arc drop and the collector work function, thus accounting for the plasma plus the collection losses. The lower this parameter, the higher is the converter performance.

**Collector**  Cool electrode in a thermionic converter that collects the electron current.

**Emitter**  Hot electrode in a thermionic converter from which the electron current is emitted.

**Fermi level**  Characteristic energy of electrons in a material under thermal equilibrium at which the probability of an allowed quantum state being occupied by an electron is $\frac{1}{2}$.

**Plasma**  Ionized gas, containing approximately equal concentrations of positive ions and electrons, that is electrically conductive and relatively field free.

**Space charge**  Negative electrostatic barrier formed as a consequence of the finite transit time of the electrons crossing from the emitter to the collector.

**Work function**  Heat of electron vaporization (and condensation) from a material given by the difference in electrostatic potential energy between its Fermi level and a point just outside its surface.

**A THERMIONIC** energy converter transforms heat into electricity by evaporating electrons from a hot emitter and condensing them on a cooler collector. This device is characterized by extremely high operating temperatures, no moving parts, modularity, and the capacity to operate inside the core of a nuclear reactor. Thermionic power generators have been built utilizing combustion, solar, radioisotope, and reactor heat sources. Although

commercial applications have yet to be realized, thermionic energy conversion has the potential to (1) produce power in space reliably using nuclear reactors, (2) cogenerate electric power for a large fraction of industry unsuited to other energy converters, and (3) increase substantially the efficiency of utility power plants. Steady technical progress is bringing these possibilities closer to reality.

## I. INTRODUCTION

### A. Elementary Description

A thermionic converter is illustrated in Fig. 1. It consists of a hot electrode, or emitter, separated from a cooler electrode or collector by an electrical insulator. The spacing between the emitter and collector is usually a fraction of a millimeter. Electrons vaporized from the emitter cross the gap, condense on the collector, and are returned to the emitter by way of the electrical load.

The converter enclosure is hermetically sealed so that the atmosphere between the electrodes can be controlled. In the conventional thermionic converter, the interelectrode space is filled with cesium vapor from a liquid reservoir at a pressure of ∼1 torr. The cesium performs two functions. First, it adsorbs on the electrodes to provide the desired electron emission properties. Second, it provides positive ions to neutralize the electron space charge so that practical current densities can be obtained from the converter.

The thermionic converter is a heat engine utilizing an electron gas as the working fluid. Thus, its efficiency cannot exceed that of the Carnot cycle. In effect, the

temperature difference between the emitter and collector drives the electrons through the load.

For a given set of electrodes, the output power of the thermionic converter is a function of the emitter temperature, collector temperature, interelectrode spacing, and cesium pressure. Characteristically, thermionic converters operate at high electrode temperatures. Typical emitter temperatures range between 1600 and 2400 K and collector temperatures vary from 800 to 1100 K to provide a high current density (say, 5–10 A/cm$^2$) at an output potential of ∼0.5 V per converter. The efficiency of converting heat to electricity is usually between 10 and 15%.

### B. Historical Background

The principle of thermionic conversion derives from Edison's discovery in 1885 that current could be made to flow between two electrodes at different temperatures in a vacuum. The analysis and experimental investigation of thermionic emission from a hot electrode were performed by O. W. Richardson in 1912. W. Schlichter in 1915 recognized this means of converting heat to electricity. A patent was submitted on this topic in 1923. I. Langmuir and his associates characterized the electron and ion emission from cesium-adsorbed films on tungsten in the 1920s.

There was some interest in thermionic conversion in Russia in the 1940s. In particular, A. F. Ioffe discussed the thermionic converter as a "vacuum thermoelement" in 1949. Analytical studies of thermionic conversion that neglected the effects of space charge and collector temperature were published in the early 1950s. The first thermodynamic analysis of a thermionic converter was given by G. N. Hatsopoulos in 1956.

In the late 1950s a remarkable metamorphosis took place when several groups in the United States and Russia (working, with few exceptions, independently and without knowledge of previous or concurrent efforts) achieved efficiencies of 5 to 10% and electrical power densities of 3 to 10 W/cm$^2$. V. C. Wilson pioneered investigations of ignited-mode converters.

During the 1960s and early 1970s, thermionic technology development in the United States and Russia was concentrated on space reactor power systems. Remarkable progress was made in both countries during this period. In 1973, the United States terminated its space reactor program. However, Russia continued its efforts and by 1977 had tested four TOPAZ thermionic reactors.

The demise of the U.S. space thermionic program coincided with the OPEC oil action. These events provided the motivation for transferring the thermionic technology that had been developed for space to potential terrestrial applications that could conserve premium fossil fuels.
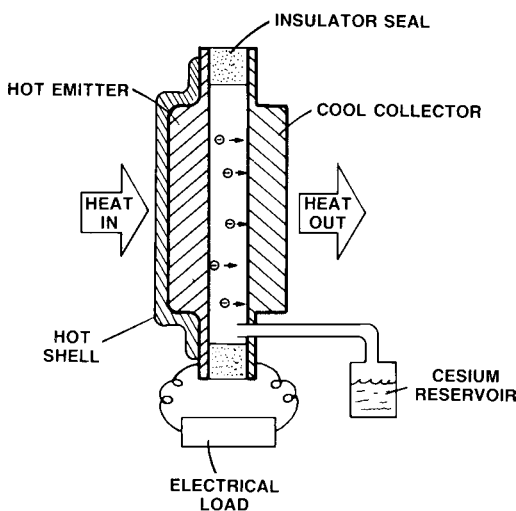


**FIGURE 1** Thermionic energy converter.

By the early 1980s, the wheel had made another cycle, and the U.S. combustion thermionic conversion program was redirected to nuclear space activities. By 1985, only a token terrestrial thermionic effort existed.

## II. FUNDAMENTAL PROCESSES

### A. Surface Phenomena

#### 1. Fermi Level

To a first approximation, the long-range forces between electrons in a metal are neutralized by the positive ion cores of the atoms. This implies that the metal lattice can be thought of as a box in which electrons are free to move. The number of free electrons is taken as the number of atoms times the valence of each atom. These valence electrons are responsible for electrical conduction by the metal; hence, they are termed conduction electrons, as distinguished from the electrons of the filled shells of the ion cores.

In the condensed state, the valence electrons are associated with the quantum states of the entire crystal rather than states of individual atoms. By applying the rules of statistical thermodynamics for spin $\frac{1}{2}$ particles (i.e., Fermi–Dirac statistics) the electron energy distribution $f(E)$ in a metal at a temperature $T$ is given by the expression for the Fermi factor,

$$f(E) = \frac{1}{1 + \exp[(E - E_F)/kT]}, \tag{1}$$

where $k$ is Boltzmann's constant ($k = 8.62 \times 10^{-5}$ eV/K) and $E_F$ is an energy characteristic of the metal called the Fermi energy. It is numerically equal to the energy for which the Fermi factor is 0.5. At $T = 0$, $E_F$ represents the highest energy of the electrons in the metal.

#### 2. Work Function

A diagram of the energy levels in a conductor and the potential energy of an electron near the surface is shown in Fig. 2. To remove an electron from the metal, it must have an energy above the Fermi level by the amount of $\phi$ eV; this parameter is defined as the work function. The work function can be thought of as the electron heat of vaporization. If an electron is introduced into the metal from the surface, a quantity of heat, $\phi$ eV, will be given up. Thus, the work function is also the electron heat of condensation. It is convenient to think of the work function as a potential barrier between the inside of the metal and a point just outside (say, 500 Å) the metal surface that must be surmounted by an electron in order to escape from the metal.

Work function values vary between 4 and 5 eV for most materials. However, they can range as widely as 1.8 eV for cesium to 5.5 eV for platinum. Considerable variation of work function values for a given material can be found in the literature. These variations are due to impurities, surface films, nonuniform surfaces, and measurement techniques.

#### 3. Thermionic Emission

Thermionic emission from a metal can be explained in terms of Fig. 2 and Eq. (1). At room temperature, few of the quantum states above the Fermi level will be filled. However, as can be seen from the Fermi factor expression, there will always be some electrons occupying the high-energy states as long as $T > 0$. Due to their random motion inside the metal, many electrons will impinge on the surface. Those electrons with kinetic energies greater than the work function barrier may escape the metal. Those that cross the boundary at the surface will transform part of their kinetic energy (i.e., $\phi$ eV) into potential energy.
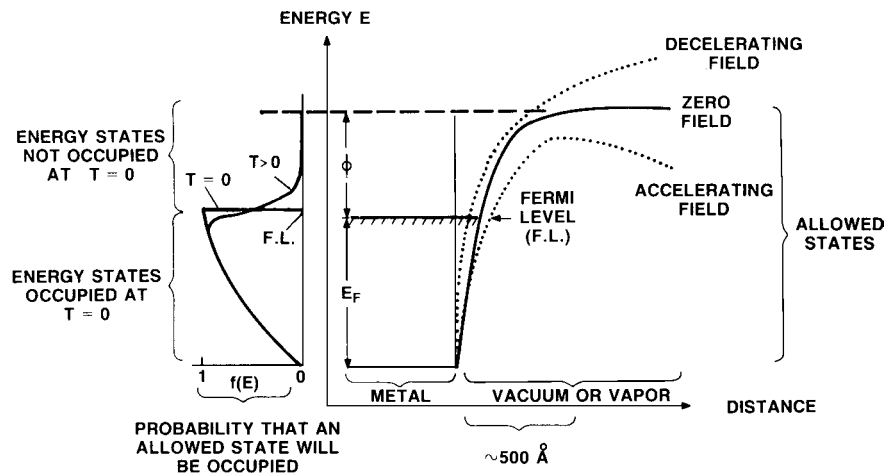


**FIGURE 2**  Energy levels near a metal surface.

It can be shown, either from thermodynamic considerations or from the application of statistical mechanics in connection with the quantum mechanics of electrons in metals, that the current density $J$ of electrons emitted from a uniform surface of a pure metal of absolute temperature $T$ can be expressed by the Richardson equation,

$$J = (1 - r)AT^2 \exp(-\phi/kT), \qquad (2)$$

where $A$ is a constant and $r$ is the electron reflection coefficient of the surface (of the order of 0.05). The coefficient $A$ is composed of a combination of fundamental physical constants,

$$A = 4\pi mk^2 e/h^3 = 120 \, \text{A/cm}^2 \, \text{K}^2, \qquad (3)$$

where $e$ is the electronic charge, $m$ is the electronic mass, and $h$ is Planck's constant. If the reflection coefficient is assumed to be zero, it is convenient to express the Richardson equation as

$$J = 120 \, T^2 \exp(-11606 \, \phi/T), \qquad (4)$$

where $J$ is the current density in amperes per (centimeter)$^2$, $\phi$ is the work function in electronvolts, and $T$ is the electrode temperature in kelvins. The value $J$ is called the saturation current density and corresponds to a zero electric field (horizontal potential distribution in Fig. 2). A strong applied electric field changes the electron emission, because its effects are superimposed on those of the image force, and alters the shape of the potential distribution outside the electrode. The dotted-line distributions in Fig. 2 correspond to electron-accelerating and -decelerating fields. Equation (4) shows that the saturation current density of a surface increases rapidly with increasing temperature and decreasing work function.

### 4. Cesiated Surfaces

Thus far, all practical thermionic converters have used cesium vapor because of the remarkably fortunate combination of three major phenomena. First, Cesium atoms are adsorbed on the emitter and collector surfaces to reduce the work functions of these electrodes to values favorable for energy conversion. No other electrodes have been found that provide long life at converter temperatures with work functions suitable for energy conversion. Cesiated electrodes are quite stable since the adsorbed coverages are due to the equilibrium between the evaporating and impinging atoms. Second, some of the adsorbed cesium atoms are evaporated as positive ions, which contribute to reducing the electron space charge between the emitter and collector. Third, electron collisions with cesium atoms in the interelectrode space provide cesium ions to neutralize the space charge. Indeed, in most thermionic converters, this is the primary ion source for this purpose.

## B. Plasma Considerations

A plasma is a relatively field-free region of an ionized gas that electrostatically shields itself in a distance that is small compared with other lengths of physical interest. The space charge or strong field regions on the boundary of the plasma are called sheaths. The distance over which the potential of a charge immersed in a plasma is reduced to a negligible amount can be shown, using Poisson's equation, to be

$$\lambda = \sqrt{\varepsilon_0 kT/2N_e e^2}, \qquad (5)$$

where $N_e$ is the electron concentration and $\varepsilon_0$ is the permittivity of free space. This distance is called the Debye shielding distance. It gives the approximate thickness of the sheath that develops when the plasma contacts an electrode. Without such a sheath, the plasma would lose electrons much more rapidly than positive ions because of the greater electron velocity.

A plasma is in thermal equilibrium when all species (electrons, ions, and atoms) have the same average energy. At high particle densities, the collision frequency is high enough for this condition to exist. However, at the low pressures (order of 1 torr) at which a conventional thermionic converter operates, the collision frequency for electrons is low enough that they can gain significant energy between collisions. Under converter conditions, the electron temperature $T_e$ may exceed the atom temperature $T_a$ by a large factor. Typically, $T_e$ is of the order of 3000 K, whereas $T_a \simeq 0.5 \, (T_E + T_C)$. A plasma cannot greatly exceed the density at which as many ions recombine locally as are produced locally. In this condition, known as local thermodynamic equilibrium, the plasma properties are equivalent to those that would exist in equilibrium with a hypothetical surface at $T_e$ and emitting a neutral plasma.

## III. BASIC PRINCIPLES OF THERMIONIC CONVERSION

An idealized potential diagram of a thermionic energy converter is shown in Fig. 3a. This diagram shows the spatial variation of the electrostatic potential perpendicular to the electrodes. Since the potential energy of the electrons in the collector is greater than that in the emitter, the collected electrons can perform work as they flow back to the emitter through the electrical load. The load voltage $V$ is given by the difference in the Fermi levels between the emitter and collector. The emitter and collector work functions are denoted by $\phi_E$ and $\phi_C$, respectively. The current density of the thermionically emitted electrons is given by the Richardson equation. The output power density of the converter is the product of $V$ times the load current
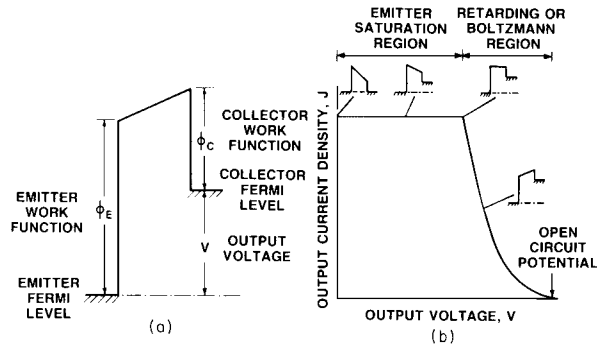
**FIGURE 3** Ideal-mode converter. (a) Motive diagram; (b) current–density voltage characteristic.

density $J$. The $J$–$V$ characteristic corresponding to the "ideal-mode" potential diagram given in Fig. 3a is shown in Fig. 3b.

An analogy can be drawn between thermionic conversion and the conversion of solar heat to hydropower. In this analogy, water plays the role of electrons, the sea surface corresponds to the emitter Fermi level, the lake surface corresponds to the collector Fermi level, and the gravitational potential corresponds to the electrical potential. As water is vaporized from the sea by heat, the water vapor migrates over the mountains to a cooler region, where it condenses and falls as rain into a lake at a high altitude. The water returns through a turbine to the sea to complete the cycle by which solar energy is converted to hydropower.

## A. Ideal Diode Model

The ideal diode thermionic converter model corresponds to a thermionic converter in which the emitter and collector are spaced so closely that no collisional or space charge effects take place. To reduce the complexity of the equations, ion emission effects will also be neglected. Although these assumptions do not strictly correspond to any thermionic converter, they do approach those of a very closely spaced diode operating in the vacuum mode. The ideal diode model defines the performance limit imposed by essential electron emission and heat transfer and provides a basis for comparison with practical converters.

In Fig. 3a, the energy $V + \phi_C$ must be supplied to an electron to remove it from the emitter and transport it across the gap to the collector. The electron gives up heat in the amount of $\phi_C$ when it is collected. The collected electron has an electrostatic potential $V$ relative to the emitter. The electron can then be run through an electrical load to provide useful work output on its route back to the emitter.

The current density $J$ through the load consists of $J_{EC}$, flowing from emitter to collector, minus $J_{CE}$, flowing from collector to emitter. Thus,

$$J = AT_E^2 \exp\left(-\frac{V + \phi_C}{kT_E}\right) - AT_C^2 \exp\left(-\frac{\phi_C}{kT_C}\right) \quad (6)$$

for $V + \phi_C > \phi_E$. As the output voltage is reduced such that $V + \phi_C < \phi_E$,

$$J = AT_E^2 \exp\left(-\frac{\phi_E}{kT_E}\right) - AT_C^2 \exp\left(-\frac{\phi_E - V}{kT_C}\right) \quad (7)$$

In Fig. 3b, the saturation region of the $J$–$V$ curve corresponds to negligible backemission. The open-circuit voltage is given by setting $J = 0$ and the short-circuit current density for the case of $V = 0$. Equations (6) and (7) can be used to construct $J$–$V$ curves for a range of $\phi_E$, $\phi_C$, $T_E$, and $T_C$ values. Usually, $1600 < T_E < 2000$ K, $600 < T_C < 1200$ K, $2.4 < \phi_E < 2.8$ eV and $\phi_C \simeq 1.6$ eV.

The electrode power density is $P = JV$, and the thermal input $Q_{IN}$ to the emitter is given by

$$Q_{IN} = Q_E + Q_R + Q_L + Q_X, \quad (8)$$

where $Q_E$ is the heat to evaporate electrons, $J(\phi_E + 2kT_E)$ (this expression assumes negligible backemission from the collector), $Q_R$ is the thermal radiation between electrodes, $Q_L$ is the heat conducted down the emitter lead, and $Q_X$ is the extraneous heat losses through structure and interelectrode vapor. Thus, the thermionic conversion efficiency is

$$\eta = JA(V - V_L)/Q_{IN}, \quad (9)$$

where $A$ is the electrode area, $V_L$ is the voltage loss in emitter lead ($= JAR_L$), and $R_L$ is the resistance of emitter lead. For a given cross-sectional area of the lead, lengthening will increase $V_L$ and decrease $Q_L$; shortening will decrease $V_L$ and increase $Q_L$. To obtain the maximum efficiency in Eq. (9), the lead dimensions must be optimized. This optimization can be performed by setting $d\eta/dR_L = 0$. Typically, $V_L \simeq 0.1$ V.

## B. Operating Modes

In practice, the ideal motive and output characteristic illustrated in Fig. 3 is never achieved because of the problems associated with the space charge barrier built up as the electrons transit from the emitter to the collector. Three of the basic approaches to circumventing the space charge problem are (1) the vacuum-mode diode, in which the interelectrode spacing is quite small; (2) the unignited-mode converter, in which the positive ions for space charge neutralization are provided by surface-contact ionization; and (3) the ignited mode, in which the positive ions are provided by thermal ionization in the interelectrode space. The potential diagrams and $J$–$V$ characteristics of these three operating modes are given in Figs. 4a, b, and c, respectively.
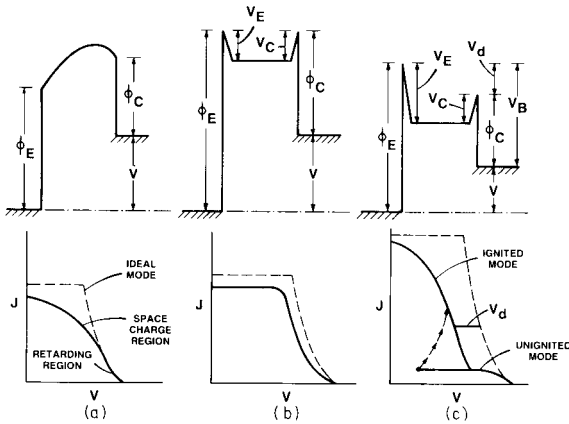
**FIGURE 4** Motive diagrams and $J$–$V$ characteristics for (a) vacuum, (b) unignited, and (c) ignited modes of operating a thermionic converter.
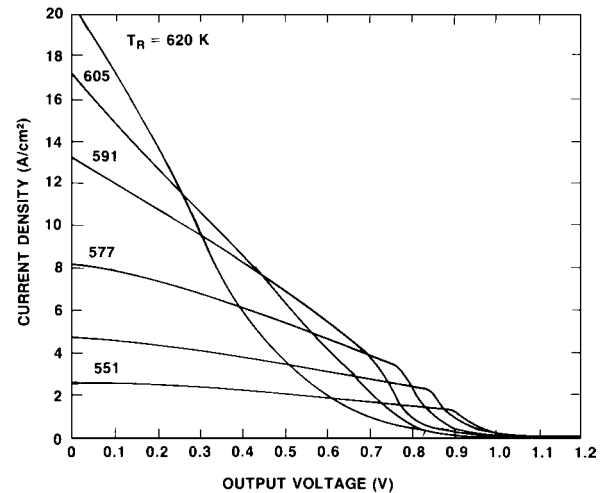


**FIGURE 5** Typical output characteristics, parametric in cesium reservoir temperature. Emitter, CVD W(110); collector, noibium; $T_E = 1800$ K; $T_C = 973$ K; $d = 0.5$ mm.

In the vacuum mode (Fig. 4a) the space charge is suppressed by making the spacing quite small—of the order of 0.01 mm. Although converters of this type have been built and operated with small-area electrodes, the practical difficulty of maintaining the necessary close spacing over large areas at high temperature differences has made the vacuum-mode diode mostly of academic interest.

Another approach to the space charge problem is to introduce low-pressure cesium vapor between the emitter and collector. This vapor is ionized when it contacts the hot emitter, provided that the emitter work function is comparable to the 3.89-eV ionization potential of cesium. A thermionic converter of this type (Fig. 4b), in which the ions for neutralizing the space charge are provided by surface-contact ionization, is said to be operating in the unignited mode. Unfortunately, surface-contact ionization is effective only at emitter temperatures that are so high (say, 2000 K) as to preclude most heat sources. To operate at more moderate temperatures, the cesium pressure must be increased to the order of 1 torr. In this case (Fig. 4c) cesium adsorption on the emitter reduces its work function so that high current densities can be achieved at significantly lower temperatures than in low-pressure unignited-mode diodes. Most of the ions for space charge neutralization are provided by electron-impact ionization in the interelectrode spacing. This ignited mode is initiated when the output voltage is lowered sufficiently. This mode of operation is that of the "conventional" thermionic converters. This ionization energy is provided at the expense of the output voltage via an "arc drop" $V_d$ across the interelectrode plasma. Because of the high cesium pressure, the electrons are scattered many times as they cross to the collector.

For a given electrode pair, emitter temperature, collector temperature, and spacing, the $J$–$V$ characteristic is a function of cesium reservoir temperature. A typical $J$–$V$ family, parametric in cesium reservoir temperature, is given in Fig. 5.

