Advances in Solar Energy

An Annual Review of Research and Development

Volume 2

ASES

Edited by
Karl W. Böer
and
John A. Duffie

Advances in Solar Energy

An Annual Review of Research and Development

Volume 2

Editorial Board

A Continuation Order Plan is available for this series. A continuation order will bring delivery of each new volume immediately upon publication. Volumes are billed only upon actual shipment. For further information please contact the publisher.

Advances in Solar Energy

An Annual Review of Research and Development

Volume 2

Edited by

Karl W. Böer
University of Delaware and SES, Inc.
Newark, Delaware

and

John A. Duffie
University of Wisconsin at Madison
Madison, Wisconsin

AMERICAN SOLAR ENERGY SOCIETY, INC.
Boulder, Colorado • Newark, Delaware
and
PLENUM PRESS
New York • London

ISBN 978-1-4613-9953-7 ISBN 978-1-4613-9951-3 (eBook)
DOI 10.1007/978-1-4613-9951-3

A Division of Plenum Publishing Corporation
233 Spring Street, New York, N.Y. 10013

FOREWORD

A number of significant changes have occurred in *Advances in Solar Energy* since Volume 1 appeared in 1982. The delays in publication of the second volume are the result of reorganization of the American Solar Energy Society, and the negotiation of a new publishing arrangement. Beginning with this volume, *Advances* is now published jointly by the Society and Plenum Press. The Editorial Board has been enlarged to be more representative of the different fields of solar energy conversion.

Production of *Advances* is being expedited through the use of modern word processing equipment and the TEX typesetting-editing program. We have gone to a single-column format to ease the problems of presenting long equations, and we expect that the user of the volume will find it easy to read. The use of TEX will make last minute updates possible. The external appearance of the volume matches that of Volume 1.

We expect that future volumes of this annual will be proceeding on schedule. We invite comments from users and correspondence from prospective authors of critical reviews.

Karl W. Böer
John A. Duffie

CONTENTS

CHAPTER 3 Principles and Technology of Biomass Gasification

Thomas B. Reed

CHAPTER 4 Biomass Pyrolysis: A Review of the Literature Part 2—Lignocellulose Pyrolysis

Michael Jerry Antal, Jr.

CHAPTER 6 Earth Contact Buildings: Applications, Thermal Analysis and Energy Benefits

John C. Carmody, George D. Meixel, Kenneth B. Labs and Lester S. Shen

ABOUT THE AUTHORS

CHAPTER 1

Ronald Stewart is a Research Professor at the Atmospheric Sciences Research Center, State University of New York at Albany. He directed: (1) The U.S. DOE sponsored Solar Energy Meteorological Research and Training Site: Region 2, (2) the operation of the photovoltaic communication system at the 1980 Winter Olympics, (3) the statewide energy audit program (8500 energy audits) including ten colleges, and (4) the analysis of an alumini house–conference center which had an active solar heating system.

In addition Mr. Stewart has published over 100 articles, reports, etc. and directed over $4 million in research projects. He is a member of ISES, ASES, and the Royal Meteorological Society.

Daniel Spencer is a Program Coordinator at the University of Wisconsin. He received his doctorate in atmospheric sciences from the State University of New York at Albany in 1983. His dissertation title was "Clear Atmospheric Effects on Insolation Illustrated by Broadband Radiometer Measurements." Spencer has published articles in several journals and presented professional papers in Europe, Australia, and the U.S. Previously he developed the computerized quality control programming for the U.S. DOE supported Solar Energy Meteorological Research and Training Site: Region 2, and studied lightning flash frequency based upon satellite photography.

Richard Perez is a Research Associate at the Atmospheric Sciences Research Center, State University of New York at Albany. He is currently Project Director of a daylight availability resource assessment program. This project involves the measurement of over 20 radiation/daylight parameters, as well as the study of their interrelationship.

As a participant to the U.S. DOE funded Solar Energy Meteorological Research and Training Sites program, he has done extensive radiation modeling work; the anisotropic diffuse model he developed is now assessed by SERI and Sandia National Laboratories for potential large scale utilization. This work also involved collector geometry optimization for photovoltaic applications.

CHAPTER 2

Richard E. Speece is Betz Chair Professor of Environmental Engineering at Drexel University. He received his Bachelor's, Master's. amd PhD. from Penn Collge, Yale University and MIT respectively.

He received the Eddy Medal from the Water Polution Control Federation and the Croes Medal from the American Society of Civil Engineers. He received the Distinguished Faculty Award, the Distinguished Lecturer Award and the Engineering Science Award from the Association of Environmental Engineering Professors. He is listed in Who's Who in Engineering, America and the World.

CHAPTER 3

Thomas B. Reed is currently principal scientist in the Thermochemical and Electrochemical research branch at the Solar Energy Research Institute (SERI) in Golden, Colorado. He obtained his PhD. in physical chemistry from the University of Minnesota in 1952 in the field of crystallography. He worked for the Linde division of Union Carbide from then until 1959 in the field of high temperature processing and chemicals. In 1960 he went to the MIT Lincoln Laboratory and worked in the field of materials science and energy until 1977, when he went to SERI to continue his work in the field of renewable energy, espcially alcohols from biomass. He has published over 200 papers in the fields of high temperature processes, materials science, renewable energy and thermodynamics. In 1965 he spent a year at Oxford University as Senior Research Fellow. He was given an IR 100 award in 1983 for invention of the high pressure oxygen biomass gasifier.

CHAPTER 4

Michael Jerry Antal, Jr. graduated Summa Cum Laude with Highest Distinction in Physics and High Distinction in Mathematics from Dartmouth College in 1969. He earned an MS in Applied Physics in 1970, and a PhD. in Applied Mathematics in 1973, both from Harvard University. Afterwards, Dr. Antal spent two years as a theorist with the Thermonuclear Weapons Physics Group of the Los Alamos Scientific Laboratory, and six years with Princeton University. At Princeton he was a member of the faculty of the Mechanical and Aerospace Engineering Department and Director of the Renewable Resources Research Laboratory (R^3L). In 1981 Dr. Antal was invited to assume the newly endowed Coral Industries Distinguished Professor of Renewable Energy Resources Chair with the University of Hawaii, where he is now engaged in research on the pyrolysis of biomass materials, and high temperature solar thermal energy utilization.

CHAPTER 5

Steven V. Szokolay is Director of the Architectural Science Unit at the University of Queensland, Brisbane, Australia. He was born and brought up in Hungary, qualified as an architect in Sydney, and has an Master's of Architecture degree of Liverpool (U.K.) and a PhD. of Queensland university. For some ten years he was

teaching at Liverpool and London. He designed the first solar heating system in the U.K. for a house at Milton Keynes new town (1973). He is co-author (with Koenigsberger) of a book on climatic design, the author of *Solar Energy and Building* (1975), *World Solar Architecture* (1980), *Environmental Science Handbook for Architects* (1980), *ANZ Solar Home Book* (1981), *Climatic Data and its Use in Design* (1982), and of innumerable papers, articles, and other contributions. Currently his work is aimed at design tool development, in order to integrate solar design with the everyday practice of architeccture.

CHAPTER 6

John Carmody is an architect and is currently the Associate Director of the Underground Space Center at the University of Minnesota. After receiving a Bachelor of Architecture degree from the University of Minnesota, he specialized in the design and construction of housing while in the Peace Corps and later in the United States. Mr. Carmody has been the co-author of five books on energy-efficient and underground building design. At the Underground Space Center, he has worked on a wide variety of research projects including underground space planning and development, determining the energy performance of housing, and developing design guidelines for energy-efficient buildings.

George Meixel, Jr. received a Bachelor of Science degree in Engineering Physics in 1967, and a PhD. in Aerospace Engineering and Plasma Physics in 1973 from Cornell University. Dr. Meixel's research experience includes hypersonic aerodynamics, controlled fusion, coal gasification, computer graphics, and energy conservation in buildings. As a research associate at the Underground Space Center at the University of Minnesota for six years, Dr. Meixel has specialized in computer-based theoretical studies and experimental measurements of the heat transfer characteristics of underground buildings. Recently, Dr. Meixel was the Principal Investigator for a multi-year research project on passive cooling through earth contact structures for the U.S. DOE.