## C. Barrier Index

The barrier index, or $V_B$, is a convenient parameter for characterizing thermionic converter development, comparing experimental data, and evaluating converter concepts. It is an inverse figure of merit because the lower the $V_B$, the higher is the converter performance. The barrier index is defined as

$$V_B = V_d + \phi_C. \tag{10}$$

Inspection of Fig. 4c shows that, the lower the $\phi_C$ and $V_d$, the higher is the output voltage $V$.

The barrier index can be defined operationally. For any given emitter temperature and output current, it is possible to adjust cesium pressure, spacing, and collector temperature to maximize the power output. The spacing envelope of the optimized performance curves is shown in Fig. 6 for a converter with an emitter temperature of 1800 K. This envelope is shifted by a constant potential difference from the Boltzmann line, which represents the ideal current–voltage characteristic. This potential difference is defined as the barrier index. In Fig. 6, the barrier index of 2.1 eV represents good performance for a thermionic converter with bare metal electrodes, corresponding to $V_d \simeq 0.5$ eV and $\phi_C \simeq 1.6$ eV. The equation for the Boltzmann line is

$$J_B = AT_E^2 \exp(-V/kT_E). \tag{11}$$

The Boltzmann line represents the idealized converter output (up to the emitter saturation current) for zero
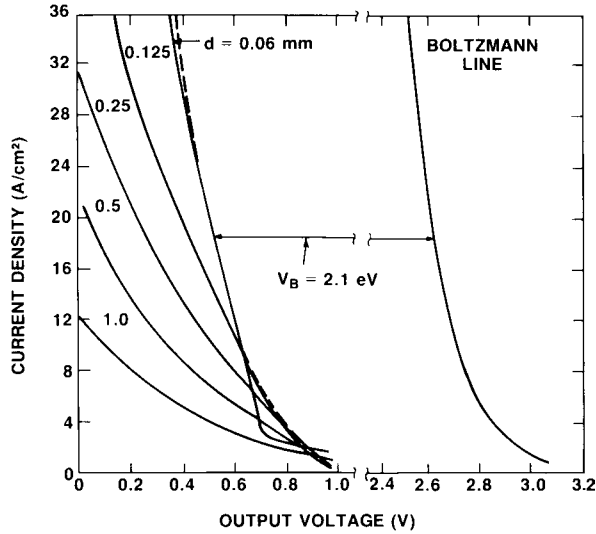
**FIGURE 6** Spacing envelope of the optimized performance curves for a converter with a barrier index of 2.1 eV. Emitter, (110)W; collector, niobium; $T_E = 1800$ K; $T_C$ and $T_R$, optimum.

collector work function and zero collector temperature. Thus, the barrier index incorporates the sum of the collector work function and the electron transport losses (due to ionization, scattering, electrode reflection, and electrode patchiness) into a single factor. For a real thermionic converter, the current density is

$$J = A T_E^2 \exp[-(V + V_B)/kT_E]. \tag{12}$$

Improvements in the barrier index can be translated into either higher efficiency at a given emitter temperature or the same efficiency at a lower temperature.

The lead efficiency (i.e., the thermionic converter efficiency, which takes into account the Joule and thermal heat losses down the electrical lead) is shown in Fig. 7 as
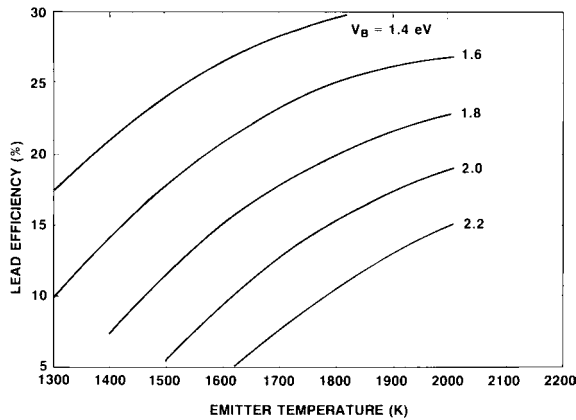


**FIGURE 7** Lead efficiency as a function of emitter temperature, parametric in barrier index ($T_C = 700$–900 K; $J = 10$ A/cm$^2$; $\phi_C = 1.3$–1.7 eV).

a calculated function of emitter temperature for a range of barrier index values. These data emphasize that the thermionic converter is an extremely high temperature device.

## IV. ELECTRODES

To operate at a practical power density and efficiency, an electrode should provide of the order of 10 A/cm$^2$ over a long period at emitter temperatures of 1600 to 1800 K without excessive evaporation. This requirement implies an emitter work function of 2.4 to 2.7 eV.

### A. Bare Electrodes

A bare electrode is defined as a surface that does not require a vapor supply from the inter-electrode space to maintain its emission properties. No bare electrode has been found that meets the foregoing criterion. The conventional (Ba, Sr, Ca)O coating on a nickel substrate, which is widely used (e.g., radio, oscilloscope, television, microwave tubes), is not stable at the desired emitter temperatures. Lanthanum hexaboride showed promise, but its high evaporation rate and the extreme difficulty of bonding it to a substrate over a large area has eliminated it from thermionic application. For short-term experimental devices, dispenser cathodes (barium compounds impregnated in a tungsten matrix) have given useful results. However, their current density and stability do not meet the needs of a practical thermionic converter.

#### 1. Cesium Pressure Relationship

In almost all vapor converters, the cesium pressure is adjusted by controlling the temperature of a liquid cesium reservoir. To a good approximation, the cesium vapor pressure $p_{cs}$ in torr can be calculated from

$$p_{cs} = 2.45 \times 10^8 T_R^{-0.5} \exp(-8910/T_R), \tag{13}$$

where the cesium reservoir temperature $T_R$ is given in kelvins. It is also possible to control the pressure using cesium-intercalated graphite compounds that shift the operating reservoir temperature to higher values, which may be advantageous in some applications.

#### 2. Rasor Plot

Experimental and theoretical results indicate that, for less than a monolayer coverage, the work function can be expressed to a good approximation as
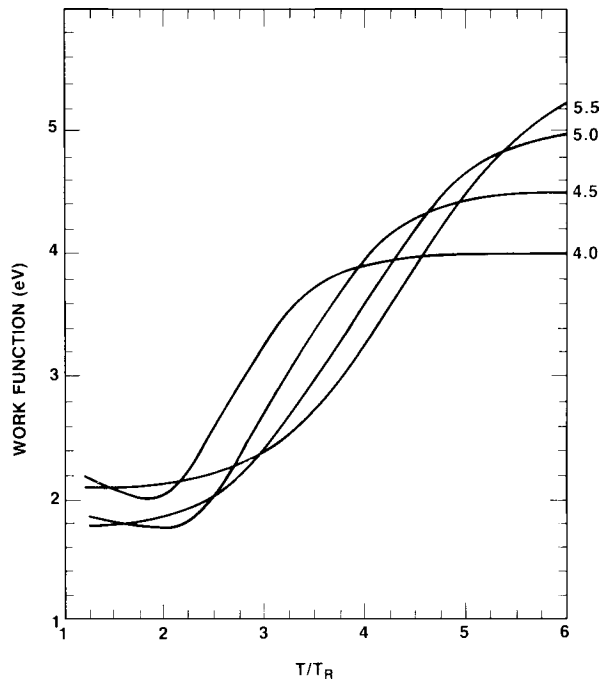
$$\phi = f(T/T_R) \tag{14}$$

**FIGURE 8** Cesiated work function of a surface versus the ratio of electrode temperature to cesium reservoir temperature, parametric in the bare work function of the surface.

for a given substrate material. This relationship is shown in Fig. 8, parametric in the bare work function $\phi_B$ of the substrate. Such a representation, which condenses a morass of electron emission data into a tractable form, is usually called a Rasor plot.

### 3. Saha–Langmuir Equation

Cesiated surfaces emit positive ions as well as electrons. The fraction of the cesium impingement rate, $g$, that evaporates as ions is given by the Saha–Langmuir equation,

$$g = \frac{1}{1 + 2\exp(I - \phi/kT)}, \tag{15}$$

where $I$ is the ionization potential of the impinging vapor atoms (3.89 eV for cesium) and $\phi$ and $T$ refer to the surface.

### B. Additives

It can be advantageous to operate thermionic converters with additives in combination with cesium. For some applications, the improved performance can be substantial.

#### 1. Electropositive Additives

The work function of a metal surface is reduced when the surface area is covered with a fraction of adsorbed

monolayer of a more electropositive metal. Such an additive (e.g., barium) could reduce the emitter work function while using cesium primarily for thermionic ion emission. Operating in this mode would require a very high emitter temperature compared with the conventional ignited mode.

#### 2. Electronegative Additives

The adsorption of a fractional layer of an electronegative element, such as oxygen, will increase the work function of the surface. However, Langmuir and his co-workers found that, when such a surface is operated in cesium vapor, the work function (for a given cesium pressure and substrate temperature) may be significantly lower than without the electronegative additive. The thermionic converter performance can be improved by a lower collector work function or a lower cesium pressure in the interelectrode space (or both).

This behavior is a result of the interaction of adsorbed layers with opposing dipole polarities. It should not be too surprising in view of the Rasor plot (see Fig. 8), which shows that, the higher the bare work function of a substrate, the lower is the cesiated work function for a given ratio $T/T_R$.

## V. EXPERIMENTAL RESULTS

This section summarizes the experimental results from several classes of thermionic converters. It should provide a background for choosing the most promising paths to improved performance.

### A. Vacuum-Mode Diodes

The $J–V$ characteristic of a close-spaced diode, indicating the space charge and retardingmode regions, is illustrated in Fig. 4a. Practically, it has not been possible to operate thermionic diodes at spacings appreciably closer than 0.01 mm. At such spacings, the output power density is usually less than 1.0 W/cm$^2$. Using dispenser cathodes for the emitter (1538 K) and collector (810 K), G. N. Hatsopoulos and J. Kaye obtained a current density of 1.0 A/cm$^2$ at an output potential of 0.7 V. They encountered output degradation due to evaporation of barium from the emitter to the collector, which increased the collector work function. Although the electrode stability problem can be avoided by using cesium vapor to adjust the electrode work functions, the practical problem of maintaining spacings close enough for vacuum-mode operation over large areas at high temperatures has eliminated close-spaced converters from practical applications.

To circumvent the problem of extremely small inter-electrode spacings in a vacuum converter, both magnetic and electrostatic triodes have been investigated. However, neither device appears to be practical.

## B. Bare Electrode Converters

### 1. Unignited-Mode Converters

Thermionic converters with bare electrodes have been operated in both the unignited and the ignited modes. As discussed previously, unignited-mode operation requires emitter temperatures above 2000 K. To approximate the ideal diode $J-V$ characteristics, electron–cesium scattering must be minimized by interelectrode spacings of the order of 0.1 mm. Under these conditions, the unignited-mode operation may be substantially more efficient than the ignited-mode operation, since the ions are supplied by thermal energy rather than by the arc drop, which subtracts from the output voltage.

Unfortunately, the high emitter temperatures required for unignited operation eliminate most heat sources because of materials limitations. For example, most nuclear heat sources cannot be used on a long-term basis at these temperatures because of either fuel swelling or fuel–emitter incompatibility.

### 2. Ignited-Mode Converters

Thus far, essentially all practical thermionic converters have operated in the ignited mode. This conventional converter has demonstrated power densities of 5 to 10 W/cm$^2$ and efficiencies of 10 to 15% for emitter temperatures between 1600 and 1800 K.

For a wide range of converter parameters, there exists an optimum pressure–spacing product (or $Pd$) of $\sim$0.5-mm torr. For higher values of this product, the plasma loss is increased because of unnecessary electron–cesium scattering losses. From a practical viewpoint, it is difficult to fabricate large-area thermionic converters with spacings of less than 0.25 mm. However, for long-lived space reactors, system studies indicate that interelectrode spacings of up to 0.5 mm may be required to accommodate emitter distortion due to fuel swelling.

The higher the bare work function, the lower is the cesium pressure required to maintain a given electron emission. For practical spacings, the reduced cesium pressure tends toward the optimum pressure–spacing product to reduce plasma losses and improve performance. The output power density as a function of emitter temperature, parametric in the bare work function of the emitter, is given in Fig. 9. It is evident that large increases in output power density can be achieved by the use of oriented surfaces with bare work functions of greater than 4.5 eV.
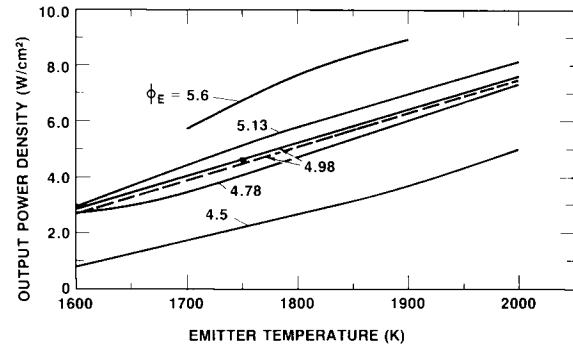


**FIGURE 9** Output power density as a function of emitter temperature, parametric in the bare work function of the emitter ($J = 10$ A/cm$^2$; $d = 0.25$ mm).

One avenue for improving thermionic converter performance is to fabricate emitters with preferred crystalline orientations with high bare work functions.

## C. Additive Diodes

Controlled additions of oxygen into thermionic converters can yield substantial improvement in power density, especially at spacings $\geq 0.5$ mm. A comparison of the $J-V$ characteristics of two converters, identical except for oxide collector coating, is shown in Fig. 10. The dramatic increase in output is clear for the converter with the tungsten oxide coating on the collector. In terms of the barrier index defined earlier, $V_B \simeq 1.9$ eV for the bare tungsten collector diode. The improved performance is due to a combination of reduced collector work function and lower potential losses in the interelectrode plasma. The oxide collector supplies oxygen to the emitter so that
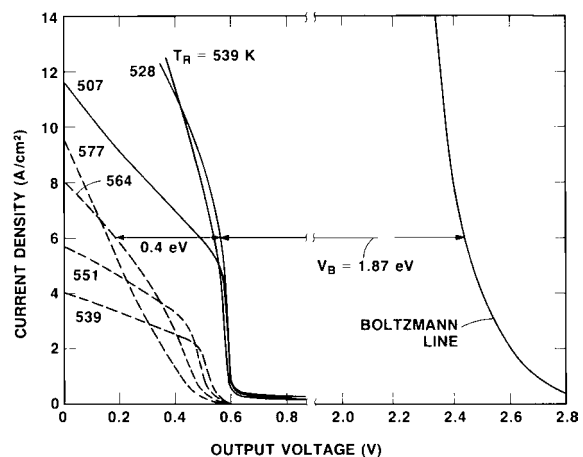


**FIGURE 10** Comparison of the current density–voltage characteristics of two thermionic converters, identical except for oxide coating on collector (——, WO$_x$, $T_C = 800$ K; - - -, W, $T_C = 975$ K, $T_E = 1600$ K; $d = 1$ mm).

a given current density can be obtained at a significantly lower cesium pressure. In effect, the added oxygen makes the emitter act as if it has an extremely high bare work function. As a result of the lower cesium pressure, the optimum interelectrode spacing increases.

## VI. APPLICATIONS

Historically, the primary motivation for the development of thermionic converters has been their potential application to nuclear power systems in space, both reactor and radioisotope. In addition, thermionic converters are attractive for terrestrial, solar, and combustion applications.

### A. Nuclear Reactors

The capacity of a thermionic converter to operate efficiently at high emitter and collector temperatures inside the core of a nuclear reactor makes it very suitable for space power applications. This combination of characteristics results in a number of fundamental and developmental advantages. For example, the necessity of radiating away all the reject heat from space conversion systems puts a premium on the thermionic converter's demonstrated capability of operating reliably and efficiently (typically, 10–15%), even at high collector temperatures (up to 1100 K). The capacity of the thermionic converter to operate inside the core of a reactor essentially eliminates the high-temperature heat transport system inherent in all out-of-core conversion systems. Thus, the thermionic reactor coolant system is at radiator temperature so that, except for the emitters, the balance of the reactor system (pumps, reflector, controls, moderator, shield, etc.) operates near the radiator temperature. Even the high-temperature emitters are distributed inside the core in "bite-sized" pellets.

A cutaway drawing of a thermionic diode that has operated inside a reactor is shown in Fig. 11. The enriched uranium oxide fuel pellets are inside the tungsten emitter, which is made by chemical vapor deposition (CVD). The outside diameter of the emitter is ∼28 mm. The collector is niobium, and the insulator seal is niobium–alumina. Thermionic diodes of this general design have operated stably inside a reactor for more than a year, and an electrically heated converter has given constant output for more than 5 years. Typically, emitter temperatures have ranged between 1700 and 2000 K.

A thermionic reactor is composed of an array of thermionic fuel elements (TFEs) in which multiple thermionic converters are connected in series inside an electrically insulating tube, much like cells in a flashlight. In principle, it is possible to design reactor systems with the thermionic converters out of core. This
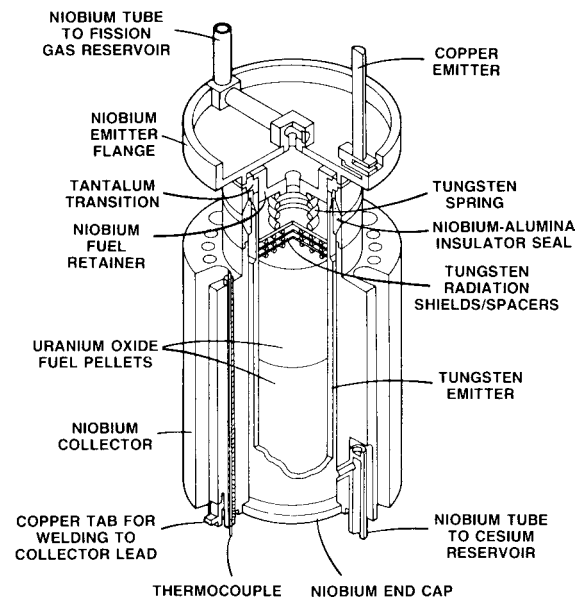


**FIGURE 11** Uranium oxide-fueled thermionic diode.

design concept presents a difficult problem of electrical insulation at emitter temperature and vitiates many of the basic advantages of the TFE reactor. It is also possible to fuel the converter externally using a fuel element with a central coolant tube, the outer surface of which serves as the collector; such a design concept appears to be more limited in system power than the flashlight TFE approach.

The feasibility of the thermionic reactor system has been demonstrated by in-core converter and TFE tests in the United States, France, and Germany. As of the mid-1970s, four TOPAZ thermionic reactors had been built in Russia and tested at outputs of up to 10 kWe (kilowatts electrical).

### B. Radioisotope Generators

A radioisotope thermionic generator uses the decay heat of a radioactive element. The design is usually modular. A module includes the radioisotope fuel, its encapsulation, a thermionic converter, heat transfer paths (for coupling heat from the fuel to the converter and from the converter to the radiator), hermetically sealed housing, and thermal insulation. A variety of design concepts have been considered (planar and cylindrical converters, heat pipe and conduction coupling, and several radioisotopes), and a few fueled systems have been built. However, no report of an operational use of a radioisotope thermionic generator has yet been made.

The high operating temperature of the radioisotope thermionic converter poses difficult radiological safety

problems of fuel encapsulation integrity relative to launch pad accidents, launch aborts, and reentry. Since these problems are more tractable at the lower operating temperatures of radioisotope thermoelectric generators, these units have been employed for remote and space missions requiring less than a kilowatt of electrical output.

A thermionic space reactor system does not suffer a radiological safety disadvantage compared with a thermoelectric space reactor system since both systems are launched cold and neither reactor is started until a nuclear safe orbit has been achieved. The radiological hazards associated with the launch of a cold reactor system without a fission product inventroy are significantly less than those of a launch of a radioisotope thermoelectric generator.

## C. Combustion Converters

The demise of the U.S. space reactor effort in 1973, along with the accompanying OPEC oil action, provided the motivation for utilizing on the ground the thermionic technology that had been developed for space. There are a number of potential terrestrial applications of thermionic conversion. These include power plant topping, cogeneration with industrial processes, and solar systems.

Although refractory metals such as tungsten and molybdenum operate stably in the vacuum of outer space at thermionic temperatures, converters operating in air or combustion atmosphere would oxidize to destruction within minutes without protection. Therefore, it is essential to develop a protective coating, or "hot shell," to isolate the refractory metals from the terrestrial environment (see Fig. 1). The thermionic converter illustrated in Fig. 12 is representative of the combustion-heated diodes that have been constructed and tested. The dome is exposed to high-temperature combustion gases. Electrons evaporated from the tungsten emitter are condensed on the mating

nickel collector, which is air-cooled. A ceramic seal provides electrical insulation between the electrodes. The electrical power output is obtained between the flange and the air tube, which also functions as the collector lead.

The key component is the hot-shell–emitter (HS–EM) structure, which operates between the emitter and collector temperatures. This trilayer composite structure (tungsten–graphite–silicon carbide) is fabricated by CVD. First, the graphite is machined to the desired configuration. Next, the inside of the graphite is coated with CVD tungsten to form the emitter and its electrical lead. Finally, the outside is coated with CVD silicon carbide to protect the emitter from the combustion atmosphere.

The properties of the materials used in the HS–EM structure complement one another. Tungsten is used because of its low vapor pressure at emitter temperatures, high bare work function, low electrical resistivity, and compatibility with cesium. Silicon carbide is chosen because of its excellent high-temperature corrosion resistance, strength, thermal shock characteristics, and close match to the thermal expansion of tungsten. The graphite is selected, to match the thermal expansions of the tungsten and silicon carbide. The graphite is essential for providing good thermal shock and thermal cycle characteristics to the composite structure.

Arrays of such converters will be connected in series to provide practical output voltage. Converters of this general design have operated stably for periods of as long as 12,500 h at emitter temperatures of up to 1730 K in a natural-gas burner. The HS–EM structure has survived severe thermal shock, thermal cycle, and pressure tests.

Although combustion thermionics has made substantial progress, it is not yet ready for commercialization. The first application would probably be a thermionic burner that could be retrofit onto industrial furnaces. A
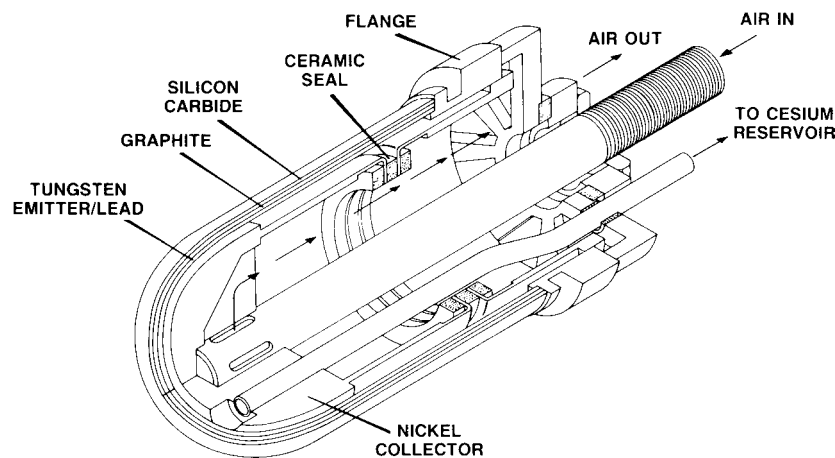


**FIGURE 12**  Combustion thermionic converter.

thermionic burner is a combustor whose walls are lined with thermionic converters. The emitters of the converters receive heat from the combustion gases and convert part of this heat into electricity while rejecting the balance of the heat from the collectors into the air for combustion. Operational experience with such units should provide the database for subsequent power plant topping use.