Kenneth Labs received his Master's of Architecture degree from Washington University in 1975, and presently maintains his own design and research practice in New Haven, Connecticut, where he is a Visiting Lecturer in Environmental Technology at the Yale School of Architecture. He is the author of over 30 publications dealing with earth coupled building design, including the award-winning report to the U.S. DOE, "Regional Analysis of Ground and Above Ground Climate." He is also co-author (with Donald Watson, FAIA), of *Climatic Design: Energy Efficient Building Principles and Practices* (McGraw-Hill, 1983).

Lester Shen received his Bachelor's degree in Chemistry at Haverford College in 1977, and a Master's of Science degree in Mechanical Engineering from the Georgia Institute of Technology. Currently at the Unversity of Minnesota, Mr. Shen is completing his PhD. in mechanical engineering, studying coupled heat and moisture flow in soils. He is presently a researcher at the Underground Space Center. Since being at Minnesota, he has been involved in research dealing with the thermal performance of earth-sheltered construction, building load simulation programs, and weatherization of residential housing.

CHAPTER 7

James E. Hill (PhD. Georgia Institute of Technology, 1968) is the Chief of the Building Equipment Division at the National Bureau of Standards. He is responsible for applied research and the development of testing and performance standards for building equipment. He was Assistant Professor of Mechanical Engineering at the University of Maryland for three years before joining NBS in 1972.

Dr. Hill is a member of ASHRAE, the American Society of Mechanical Engineers, the Society of Sigma Xi, and ISES. He has authored over 50 technical papers in the building research/solar energy areas.

Byard D. Wood is a Professor of Mechanical Engineering at Arizona State University, Tempe, Arizona. He received his Bachelor of Science and Master's of Science degrees from Utah State University and his PhD. from the University of Minnesota. A registered professional engineer, he is an active member of several professional societies.

Dr. Wood has served on a number of national and international committees concerned with the development and evaluation of test methods for solar collectors and solar systems. He is chairman of the revision committee for ASHRAE Standard 93-77, chairman of the Technical Compliance Committee for Solar Rating and Certification Corporation, member of the U.S. delegation to International Standards Organization TC 180 Solar Energy. He is an Associate Editor of the *Journal of Solar Energy Engineering*.

Kent A. Reed (PhD. The University of Chicago, 1976) is the Leader of the Computer Integrated Construction Group at the National Bureau of Standards. He is responsible for planning and executing projects. For three and half years, Dr. Reed was the leader of the Solar Equipment Group at NBS. Before joining NBS, Dr. Reed was staff scientist in the Solar Energy Group at Argonne National Laboratory for four years.

Dr. Reed is a member of the American Society for Testing and Materials (ASTM), ISES, ASES, the American Physical Society, and the American Association for Artificial Intelligence.

CHAPTER 8

Ari Rabl is research scientist at the Center for Energy and Environmental Studies of Princeton University, where he has been since 1980. A recipient of a Fulbright grant, he obtained his PhD. in theoretical high energy physics from the University of California at Berkeley in 1969. He continued in this field during postdoctoral appointments at the Weizmann Institute and Ohio State University until 1973 where he began work in solar energy at the Argonne National Laboratory, the University of Chicago, and the Solar Energy Research Institute.

He has published over fifty journal articles on high energy physics, applied optics, heat transfer, and solar energy; and he holds five patents on solar collectors. He is author of *Active Solar Collectors and their Applications* (Oxford University Press, 1985). He has lectured and consulted extensively, both in the U.S. and abroad; and at Princeton he has taught courses on energy and environment and on solar energy.

CHAPTER 1

THE MEASUREMENT OF SOLAR RADIATION

Ronald Stewart
Daniel W. Spencer
Richard Perez

1.1 ABSTRACT

The measurement of solar radiation has run the gamut from casual interest recordings to intensive scientific/technological experiments on satellites. In these situations, the interest has always been to obtain reliable, usable date. The choice of measurement technique has often depended upon availability of instrumentation and funding to design a proper data acquisition system and analysis program. The following monograph attempts to deal with some of the problems associated with the choice of instruments for such a program and with some of the questions of the techniques for the analysis of the data acquired.

The first section, adapted from the International Energy Agency Conference on Pyranometer Measurements,[1] provides an analysis of various characteristics of pyranometers. We acknowledge the advice of Michael R. Riches, Office of Energy Research, U.S. DOE, on the use of this material.

The next section outlines some of the commonly noted problems with the measurement of solar radiation, as well as possible causes and solutions. Much of this information is based upon the recent experience of the U. S. DOE-sponsored Solar Energy Meteorological Research and Training Sites.

The last section is adapted from *Sunworld* (Solar Radiation Instruments by Kinsell L. Coulson and Yvonne Howell) and is reprinted with the kind permission of the publishers, Pergamon Press, Ltd. Additional information has been added from recent publications.

Appendix 1.1, Historical Perspective, is simply a compilation of dates and events in the history of solar radiation measurement. It provides the time perspective

which has led us to today's solar-oriented society. Much of the information is adapted from Coulson[2] and Drummond,[3] with input from EPLAB, Inc., Hollis Geosystems Corp., Casella London Ltd., LI-COR, Inc., and Kipp & Zonen.

1.2 CHARACTERISTICS OF PYRANOMETERS

Pyranometers are instruments used to measure global solar radiation.[4] These instruments combine the components of the incoming direct-beam and the diffuse-sky solar radiation, as received from a 2 πsr solid angle above the plane of the instrument's sensing surface. The instrument is generally used to measure radiation over the solar spectrum wavelength range of about 0.13 to 3.0 micrometers. Webster[5] defines this instrument as "an instrument for measuring radiation from the sky by comparing the heating effect of such radiation upon two blackened metallic strips with that produced in the same strips when heated by means of an electric current." This description of operation fits only one of several possible designs, in this case, the Robitzsch bimetallic pyranometer or actinograph,[6] but does illustrate the concept of equating electrical energy, which can be measured directly, with solar radiation intensity.

The ideal pyranometer would be characterized as having an output signal, S, which is directly proportional to the sum of the vertical component of direct normal radiation (the beam intensity, I, multiplied by the cosine of the incidence angle, θ, or zenith angle for horizontally mounted instruments) and the diffuse sky radiation, D:

$$S \propto I \cdot \cos\theta + D \tag{1.1}$$

Pyranometers available today are simple in concept, though complex in their true microscopic behavior. They are adequate for most meteorological measurement applications with the use of a single calibration factor, C_f, to convert the output signal into units of irradiance, that is, W/m^2, for global solar radiation, $K\downarrow$:

$$K\downarrow = C_f \cdot S \tag{1.2}$$

A more recent application of pyranometry has been for solar collector performance testing. Here, the pyranometer measurements obtained with a single calibration factor are not sufficiently accurate to meet the needs of the solar test engineer in determining the precise amounts of solar energy incident to the collector. In fact, it remains to be proven that sufficient accuracy can be achieved for these purposes using the best of present methods for determining and applying corrections to the measurements.

The following information is placed in this report to aid the reader in understanding pyranometry, specifically those concepts discussed at the International Energy Agency Conference on Pyranometer Measurements, and the principles underlying the experimental work embodied in the Statement of Work which was outlined during the meeting.