The most recent cost estimate for an installed thermionic cogeneration system was approximately $1600/kW. Achievement of the corresponding converter cost of $540/kW would require additional investment in converter and manufacturing development.

Relative to other advanced combustion conversion systems, such as magnetohydrodynamics (MHD), ceramic blade turbine gas turbine, and potassium Rankine cycle, thermionic development costs should be substantially lower. The cost effectiveness is a result of the modularity of thermionics, which makes it possible to perform meaningful experiments with small equipment. Thus, large investments should not be required until there is a high probability of success.

### D. Solar Converters

Although the CVD silicon carbide converter was developed for combustion applications, it is very suitable for solar systems. Previous solar tests of terrestrial thermionic converters had to utilize a window so that the high-temperature refractory components could operate in an inert atmosphere. In addition to transmission losses, such windows are subject to problems of overheating and leakage.

A CVD converter has been solar-tested in a central receiver heliostat array at the Georgia Institute of Technology. The test examined heat flux cycling, control of the operating point, and mounting arrangements. The converter was mounted directly in the solar image with no cavity. The converter performance was comparable with combustion measurements made on the same diode. Thermal cycling caused no problems, and the converter showed no degradation after testing.

## SEE ALSO THE FOLLOWING ARTICLES

ELECTRIC PROPULSION ● NUCLEAR POWER REACTORS ● PLASMA SCIENCE AND ENGINEERING

## BIBLIOGRAPHY

Angrist, S. W. (1965). "Direct Energy Conversion," Allyn and Bacon, Rockleigh, New Jersey.

Baksht, F. G., *et al.* (1978). "Thermionic Converters and Low Temperature Plasma" (Engl. transl.). Technical Information Center/U.S. Department of Energy, Washington, DC.

Hatsopoulos, G. N., and Gyftopoulos, E. P. (1974, 1979)."Thermionic Energy Conversion," Vols. 1 and 2. MIT Press, Cambridge, Massachusetts.

Nottingham, W. B. (1956). Thermionic emission. *In* "Handbuch der Physik," Vol. 21. Springer-Verlag, New York.

Rasor, N. S. (1982). Thermionic energy conversion. *In* "Applied Atomic Collision Physics," Ch. 5, Vol. 5. Academic Press, New York.

Ure, R. W., and Huffman, F. N. (1987). Thermoelectric and thermionic conversion. *In* "Standard Handbook for Electrical Engineers," 12th ed. McGraw-Hill, New York.

# Thermoeconomics

**George Tsatsaronis**
**Frank Cziesla**
*Technical University of Berlin*

## GLOSSARY

**Exergy** Useful energy or a measure for the quality of energy.
**Exergy destruction** Thermodynamic loss (inefficiency) due to irreversibilities within a system.
**Exergetic efficiency** Characterizes the performance of a component or a thermal system from the thermodynamic viewpoint.
**Exergy loss** Thermodynamic loss due to exergy transfer to the environment.
**Thermoeconomics** Exergy-aided cost minimization.

**THERMOECONOMICS** is the branch of thermal sciences that combines a thermodynamic (exergy) analysis with economic principles to provide the designer or operator of an energy-conversion system with information which is not available through conventional thermodynamic analysis and economic evaluation but is crucial to the design and operation of a cost-effective system. Thermoeconomics rests on the notion that exergy (available energy) is the only rational basis for assigning monetary costs to the interactions of an energy conversion system with its surroundings and to the sources of thermodynamic inefficiencies within it. Since the thermodynamic considerations of thermoeconomics are based on the exergy concept, the terms *exergoeconomics* and *thermoeconomics* can be used interchangeably.

A complete thermoeconomic analysis consists of (1) an exergy analysis, (2) an economic analysis, (3) exergy costing, and (4) a thermoeconomic evaluation. A thermoeconomic analysis is usually conducted at the system component level and calculates the costs associated with all material and energy streams in the system and the thermodynamic inefficiencies (exergy destruction) within each component. A comparison of the cost of exergy destruction with the investment cost for the same component provides useful information for improving the cost effectiveness of the component and the overall system by pinpointing the required changes in structure and parameter values. An iterative thermoeconomic evaluation and optimization is, among the currently available methods, the most effective approach for optimizing the design of a

system when a mathematical optimization procedure cannot be used due to the lack of information (e.g., cost functions) and our inability to appropriately consider important factors such as safety, availability, and maintainability of the system in the modeling process.

The objectives of thermoeconomics include

- Calculating the cost of each product stream generated by a system having more than one product
- Optimizing the overall system or a specific component
- Understanding the cost-formation process and the flow of costs in a system

Thermoeconomics uses results from the synthesis, cost analysis, and simulation of thermal systems and provides useful information for the evaluation and optimization of these systems as well as for the application of artificial intelligence techniques to improve the design and operation of such systems.

## I. EXERGY ANALYSIS

An exergy analysis identifies the location, the magnitude, and the sources of thermodynamic inefficiencies in a thermal system. This information, which cannot be provided by other means (e.g., an energy analysis), is very useful for improving the overall efficiency and cost effectiveness of a system or for comparing the performance of various systems (Bejan, Tsatsaronis, and Moran, 1996; Moran and Shapiro, 1998).

An exergy analysis provides, among others, the exergy of each stream in a system as well as the real "energy waste," i.e., the thermodynamic inefficiencies (exergy destruction and exergy loss), and the exergetic efficiency for each system component.

In the sign convention applied here, work done on the system and heat transferred to the system are positive. Accordingly, work done by the thermal system and heat transferred from the system are negative.

### A. Exergy Components

*Exergy* is the maximum theoretical useful work (shaft work or electrical work) obtainable from a thermal system as it is brought into thermodynamic equilibrium with the environment while interacting with the environment only. Alternatively, exergy is the minimum theoretical work (shaft work or electrical work) required to form a quantity of matter from substances present in the environment and to bring the matter to a specified state. Hence, exergy is a measure of the departure of the state of the system from the state of the environment.

The environment is a large equilibrium system in which the state variables ($T_0$, $p_0$) and the chemical potential of the chemical components contained in it remain constant, when, in a thermodynamic process, heat and materials are exchanged between another system and the environment. It is important that no chemical reactions can take place between the environmental chemical components. The environment is free of irreversibilities, and the exergy of the environment is equal to zero. The environment is part of the surroundings of any thermal system.

In the absence of nuclear, magnetic, electrical, and surface tension effects, the *total exergy* of a system $E_{sys}$ can be divided into four components: *physical exergy* $E_{sys}^{PH}$, *kinetic exergy* $E^{KN}$, *potential exergy* $E^{PT}$, and *chemical exergy* $E^{CH}$:

$$E_{sys} = E_{sys}^{PH} + E^{KN} + E^{PT} + E^{CH}. \qquad (1)$$

The subscript *sys* distinguishes the total exergy and physical exergy of *a system* from other exergy quantities, including transfers associated with streams of matter. The total specific exergy on a mass basis $e_{sys}$ is

$$e_{sys} = e_{sys}^{PH} + e^{KN} + e^{PT} + e^{CH}. \qquad (2)$$

The *physical exergy* associated with a thermodynamic system is given by

$$E_{sys}^{PH} = (U - U_0) + p_0(V - V_0) - T_0(S - S_0), \qquad (3)$$

where $U$, $V$, and $S$ represent the internal energy, volume, and entropy of the system, respectively. The subscript 0 denotes the state of the same system at temperature $T_0$ and pressure $p_0$ of the environment.

The rate of physical exergy $\dot{E}_{ms}^{PH}$ associated with a material stream (subscript *ms*) is

$$\dot{E}_{ms}^{PH} = (\dot{H} - \dot{H}_0) - T_0(\dot{S} - \dot{S}_0), \qquad (4)$$

where $\dot{H}$ and $\dot{S}$ denote the rates of enthalpy and entropy, respectively. The subscript 0 denotes property values at temperature $T_0$ and pressure $p_0$ of the environment.

*Kinetic* and *potential exergy* are equal to kinetic and potential energy, respectively.

$$E^{KN} = \frac{1}{2}m\vec{v}^2 \qquad (5)$$

$$E^{PT} = mgz \qquad (6)$$

Here, $\vec{v}$ and $z$ denote velocity and elevation relative to coordinates in the environment ($\vec{v}_0 = 0$, $z_0 = 0$). Equations (5) and (6) can be used in conjunction with both systems and material streams.

The *chemical exergy* is the maximum useful work obtainable as the system at temperature $T_0$ and pressure $p_0$ is brought into chemical equilibrium with the environment. Thus, for calculating the chemical exergy,

not only do the temperature and pressure have to be specified, but the chemical composition of the environment also has to be specified. Since our natural environment is not in equilibrium, there is a need to model an exergy-reference environment (Ahrendts, 1980; Bejan, Tsatsaronis, and Moran, 1996; Szargut, Morris, and Steward, 1988). The use of tabulated *standard chemical exergies* for substances contained in the environment at standard conditions ($T_{ref} = 298.15$ K, $p_{ref} = 1.013$ bar) facilitates the calculation of exergy values. Table I shows values of the standard chemical exergies of selected substances in two alternative exergy-reference environments. The effect of small variations in the values of $T_0$ and $p_0$

on the chemical exergy of reference substances might be neglected in practical applications.

The chemical exergy of an ideal mixture of $N$ ideal gases $\bar{e}_{M,ig}^{CH}$ is given by

$$\bar{e}_{M,ig}^{CH} = \sum_{k=1}^{N} x_k \bar{e}_k^{CH} + \bar{R}T_0 \sum_{k=1}^{N} x_k \ln x_k, \qquad (7)$$

where $T_0$ is the ambient temperature, $\bar{e}_k^{CH}$ is the standard molar chemical exergy of the $k$th substance, and $x_k$ is the mole fraction of the $k$th substance in the system at $T_0$. For solutions of liquids, the chemical exergy $\bar{e}_{M,l}^{CH}$ can be obtained if the activity coefficients $\gamma_k$ are known (Kotas, 1985):

$$\bar{e}_{M,l}^{CH} = \sum_{k=1}^{N} x_k \bar{e}_k^{CH} + \bar{R}T_0 \sum_{k=1}^{N} x_k \ln(\gamma_k x_k). \qquad (8)$$

The standard chemical exergy of a substance not present in the environment can be calculated by considering a reversible reaction of the substance with other substances for which the standard chemical exergies are known. For energy-conversion processes, calculation of the exergy of fossil fuels is particularly important.

Figure 1 shows a hypothetical reversible isotherm-isobaric reactor where a fuel reacts completely at steady state with oxygen to form $CO_2$, $SO_2$, $H_2O$, and $N_2$. All substances are assumed to enter and exit unmixed at $T_0$, $p_0$. The chemical exergy of a fossil fuel $\bar{e}_f^{CH}$ on a molar basis can be derived from exergy, energy, and entropy balances for the reversible reaction (Bejan, Tsatsaronis, and Moran, 1996):

$$\bar{e}_f^{CH} = -(\Delta\bar{h}_R - T_0\Delta\bar{s}_R) + \Delta\bar{e}^{CH} = -\Delta\bar{g}_R + \Delta\bar{e}^{CH} \quad (9)$$

with

$$\Delta\bar{h}_R = \sum_i \nu_i \bar{h}_i = -\bar{h}_f + \sum_k \nu_k \bar{h}_k = -\overline{HHV}$$

$$\Delta\bar{s}_R = \sum_i \nu_i \bar{s}_i = -\bar{s}_f + \sum_k \nu_k \bar{s}_k$$

**TABLE I  Standard Molar Chemical Exergy $\bar{e}^{CH}$ of Various Substances at 298.15 K and $p_{ref}$**

| | | $\bar{e}^{CH}$ (kJ/kmol) | |
|---|---|---|---|
| **Substance** | **Formula[a]** | **Model I[b]** | **Model II[c]** |
| Ammonia | $NH_3$(g) | 336,684 | 337,900 |
| n-Butane | $C_4H_{10}$(g) | — | 2,805,800 |
| Calcium oxide | CaO(s) | 120,997 | 110,200 |
| Calcium hydroxide | $Ca(OH)_2$(s) | 63,710 | 53,700 |
| Calcium carbonate | $CaCO_3$(s) | 4,708 | 1,000 |
| Calcium sulfate (gypsum) | $CaSO_4 \cdot 2H_2O$(s) | 6,149 | 8,600 |
| Carbon (graphite) | C(s) | 404,589 | 410,260 |
| Carbon dioxide | $CO_2$(g) | 14,176 | 19,870 |
| Carbon monoxide | CO(g) | 269,412 | 275,100 |
| Ethane | $C_2H_6$(g) | 1,482,033 | 1,495,840 |
| Hydrogen | $H_2$(g) | 235,249 | 236,100 |
| Hydrogen peroxide | $H_2O_2$(g) | 133,587 | — |
| Hydrogen sulfide | $H_2S$(g) | 799,890 | 812,000 |
| Methane | $CH_4$(g) | 824,348 | 831,650 |
| Methanol (g) | $CH_3OH$(g) | 715,069 | 722,300 |
| Methanol (l) | $CH_3OH$(l) | 710,747 | 718,000 |
| Nitrogen | $N_2$(g) | 639 | 720 |
| Nitrogen monoxide | NO(g) | 88,851 | 88,900 |
| Nitrogen dioxide | $NO_2$(g) | 55,565 | 55,600 |
| Octane | $C_8H_{18}$(l) | — | 5,413,100 |
| Oxygen | $O_2$(g) | 3,951 | 3,970 |
| n-Pentane | $C_5H_{12}$(g) | — | 3,463,300 |
| Propane | $C_3H_8$(g) | — | 2,154,000 |
| Sulfur | S(s) | 598,158 | 609,600 |
| Sulfur dioxide | $SO_2$(g) | 301,939 | 313,400 |
| Sulfur trioxide (g) | $SO_3$(g) | 233,041 | 249,100 |
| Sulfur trioxide (l) | $SO_3$(l) | 235,743 | — |
| Water (g) | $H_2O$(g) | 8,636 | 9,500 |
| Water (l) | $H_2O$(l) | 45 | 900 |

[a] (g): gaseous, (l): liquid, (s): solid.
[b] Ahrendts, 1980. In this model, $p_{ref} = 1.019$ atm.
[c] Szargut, Morris, and Stewart, 1988. In this model, $p_{ref} = 1.0$ atm.
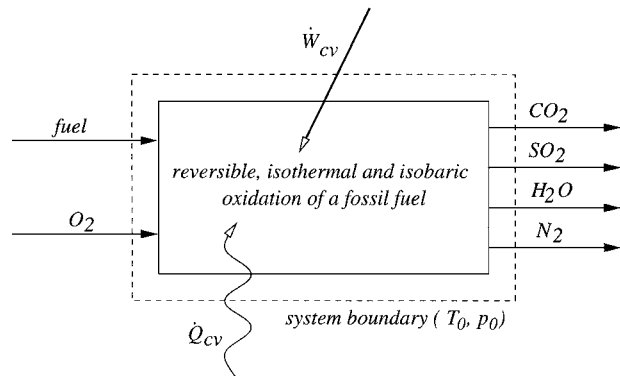


**FIGURE 1**  Device for evaluating the chemical exergy of a fuel.

$$\Delta \bar{g}_R = \Delta \bar{h}_R - T_0 \Delta \bar{s}_R$$

$$\Delta \bar{e}^{CH} = \sum_k \nu_k \bar{e}_k^{CH}$$

$$i = f, O_2, CO_2, H_2O, SO_2, N_2$$

$$k = O_2, CO_2, H_2O, SO_2, N_2$$

$$\nu_i \text{ and } \nu_k \geq 0 : CO_2, H_2O, SO_2, N_2$$

$$\nu_i \text{ and } \nu_k < 0 : f, O_2$$

Here, $\Delta \bar{h}_R$, $\Delta \bar{s}_R$, and $\Delta \bar{g}_R$ denote the molar enthalpy, entropy, and Gibbs function, respectively, of the reversible combustion reaction of the fuel with oxygen. $\overline{HHV}$ is the molar higher heating value of the fuel, and $\nu_k(\nu_i)$ is the stoichiometric coefficient of the $k$th ($i$th) substance in this reaction.

The higher heating value is the primary contributor to the chemical exergy of a fossil fuel. For back-of-the-envelope calculations, the molar chemical exergy of a fossil fuel $\bar{e}_f^{CH}$ may be estimated with the aid of its molar higher heating value $\overline{HHV}$:

$$\frac{\bar{e}_f^{CH}}{\overline{HHV}} \approx \begin{cases} 0.95 - 0.985 & \text{for gaseous fuels (except} \\ & H_2 \text{ and } CH_4) \\ 0.98 - 1.00 & \text{for liquid fuels} \\ 1.00 - 1.04 & \text{for solid fuels} \end{cases} \quad (10)$$

For hydrogen and methane, this ratio is 0.83 and 0.94, respectively.

The standard molar chemical exergy $\bar{e}_s^{CH}$ of any substance not present in the environment can be determined using the change in the specific Gibbs function $\Delta \bar{g}$ for the reaction of this substance with substances present in the environment (Bejan, Tsatsaronis, and Moran, 1996; Moran and Shapiro 1998):

$$\bar{e}_s^{CH} = -\Delta \bar{g} + \Delta \bar{e}^{CH} = -\sum_i \nu_i \bar{g}_i + \sum_{i \neq s} \nu_i \bar{e}_i^{CH}, \quad (11)$$

where $\bar{g}_i$, $\nu_i$, and $\bar{e}_i^{CH}$ denote, for the $i$th substance, the Gibbs function at $T_0$ and $p_0$, the stoichiometric coefficient in the reaction, and the standard chemical exergy, respectively.

In solar energy applications, the ratio of the exergy of solar radiation to the total energy flux at the highest layer of the earth's atmosphere is 0.933 (Szargut, Morris, and Steward, 1988).

## B. Exergy Balance and Exergy Destruction

Thermodynamic processes are governed by the laws of conservation of mass and energy. These conservation laws state that the total mass and total energy can neither be created nor destroyed in a process. However, exergy is not generally conserved but is destroyed by irreversibilities within a system. Furthermore, exergy is lost, in general, when the energy associated with a material or energy stream is rejected to the environment.

### 1. Closed System Exergy Balance

The change in total exergy ($E_{sys,2} - E_{sys,1}$) of a closed system caused through transfers of energy by work and heat between the system and its surroundings is given by

$$E_{sys,2} - E_{sys,1} = E_q + E_w - E_D. \quad (12)$$

The exergy transfer $E_q$ associated with heat transfer $Q$ is

$$E_q = \int_1^2 \left(1 - \frac{T_0}{T_b}\right) \delta Q, \quad (13)$$

where $T_b$ is the temperature at the system boundary at which the heat transfer occurs.

The exergy transfer $E_w$ associated with the transfer of energy by work $W$ is given by

$$E_w = W + p_0(V_2 - V_1). \quad (14)$$

In a process in which the system volume increases ($V_2 > V_1$), the work $p_0(V_2 - V_1)$ being done on the surroundings is unavailable for use, but it can be recovered when the system returns to its original volume $V_1$.

A part of the exergy supplied to a real thermal system is destroyed due to irreversibilities within the system. The exergy destruction $E_D$ is equal to the product of entropy generation $S_{gen}$ within the system and the temperature of the environment $T_0$.

$$E_D = T_0 S_{gen} \geq 0. \quad (15)$$

Hence, exergy destruction can be calculated either from the entropy generation using an entropy balance or directly from an exergy balance. $E_D$ is equal to zero only in ideal processes.

### 2. Control Volume Exergy Balance

Exergy transfer across the boundary of a control volume system can be associated with material streams and with energy transfers by work or heat. The general form of the exergy balance for a control volume involving multiple inlet and outlet streams is

$$\frac{dE_{cv,sys}}{dt} = \sum_j \underbrace{\left(1 - \frac{T_0}{T_j}\right) \dot{Q}_j}_{\dot{E}_{q,j}} + \underbrace{\left(\dot{W}_{cv} + p_0 \frac{dV_{cv}}{dt}\right)}_{\dot{E}_w}$$

$$+ \sum_i \dot{E}_i - \sum_e \dot{E}_e - \dot{E}_D, \quad (16)$$

where $\dot{E}_i$ and $\dot{E}_e$ are total exergy transfer rates at inlets and outlets [see Eq. (4) for the physical exergy associated

with these transfers]. The term $\dot{Q}_j$ represents the time rate of heat transfer at the location on the boundary where the temperature is $T_j$. The associated rate of exergy transfer $\dot{E}_{q,j}$ is given by

$$\dot{E}_{q,j} = \left(1 - \frac{T_0}{T_j}\right)\dot{Q}_j. \qquad (17)$$

The exergy transfer $\dot{E}_w$ associated with the time rate of energy transfer by work $\dot{W}_{cv}$ other than flow work is

$$\dot{E}_w = \dot{W}_{cv} + p_0 \frac{dV_{cv}}{dt}. \qquad (18)$$

Finally, $\dot{E}_D$ accounts for the time rate of exergy destruction due to irreversibilities within the control volume. Either the exergy balance [Eq. (16)] or $\dot{E}_D = T_0\dot{S}_{gen}$ can be used to calculate the exergy destruction in a control volume system.

Under steady state conditions, Eq. (16) becomes

$$0 = \sum_j \dot{E}_{q,j} + \dot{W}_{cv} + \sum_i \dot{E}_i - \sum_e \dot{E}_e - \dot{E}_D. \qquad (19)$$

## 3. Exergy Destruction

The real thermodynamic inefficiencies in a thermal system are related to exergy destruction and exergy loss. All real processes are irreversible due to effects such as chemical reaction, heat transfer through a finite temperature difference, mixing of matter at different compositions or states, unrestrained expansion, and friction. An exergy analysis identifies the system components with the highest thermodynamic inefficiencies and the processes that cause them.

In general, inefficiencies in a component should be eliminated or reduced if they do not contribute to a reduction in capital investment for the overall system or a reduction of fuel costs in another system component. Efforts for reducing the use of energy resources should be centered on components with the greatest potential for improvement. Owing to the present state of technological development, some exergy destructions and losses in a system component are unavoidable (Tsatsaronis and Park, 1999). For example, the major part of exergy destruction in a combustion processes cannot be eliminated. Only a small part can be reduced by preheating the reactants and reducing the amount of excess air.

The objective of a thermodynamic optimization is to minimize the inefficiencies, whereas the objective of a thermoeconomic optimization is to estimate the cost-optimal values of the thermodynamic inefficiencies.