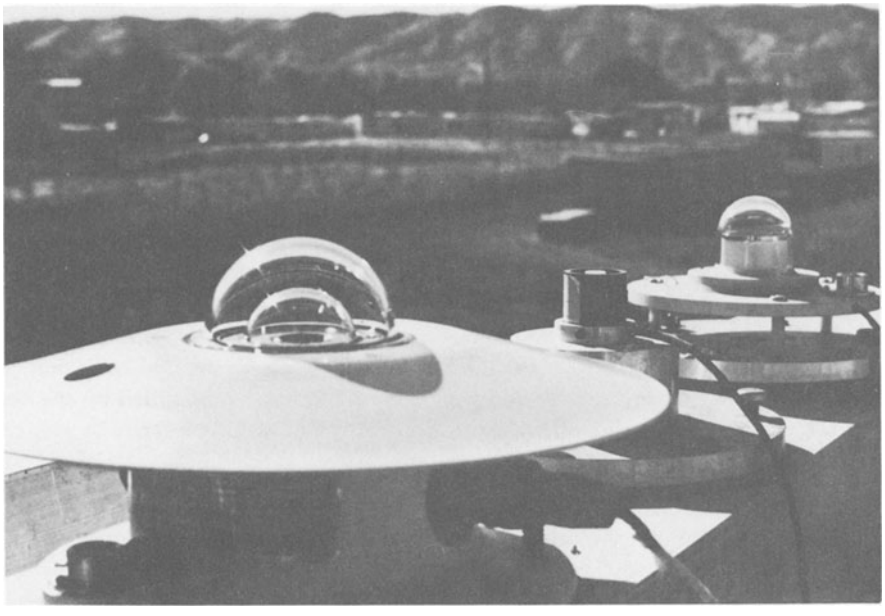


Fig. 1.1: Examples of thermopile (Eppley PSP), photodiode (Li-Cor LI-2000S), and silicon cell (Matrix MK-1G) pyranometer (photo by Tom Stoffel).

1.3 GENERAL FEATURES OF A PYRANOMETER

A pyranometer consists of the following basic components:

(1) A detector or sensing element protected by glass dome(s), Teflon envelopes, or a solid acrylic diffuser;

(2) An instrument case (body) with a spirit level, adjustable leveling screws, and a desiccant chamber;

(3) Some type of radiation shield which protects the case of the instrument from direct sunlight (a requirement for thermopile designs using the case as the reference junction);

(4) An electrical connector or attached cable for the output signal.

The physical design of the detecting surface or sensing element can be based upon the principles of either a thermocouple or photoelectric effect. This results in the commercial availability of multijunction thermocouples (thermopile) and silicon cell or photodiode pyranometers (see Fig. 1.1).

A thermopile-type pyranometer is typically 15 to 30 cm in diameter overall, about 15 cm high, and weighs 2 to 3 kg. The sensitive area is often less than 6 cm in diameter, with some surface coating or treatment (for example, Parsons Optical Black lacquer or 3M Black Velvet paint). The shape of the sensor surface varies, as does the shape of the thermopile. The Eppley Precision Spectral Pyranometer

(PSP), for example, utilizes a wire-wound rectangular thermopile under a circular, black film covering. The spirit level used to set the sensor surface (actually the attachment point on the case) to a horizontal plane usually has an accuracy better than $\pm$ 0.3°.

The silicon-based pyranometer is typically 1 to 10 cm in diameter, stands 2 to 10 cm high, and weighs 0.1 to 0.5 kg. The detecting surface is generally less than 1.0 cm in diameter for photodiodes and 2.5 cm on a side for exposed solar cells.

1.3.1 Instrument Sensitivity

In the case of an ideal pyranometer, mounted in a horizontal plane, the output signal is proportional to the vertical component of the direct normal radiation (that is, direct-beam radiation as measured with a pyrheliometer multiplied by the cosine of the zenith angle) plus the diffuse sky radiation, without interference by any other parameters, as in Eq. (1.1).

In practice, however, all pyranometers show deviations from the ideal due to the manner in which complicating influences affect the measurement and are accounted for in the final analysis. A pyranometer's sensitivity is defined as the ratio of the output signal to the received irradiance. It can be a function of several factors, including the magnitude and direction of the irradiance vector(s), position of the sensor, environmental conditions, time, and so on. The text which follows describes those factors that influence pyranometer measurements. The order of their appearance coincides with the suggested characterization procedure, which avoids compounding effects.

1.3.2 Response with Time

The time response of an instrument can be defined in terms of its response to a step input. The response time of a pyranometer is the time for the output signal to fall (rise) to 10% (90%) of the final steady-state value change following an abrupt decrease (increase) in irradiance. The so-called rise and fall times for the instrument are often unequal. The time constant is defined as the time in seconds for the transient signal to decay (rise) to $1/e$ $(1 - 1/e)$ of the total change.

1.3.3 Sensitivity

Sensitivity, R, is simply the ratio of the output signal of the pyranometer, S, to received irradiance, E_g:

$$R = \frac{S}{E_g} \tag{1.3}$$

In general, a single sensitivity number represents the mean value derived from a range of test conditions, that is, from integrated output signals over varying time scales—typically ranging from minutes to weeks. A single number may also represent a value generated under a specified set of test conditions. The single sensitivity number is often referred to as the calibration factor, C_f.