Heat transfer through a finite temperature difference is irreversible. Figure 2 shows the temperature profiles for two streams passing through an adiabatic heat exchanger. The following expression for the exergy destruction $\dot{E}_{D,q}$ due to heat transfer from the hot stream 3 to the cold
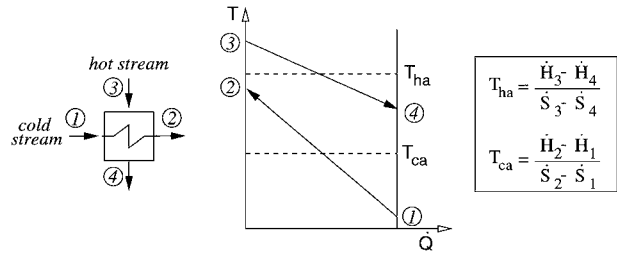


**FIGURE 2** Temperature profiles and thermodynamic average temperatures for two streams passing through an adiabatic heat exchanger at constant pressure.

stream 1 can be derived (Bejan, Tsatsaronis, and Moran, 1996):

$$\dot{E}_{D,q} = T_0\dot{Q}\frac{T_{ha} - T_{ca}}{T_{ha}T_{ca}}, \qquad (20)$$

where the thermodynamic average temperatures $T_{ha}$ and $T_{ca}$ of the hot stream and the cold stream are given by

$$T_a = \frac{\int_i^e T\,ds}{s_e - s_i} = \frac{h_e - h_i - \int_i^e v\,dp}{s_e - s_i}. \qquad (21)$$

Here, the subscripts $i$ and $e$ denote inlet and exit, respectively. For applications where the heat transfer occurs at constant pressure, Eq. (21) reduces to

$$T_a = \frac{h_e - h_i}{s_e - s_i} \qquad (22)$$

Equation (20) shows that the difference in the average thermodynamic temperatures $(T_{ha} - T_{ca})$ is a measure for the exergy destruction. Mismatched heat capacity rates of the two streams, i.e., $(\dot{m}c_p)_h/(\dot{m}c_p)_c \neq 1$, and a finite minimum temperature difference $(\Delta T_{min} = (T_3 - T_2)$ in Fig. 2) are the causes of the thermodynamic inefficiencies. Matching streams of significantly different heat capacity rates $\dot{m}c_p$ in a heat exchanger should be avoided. Furthermore, the lower the temperature levels $T_{ha}$ and $T_{ca}$, the greater the exergy destruction at the same temperature difference $(T_{ha} - T_{ca})$.

The rate of exergy destruction associated with friction $\dot{E}_{D,fr}$ can be expressed as

$$\dot{E}_{D,fr} = -\frac{T_0}{T_a} \cdot \dot{m} \cdot \int_i^e v\,dp, \qquad (23)$$

where $T_a$ is the thermodynamic average temperature of the working fluid and $\int_i^e v\,dp$ is the head loss (Bejan, Tsatsaronis, and Moran, 1996). The effect of friction is more significant at higher mass flow rates and lower temperature levels. Although exergy destruction related to friction is usually of secondary importance in thermal systems, the costs associated with the exergy destruction might be significant. Since the unit cost of electrical or mechanical power required to feed a pump, compressor, or

fan is significantly higher than the unit cost of a fossil fuel, each unit of exergy destruction by frictional dissipation is relatively expensive.

Examples for the exergy loss include heat transfer to the environment ("heat losses"), dissipation of the kinetic energy of an exhaust gas, and irreversibilities due to mixing of an exhaust gas with the atmospheric air.

Guidelines for improving the use of energy resources in thermal systems by reducing the sources of thermodynamic inefficiencies are discussed by Bejan, Tsatsaronis, and Moran (1996) and by Sama (1995). However, the major contribution of an exergy analysis to the evaluation of a system comes through a thermoeconomic evaluation that considers not only the inefficiencies, but also the costs associated with these inefficiencies and the investment expenditures required to reduce the inefficiencies.

## C. Exergetic Variables

The performance evaluation and the design optimization of thermal systems require a proper definition of the exergetic efficiency and a proper costing approach for each component of the system (Lazzaretto and Tsatsaronis, November 1999). The exergetic efficiency of a component is defined as the ratio between product and fuel. The product and fuel are defined by considering the desired result produced by the component and the resources expended to generate the result.

$$\varepsilon_k \equiv \frac{\dot{E}_{P,k}}{\dot{E}_{F,k}} = 1 - \frac{\dot{E}_{D,k} + \dot{E}_{L,k}}{\dot{E}_{F,k}} \qquad (24)$$

An appropriately defined exergetic efficiency is the only variable that unambiguously characterizes the performance of a component from the thermodynamic viewpoint (Tsatsaronis, 1999). Table II shows the definition of fuel and product for selected system components at steady state operation.

The rate of exergy destruction in the $k$th component is given by

$$\dot{E}_{D,k} = \dot{E}_{F,k} - \dot{E}_{P,k} - \dot{E}_{L,k}. \qquad (25)$$

Here, $\dot{E}_{L,k}$ represents the exergy loss in the $k$th component, which is usually zero when the component boundaries are at $T_0$. For the overall system, $\dot{E}_L$ includes the exergy flow rates of all streams leaving the system (see, for example, the different definitions of $\dot{E}_F$ for a fuel cell in Table II).

In addition to $\varepsilon_k$ and $\dot{E}_{D,K}$, the thermodynamic evaluation of a system component is based on the exergy destruction ratio $y_{D,k}$, which compares the exergy destruction in the $k$th component with the fuel supplied to the overall system $\dot{E}_{F,tot}$:

$$y_{D,k} \equiv \frac{\dot{E}_{D,k}}{\dot{E}_{F,tot}}. \qquad (26)$$

This ratio expresses the percentage of the decrease in the overall system exergetic efficiency due to the exergy destruction in the $k$th system component:

$$\varepsilon_{tot} = \frac{\dot{E}_{P,tot}}{\dot{E}_{F,tot}} = 1 - \sum_k y_{D,k} - \frac{\dot{E}_{L,tot}}{\dot{E}_{F,tot}}. \qquad (27)$$

$\dot{E}_{D,k}$ is an absolute measure of the inefficiencies in the $k$th component, whereas $\varepsilon_k$ and $y_{D,k}$ are relative measures of the same inefficiencies. In $\varepsilon_k$ the exergy destruction within a component is related to the fuel for the same component, whereas in $y_{D,k}$ the exergy destruction in a component is related to the fuel for the overall system.

## II. REVIEW OF AN ECONOMIC ANALYSIS

In the evaluation and cost optimization of an energy-conversion system, we need to compare the annual values of capital-related charges (carrying charges), fuel costs, and operating and maintenance expenses. These cost components may vary significantly within the plant economic life. Therefore, levelized annual values for all cost components should be used in the evaluation and cost optimization.

The following sections illustrate the total revenue requirement method (TRR method) which is based on procedures adopted by the Electric Power Research Institute (EPRI; 1993). This method calculates all the costs associated with a project, including a minimum required return on investment. Based on the estimated total capital investment and assumptions for economic, financial, operating, and market input parameters, the total revenue requirement is calculated on a year-by-year basis. Finally, the nonuniform annual monetary values associated with the investment, operating (excluding fuel), maintenance, and fuel costs of the system being analyzed are levelized; that is, they are converted to an equivalent series of constant payments (annuities).

An economic analysis can be conducted in current dollars by including the effect of inflation or in constant dollars by not including inflation. In general, studies involving the near term (the next 5–10 years) values are best presented in current dollars. The results of longer term studies (20–40 years) may be best presented in constant dollars.

## A. Estimation of Total Capital Investment

The best cost estimates for purchased equipment can be obtained directly through vendors' quotations. The next

**TABLE II   Exergy Rates Associated with Fuel and Product for Defining Exergetic Efficiencies of Selected Components at Steady State Operation**

| Component | Schematic | Exergy rate of product $\dot{E}_P$ | Exergy rate of fuel $\dot{E}_F$ |
|---|---|---|---|
| Compressor, pump, quad or fan |  | $\dot{E}_2 - \dot{E}_1$ | $\dot{E}_3$ |
| Turbine or expander |  | $\dot{E}_4$ | $\dot{E}_1 - \dot{E}_2 - \dot{E}_3$ |
| Heat exchanger[a] |  | $\dot{E}_2 - \dot{E}_1$ | $\dot{E}_3 - \dot{E}_4$ |
| Mixing unit |  | $\dot{m}_1(e_3 - e_1)$ | $\dot{m}_2(e_2 - e_3)$ |
| Combustion chamber |  | $\dot{E}_3 - \dot{E}_2$ | $\dot{E}_1 - \dot{E}_4$ |
| Steam generator |  | $(\dot{E}_6 - \dot{E}_5) + (\dot{E}_8 - \dot{E}_7)$ | $(\dot{E}_1 + \dot{E}_2) - (\dot{E}_3 + \dot{E}_4)$ |
| Deaerator |  | $\dot{m}_2(e_3 - e_2)$ | $\dot{m}_1 e_1 - (\dot{m}_3 - \dot{m}_2)e_1 - \dot{m}_4 e_4$ |
| Compressor with cooling air extractions |  | $\dot{E}_3 + \dot{E}_4 + \dot{E}_5 - \dot{E}_2$ | $\dot{E}_1$ |

*continues*

**TABLE II**   (*Continued*)

| Component | Schematic | Exergy rate of product $\dot{E}_P$ | Exergy rate of fuel $\dot{E}_F$ |
|---|---|---|---|
| Expander with cooling air supply |  | $\dot{E}_3$ | $\dot{E}_1 + \dot{E}_4 + \dot{E}_5 - \dot{E}_2$ |
| Fuel cell |  | $\dot{W} + (\dot{E}_6 - \dot{E}_5)$ | Fuel cell as part of a system: $(\dot{E}_1 + \dot{E}_2) - (\dot{E}_3 + \dot{E}_4)$ Stand-alone fuel cell: $\dot{E}_1 + \dot{E}_2$ |
| Ejector |  | $\dot{m}_1(e_3 - e_1)$ | $\dot{m}_2(e_2 - e_3)$ |
| Distillation column[b] |  | $\dot{E}_2^{CH} + \dot{E}_3^{CH} + \dot{E}_6^{CH} + \dot{E}_7^{CH}$ $- \dot{E}_1^{CH} + \dot{m}_6(e_6^{PH} - e_1^{PH})$ $+ \dot{m}_2(e_2^{PH} - e_1^{PH})$ | $(\dot{E}_4 - \dot{E}_5) + \dot{m}_7(e_1^{PH} - e_7^{PH})$ $+ \dot{m}_3(\dot{e}_1^{PH} - e_3^{PH})$ |
| Gasifier |  | $\dot{E}_3^{CH} + (\dot{E}_3^{PH} + \dot{E}_4^{PH}$ $- \dot{E}_2^{PH} - \dot{E}_1^{PH})$ | $\dot{E}_1^{CH} + \dot{E}_2^{CH} - \dot{E}_4^{CH}$ |
| Steam reformer |  | $\dot{E}_3 - \dot{E}_1 - \dot{E}_2$ | $\dot{E}_4 + \dot{E}_5 - \dot{E}_6$ |
| Evaporator including steam drum |  | $\dot{E}_2 + \dot{E}_5 - \dot{E}_1$ | $\dot{E}_3 - \dot{E}_4$ |

[a] These definitions assume that the purpose of the heat exchanger is to heat the cold steam ($T_1 > T_0$). If the purpose of the heat exchanger is to provide cooling ($T_3 < T_0$), then the following relations should be used: $\dot{E}_P = \dot{E}_4 - \dot{E}_3$ and $\dot{E}_F = \dot{E}_1 - \dot{E}_2$.

[b] Here, it is assumed that $e_j^{CH} > e_1^{CH}$ ($j = 2, 3, 6, 7$), $e_2^{PH} > e_1^{PH}$, $e_6^{PH} > e_1^{PH}$, $e_3^{PH} < e_1^{PH}$, $e_7^{PH} < e_1^{PH}$.

best source of cost estimates are cost values from past purchase orders, quotations from experienced professional cost estimators, or calculations using the extensive cost databases often maintained by engineering companies or company engineering departments. In addition, some commercially available software packages can assist with cost estimation.

When vendor quotations are lacking, or the cost or time requirements to prepare cost estimates are unacceptably high, the purchase costs of various equipment items can be obtained from the literature where they are usually given in the form of estimating charts. These charts have been obtained through a correlation of a large number of cost and design data. In a typical cost-estimating chart, when all available cost data are plotted versus the equipment size on a double logarithmic plot, the data correlation results in a straight line within a given capacity range. The slope of this line $\alpha$ represents an important cost-estimating parameter (scaling exponent) as shown by the relation

$$C_{P,Y} = C_{P,W} \left( \frac{X_Y}{X_W} \right)^{\alpha}, \qquad (28)$$

where $C_{P,Y}$ is the purchase cost of the equipment in question, which has a size or capacity $X_Y$; and $C_{P,W}$ is the purchase cost of the same type of equipment in the same year but of capacity or size $X_W$.

In practice, the scaling exponent $\alpha$ does not remain constant over a large size range. For very small equipment items, size has almost no effect on cost, and the scaling exponent $\alpha$ is close to zero. For very large capacities, the difficulties of building and transporting the equipment may increase the cost significantly, and the scaling exponent approaches unity. Examples of scaling exponents for different plant components are provided by Bejan, Tsatsaronis, and Moran (1996). In the absence of other cost information, an exponent value of 0.6 may be used (six-tenths rule).

Predesign capital cost estimates are very often assembled from old cost data. These reference data $C_{ref}$ can be updated to represent today's cost $C_{new}$ by using an appropriate cost index.

$$C_{new} = C_{ref} \left( \frac{I_{new}}{I_{ref}} \right) \qquad (29)$$

Here, $I_{new}$ and $I_{ref}$ are the values of the cost index today and in the reference year, respectively. Cost indices for special types of equipment and processes are published frequently in several journals (e.g., *Chemical Engineering*, *The Oil and Gas Journal*, and *Engineering News Record*).

## B. Calculation of Revenue Requirements

The annual total revenue requirement (*TRR*, total product cost) for a system is the revenue that must be collected in a given year through the sale of all products to compensate the system operating company for all expenditures incurred in the same year and to ensure sound economic plant operation. It consists of two parts: carrying charges and expenses. Carrying charges are a general designation for charges that are related to capital investment, whereas expenses are used to define costs associated with the operation of a plant. Carrying charges (*CC*) include the following: total capital recovery; return on investment for debt, preferred stock, and common equity; income taxes; and other taxes and insurance. Examples for expenses are fuel cost (*FC*) and operating and maintenance costs (*OMC*). All annual carrying charges and expenses have to be estimated for each year over the entire economic life of a plant. Detailed information about the TRR method can be found in Bejan, Tsatsaronis, and Moran (1996) and in EPRI (1993).

## C. Levelized Costs

The series of annual costs associated with carrying charges $CC_j$ and expenses ($FC_j$ and $OMC_j$) for the $j$th year of plant operation is not uniform. In general, carrying charges decrease while fuel costs increase with increasing years of operation (Bejan, Tsatsaronis, and Moran, 1996). A levelized value $TRR_L$ for the total annual revenue requirement can be computed by applying a discounting factor and the capital-recovery factor CRF:

$$TRR_L = CRF \sum_{1}^{n} \frac{TRR_j}{(1 + i_{eff})^j}, \qquad (30)$$

where $TRR_j$ is the total revenue requirement in the $j$th year of plant operation, $i_{eff}$ is the average annual effective discount rate (cost of money), and $n$ denotes the plant economic life expressed in years. For Eq. (30), it is assumed that each money transaction occurs at the end of each year. The capital-recovery factor *CRF* is given by

$$CRF = \frac{i_{eff}(1 + i_{eff})^n}{(1 + i_{eff})^n - 1}. \qquad (31)$$

If the series of payments for the annual fuel cost $FC_j$ is uniform over time except for a constant escalation $r_{FC}$ (i.e., $FC_j = FC_0(1 + r_{FC})^j$), then the levelized value $FC_L$ of the series can be calculated by multiplying the fuel expenditure $FC_0$ at the beginning of the first year by the constant-escalation levelization factor *CELF*.

$$FC_L = FC_0 CELF = FC_0 \frac{k_{FC}(1 - k_{FC}^n)}{(1 - k_{FC})} CRF \qquad (32)$$

with

$$k_{FC} = \frac{1 + r_{FC}}{1 + i_{eff}} \quad \text{and} \quad r_{FC} = \text{constant.}$$

The terms $r_{FC}$ and $CRF$ denote the annual escalation rate for the fuel cost and the capital-recovery factor [Eq. (31)], respectively.

Accordingly, the levelized annual operating and maintenance costs $OMC_L$ are given by

$$OMC_L = OMC_0 CELF$$

$$= OMC_0 \frac{k_{OMC}(1 - k_{OMC}^n)}{(1 - k_{OMC})} CRF \qquad (33)$$

with

$$k_{OMC} = \frac{1 + r_{OMC}}{1 + i_{eff}} \quad \text{and} \quad r_{OMC} = \text{constant.}$$

The term $r_{OMC}$ is the nominal escalation rate for the operating and maintenance costs.

Finally, the levelized carrying charges $CC_L$ are obtained from

$$CC_L = TRR_L - FC_L - OMC_L. \qquad (34)$$

The major difference between a conventional economic analysis and an economic analysis conducted as part of a thermoeconomic analysis is that the latter is done at the plant component level. The annual carrying charges (superscript $CI$ = capital investment) and operating and maintenance costs (superscript $OM$) of the total plant can be apportioned among the system components according to the contribution of the $k$th component to the purchased-equipment cost $PEC_{tot} = \sum_k PEC_k$ for the overall system:

$$\dot{Z}_k^{CI} = \frac{CC_L}{\tau} \frac{PEC_k}{\sum_k PEC_k}; \qquad (35)$$

$$\dot{Z}_k^{OM} = \frac{OMC_L}{\tau} \frac{PEC_k}{\sum_k PEC_k}. \qquad (36)$$

Here, $PEC_k$ and $\tau$ denote the purchased-equipment cost of the $k$th plant component and the total annual time (in hours) of system operation at full load, respectively. The term $\dot{Z}_k$ represents the cost rate associated with capital investment and operating and maintenance expenses:

$$\dot{Z}_k = \dot{Z}_k^{CI} + \dot{Z}_k^{OM}. \qquad (37)$$

The levelized cost rate of the expenditures for fuel $\dot{C}_F$ supplied to the overall system is given by

$$\dot{C}_F = \frac{FC_L}{\tau}. \qquad (38)$$

Levelized costs, such as $\dot{Z}_k^{CI}$, $\dot{Z}_k^{OM}$, and $\dot{C}_F$, are used as input data for the thermoeconomic analysis.

## D. Sensitivity Analysis

An economic analysis generally involves more uncertainties than a thermodynamic analysis. In the above discussion, it has been assumed that each variable in the economic analysis is known with certainty. However, many values used in the calculation are uncertain. A sensitivity analysis determines by how much a reasonable range of uncertainty assumed for each uncertain variable affects the final decision. Sensitivity studies are recommended to investigate the effect of major assumptions about values referring to future years (e.g., cost of money, inflation rate, and escalation rate of fuels) on the results of an economic analysis.

## III. THERMOECONOMIC ANALYSIS

The exergy analysis yields the desired information for a complete evaluation of the design and performance of an energy system from the thermodynamic viewpoint. However, we still need to know how much the exergy destruction in a system component costs the system operator. Knowledge of this cost is very useful in improving the cost effectiveness of the system.

In a thermoeconomic analysis, evaluation and optimization, the cost rates associated with each material and energy stream are used to calculate component-related thermoeconomic variables. Thermoeconomic variables deal with investment costs and the corresponding costs associated with thermodynamic inefficiencies. Based on rational decision criteria, the required changes in structure and parameter values are identified. Thus, the cost minimization of a thermal system involves finding the optimum trade-offs between the cost rates associated with capital investment and exergy destruction.

### A. Exergy Costing

In exergy costing a cost is assigned to each exergy stream. The cost rate associated with the $j$th material stream is expressed as the product of the stream exergy rate $\dot{E}_j$ and the average cost per exergy unit $c_j$:

$$\dot{C}_j = \dot{E}_j \cdot c_j = \dot{m}_j \cdot e_j \cdot c_j, \qquad (39)$$

where $e_j$ is the specific exergy on a mass basis. A cost is also assigned to the exergy transfers associated with heat or work:

$$\dot{C}_q = c_q \cdot \dot{Q}; \qquad (40)$$

$$\dot{C}_w = c_w \cdot \dot{W}. \qquad (41)$$

Here, $c_j$, $c_q$, and $c_w$ denote average costs per unit of exergy in dollars per gigajoule exergy. The costs associated

with each material and energy stream in a system are calculated with the aid of cost balances and auxiliary cost equations.

Sometimes it is appropriate to consider the costs of physical and chemical exergy associated with a material stream separately. By denoting the average costs per unit of physical and chemical exergy by $c_j^{PH}$ and $c_j^{CH}$, respectively, the cost rate associated with stream $j$ becomes

$$\dot{C}_j = c_j E_j = c_j^{PH} \dot{E}_j^{PH} + c_j^{CH} \dot{E}_j^{CH}$$
$$= \dot{m}_j \left( c_j^{PH} e_j^{PH} + c_j^{CH} e_j^{CH} \right). \qquad (42)$$

## B. Cost Balance

Exergy costing involves cost balances usually formulated for each system component separately. A cost balance applied to the $k$th system component expresses that the total cost of the exiting streams equals the total cost of the entering streams plus the appropriate charges due to capital investment and operating and maintenance expenses $\dot{Z}$ (Fig. 3).

$$\sum_{j=1}^{n} \dot{C}_{j,k,in} + \underbrace{\dot{Z}_k^{CI} + \dot{Z}_k^{OM}}_{\dot{Z}_k} = \sum_{j=1}^{m} \dot{C}_{j,k,out} \qquad (43)$$

Cost balances are generally written so that all terms are positive. The rates $\dot{Z}_k^{CI}$ and $\dot{Z}_k^{OM}$ are calculated from Eqs. (35) and (36).

Introducing the cost rate expressions from Eqs. (39)–(41), we obtain

$$\sum_{j=1}^{n} (c_j \dot{E}_j)_{k,in} + \dot{Z}_k^{CI} + \dot{Z}_k^{OM} = \sum_{j=1}^{m} (c_j \dot{E}_j)_{k,out}. \qquad (44)$$

The exergy rates $\dot{E}_j$ entering and exiting the $k$th system component are calculated in an exergy analysis. In a thermoeconomic analysis of a component, we may assume that the costs per exergy unit $c_j$ are known for all entering streams. These costs are known either from the components they exit or, if they are streams entering the overall system, from their purchase costs. Consequently, the unknown variables that need to be calculated with the
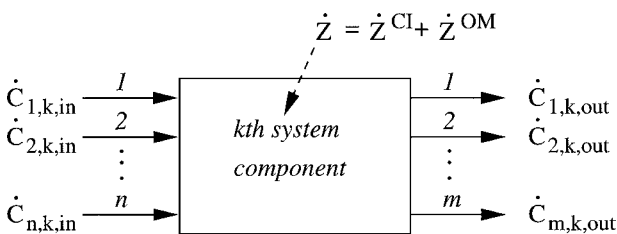


**FIGURE 3** Schematic of a system component to illustrate a cost balance.

aid of the cost balance are the costs per exergy unit of the exiting streams. Usually, some auxiliary relations are required to calculate these costs.

In the exergetic evaluation, a fuel and a product were defined for each component of a system (see Table II). The cost flow rates associated with the fuel $\dot{C}_F$ and product $\dot{C}_P$ of a component are calculated in a similar way to the exergy flow rates $\dot{E}_F$ and $\dot{E}_P$. Then, the cost balance becomes

$$\dot{C}_P = \dot{C}_F + \dot{Z} - \dot{C}_L. \qquad (45)$$

The term $\dot{C}_L$ represents the monetary loss associated with the rejection of exergy to the environment. Table III shows the definitions of $\dot{C}_P$ and $\dot{C}_F$ for selected components at steady state.

## C. Auxiliary Costing Equations

When the costs of entering streams are known, a cost balance is generally not sufficient to determine the costs of the streams exiting a component because the number of exiting streams is usually larger than one. Thus, in general, it is necessary to formulate some auxiliary equations for a component, with the number of these equations being equal to the number of exiting streams minus one. The auxiliary equations are called either $F$ equations or $P$ equations depending on whether the exergy stream being considered is used in the calculation of fuel or product for the component (Lazzaretto and Tsatsaronis, November 1999).

$F$ Equations. The total cost associated with the removal of exergy from an exergy stream in a component must be equal to the cost at which the removed exergy was supplied to the same stream in upstream components. The exergy difference of this stream between inlet and outlet is considered in the definition of fuel for the component.

$P$ Equations. Each exergy unit is supplied to any stream associated with the product of a component at the same average cost $c_{P,k}$. This cost is calculated from the cost balance and the $F$ equations.

The following two examples illustrate the development of the $F$ and $P$ equations.

**Example 1.** Calculate the costs of the streams exiting the adiabatic turbine (subscript $T$) or expander in Table III. The cost balance for this component is

$$\dot{C}_1 + \dot{Z}_T = \dot{C}_2 + \dot{C}_3 + \dot{C}_4. \qquad (46)$$

By grouping together the terms associated with fuel and product, we obtain

**TABLE III   Cost Rates Associated with Fuel and Product for Selected Components at Steady State Operation**

| Component | Schematic | Cost rate of product $\dot{C}_P$ | Cost rate of fuel $\dot{C}_F$ |
|---|---|---|---|
| Compressor pump, or fan |  | $\dot{C}_2 - \dot{C}_1$ | $\dot{C}_3$ |
| Turbine or expander |  | $\dot{C}_4$ | $\dot{C}_1 - \dot{C}_2 - \dot{C}_3$ |
| Heat exchanger[a] |  | $\dot{C}_2 - \dot{C}_1$ | $\dot{C}_3 - \dot{C}_4$ |
| Mixing unit |  | $\dot{m}_1(c_{3,1}e_3 - c_1 e_1)$ with $c_{3,1} = c_3 + \dfrac{\dot{m}_2}{\dot{m}_1}(c_3 - c_2)$ | $\dot{m}_2 c_2(e_2 - e_3)$ |
| Combustion chamber |  | $\dot{C}_3 - \dot{C}_2$ | $\dot{C}_1 - \dot{C}_4$ |
| Steam generator |  | $(\dot{C}_6 - \dot{C}_5) + (\dot{C}_8 - \dot{C}_7)$ | $(\dot{C}_1 + \dot{C}_2) - (\dot{C}_3 + \dot{C}_4)$ |
| Deaerator |  | $\dot{m}_2(c_{3,2}e_3 - c_2 e_2)$ with $c_{3,2} = c_3 + \dfrac{\dot{m}_3 - \dot{m}_2}{\dot{m}_2}(c_3 - c_1)$ | $(\dot{m}_3 - \dot{m}_2)c_1(e_1 - e_3)$ $+ \dot{m}_4 c_1(e_1 - e_4)$ |
| Compressor with cooling air extractions |  | $\dot{C}_3 + \dot{C}_4 + \dot{C}_5 - \dot{C}_2$ | $\dot{C}_1$ |
| Expander with cooling air supply |  | $\dot{C}_3$ | $\dot{C}_1 + \dot{C}_4 + \dot{C}_5 - \dot{C}_2$ |

*continues*

**TABLE III** (*Continued*)

| Component | Schematic | Cost rate of product $\dot{C}_P$ | Cost rate of fuel $\dot{C}_F$ |
|---|---|---|---|
| Fuel cell |  | Fuel cell as part of a system:<br>$\dot{C}_7 + (\dot{C}_6 - \dot{C}_5) + (\dot{C}_3^{PH} - \dot{C}_1^{PH})$<br>$+ (\dot{C}_4^{PH} - \dot{C}_2^{PH})$<br>Stand-alone fuel cell:<br>$\dot{C}_7 + (\dot{C}_6 - \dot{C}_5)$ | Fuel cell as part of a system:<br>$(\dot{C}_1^{CH} + \dot{C}_2^{CH}) - (\dot{C}_3^{CH} + \dot{C}_4^{CH})$<br>Stand-alone fuel cell:<br>$\dot{C}_1 + \dot{C}_2$ |
| Ejector |  | $\dot{m}_1(c_{3,1}e_3 - c_1 e_1)$<br>where<br>$c_{3,1} = c_3 + \frac{\dot{m}_2}{\dot{m}_1}(c_3 - c_2)$ | $\dot{m}_2 c_2 (e_2 - e_3)$ |
| Distillation column[b] |  | $\dot{C}_2^{CH} + \dot{C}_3^{CH} + \dot{C}_6^{CH} + \dot{C}_7^{CH} - \dot{C}_1^{CH}$<br>$+ \dot{m}_6(c_6^{PH}e_6^{PH} - c_1^{CH}e_1^{PH})$<br>$+ \dot{m}_2(c_2^{PH}e_2^{PH} - c_1^{CH}e_1^{PH})$ | $(\dot{C}_4 - \dot{C}_5) + \dot{m}_7(c_1^{PH}e_1^{PH}$<br>$- c_7^{PH}e_7^{PH}) + \dot{m}_3(c_1^{PH}e_1^{PH}$<br>$- c_3^{PH}e_3^{PH})$ |
| Gassifier |  | $\dot{C}_3^{CH} + (\dot{C}_3^{PH} + \dot{C}_4^{PH}$<br>$- \dot{C}_2^{PH} - \dot{C}_1^{PH})$ | $\dot{C}_1^{CH} + \dot{C}_2^{CH} - \dot{C}_4^{CH}$ |
| Steam reformer |  | $\dot{C}_3 - \dot{C}_1 - \dot{C}_2$ | $\dot{C}_4 + \dot{C}_5 - \dot{C}_6$ |
| Evaporator including steam drum |  | $\dot{C}_2 + \dot{C}_5 - \dot{C}_1$ | $\dot{C}_3 - \dot{C}_4$ |

[a] These definitions assume that the purpose of the heat exchanger is to heat the cold steam ($T_1 > T_0$). If the purpose of the heat exchanger is to provide cooling ($T_3 < T_0$), then the following relations should be used: $\dot{C}_P = \dot{C}_4 - \dot{C}_3$ and $\dot{C}_F = \dot{C}_1 - \dot{C}_2$.

[b] Here, it is assumed that $e_j^{CH} > e_1^{CH}$ ($j = 2, 3, 6, 7$), $e_2^{PH} > e_1^{PH}$, $e_6^{PH} > e_1^{PH}$, $e_3^{PH} < e_1^{PH}$, $e_7^{PH} < e_1^{PH}$.

$$c_p \cdot \dot{E}_P = \underbrace{\dot{C}_4}_{\dot{C}_P} = \underbrace{(\dot{C}_1 - \dot{C}_2 - \dot{C}_3)}_{\dot{C}_F} + \dot{Z}_T. \qquad (47)$$

The $F$ equation for this component states that the cost associated with each removal of exergy from the entering material stream (stream 1) must be equal to the average cost at which the removed exergy ($\dot{E}_1 - \dot{E}_2 - \dot{E}_3$) was supplied to the same stream in upstream components.

$$\dot{C}_1 - \dot{C}_2 - \dot{C}_3 = c_1 \cdot (\dot{E}_1 - \dot{E}_2 - \dot{E}_3) \qquad (48)$$

Equation (48) is equivalent to

$$c_2 = c_3 = c_1. \qquad (49)$$

The unknown cost rate $\dot{C}_4$ can be calculated with the aid of Eq. (48) or (49) and the cost balance [Eq. (47)].

**Example 2.** Calculate the costs of the streams exiting the adiabatic steam generator (subscript *SG*) shown in Table III, if streams 3 and 4 are (case a) rejected to the environment without additional expenses and (case b) discharged to the environment after using ash handling and gas cleaning equipment.

The cost balance for this component is

$$\dot{C}_1 + \dot{C}_2 + \dot{C}_5 + \dot{C}_7 + \dot{Z}_{SG} = \dot{C}_3 + \dot{C}_4 + \dot{C}_6 + \dot{C}_8. \quad (50)$$

By grouping together the terms associated with fuel and product, we obtain

$$c_p \cdot \dot{E}_P = \underbrace{(\dot{C}_6 - \dot{C}_5) + (\dot{C}_8 - \dot{C}_7)}_{\dot{C}_P}$$

$$= \underbrace{(\dot{C}_1 + \dot{C}_2) - (\dot{C}_3 + \dot{C}_4)}_{\dot{C}_F} + \dot{Z}_{SG}. \quad (51)$$

• Case a: The costs associated with streams 3 and 4 are calculated from *F* equations: Each exergy unit in the ash (stream 3) and in the flue gas (stream 4) is supplied by coal and oxygen at the same average cost.

$$\frac{\dot{C}_3}{\dot{E}_3} = \frac{\dot{C}_4}{\dot{E}_4} = \frac{\dot{C}_1 + \dot{C}_2}{\dot{E}_1 + \dot{E}_2} \quad (52)$$

or

$$c_3 = c_4 = \frac{\dot{C}_1 + \dot{C}_2}{\dot{E}_1 + \dot{E}_2} \quad (53)$$

The *F* equations [Eq. (52)] and the cost balance [Eq. (51)] may be used to calculate the $c_P$ value in Eq. (51). Then, the *P* equations supply the costs associated with streams 6 and 8: each exergy unit is supplied to the main steam stream and to the reheat stream at the same average cost.

$$c_P = \frac{\dot{C}_6 - \dot{C}_5}{\dot{E}_6 - \dot{E}_5} = \frac{\dot{C}_8 - \dot{C}_7}{\dot{E}_8 - \dot{E}_7} \quad (54)$$

• Case b: All costs associated with the final disposal of streams 3 and 4 must be charged to the useful streams exiting the boiler, that is, to the main steam and the hot reheat.

The following equations are obtained from the overall system (Fig. 4) and the cost balances for ash handling (subscript *ah*) and gas cleaning (subscript *gc*):

$$\frac{\dot{C}_9}{\dot{E}_9} = \frac{\dot{C}_{10}}{\dot{E}_{10}} = \frac{\dot{C}_1 + \dot{C}_2}{\dot{E}_1 + \dot{E}_2}, \quad (55)$$

$$\dot{C}_3 = \dot{C}_9 - \dot{Z}_{ah}, \quad (56)$$

$$\dot{C}_4 = \dot{C}_{10} - \dot{Z}_{gc}. \quad (57)$$

The cost rates $\dot{C}_3$ and $\dot{C}_4$ are usually negative in case b. The unknown cost rates $\dot{C}_3$, $\dot{C}_4$, $\dot{C}_6$, $\dot{C}_8$, $\dot{C}_9$, and $\dot{C}_{10}$ are calculated from Eqs. (51) and (54)–(57).
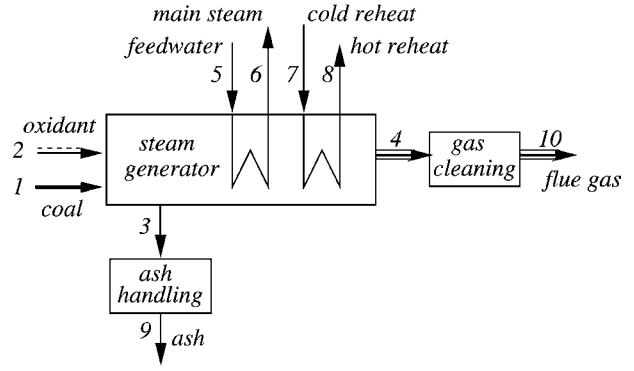


**FIGURE 4** Schematic of a steam generator to illustrate case b in Example 2.

For the system components having only one exiting stream, such as compressors without extractions, pumps, fans, mixing units, and ejectors, the cost associated with this exiting stream can be calculated from the cost balance Eq. (43). When the costs of physical and chemical exergy are considered separately (e.g., for a distillation column and a gasifier in Table III), and when more than one exergy stream exits a system component, *F* or *P* equations are required, depending on whether the exiting stream being considered belongs to the fuel or to the product.

The costing equations for the turbine and the steam generator are discussed above in Examples 1 and 2. In the following, the cost balances and the required *F* and *P* equations are given for the remaining devices shown in Table III.

• Heat exchanger (*HX*):

$$(\dot{C}_2 - \dot{C}_1) = \dot{Z}_{HX} + (\dot{C}_3 - \dot{C}_4) \quad (58)$$

$$\frac{\dot{C}_4}{\dot{E}_4} = \frac{\dot{C}_3}{\dot{E}_3} \quad (59)$$

• Combustion chamber (*CC*):

$$(\dot{C}_3 - \dot{C}_2) = \dot{Z}_{CC} + (\dot{C}_1 - \dot{C}_4) \quad (60)$$

$$\frac{\dot{C}_4}{\dot{E}_4} = \frac{\dot{C}_1}{\dot{E}_1} \quad (61)$$

• Deaerator (*DA*):

$$\dot{C}_4 + \dot{C}_3 = \dot{C}_1 + \dot{C}_2 + \dot{Z}_{DA} \quad (62)$$

$$\frac{\dot{C}_4}{\dot{E}_4} = \frac{\dot{C}_1}{\dot{E}_1} \quad (63)$$

• Compressor with cooling air extractions (*C*):

$$c_3 \dot{E}_3 + c_4 \dot{E}_4 + c_5 \dot{E}_5 - c_2 \dot{E}_2 = c_1 \dot{E}_1 + \dot{Z}_C \quad (64)$$

$$\frac{c_3 e_3 - c_2 e_2}{e_3 - e_2} = \frac{c_4 e_4 - c_2 e_2}{e_4 - e_2} = \frac{c_5 e_5 - c_2 e_2}{e_5 - e_2} \quad (65)$$

- Expander with cooling air supply ($T$):

$$\dot{C}_3 = (\dot{C}_1 + \dot{C}_4 + \dot{C}_5 - \dot{C}_2) + \dot{Z}_T \qquad (66)$$

$$\frac{\dot{C}_2}{\dot{E}_2} = \frac{\dot{C}_1 + \dot{C}_4 + \dot{C}_5}{\dot{E}_1 + \dot{E}_4 + \dot{E}_5} \qquad (67)$$

- Fuel cell ($FC$): The purpose of the fuel cell shown in Table III is to generate electricity and to supply useful thermal energy.

$$\dot{C}_7 + (\dot{C}_6 - \dot{C}_5) = (\dot{C}_1 + \dot{C}_2) - (\dot{C}_3 + \dot{C}_4) + \dot{Z}_{FC} \quad (68)$$

The $F$ and $P$ equations for a fuel cell as part (component) of a system are

$$\frac{\dot{C}_3^{CH}}{\dot{E}_3^{CH}} = \frac{\dot{C}_1^{CH}}{\dot{E}_1^{CH}}; \qquad \frac{\dot{C}_4^{CH}}{\dot{E}_4^{CH}} = \frac{\dot{C}_2^{CH}}{\dot{E}_2^{CH}} \qquad (69)$$

$$\frac{\dot{C}_3^{PH} - \dot{C}_1^{PH}}{\dot{E}_3^{PH} - \dot{E}_1^{PH}} = \frac{\dot{C}_4^{PH} - \dot{C}_2^{PH}}{\dot{E}_4^{PH} - \dot{E}_2^{PH}} = \frac{\dot{C}_6 - \dot{C}_5}{\dot{E}_6 - \dot{E}_5} = \frac{\dot{C}_7}{\dot{E}_7}. \quad (70)$$

For a stand-alone fuel cell (system), we obtain

$$\frac{\dot{C}_3}{\dot{E}_3} = \frac{\dot{C}_1}{\dot{E}_1}; \qquad \frac{\dot{C}_4}{\dot{E}_4} = \frac{\dot{C}_2}{\dot{E}_2} \qquad (71)$$

and

$$\frac{\dot{C}_6 - \dot{C}_5}{\dot{E}_6 - \dot{E}_5} = \frac{\dot{C}_7}{\dot{E}_7}. \qquad (72)$$

For a stand-alone fuel cell, the exergies of streams 3 and 4 are exergy losses.

- Distillation column ($DC$): The purpose of a distillation column is to separate the chemical components in the feed, i.e., to increase the chemical exergy of the substances in the feed at the expense of the physical exergy supplied to the system with the feed and in the reboiler. For this component, it is appropriate to consider the costs of physical and chemical exergy for streams 1, 2, 3, 6, and 7 separately.

The cost balance for the distillation column is given by

$$
\begin{aligned}
& c_2^{CH} \dot{E}_2^{CH} + c_3^{CH} \dot{E}_3^{CH} + c_6^{CH} \dot{E}_6^{CH} + c_7^{CH} \dot{E}_7^{CH} \\
& \quad - c_1^{CH} \dot{E}_1^{CH} + \dot{m}_6\left(c_6^{PH} e_6^{PH} - c_1^{PH} e_1^{PH}\right) \\
& \quad + \dot{m}_2\left(c_2^{PH} e_2^{PH} - c_1^{PH} e_1^{PH}\right) + (c_9 \dot{E}_9 - c_8 \dot{E}_8) \\
& = (c_4 \dot{E}_4 - c_5 \dot{E}_5) + \dot{m}_7\left(c_1^{PH} e_1^{PH} - c_7^{PH} e_7^{PH}\right) \\
& \quad + \dot{m}_3\left(c_1^{PH} e_1^{PH} - c_3^{PH} e_3^{PH}\right) + \dot{Z}_{DC}. \quad (73)
\end{aligned}
$$

The $F$ equations are

$$c_5 = c_4, \qquad (74)$$

$$c_7^{PH} = c_3^{PH} = c_1^{PH}. \qquad (75)$$

Each exergy unit is supplied to the product at the same average cost ($P$ equations):

$$
\begin{aligned}
\frac{\dot{m}_6\left(c_6^{PH} e_6^{PH} - c_1^{PH} e_1^{PH}\right)}{\dot{m}_6\left(e_6^{PH} - e_1^{PH}\right)} &= \frac{\dot{m}_2\left(c_2^{PH} e_2^{PH} - c_1^{PH} e_1^{PH}\right)}{\dot{m}_2\left(e_2^{PH} - e_1^{PH}\right)} \\
&= \frac{\dot{m}_2\left(c_2^{CH} e_2^{CH} - c_1^{CH} e_1^{CH}\right)}{\dot{m}_2\left(e_2^{CH} - e_1^{CH}\right)} = \frac{\dot{m}_3\left(c_3^{CH} e_3^{CH} - c_1^{CH} e_1^{CH}\right)}{\dot{m}_3\left(e_3^{CH} - e_1^{CH}\right)} \\
&= \frac{\dot{m}_6\left(c_6^{CH} e_6^{CH} - c_1^{CH} e_1^{CH}\right)}{\dot{m}_6\left(e_6^{CH} - e_1^{CH}\right)} = \frac{\dot{m}_7\left(c_7^{CH} e_7^{CH} - c_1^{CH} e_1^{CH}\right)}{\dot{m}_7\left(e_7^{CH} - e_1^{CH}\right)}.
\end{aligned}
$$
$$(76)$$

The exergy increase of the cooling water (stream 8) is an exergy loss.

$$c_9 = c_8 \qquad (77)$$

- Gasifier ($GS$):

The purpose of a gasifier is to convert the chemical exergy of a solid fuel into the chemical exergy of a gaseous fuel. Due to the chemical reactions involved, the physical exergy of the material streams increase between the inlet and the outlet.

$$
\begin{aligned}
\dot{C}_3^{CH} + \left(\dot{C}_3^{PH} + \dot{C}_4^{PH} - \dot{C}_1^{PH} - \dot{C}_2^{PH}\right) \\
= \dot{C}_1^{CH} + \dot{C}_2^{CH} - \dot{C}_4^{CH} + \dot{Z}_{GS} \quad (78)
\end{aligned}
$$

When the chemical and physical exergies are considered separately, the following $P$ equation is obtained:

$$\frac{\dot{C}_3^{CH}}{\dot{E}_3^{CH}} = \frac{\dot{C}_3^{PH} + \dot{C}_4^{PH} - \dot{C}_1^{PH} - \dot{C}_2^{PH}}{\dot{E}_3^{PH} + \dot{E}_4^{PH} - \dot{E}_1^{PH} - \dot{E}_2^{PH}}, \qquad (79)$$

$$\frac{\dot{C}_4^{CH}}{\dot{E}_4^{CH}} = \frac{\dot{C}_1^{CH}}{\dot{E}_1^{CH}}. \qquad (80)$$

- Steam reformer ($SR$):

$$\dot{C}_3 - \dot{C}_1 - \dot{C}_2 = \dot{C}_4 + \dot{C}_5 - \dot{C}_6 + \dot{Z}_{SR}, \qquad (81)$$

$$\frac{\dot{C}_6}{\dot{E}_6} = \frac{\dot{C}_4 + \dot{C}_5}{\dot{E}_4 + \dot{E}_5}. \qquad (82)$$

- Evaporator including steam drum ($EV$):

$$\dot{C}_2 + \dot{C}_5 - \dot{C}_1 = \dot{C}_3 - \dot{C}_4 + \dot{Z}_{EV}, \qquad (83)$$

$$\frac{c_5 e_5 - c_1 e_1}{e_5 - e_1} = \frac{c_2 e_2 - c_1 e_1}{e_2 - e_1}. \qquad (84)$$

## IV. THERMOECONOMIC EVALUATION

In a thermoeconomic evaluation, the cost per unit of exergy and the cost rates associated with each material and exergy stream are used to calculate thermoeconomic variables for

each system component. From the exergy analysis we already know the exergetic variables: rate of exergy destruction $\dot{E}_{D,k}$; exergetic efficiency $\varepsilon_k$; and exergy destruction ratio $y_{D,k}$.