The conditions under which the pyranometer sensitivity was measured must be reported to the user in order to correctly apply the value to the measured output

signal and convert it into units of irradiation. This concept forms the basis of the sensitivity function hypothesis, which proposes that the sensitivity of a pyranometer is a variable quantity, depending upon the individual or combined effects of the aforementioned outside influences.

1.3.4 Responsivity

Responsivity, a term closely related to sensitivity, is usually defined as the ratio of the output signal, S, to the radiant power, P_i, incident upon the detector:

$$\text{Responsivity} = \frac{S}{P_i} \tag{1.4}$$

typically expressed in terms of volts/watt. For pyranometry, the radiant power per unit area, or irradiance (W/m^2), is desired. Responsivity is a widely used term in the field of radiometry and photometry.[7-9]

1.3.5 Temperature Coefficient of Sensitivity

Radiometers exhibit a change of sensitivity with variations of instrument temperature. This temperature dependence is usually specified as the ratio

$$C_T = \frac{\Delta R/R}{\Delta T} \tag{1.5}$$

where C_T = temperature coefficient; $\Delta R/R$ = relative change in sensitivity; and ΔT = change in case temperature.

Pyranometers have been designed with resistive networks which compensate for nearly all of the instrument's temperature dependence. Some models, especially the earlier designs, have been tested by the manufacturers, who then provide a value for C_T, usually in terms of percentage change in sensitivity per degree of temperature departure from a reference or calibration value. Figure 1.2 shows data from three different tests for both compensated and uncompensated instruments. Table 1.1 summarizes the manufacturers' specifications for this and other characteristics.

1.3.6 Thermal Transient Response

The time rate of change in the temperature coefficient is a function of the magnitude and nature of the forcing function or temperature differences, and of the instrument's physical properties.

1.3.7 Linearity

The ideal pyranometer should provide an output signal that is directly proportional to the radiation received over a normal range of irradiance levels. As shown in Table 1.2, most instruments have a sensitivity which varies within $\pm 2\%$ up to an irradiance of one solar constant (1377 W/m^2).

Table 1.1

Manufacturer specifications for a sample of pyranometers

Manufacturer	Model	Detector Type	Temperature Dependence	Linearity	Cosine Response	Time Constant $\frac{1}{m}$ Response Time	Effects of Tilt
Eppley Laboratory	PSP	thermopile	±1% (-20°to 40°C)	±0.5% (O–2800 W/m^2)	±1% 0°–70° ±3% 70°–80°	1 s (1/e) of signal	unaffected
Kip and Zonen	Cm-5	thermopile	0.15% per °C	1%	$\frac{5}{M}$	70% of final value in 3 s	$\frac{5}{M}$
Phillipp Schenk	Star 8101	thermopile	+3% per °C	1% (80–1300 W/m^2)	1% of 0°–60° 1% 60°–80°	95% of final value in 20 s	±1% for 0°–180°
Lintronic Ltd.	Dome 615	thermopile	−0.2% per °C	$\frac{5}{M}$	±2% 0°–65° ±4% 65°–80°	66% of final value in 20 s 99% of final value in 30 min	$\frac{5}{M}$
Hollis Observatory	MR-5	silicon diode	±1.5% (−20°to 40°C)	±1% (0–1400 W/m^2)	±1.5% 0°–80°	$\frac{5}{M}$	$\frac{5}{M}$
LiCor, Inc.	LI-200s	silicon diode	±15% per °C (maximum)	1% max (0–3000 W/m^2)	corrected 0°–80°	10–90% in 10 microsec	$\frac{5}{M}$
Matrix, Inc.	MK-1G	silicon cell	compensated 4.5°to 60°C	$\frac{5}{M}$	$\frac{5}{M}$	100% in less than 1 millsec	$\frac{5}{M}$