## A. Thermoeconomic Variables

The definition of fuel and product for the exergetic efficiency calculation in a component leads to the cost flow rates associated with the fuel $\dot{C}_{F,k}$ and product $\dot{C}_{P,k}$ of the the $k$th component. $\dot{C}_{F,k}$ represents the cost flow rate at which the fuel exergy $\dot{E}_{F,k}$ is provided to the $k$th component. $\dot{C}_{P,k}$ is the cost flow rate associated with the product exergy $\dot{E}_{P,k}$ for the same component.

The average cost of fuel for the $k$th system component $c_{F,k}$ expresses the average cost at which each exergy unit of fuel (as defined in the exergetic efficiency) is supplied to the $k$th system component.

$$c_{F,k} = \frac{\dot{C}_{F,k}}{\dot{E}_{F,k}}. \tag{85}$$

The value of $c_{F,k}$ depends on the relative position of the $k$th component in the system and on the interconnections between the $k$th component and the remaining components. As a general rule, the closer the $k$th component is to the product (fuel) stream of the overall system, usually the larger (smaller) the value of $c_{F,k}$.

Similarly, the unit cost of product $c_{P,k}$ is the average cost at which each exergy unit was supplied to the product for the $k$th component:

$$c_{P,k} = \frac{\dot{C}_{P,k}}{\dot{E}_{P,k}}. \tag{86}$$

Using Eqs. (85) and (86), the cost balance for the $k$th system component can be written as

$$\dot{c}_{P,k} \dot{E}_{P,k} = c_{F,k} \dot{E}_{F,k} + \dot{Z}_k - \dot{C}_{L,k}. \tag{87}$$

One of the most important aspects of exergy costing is calculating the cost of exergy destruction in each component of the energy system being considered. The cost rate $\dot{C}_{D,k}$ associated with exergy destruction $\dot{E}_{D,k}$ in the $k$th system component is a "hidden" cost that can be revealed only through a thermoeconomic analysis. It can be approximated by the cost of the additional fuel that needs to be supplied to this component to cover the exergy destruction and to generate the same exergy flow rate of product $\dot{E}_{P,k}$:

$$\dot{C}_{D,k} = c_{F,k} \dot{E}_{D,k}, \quad \text{when } \dot{E}_{P,k} = \text{const.} \tag{88}$$

For most well-designed components, as the exergy destruction decreases or efficiency increases, the cost of exergy destruction $\dot{C}_{D,k}$ decreases while the capital investment $\dot{Z}_k^{CI}$ increases. The higher the $\dot{C}_{D,k}$ value is,

the lower the $\dot{Z}_k$ value. It is one of the most interesting features of thermoeconomics that the exergy destruction costs associated with a component are estimated and compared with the investment costs of the same component. This comparison facilitates the decision about the design changes that might improve the cost effectiveness of the overall system. In a thermoeconomic optimization we try to find the appropriate trade-offs between $\dot{C}_{D,k}$ and $\dot{Z}_k$. The $\dot{C}_{D,k}$ values cannot be added to calculate the $\dot{C}_D$ value for a group of components because each $\dot{C}_{D,k}$ already contains information related to the interactions among the cost of exergy destruction in upstream components. The $\dot{C}_D$ value for the group should be calculated using the average cost per exergy unit of fuel $c_F$ for the group and the exergy destruction rate within the group.

Figure 5 shows the relation between investment cost per unit of product exergy $\dot{Z}_k^{CI}/\dot{E}_{P,k}$ and exergy destruction per unit of product exergy $\dot{E}_{D,k}/\dot{E}_{P,k}$ for the $k$th component. The shaded area illustrates the range of variation of the investment cost due to uncertainty and to multiple technical design solutions that might be available. The investment cost per unit of product exergy $c_{P,k}^Z$ increases with decreasing exergy destruction per unit of product exergy or with increasing exergetic efficiency. This cost behavior is exhibited by most components. The components that exhibit a decrease of $c_{P,k}^Z$ with increasing efficiency do not need to be considered in a thermoeconomic evaluation since for these components no optimization dilemma exists: Among all available solutions we would use the most efficient component that has both the lowest specific fuel expenses and the lowest specific investment cost $c_{P,k}^Z$ and, thus, the minimum $c_{P,k}$ value (Tsatsaronis and Park, 1999).

The relative cost difference $r_k$ between the average cost per exergy unit of product and average cost per exergy unit of fuel is given by

$$r_k \equiv \frac{c_{P,k} - c_{F,k}}{c_{F,k}} = \frac{c_{F,k}(\dot{E}_{D,k} + \dot{E}_{L,k}) + \left(\dot{Z}_k^{CI} + \dot{Z}_k^{OM}\right)}{c_{F,k} \dot{E}_{P,k}}$$

$$= \frac{1 - \varepsilon_k}{\varepsilon_k} + \frac{\dot{Z}_k^{CI} + \dot{Z}_k^{OM}}{c_{F,k} \dot{E}_{P,k}}. \tag{89}$$

This equation reveals the real cost sources in the $k$th component, which are the capital cost $\dot{Z}_k^{CI}$, the exergy destruction within the component as expressed by $\dot{C}_{D,k}$, the operating and maintenance costs, and the exergy loss. Among these sources the first two are the most significant ones and are used for calculating the exergoeconomic factor.

The exergoeconomic factor expresses the contribution of the capital cost to the sum of capital cost and cost of exergy destruction:
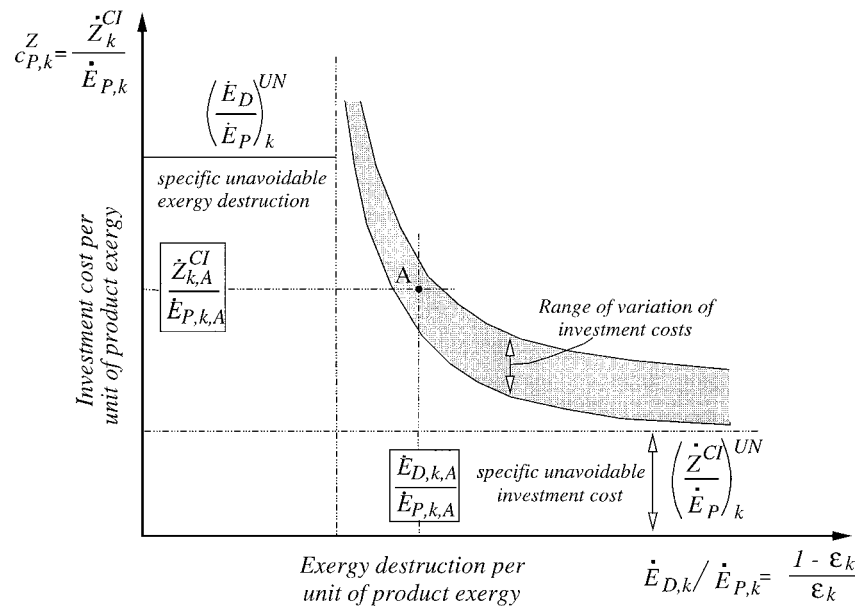
**FIGURE 5** Expected relationship between investment cost and exergy destruction (or exergetic efficiency) for the $k$th component of a thermal system [Based on Fig. 1 in Tsatsaronis, G., and Park, M. H. (1999). "ECOS'99, Efficiency Costs, Optimization, Simulation and Environmental Aspects of Energy Systems," pp. 161–121, Tokyo, Japan, June 8–10.]

$$f_k \equiv \frac{\dot{Z}_k^{CI}}{\dot{Z}_k^{CI} + \dot{C}_{D,k}} = \frac{\dot{Z}_k^{CI}}{\dot{Z}_k^{CI} + c_{F,k}\dot{E}_{D,k}}. \qquad (90)$$

The thermoeconomic variables $\dot{Z}_k^{CI}$ and $\dot{C}_{D,k}$ provide absolut measures of the importance of the $k$th component, whereas the variables $r_k$ and $f_k$ provide relative measures of the component cost effectiveness.

## B. Design Evaluation

A detailed thermoeconomic evaluation of the design of a thermal system is based on the following exergetic and thermoeconomic variables, each calculated for the $k$th system component:

| | | |
|---|---|---|
| $\dot{E}_{D,k}$ | rate of exergy destruction | Eq. (25) |
| $\varepsilon_k$ | exergetic efficiency | Eq. (24) |
| $y_{D,k}$ | exergy destruction ratio | Eq. (26) |
| $\dot{Z}_k^{CI}$ | cost rate associated with capital investment | Eqs. (35) and (37) |
| $\dot{C}_{D,k}$ | cost rate associated with exergy destruction | Eq. (88) |
| $r_k$ | relative cost difference | Eq. (89) |
| $f_k$ | exergoeconomic factor | Eq. (90) |

The values of all exergetic and thermoeconomic variables depend on the component type (e.g., turbine, heat exchanger, chemical reactor). For instance, the value of the exergoeconomic factor is typically between 25 and 65% for compressors and turbines.

The following guidelines may be applied to the evaluation of the $k$th system component to improve the cost effectiveness of the entire system:

1. Rank the components in descending order of cost importance using the sum $\dot{Z}_k^{CI} + \dot{C}_{D,k}$.
2. Consider initially design changes for the components for which the value of this sum is high.
3. Pay particular attention to components with a high relative cost difference $r_k$, especially when the cost rates $\dot{Z}_k^{CI}$ and $\dot{C}_{D,k}$ are high.
4. Use the exergoeconomic factor $f_k$ to identify the major cost source (capital investment or cost of exergy destruction).
   a. If the $f_k$ value is high, investigate whether it is cost effective to reduce the capital investment for the $k$th component at the expense of the component efficiency.
   b. If the $f_k$ value is low, try to improve the component efficiency by increasing the capital investment.
5. Eliminate any subprocesses that increase the exergy destruction or exergy loss without contributing to the reduction of capital investment or of fuel costs for other components.
6. Consider improving the exergetic efficiency of a component if it has a relatively low exergetic efficiency or a relatively large value for the rate of exergy destruction, or the exergy destruction ratio.

# V. ITERATIVE OPTIMIZATION OF A THERMAL SYSTEM

Design optimization of a thermal system means the modification of the structure and the design parameters of a system to minimize the total levelized cost of the system products under boundary conditions associated with available materials, financial resources, protection of the environment, and government regulation, together with safety, reliability, operability, availability, and maintainability of the system (Bejan, Tsatsaronis, and Moran, 1996). A truly optimized system is one for which the magnitude of every significant thermodynamic inefficiency (exergy destruction and exergy loss) is justified by considerations related to costs or is imposed by at least one of the above boundary conditions.

The iterative thermoeconomic optimization technique discussed here is illustrated with the aid of a simplified cogeneration system as shown in Fig. 6. The purpose of the cogeneration system is to generate 30 MW net electric power and to provide 14 kg/s saturated steam at 20 bar (stream 7). A similar cogeneration system is used in Penner and Tsatsaronis (1994) for comparing different thermoeconomic optimization techniques. It should be emphasized that the cost minimization of a real system according to Fig. 6 is significantly simpler compared with the case considered here, since in a real system engineers have to select one of the few gas turbine systems available in the market for a given capacity and then have to optimize the design of only the heat-recovery steam generator. In this example, however, to demonstrate application of the iterative thermoeconomic optimization technique with the aid of a simple system containing different components, we assume that we can freely decide about the design of each component included in the gas turbine system.

The decision variables selected for the optimization are the compressor pressure ratio $p_2/p_1$, the isentropic compressor efficiency $\eta_{sc}$, the isentropic turbine efficiency $\eta_{st}$, and the temperature $T_3$ of the combustion products entering the gas turbine. All other thermodynamic variables can be calculated as a function of the decision variables. All costs associated with owning and operating this system are charged to the product streams 7 and 10. This includes the cost rate associated with the exergy loss from the overall system (stream 5).

The total cost rate associated with the product for the overall system $\dot{C}_{P,tot}$ is given by (Bejan, Tsatsaronis, and Moran, 1996)

$$\dot{C}_{P,tot} = (\dot{C}_7 - \dot{C}_6) + \dot{C}_{10} + \dot{C}_5 + \dot{Z}_{other}$$

$$= \dot{C}_1 + \dot{C}_8 + \sum_k \dot{Z}_k + \dot{Z}_{other} \overset{!}{=} min \quad (91)$$

$$k = AC, CC, GT, HRSG$$

Here, $\dot{Z}_{other}$ denotes the levelized cost associated with the capital investment for other system components not shown in Fig. 6. The subscripts AC, CC, GT, and HRSG refer to the components air compressor, combustion chamber, gas turbine, and heat-recovery steam generator, respectively. Equation (91) represents the objective function for the cost minimization.

The iterative thermoeconomic optimization of the cogeneration system is subject to the following constraints: $6.0 \le p_2/p_1 \le 14.0$; $0.75 \le \eta_{sc} \le 0.89$; $0.8 \le \eta_{st} \le 0.91$; and 900 K $\le T_3 \le$ 1600 K. The minimum temperature difference $\Delta T_{min}$ in the heat-recovery steam generator must be at least 15 K. A first workable design (base-case design) was developed using the following values of the decision variables:

$$p_2/p_1 = 14, \quad T_3 = 1260 \text{ K}, \quad \eta_{sc} = 0.86, \quad \eta_{st} = 0.87.$$

Relevant thermodynamic and economic data as well as the values of the thermoeconomic variables for the base-case design are shown in Tables IV and V. The thermoeconomic variables are used to determine the changes in the design of each component to be implemented in the next step of the iterative optimization procedure. Knowledge about how the decision variables qualitatively affect the exergetic efficiency and the costs associated with each system component is required to determine the new values of the decision variables that improve the cost effectiveness of the overall system. Engineering judgment and critical evaluations must be used when deciding on the changes to be made from one iteration step to the next.

## A. Evaluation of the First Design Case (Base-Case Design)

Table V summarizes the thermoeconomic variables and the purchased equipment costs $PEC$ calculated for each
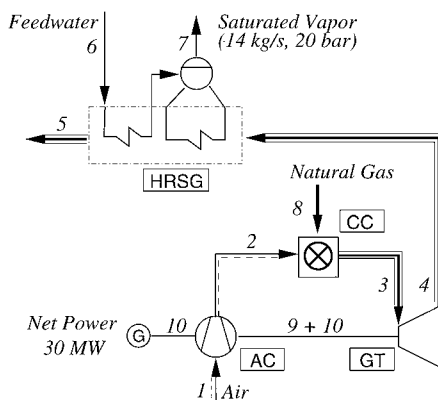


**FIGURE 6** A simple cogeneration system.

**TABLE IV  Mass Flow Rate, Temperature, Pressure, Exergy Flow Rates, Cost Per Exergy Unit, and Cost Flow Rate for Each Stream in the Base-Case Design of the Cogeneration System**

| Nr. | $\dot{m}_j$ (kg/s) | T (K) | p (bar) | $\dot{E}_j^{PH}$ (MW) | $\dot{E}_j^{CH}$ (MW) | $\dot{E}_j$ (MW) | $c_j$ ($/GJ) | $\dot{C}_j$ ($/h) |
|---|---|---|---|---|---|---|---|---|
| 1 | 132.960 | 298.15 | 1.013 | 0.000 | 0.000 | 0.000 | 0.00 | 0 |
| 2 | 132.960 | 674.12 | 14.182 | 49.042 | 0.000 | 49.042 | 32.94 | 5815 |
| 3 | 134.943 | 1260.00 | 13.473 | 115.147 | 0.351 | 115.498 | 18.11 | 7532 |
| 4 | 134.943 | 754.35 | 1.066 | 27.518 | 0.351 | 27.869 | 18.11 | 1817 |
| 5 | 134.943 | 503.62 | 1.013 | 7.305 | 0.351 | 7.656 | 18.11 | 499 |
| 6 | 14.000 | 298.15 | 20.000 | 0.027 | 0.035 | 0.062 | 0.00 | 0 |
| 7 | 14.000 | 485.52 | 20.000 | 12.775 | 0.035 | 12.810 | 34.25 | 1579 |
| 8 | 1.983 | 298.15 | 20.000 | 0.904 | 101.896 | 102.800 | 4.56 | 1689 |
| 9 | | | | | | 52.479 | 21.47 | 4057 |
| 10 | | | | | | 30.000 | 21.47 | 2319 |

component of the cogeneration system. According to the methodology described in Section IV.B, the components are listed in order of descending value of the sum $\dot{Z}^{CI} + \dot{C}_D$. The compressor and the turbine have the highest values of the sum $\dot{Z}^{CI} + \dot{C}_D$ and are the most important components from the thermoeconomic viewpoint.

The value of the exergoeconomic factor $f$ for the air compressor (AC) is high compared with a target value of below 65%. The $f$ factor indicates that most of the costs associated with the air compressor are due to capital investment. Although the cost per exergy unit of fuel $c_F$ supplied to the air compressor is the highest among all system components, the cost rate $\dot{C}_D$ associated with exergy destruction $\dot{E}_D$ in this component is relatively low. Only 6.46% of the total exergy destruction occurs in the air compressor. A decrease of the capital investment costs for the air compressor might be cost effective for the entire system even if this would decrease the exergetic efficiency $\varepsilon_{AC}$. These observations suggest a significant decrease in the values of $p_2/p_1$ and $\eta_{sc}$. Compared to the isentropic efficiency, the influence of the pressure ratio on the compressor exergetic efficiency is low. Here, a reduction of $p_2/p_1$ aims at a decrease of $\dot{Z}_{AC}$ only.

The value of the exergoeconomic factor $f$ for the gas turbine expander (GT) is within the range of the target values (between 25 and 65%, Section IV.B). Therefore, no recommendations can be derived from the thermoeconomic evaluation of the gas turbine expander.

The heat-recovery steam generator (HRSG) has the highest value of the relative cost difference $r$. The exergoeconomic factor $f$ indicates that most of the costs for this component are related to the cost rate associated with exergy destruction $\dot{C}_D$ [Eq. (89)]. The difference in the average thermodynamic temperature between the hot and the cold streams is a measure for the exergy destruction in the heat-recovery steam generator [Eq. (20)]. A reduction in the value of the temperature $T_3$ and/or an increase in the isentropic turbine efficiency $\eta_{st}$ or the pressure ratio $p_2/p_1$ lead to a decrease in the temperature $T_4$. Since the temperature profile of the cold stream and the heating duty in the heat-recovery steam generator are fixed, both the average temperature difference and the temperature $T_5$ of the exhaust gas would decrease. These changes should result in an increase in both the exergetic efficiency and the capital investment in the heat-recovery steam generator, as well as in a decrease in the exergy loss from the overall system.

**TABLE V  First Design Case (Base-Case Design)**[a]

| Name | PEC ($10^3$\$) | $\epsilon$ (%) | $\dot{E}_F$ (MW) | $\dot{E}_P$ (MW) | $\dot{E}_D$ (MW) | $y_D$ (%) | $c_F$ ($/GJ) | $c_P$ ($/GJ) | $\dot{C}_D$ ($/h) | $\dot{Z}^{CI}$ ($/h) | $r$ (%) | $f$ (%) | $\dot{Z}^{CI} + \dot{C}_D$ ($/h) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AC | 8732 | 93.56 | 52.48 | 49.10 | 3.38 | 3.29 | 21.47 | 32.90 | 261 | 1112 | 53.2 | 81.0 | 1374 |
| GT | 3282 | 94.12 | 87.63 | 82.48 | 5.15 | 5.01 | 18.11 | 21.47 | 335 | 418 | 18.5 | 55.4 | 754 |
| HRSG | 1297 | 63.07 | 20.21 | 12.75 | 7.46 | 7.26 | 18.11 | 34.41 | 486 | 165 | 90.0 | 25.3 | 652 |
| CC | 139 | 64.65 | 102.80 | 66.46 | 36.34 | 35.35 | 4.56 | 7.18 | 597 | 17 | 57.3 | 2.9 | 615 |

[a] $T_3 = 1260$ K, $p_2/p_1 = 14$, $\eta_{sc} = 0.86$, $\eta_{st} = 0.87$; overall system: $\varepsilon_{tot} = 41.58\%$; $\dot{C}_{P,tot} = \$4587$/h; $\dot{C}_{L,tot} = \dot{C}_5 = \$499$/h; HRSG: $\Delta T_{min} = 88$ K.

**TABLE VI   Sample of a Qualitative Decision Matrix for the First Design Case (Base-Case Design)[a]**

| System component | Objective $Z_k^{CI}$ or $\varepsilon_k$ | Decision variables | | | | | |
|---|---|---|---|---|---|---|---|
| | | $T_3$ | $p_2/p_1$ | $\eta_{sc}$ | $\eta_{st}$ | $\dot{Z}_k^{CI} + \dot{C}_{D,k}$ | |
| | | 1260 | 14 | 0.86 | 0.87 | | Initial values |
| AC | ↓ | — | ↓ | ↓ | — | 1374 | |
| GT | — | — | — | — | — | 753 | |
| HRSG | ↑ | ↓ | ↑ | — | ↑ | 652 | |
| CC | ↑ | ↑ | ↑ | ↓ | — | 615 | |
| | | — | ↓ | ↓ | ↑ | | Suggestions |
| | | 1260 | 12 | 0.84 | 0.88 | | New values |

[a] The symbols ↓, ↑, and — mean a decrease, an increase, or no change, respectively, in the value of a decision variable.

The combustion chamber (CC) has the largest exergy destruction rate. However, according to the sum $\dot{Z}^{CI} + \dot{C}_D$, the economic importance of this component is rather small. Each exergy unit of the fuel is supplied to the combustion chamber at a relatively low cost. Hence, each unit of exergy destruction can be covered at the same low cost. The low value of the exergoeconomic factor $f$ shows that $\dot{C}_{D,CC}$ is the dominant cost rate. However, only a part of the exergy destruction can be avoided by preheating the reactants or by reducing the excess air. A higher value of $p_2/p_1$ and/or a lower value of $\eta_{sc}$ lead to an increase in the temperature $T_2$ of the air entering the combustion chamber, whereas a lower temperature of the combustion products $T_3$ reduces the amount of excess air.

Table VI shows a sample of a qualitative decision matrix for the base-case design which summarizes the suggestions from the thermoeconomic evaluation of each component. Decreasing the values of the pressure ratio $p_2/p_1$ and the isentropic compressor efficiency $\eta_{sc}$ as well as increasing the isentropic turbine efficiency $\eta_{st}$ are expected to improve the cost effectiveness of the cogeneration system. Note, that the decrease in the $p_2/p_1$ value contradicts the corresponding suggestions from the heat-recovery steam generator and the combustion chamber. However, changes suggested by the evaluation of a component should only be considered if they do not contradict

changes suggested by components with a significantly higher value of the sum $\dot{Z}^{CI} + \dot{C}_D$. The temperature $T_3$ remains unchanged, since contradictory indications are obtained from the evaluations of the heat-recovery steam generator and the combustion chamber. The values of the sum $\dot{Z}^{CI} + \dot{C}_D$ for both components are in the same range. After summarizing the results from the evaluation of the base-case design, the following new values are selected for the decision variables in the second design case: $T_3 = 12604$ K (unchanged), $p_2/p_1 = 12$, $\eta_{sc} = 0.84$, and $\eta_{st} = 0.88$.

## B. Evaluation of the Second Design Case (1. Iteration)

Through the changes in the decision variables, the value of the objective function $\dot{C}_{P,tot}$ is reduced from \$4587/h to \$3913/h, and the cost rate associated with the exergy loss $\dot{C}_5$ decreased from \$499/h to \$446/h. The new values of the thermoeconomic variables are summarized in Table VII. The sum $\dot{Z}^{CI} + \dot{C}_D$ shows that the air compressor and the gas turbine expander are still the most important components from the thermoeconomic viewpoint. The importance of both components is due to the relatively high investment cost rate $\dot{Z}^{CI}$ and, to a lesser extent, to the high fuel cost $c_F$ for these components. The

**TABLE VII   Second Design Case (1. Iteration)[a]**

| Name | PEC ($10^3$\$) | $\epsilon$ (%) | $\dot{E}_F$ (MW) | $\dot{E}_P$ (MW) | $\dot{E}_D$ (MW) | $y_D$ (%) | $c_F$ (\$/GJ) | $c_P$ (\$/GJ) | $\dot{C}_D$ (\$/h) | $\dot{Z}^{CI}$ (\$/h) | $r$ (%) | $f$ (%) | $\dot{Z}^{CI} + \dot{C}_D$ (\$/h) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AC | 4633 | 92.37 | 48.70 | 44.98 | 3.72 | 3.55 | 18.00 | 25.31 | 241 | 601 | 40.63 | 71.4 | 841 |
| GT | 3802 | 94.77 | 83.04 | 78.70 | 4.34 | 4.14 | 14.47 | 18.00 | 226 | 493 | 24.40 | 68.5 | 719 |
| CC | 138 | 64.26 | 104.85 | 67.37 | 37.48 | 35.74 | 4.56 | 7.22 | 616 | 18 | 58.15 | 2.8 | 634 |
| HRSG | 1245 | 61.62 | 20.69 | 12.75 | 7.94 | 7.57 | 14.47 | 29.00 | 414 | 161 | 100.45 | 28.1 | 575 |

[a] $T_3 = 1260$ K, $p_2/p_1 = 12$, $\eta_{sc} = 0.84$, $\eta_{st} = 0.88$; overall system: $\varepsilon_{tot} = 40.77\%$; $\dot{C}_{P,tot} = \$3913/h$; $\dot{C}_{L,tot} = \dot{C}_5 = \$446/h$; HRSG: $\Delta T_{min} = 105$ K.

cost effectiveness of these components may be improved by lowering the investment cost at the expense of the exergetic efficiency. This can be achieved by a decrease in the values of the isentropic efficiencies $\eta_{sc}$ and $\eta_{st}$. For the third design, $\eta_{sc}$ and $\eta_{st}$ are reduced to 0.82 and 0.87, respectively.

Compared to the base-case design, the combustion chamber has now a higher relative cost importance than the heat-recovery steam generator. To reduce the cost of exergy destruction in the combustion chamber, the value of $T_3$ is increased to 1320 K.

Although the value of the relative cost difference $r$ increased for the heat-recovery steam generator from the base-case design to the second design, the sum $\dot{Z}^{CI} + \dot{C}_D$ for the heat-recovery steam generator decreased. The modifications in the upstream components (AC and CC) and the interactions among the system components lead to a decrease in the cost per exergy unit of the fuel $c_{F,HRSG}$. Therefore, the cost associated with exergy destruction is reduced even if the exergy destruction rate for the heat-recovery steam generator increased in the second design. The reduced pressure ratio $p_2/p_1$ outweighed the effect of the increased isentropic turbine efficiency $\eta_{st}$ on the temperature $T_4$. Hence, $T_4$ increased in the second design instead of an anticipated decrease. These observations show that the suggested changes based on the thermoeconomic evaluation of the other components may affect the cost effectiveness of the heat-recovery steam generator negatively. The pressure ratio $p_2/p_1$ remains unchanged in the third design, since the suggested changes in $\eta_{st}$ and $T_3$ already lead to a higher temperature $T_4$, which results in a larger exergy destruction rate in the heat-recovery steam generator.

## C. Third Design Case (2. Iteration)

As a result of the changes in the decision variables, the value of the objective function $\dot{C}_{P,tot}$ decreased to $3484/h. Table VIII shows the new values of the thermoeconomic variables for each component. The costs associated with the compressor and the turbine are reduced significantly, whereas the sum $\dot{Z}^{CI} + \dot{C}_D$ is almost un-

changed for the combustion chamber which now has the second highest value. The values of the exergoeconomic factor $f_{AC}$ and $f_{GT}$ are close to their target values. Further improvements in the cost effectiveness may be achieved by slightly decreasing the isentropic efficiencies and significantly increasing the temperature $T_3$, even if the exergy destruction rate in the heat-recovery steam generator and the cost rate associated with the exergy loss of the overall system $\dot{C}_5$ increase.

## VI. RECENT DEVELOPMENTS IN THERMOECONOMICS

Complex thermal systems cannot usually be optimized using mathematical optimization techniques. The reasons include system complexity; opportunities for structural changes not identified during model development; incomplete cost models; and inability to consider in the model additional important factors such as plant availability, maintenability, and operability. Even if mathematical techniques are applied, the process designer gains no insight into the real thermodynamic losses, the cost formation process within the thermal system, or on how the solution was obtained.

As an alternative, thermoeconomic techniques provide effective assistance in identifying, evaluating, and reducing the thermodynamic inefficiencies and the costs in a thermal system. They improve the engineer's understanding of the interactions among the system components and variables and generally reveal opportunities for design improvements that might not be detected by other methods. Therefore, the interest in applying thermoeconomics has significantly increased in the last few years. In the following, some recent developments in thermoeconomics are briefly mentioned.

To evaluate the thermodynamic performance and cost effectiveness of thermal systems and to estimate the potential for improvements it is always useful to know for the most important system components the avoidable part of exergy destruction, the cost associated with this avoidable part, and the avoidable investment cost

**TABLE VIII Third Design Case (2. Iteration)**[a]

| Name | $PEC$ ($10^3$\$) | $\epsilon$ (%) | $\dot{E}_F$ (MW) | $\dot{E}_P$ (MW) | $\dot{E}_D$ (MW) | $y_D$ (%) | $c_F$ (\$/GJ) | $c_P$ (\$/GJ) | $\dot{C}_D$ (\$/h) | $\dot{Z}^{CI}$ (\$/h) | $r$ (%) | $f$ (%) | $\dot{Z}^{CI}+\dot{C}_D$ (\$/h) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AC | 3251 | 91.47 | 46.67 | 42.69 | 3.98 | 3.70 | 15.18 | 20.95 | 218 | 430 | 38.1 | 66.4 | 648 |
| CC | 135 | 65.30 | 107.68 | 70.32 | 37.36 | 34.70 | 4.56 | 7.10 | 614 | 18 | 55.5 | 2.8 | 632 |
| GT | 2852 | 94.59 | 81.06 | 76.67 | 4.38 | 4.07 | 12.34 | 15.18 | 195 | 377 | 23.0 | 66.0 | 572 |
| HRSG | 1134 | 58.78 | 21.69 | 12.75 | 8.94 | 8.30 | 12.34 | 26.09 | 397 | 150 | 111.4 | 27.4 | 547 |

[a] $T_3 = 1320$ K, $p_2/p_1 = 12$, $\eta_{sc} = 0.82$, $\eta_{st} = 0.87$; overall system: $\varepsilon_{tot} = 39.70\%$; $\dot{C}_{P,tot} = $3484/h; $\dot{C}_{L,tot} = \dot{C}_5 = $453/h; HRSG: $\Delta T_{min} = 142$ K.

associated with each system component. Improvement efforts should then focus only on these avoidable parts of inefficiencies and costs. Figure 5 illustrates the assessment of specific unavoidable exergy destruction $\dot{E}_{D,k}^{UN}/\dot{E}_{P,k}$ and specific unavoidable investment cost $\dot{Z}_{k}^{UN}/\dot{E}_{P,k}$ for the $k$th component. The avoidable parts are obtained as the difference between total value and avoidable part. More information about this topic is provided by Tsatsaronis and Park (1999).

The design and improvement of a thermal system often involve application of heuristic rules. Due to the complexity of energy conversion systems, as well as to the uncertainty involved in some design decisions, computer programs using principles from the field of artificial intelligence and soft computing are useful tools for the process designer in improving a given design and in developing a new cost-effective thermal system. The benefits of combining knowledge-based and fuzzy approaches with an iterative thermoeconomic optimization technique are discussed together with some applications by Cziesla (2000).

Calculating the cost of each product stream generated by a thermal plant having more than one product is an important subject for which several approaches have been developed in the past. Some of these approaches use exergy-based or thermoeconomic methods; however, the results obtained by different methods may vary within a wide range. Recently, a new exergy-based approach was developed (Erlach, Tsatsaronis, and Cziesla, 2001) for (a) assigning the fuel(s) used in the overall plant to the product streams and (b) calculating the costs associated with each product stream with the aid of a thermoeconomic evaluation. This new approach is general, more objective than previous approaches, and flexible, i.e., it allows designers and operators of the plant to actively participate in the fuel and cost allocation process.

## SEE ALSO THE FOLLOWING ARTICLES

ENERGY EFFICIENCY COMPARISONS AMONG COUNTRIES
• ENERGY FLOWS IN ECOLOGY AND IN THE ECONOMY
• ENERGY RESOURCES AND RESERVES • GEOTHERMAL POWER STATIONS • SOLAR THERMAL POWER STATIONS
• THERMODYNAMICS

## BIBLIOGRAPHY

Ahrendts, J. (1980). "Reference states," *Energy—Int. J.* **5,** 667–677.

Bejan, A., Tsatsaronis, G., and Moran, M. (1996). "Thermal Design and Optimization," Wiley, New York.

Cziesla, F. (2000). "Produktkostenminimierung beim Entwurf komplexer Energieumwandlungsanlagen mit Hilfe von wissensbasierten Methoden," No. 438, Fortschr.-Ber. VDI, Reihe 6, VDI Verlag, Düsseldorf.

Erlach, B., Tsatsaronis, G., and Cziesla, F., A new approach for assigning costs and fuels to cogeneration products. *In* "ECOS'o1, Efficiency, Costs, Optimization, Simulation and Environmental Aspects of Energy Systems," pp. 759–770, Istanbul, Turkey, July 4–6.

Kotas, T. J. (1985). "The Exergy Method of Thermal Plant Analysis," Butterworths, London.

Lazzaretto, A., and Tsatsaronis, G. (November 1999). On the calculation of efficiencies and costs in thermal systems. *In* "Proceedings of the ASME Advanced Energy Systems Division" (S. M. Aceves, S. Garimella, and R. Peterson, eds.), AES-Vol. 39, pp. 421–430, ASME, New York.

Moran, M. J., and Shapiro, H. N. (1998). "Fundamentals of Engineering Thermodynamics," Wiley, New York.

Penner, S. S., and Tsatsaronis, G., eds. (1994). "Invited papers on exergoeconomics," *Energy—Int. J.* **19**(3), 279–318.

Sama, D. A. (September 1995). "The use of the second law of thermodynamics in process design," *J. Eng. Gas Turbines Power* **117,** 179–185.

Szargut, J., Morris, D. R., and Steward, F. R. (1988). "Exergy Analysis of Thermal, Chemical, and Metallurgical Processes," Hemisphere, New York.

Electric Power Research Institute (1993). "Technical assessment guide (TAG$^{TM}$)," Vol. 1, TR-102276-V1R7, Electric Power Research Institute, Palo Alto, CA.

Tsatsaronis, G. (1999). Strengths and limitations of exergy analysis. *In* "Thermodynamic Optimization of Complex Energy Systems" (A. Bejan, and E. Mamut, eds.), Vol. 69, pp. 93–100, Nato Science Series, Kluwer Academic, Dordrecht/Normell, MA.

Tsatsaronis, G., and Park, M.H. (1999). On avoidable and unavoidable exergy destructions and investment costs in thermal systems. *In* "ECOS'99, Efficiency, Costs, Optimization, Simulation and Environmental Aspects of Energy Systems," pp. 116–121, Tokyo, Japan, June 8–10.

# Waste-to-Energy (WTE) Systems

## S. S. Penner

*University of California, San Diego*

## GLOSSARY

Combustor (or Burner or Incinerator) Type[1]

**AFBC** Atmospheric fluidized-bed combustor.
**AFBCRW** Atmospheric fluidized-bed combustor with refractory wall.
**CFBC** Circulating fluidized-bed combustor.
**CFBCWW** Circulating fluidized-bed combustor with a water wall.
**MB** Mass burner (a system that combusts all of the incoming material).
**MBRW** Mass burner with refractory wall.
**MCU** Modular combustion unit.
**MWC or MWI** Municipal-waste combustor or municipal-waste incinerator.
**RBCCRW** Rotary bed combustion chamber with refractory wall.

---

[1]The letter for combustor (C) may be replaced by B for burner or I for incinerator.

**RDF** Refuse-derived fuel.
**RDFC** Refuse-derived-fuel combustor.
**RDFP** Refuse-derived-fuel processor (other than a combustor).
**RWWC** Rotary water-wall combustor.
**SSCWW** Spreader and stoker combustor with a water wall.
**WTE** Waste to energy.

APC (Air-Pollution Control) System

**CI** (Activated) carbon injection.
**CyS** Cyclone separator.
**DSI** Duct sorbent (dry) injection (downstream of the furnace).
**ESP** Electrostatic precipitator.
**FF** Fabric filter.
**FSI** (Dry) furnace sorbent injection.
**SDA** Spray dryer absorber.
**SDS** Spray-dryer scrubber (used interchangeably with SDA).

**SNCR** Selective noncatalytic reduction (used for NO$_x$ control).
**WS** Wet scrubber.


## Energy Output and Material Recovery

**COG** Cogenerators (produce both electricity and steam).
**MW$_e$** Megawatts of electrical power (gross).
**STM/hr** Steam production per hour.
**TPY of FeX** Tons per year of ferrous metals recovered.
**TPY of O** Tons per year of materials other than ferrous metals recovered.


## I. HISTORICAL OVERVIEW

About 25 years ago, augmented construction of WTE systems enjoyed wide support in the United States as a preferred solution to two major problems by (i) providing for the disposal of rapidly increasing sources of municipal wastes while (ii) supplying low-cost, mostly renewable fuels for electric-power and steam generation. The important contribution to waste management is the result of a 90 to 95% reduction in the remaining waste volume, which is accompanied by the production of stable, nonleachable, nonhazardous ash that is usable in light aggregates for structural fill, road base, and subbase and as an additive to concrete or asphalt.

With the recognition of possible health effects that may be caused by emissions of even minute amounts of metals such as mercury or cadmium and of toxic compounds of the type belonging to the 210 congeners of chlorinated dioxins and furans, opposition to waste incineration became widespread, especially in the United States and Germany. Because of the availability of large areas of open and unused land in the United States, except in the vicinity of many of our urban population centers, there was renewed development of now "sanitary" landfills (i.e., landfills with impoundments, seals against leakage to the local water table, etc.). The preferred environmental agenda may be summarized by the goal of the "3 Rs" standing for "Reduce, Reuse, Recycle." *Complete* success of this agenda would obviously eliminate wastes and remove the necessity for waste management altogether. Although considerable reductions in waste generation were achieved for a large number of manufactured products, these reductions were accompanied by the creation of new sources of waste associated with the development of new technologies. Unfortunately, the net result has not been a significant reduction in *per capita* waste generation. EPA estimates have remained at about 1500 lbs of waste per person per year in the United States.

It should be noted that WTE has one important and environmentally attractive feature. Since a large percentage of combustible municipal waste is derived from biomass, WTE represents in effect one of a very small number of economically viable technologies for plant-derived-product utilization in combustors. These have the distinguishing feature of allowing energy generation without producing a net addition of carbon dioxide to the atmosphere.

New WTE plants are now in the planning stage or under construction in Southeast Asia, Eastern Europe, Latin America, and elsewhere. In the developed countries, especially the United States, Japan, and Western Europe, stringent emission controls are mandated by regulatory agencies. The complex assessments of possible health effects fall beyond the scope of our deliberations and should properly be provided by toxicologists and epidemiologists. Here, we content ourselves with descriptive material on system design, performance, and EPA-imposed emission regulations, as well as brief remarks on the economic aspects of technology utilization. WTE systems have probably become the most severely regulated of all combustion technologies with respect to allowed emissions of potentially hazardous compounds and elements.

The primary competition for waste incineration is represented worldwide by waste disposal in landfills. Sanitary landfills generally have disposal costs that are not much less than those for incineration, when allowance is made for energy recovery. Old-fashioned landfills can be constructed at much lower cost than either modern incinerators or sanitary landfills. For this reason, they have remained the preferred option in countries with large open spaces and low wast-disposal budgets. The use of WTEs made strong gains in the United States while methods for land disposal were improved and sanitary landfills became the norm. At this point, the use of WTEs became a less compelling option. In the EPA hierarchy of preferred methods for municipal waste disposal, the use of landfills and of municipal waste incinerators occupy the bottom two positions because of presumed or demonstrated undesirable health effects. Of these least desirable two options, which is the worst? The author of at least one study (Jones, 1997) awards the bottom prize to landfills after considering a wide range of estimable environmental impacts for the durations of the service lives of the two systems.


## II. INTRODUCTORY REMARKS ON WTE PLANTS

WTE plants are constructed and sold by a number of vendors and have some common design features (Penner

*et al.*, 1987). The operational sequence involves trash delivery and storage in a large pit, often followed by shredding, removal of magnetic materials, and drying before loading onto a stoker-grate bed. An average heat content for the incoming dry waste in the United States is around 11,600 kJ/kg (5000 Btu/lb) with considerable variability depending on the location (e.g., urban or rural sources) and time of year. The worldwide variability of heat content is even greater because waste generation reflects ethnic materials-use patterns, as well as preincineration separations such as removals of plant wastes, papers, plastics, glass, metals (aluminum and iron), etc., which may contribute significantly to heat-content reductions or increases.

Although different MWIs have different designs, there are some common performance features. Grate systems transport trash through the combustion zones with sequential drying, devolatilization, and burning. Primary (underfire) air is forced upward through the grate. The combined action of forced air and grate movement facilitates mixing and combustion. Grate residence times vary but fall generally in the range of 20 to 40 min. The ash-handling system controls ash movement. Hot gases rise from the grate into the furnace where secondary air injection causes complete turbulent burnout of combustibles. Overall air-to-fuel ratios are rich in oxidizer and are generally about 2 times stoichiometric.

MWIs have water-walled furnaces. The boiler feedwater is preheated by radiative and convective heat transfer to the walls before superheating by flue gases downstream. If desired, an underfire preheater may be used or the furnace feedwater may be preheated in an economizer downstream of the superheater. Removal of particulates from the gaseous products is now generally accomplished in baghouses. These have largely replaced electrostatic precipitators, which have been found to be less effective than baghouses for the removal of trace metals, dioxins, and furans. Furthermore, wet and dry scrubbers for flue gas are employed in specially designed systems to lower the effluent-gas concentrations of toxic materials such as mercury, cadmium, and congeners of dioxins and furans.

A summary of operational WTE plants for 1996 in the United States is given in Table I.

## III. INITIAL GROWTH RATES FOR MWI USE IN THE UNITED STATES

Before the normal growth rate of MWIs was seriously interrupted in the United States around 1988–89 because of environmental concerns involving especially dioxin and furan emissions, a logistics curve was used to esti-

**TABLE I   Operational WTE Plants in the United States in May 1996[a]**

| Technology type | Number of operating plants | Design capacity (mt/day) | Annual capacity ($10^6$ mt) |
|---|---|---|---|
| Mass-burn plants | 70 | 72,800 | 22.6 |
| Modular plants | 21 | 2,530 | 0.8 |
| Incinerators | 21 | 2,880 | 0.9 |
| RDF production and combustion | 15 | 22,600 | 7.0 |
|    Processing only | 11 | 4,530 | 1.4 |
|    Combustion only | 8 | 3,000 | 0.9 |
| Totals | 146 | 108,000 | 33.6 |
| MWCs without RDFs | 135 | 104,000 | 32.2 |
| WTE plants | 114 | 101,000 | 31.3 |

[From Taylor, A., and Zannes, M. (1996). Integrated Waste Services Association, Washington, D.C.]

[a] See Glossary for brief definitions of technology types.

mate the rate of market penetration as well as the ultimate value for market saturation (Penner and Richards, 1989). The latter was based on an assumed ultimate per capita municipal waste generation rate of 1200 lbs/person (p)-yr in view of stringent conservation efforts at the input end. With the optimistic value of 30% of total waste recycled, ultimate disposal would involve 840 lbs per person per year dedicated for incineration assuming that landfilling would become unacceptable. With this scenario and a stable U.S. population of $300 \times 10^6$, technology saturation would correspond to $252 \times 10^9$ lbs/yr of waste incineration. At $\$10^5/(t/d)$ as the required capital investment, the total potential capital investment in incineration facilities becomes $[\$10^5/(t/d)] \times (252 \times 10^9 \text{ lbs/yr}) \times (1 \text{ t}/2000 \text{ lbs}) \times (1 \text{ yr}/365\text{d}) \simeq \$36 \times 10^9$. Because new construction of MWIs in the United States was essentially terminated in 1993, only about one-fifth of this investment had actually been made by the end of 1998.

Technology growth rates according to the logistics curve may be estimated from the Verhulst equation,

$$F(t)/[1 - F(t)] = \{F(t_1)/G - (t_1)\} \exp k(t - t_1), \quad (1)$$

where $F(t)$ is the ratio of technology utilization at time $t$ to technology utilization at market saturation ($t \to \infty$). If we assume that $F(t_1) = 0.01$ for $t_1 = 1978$ and determine $F(t_2 = 1988)$ from the known 1988 U.S. incineration capacity of $5.09 \times 10^4$ t/d, then $k = 0.297$ yr$^{-1}$ and

$$F(t)/[1 - F(t)] = 0.196 \exp(0.297\Delta t), \quad \Delta t = t - 1988 \quad (2)$$

Although Eq. (2) was followed reasonably well during the early and middle 1980s, investments in MWIs

**TABLE II   Management of Municipal Solid Wastes[a] in Selected Countries (1996)**

| Country | Percentage recycled | Waste-to-energy (%) | Waste to landfill (%) | Percentage of WTE Ash Used[b] |
|---|---|---|---|---|
| Canada | 21 | 5 | 74 | Very small |
| Denmark | 23 | 48 | 29 | 60–90 (bottom ash) |
| France | 13 | 42 | 42 | 45 (bottom ash) |
| Germany | 18 | 36 | 46 | 60 (bottom ash) |
| Japan | 5 | 72 | 23 | 10 (bottom ash) |
| Netherlands | 20 | 35 | 45 | 40 (fly ash), 90 + bottom ash |
| Sweden | 19 | 47 | 34 | Under consideration |
| United States | 23 | 15 | 62 | 6.4 (fly + bottom ash) |

[a] Dogett *et al.* (1980).
[b] Bottom ash is discharged from the furnace and used for pavement construction, fly ash is discharged from the pollution-control equipment and may be used in asphaltic materials. Unused ash is landfilled.

decreased greatly around the end of the decade and further increases in $F(t)$ ceased altogether around 1993.