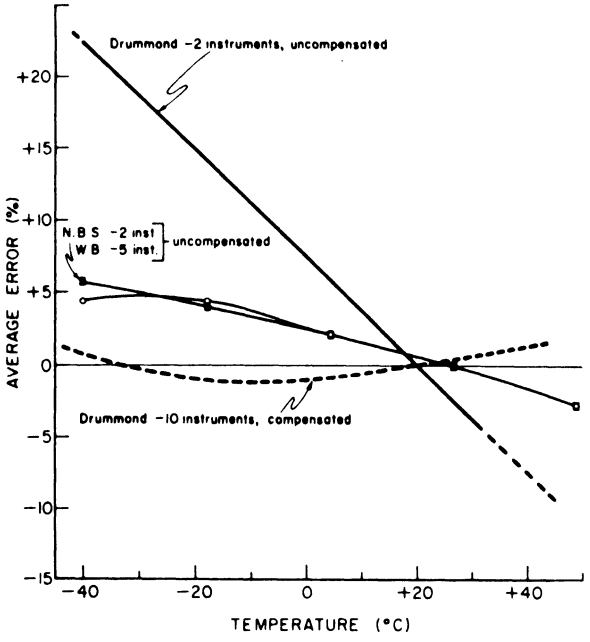


Fig. 1.2: Average error as a function of temperature for Eppley pyranometers which are compensated and uncompensated for temperature effects.

1.3.8 Angular Dependence of Sensitivity

Global radiation, as measured by a pyranometer, requires an integration over the entire hemisphere above the plane of the sensor. This angular integration imposes stringent requirements on both materials and basic design of the instrument if its sensitivity is to be independent of the angle of incidence of the radiation.[6] Three angular dependence errors are common to pyranometer measurements:

• *Cosine error* is the result of directional dependence of the pyranometer sensitivity to solar elevation (for horizontally mounted instruments) or the incidence angle defined by the radiation vector and the unit normal to the sensing surface. Ideally, the vertical component of the radiation is accepted by the detector according to the Lambert cosine law. In fact, the reflectance/absorptance of any surface is dependent on the angle at which the radiation strikes the surface. Additionally, striations or optical defects in the glass hemispherical envelope(s), curvature of the receiver surface, or internal reflections in the pyranometer may contribute to this error. By calibrating instrument sensitivity versus angle of incidence of the direct-beam radiation, it is possible to correct the data (see Fig. 1.3).

• *Azimuthal error* is the result of directional dependence of the pyranometer sensitivity to solar azimuth or the azimuthal orientation of the detector with respect to the radiation. This error is due to the surface irregularities or asymmetrical design of the sensing element. Common practice is to position the pyranometer signal cable

Table 1.2

Correction factors for Albany: $z\frac{1}{m}\cos(\text{zen})$ interval [1 ≡ .1 − .2, 2 ≡ .2 − .3, etc.]; $m\frac{1}{m}$ month; drm $\frac{1}{m}$ Drummond factors for Albany

$\frac{z}{m}$	J D	F N	M O	A S	M A	J J
1	1.10	1.10	1.15	1.10	0.95	0.90
2	1.15	1.20	1.20	1.15	1.05	0.95
3	1.20	"	"	1.18	1.15	1.05
4	"	"	"	1.20	1.18	1.15
5	$\frac{3}{m}$	"	"	"	1.20	1.20
6	$\frac{3}{m}$	"	"	"	1.21	1.22
7	–	–	"	"	1.22	"
8	–	–	–	"	"	"
9	–	–	–	–	"	"
	–	–	–	–	–	–
drm	1.09	1.10	1.15	1.10	0.95	0.90
	1.08	1.10	1.15	1.10	0.95	0.90

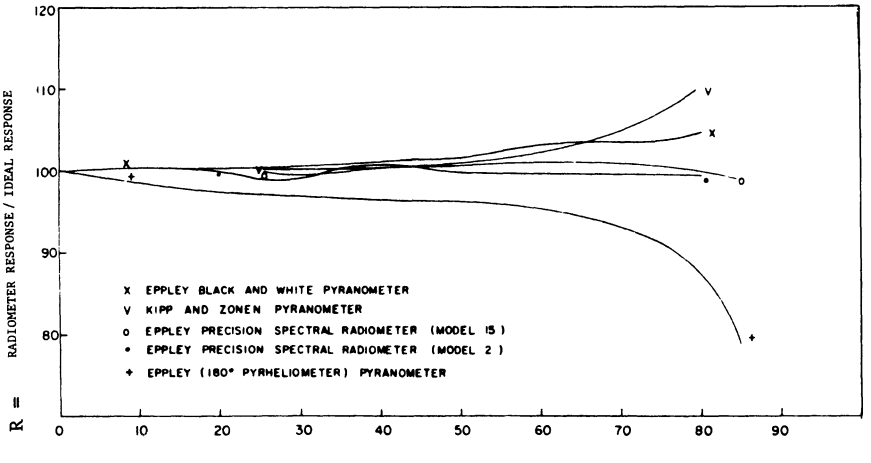


Fig. 1.3: Typical cosine response of a number of radiometers.