## IV. INTEGRATED WASTE MANAGEMENT

Integrated Waste Management (IWM) refers to the coherent composite approach used in waste management with first preference generally accorded in decreasing order to the 3 Rs: *reduce*, *reuse*, *recycle*. Reducing waste has proved to be an elusive goal because population and per capita income growths have tended to counteract more efficient packaging and manufacturing. In 1988, per capita solid-waste production was estimated at 1000 to 1500 lbs/yr. An EPA estimate for the year 2000 is $215 \times 10^6$ mt/yr or, for a population of $300 \times 10^6$ people, per capita solid waste production of 1600 lbs/yr per person, which is uncomfortably close to the high estimate made 12 years earlier (Dogett *et al.*, 1980). The EPA expects that one-third of this waste will be recycled, which is a high estimate and about 50% above the levels achieved (Dogett *et al.*, 1980) during 1996 in any developed country (see Table II).

It is evident from Table II that the ratio of waste-to-energy (WTE) to waste-to-landfill (WTL) varies widely for the listed sample of developed countries from a low of 0.07 in Canada to 0.24 in the United States to 0.78 in Germany and 3.1 in Japan. Large ratios of unpopulated to populated land areas are important factors in allowing widespread use of landfills. As probably the most severely regulated industry in terms of pollutant emissions with regard to health and environmental impacts, MWIs have served to teach us a great deal about monitoring, controlling, and coping with minuscule emissions of toxic materials.

## V. THE 1999 INTEGRATED WASTE SERVICES ASSOCIATION WTE DIRECTORY OF U.S. PLANTS

There were 111 operating plants in the United States in 1999, of which 103 (64 operated by IWSA members[2]) were WTE plants often consisting of a number of furnace units with capacities ranging from about 1120 for modules (individual combustors) to 73,920 tpd (tons per day) for complete mass burn units and an aggregate annual capacity of $32.7 \times 10^6$ t of which $30.7 \times 10^6$ t ($26 \times 10^6$ t by IWSA members) were used for energy production (Kiser and Zannes, 1999). This waste processed in WTE facilities represents about 15% ($\sim$13% by IWSA members) of the total 1999 U.S. waste disposal, serves 39 (33 for IWSA members) million people, is used in 31 (23 for IWSA members) states, involves the recovery of 773,000 tpy of ferrous metals, on-site recycling of paper, ash, nonferrous metals, etc., for a total of 462,000 tpy, generates 2770 $MW_e$ of which 2,570 (2240 for ISWA members) are $MW_e$ and 200 $MW_e$ equivalent are generated as steam.

Air-pollution control devices are widely but not universally used as follows: Of 69 mass-burn units, 51 have SDAs (spray dryer absorbers or scrubbers), 1 has a WS (wet scrubber), 0 have a C (cyclone), 9 have LIs (lime injection), 49 FFs (fabric filters), 22 ESPs (electrostatic precipitators), 25 SNCRs (selective noncatalytic reduction for $NO_x$ control), and 28 CIs (activated carbon injections); of 13 modular systems, 2 have SDAs, 3 WSs, 1 has a C, 2 have LIs, 3 FFs, 7 ESPs, none has an SNCR, and 2 have CIs; of 13 RDFs (refuse-derived-fuel) facilities

[2] The IWSA included in 1999 the following companies: American Ref-Fuel Co., Constellation Power Inc., Energy Answers Corp., Foster Wheeler Power Systems Inc., Katy-Seghes Inc., Montenay Power Corp., Ogden Energy Group Inc., Westinghouse Electric Corp., and Wheelabrator Technologies Inc.

processing and burning the fuel, 9 have SDAs, 2 WSs, 1 has a C, 1 an LI, 8 have FFs, 4 ESPs, 3 SNCRs, and 2 CIs; of 8 RDF-Cs (fuel-burning only), 2 have SDAs, none has a WS, 1 has a C, 2 have LIs, 5 FFs, 2 ESPs, none has an SNCR or CI. According to EPA standards for the end of December 2000, units processing more than 250 tpd must meet R-MACT (revised maximum achievable control technology) standards for acid-gas control [by using SDA and DSI (dry duct sorbent injection downstream of the furnace), FSI (dry furnace sorbent injection), CYC (cyclone separation), or any combination of these], $NO_x$ control (e.g., with SNCR), controls for Hg, total tetra-to-octachloro dioxins and furans, and other specified pollutants (e.g., by using CI). Meeting revised MACT limits will require pollution-control retrofits at many plants, although the R-MACT limits are already exceeded at some of the large modern plants as is illustrated by the following data: for tetra-to-octachloro dioxins and furans, optimally achieved performance is 1 to 12 nanogram (ng)/dscm (dry standard cubic meter) vs remaining R-MACT limits of 30 ng/dscm for systems using SD/FF and 60 ng/dscm for SD/ESP systems; for Hg, the R-MACT limit is 0.080 mg/dscm vs 0.008 to 0.035 mg/dscm achieved; for Pb, the R-MACT limit is 0.44 mg/dscm vs 0.001 to 0.027 mg/dscm achieved; for Cd, the R-MACT limit is 0.040 mg/dscm vs 0.001 to 0.004 mg/dscm achieved; for particulate matter (PM), the R-MACT limit remains 27 mg/dscm$^3$ vs 1 to 14 mg/dscm$^3$ achieved; for HCl, the R-MACT limit is 31 ppmv (parts per million by volume) vs 2 to 27 ppmv achieved; for $SO_2$, the R-MACT limit is 29 ppmv vs 1 to 21 achieved; for $NO_x$, the new R-MACT limit is 205 ppmv for MB (mass burn)/WW (water wall), remains at 250 ppmv for MB/RWW (rotary water-wall combustor), remains at 250 ppmv for RDF, is reduced to 180 ppmv for FBC (fluidized-bed com-

fractory walls) vs 36 to 166 ppmv achieved in the best units.

It should be noted that the R-MACT limits represent the average of the best 12% of permitted and operating systems rather than the best achieved values for any system. These perform at the lower limits in the overview given in the preceding paragraph.

## VI. LIFE CYCLE COMPARISON OF WTE WITH SELECTED FOSSIL-FUEL BURNERS FOR ELECTRICITY PRODUCTION

A life cycle (LC) comparison was prepared (Ecobalance, Inc., 1997) for IWSA in 1997 by Ecobalance Inc. The primary results in relative units per $kW_eh$ produced are

the following: acidification potential in terms of g of $H^+$ equivalent for WTE $= 0.056$ vs 0.37 for coal, 0.21 for fuel oil, and 0.061 for NG; g of $CO_2$ produced $= 554$ for WTE vs 1333 for coal, 871 for fuel oil, and 665 for NG; natural resources depletion index $= -0.69$ for WTE vs 1.0 for coal, 14.3 for fuel oil, and 28.6 for NG. These estimates depend among many other factors on the heats of combustion and compositions of MSW, emission species and concentrations for the three fuels, source compositions and production methods for the fossil fuels, designs and construction of plants and auxiliaries, ash composition and management, etc.

## VII. CURRENT AND FUTURE APPLICATIONS OF WASTE INCINERATION WORLDWIDE

An extensive assessment (H. Kaiser Consultancy, 1996) of current and future applications of municipal waste incinerations contains the following estimates as of 1996: In spite of recent selective shutdowns (e.g., 2 out of 3 operating units in Hong Kong), 2400 large-scale combustors remain in operation, 150 units are under construction, and 250 additional units are expected to be built by 2005, i.e., about 2800 units should be in service by 2005. The dominant technology will remain grate incineration with growth especially for fluidized-bed combustion and, to a lesser extent, for pyrolysis units. Small- and medium-size plants (up to 300 mt/day) will be preferred. WTE plants will represent the systems of choice because of the economic values of electricity generation and steam production. System utilization will grow preferentially in Asia with rapid urbanization and also in Eastern Europe.

New facilities will ultimately have to meet rigorous environmental controls everywhere and will therefore become more costly with time as R-MACT cleanup and possibly later tightening of these standards become universal requirements. A rough estimate of a typical plant cost for a 300 mt/day plant with R-MACT cleanup system is about \$100 to \$150 $\times 10^6$ in 1999 U.S. dollars, corresponding to \$12.5 $\times 10^9$ for 250 new units, i.e., a market for new systems of about \$3 $\times 10^9$/year to the year 2005.

## VIII. POTENTIAL ELECTRICITY PRODUCTION FROM WTE SYSTEMS IN THE UNITED STATES

The potential for electricity production may be estimated for the United States by using a procedure described in

1988 (Penner *et al.*, 1998). The method is applicable to countries other than the United States.

With stringent implementation of energy-conservation measures, steady-state generation of municipal wastes in the United States may possibly be restricted to 1200 lb/yr-p. This waste has roughly (Penner *et al.*, 1998) an energy content of $10^4$ kJ/kg. By the year 2050, a steady-state population of about $325 \times 10^6$ will then produce waste leading to an electric power output from incineration of

$$P_{MSW1} \simeq [(1200 \text{ lb/yr p}) \times (0.7)] \times (325 \times 10^6 \text{ p})$$
$$\times (10^4 \text{ kJ/kg}) \times (1 \text{ kg/2.2 lb})$$
$$\times (2.78 \times 10^{-4} \text{ kWh/kJ}) \times (1 \text{ yr/8.76} \times 10^3 \text{ h})$$
$$\times (0.2 \text{ kWe/kW}) \times (1 \text{ MWe/}10^3 \text{ kWe})$$
$$= 7.9 \times 10^3 \text{ MWe}$$

if 70% of the waste is incinerated while 30% is recycled or reused and 20% conversion efficiency of the heat release to electricity is achieved. The use of similar procedures for

hazardous and biomedical wastes led (Penner *et al.*, 1998) to estimates of $3.9 \times 10^3$ MWe and $0.46 \times 10^3$ MWe, respectively.

## IX. REPRESENTATIVE MWI SYSTEMS

A widely used system was developed by the Martin Company of Munich, Germany. The U.S. and Japanese licensees of this technology are, respectively, Ogden Corp. and Mitsubishi, both of which have important market shares in their home countries and elsewhere. A schematic of the Martin operating system is shown in Fig. 1. The novel feature of the process is the reverse-acting stoker grate, which is inclined at an angle of about 26° with respect to the horizontal to allow the trash to move forward by gravity while it is not pushed. The grate has sequential fixed and moving grate steps, with the latter stirring the waste against the downward direction. As the result, the burning refuse layer is continuously rotated
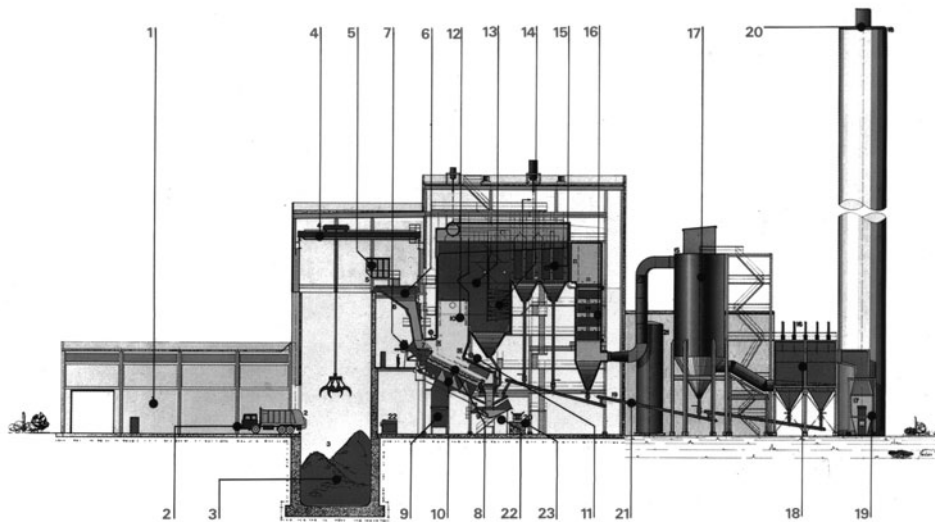


**FIGURE 1** Schematic of the MARTIN-System Technology (A) and furnace details (B), as implemented at the Bristol, CT (USA) resource-recovery facility. After weighing in the tipping hall (1), collection vehicles (2) dump unsorted waste into the storage pit (3). Overhead cranes (4) lift the waste to a bin (5) from which it enters the feed hopper (6) leading to the feed chute. Hydraulic ram feeders (7) provide controlled charging of the proprietary Martin reverse-acting stoker grate (8). A forced-draft fan (9) supplies primary underfire air below the grate (10) to the refuse burning on the stoker grate. Secondary above-fire air is injected at the front and rear walls above the grate to induce turbulence and cause burnout of the primary combustion products formed at and above the grate and in the furnace (12). The furnace walls and dividers between the boiler sections (13) are welded solid membranes. In the third pass (14) of the multipass boiler, the flue gases are cooled as they enter the horizontal superheater, which contains vertical superheater tubes (15). Economizer tubes (16) are located in the fifth pass of the boiler. During operation, the boiler surfaces are cleaned by sootblowers in the third, fourth, and fifth passes. A lime slurry is injected in the dry scrubber (17) to neutralize and capture acidic components of the flue gases. Spent lime is collected in the fabric-filter baghouse (18) together with fly ash. An induced-draft fan (19) moves the cleaned flue gases to the exhaust stack (20). Particulates collected in the dry scrubber (17) and baghouse (18) are moved via fly-ash conveyers (21) to the proprietary discharger (22) and then transferred by conveyors (23) to residue storage before disposal in a landfill (or, at other locations, use in the manufacture of aggregate construction materials). The high-temperature steam produced in the boiler may be used for power generation on passage through a turbine generator or as input to a district-heating system.
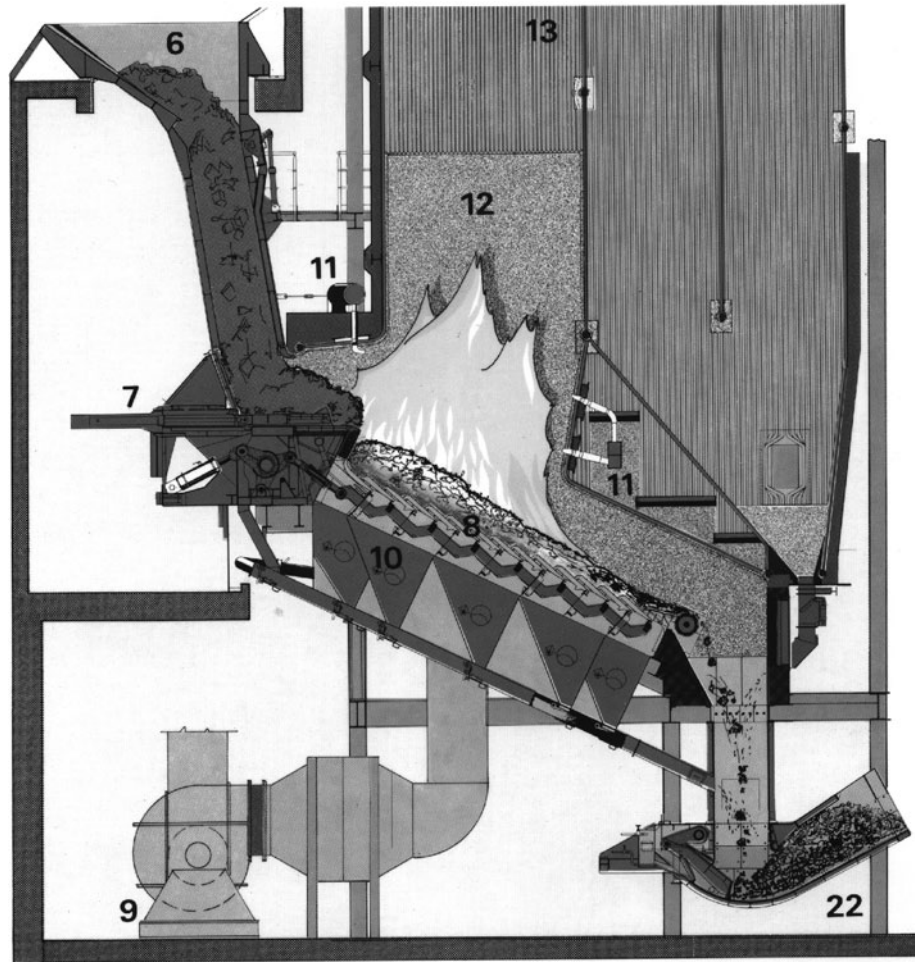
**FIGURE 1** (*continued*)

and mixed at a constant bed depth while red-hot material continues to move back in the direction of the feeder end. The net result is that vigorous refuse drying, ignition, and combustion occur along the entire grate length. These processes are properly controlled by dividing the entire grate into distinct sequential segments that are separately supplied with computer-monitored and controlled air for primary combustion. Partially burned combustion products rise above the grate and flow upward parallel to water-cooled furnace walls. Secondary air injectors provide the needed oxygen for complete combustion of fuel molecules. The total injected air-to-fuel ratio is about 180% of stoichiometric to ensure complete oxidation.

A *forward-acting* grate has been used by a number of vendors including Deutsche Babcock Anlagen (DBA), its American licensee American Ref-Fuel, Combustion Engineering, Vølund, von Roll, and others. The grate angle with respect to the horizontal is now greatly reduced below 26° since the primary force acting in the forward direction

is provided by the moving grate. Underfire and secondary air are injected in a manner analogous to that used in the Martin system and the furnace is again constructed with water-containing walls.

DBA developed the Düsseldorf grate with large rotating drums for stirring the incoming trash flow and facilitating intimate contact between primary underfire air and the flowing trash. Secondary air injection and burnout are achieved in a manner resembling the procedure sketched in Fig. 1.

Vølund has designed a system in which a rotary kiln is located downstream of a forward-acting grate. Gas-phase conversions have been implemented in furnaces with parallel flows of the primary combustion products and secondary air flows, as well as with counterflows and mixed flows of these gases.

We refer to the literature for such topics as incinerator modeling, temperature-profile estimations and verification, fouling of chamber walls, and formations of toxic compounds such as furans and dioxins in the gas phase

and on fly ash (which is probably the dominant process), as well as designs of cleanup systems that have led to the R-MACT standards described previously.

## X. ELECTRICITY AND TIPPING-FEE REVENUES

As an example, we consider a city with 1 million people generating 1.5 kg/day-p of municipal waste. The total waste production is then $(1.5 \text{ kg/day-p}) \times (10^6 \text{ p}) \times (1 \text{ mt}/10^3 \text{ kg}) = 1.5 \times 10^3$ mt/d. For an assumed heat of combustion of 10 MJ/kg and a fossil-energy-to-electricity conversion efficiency of 0.2, electricity generation per mt of refuse is then $(10^4 \text{ kJ/kg}) \times (10^3 \text{ kg/mt}) \times (1 \text{ kW}_e\text{h}/3600 \text{ kJ}) \times 0.2 = 560 \text{ kW}_e\text{h/mt}$. With refuse generation of $1.5 \times 10^3$ mt/d, the total electricity production is then $(560 \text{ kW}_e\text{h/mt}) \times (1.5 \times 10^3 \text{ mt/d}) \times (1 \text{ d}/24 \text{ h}) = 35 \text{ MW}_e$. A competitive price for electricity sales in 1995 was about $\$20 \times 10^{-3}/\text{kW}_e\text{h}$ so that the total value of the electricity generated per year becomes $35 \text{ MW}_e \times (\$20 \times 10^{-3}/\text{kW}_e\text{h}) \times (10^3 \text{ kW}_e/\text{MW}_e) \times (24\text{h/d}) = \$16,800/\text{d}$. If the refuse tipping fees are $20/mt, then the return from tipping fees is $(\$20/\text{mt}) \times (1.5 \times 10^3 \text{ mt/d}) = \$30 \times 10^3/\text{d}$, i.e. of the total return of $46,800/d, about 36% is derived from electricity generation for the specified costs.

## XI. ENVIRONMENTAL ISSUES

The stringent R-MACT control specifications have been developed over a long period of time (more than 15 years) and reflect a compromise between what is possible and what is desirable in order to protect public health while allowing improvements of valuable technologies. The emission specifications for each component and especially for the 210 congeners of chlorinated dioxins and furans were arrived at after completion of a great deal of research and many medical and epidemiological assessments of possible harm done by emissions to the air and/or water by these compounds. These studies have not yet been concluded.

Municipal waste incinerators have figured prominently in the debate about environmental equity. However, a 1990 census showed that modern WTE facilities are predominantly located in white, middle-class neighborhoods with 6% more whites than corresponds to the national average and a median household income 7% above the national average (Carr, 1996).

## SEE ALSO THE FOLLOWING ARTICLES

ENERGY FLOWS IN ECOLOGY AND IN THE ECONOMY • ENERGY RESOURCES AND RESERVES • HAZARDOUS WASTE INCINERATION • RADIOACTIVE WASTE DISPOSAL • RENEWABLE ENERGY FROM BIOMASS • WASTEWATER TREATMENT AND WATER RECLAMATION

## BIBLIOGRAPHY

Carr, G. L. (1996). "Environmental equity: Does it play a role in WTE siting?" *J. Hazardous Mat.* **47,** 303–312.

Dogett, R. M., O'Farrell, M. K., and Watson, A. L. (1980). "Forecasts of the Quantity and Composition of Solid Waste," 171 pp., EPA-600/5-80-001, EPA, Cincinnati, OH.

Ecobalance, Inc. (May 1997). "Life Cycle Comparison of Energy Production of Waste-to-Energy Facility to Other Major Fuel Sources."

Jones, K. H. (March–April 1997). "Comparing air emissions from landfills and WTE plants," *Solid Waste Technol.*, 28–39.

H. Kaiser Consultancy (1996). Philosophenweg 2, D-72076, Tübingen, Germany.

Kiser, J. V. L., and Zannes, M. (February 1999). IWSA, 1401 H Street NW, Washington, D.C.

NREL/BR-430-21437 (December 1996). Published by the National Renewable Energy Laboratory (NREL), 1617 Cole Boulevard, Golden, CO 80401-3393.

Penner, S. S., and Richards, M. B. (1989). "Estimates of growth rates for municipal waste incineration and environmental control costs for coal utilization in the U.S.," *Energy—Int. J.* **14,** 961–963.

Penner, S. S., Wiesenhahn, D. F., and Li, C. P. (1987). "Mass burning of municipal wastes," *Ann. Rev. Energy* **12,** 415–444.

Penner, S. S., Chang, D. P. Y., Goulard, R., and Lester, T. (1988). "Waste incineration and energy recovery," *Energy—Int. J.* **13,** 845–851.

Taylor, A., and Zannes, M. (May 1996). Integrated Waste Services Association (IWSA), 1401 H Street NW, Washington, D.C.