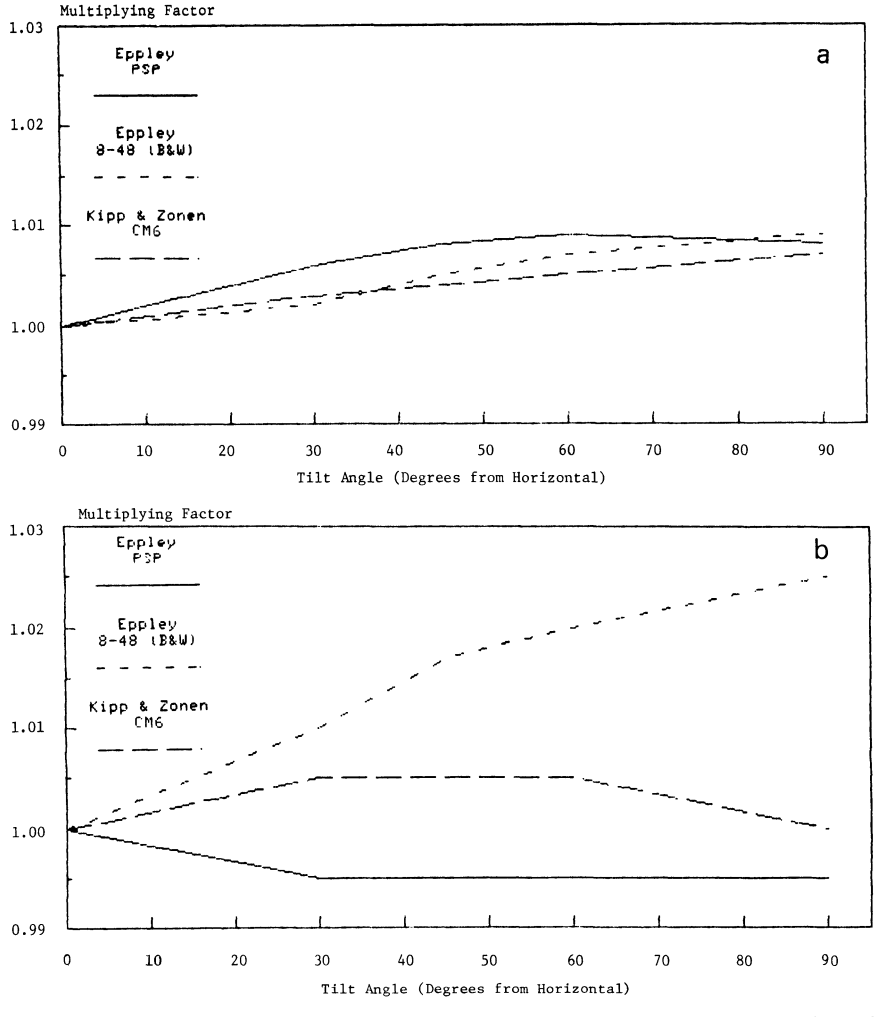


Fig. 1.4: A summary of test results evaluating the tilt effects on three models of pyranometers.

to the north in the northern hemisphere or other reference direction to reduce the possible discrepancies between the instruments under test.

• *Tilt effects* are known to exist in some pyranometers. The sensitivity of the instrument can change depending on the orientation of the detector with respect to the horizontal. Figure 1.4a and b illustrates this effect as determined by two laboratories.[10] Convective air currents above the sensing surface of the dome-design pyranometers contribute to this error, which is a function of tilt angle and irradiance level. Obviously, measurements of global radiation on inclined surfaces would have errors introduced due to a combination of tilt effects and cosine errors associated with the changing incidence angles of the radiation.

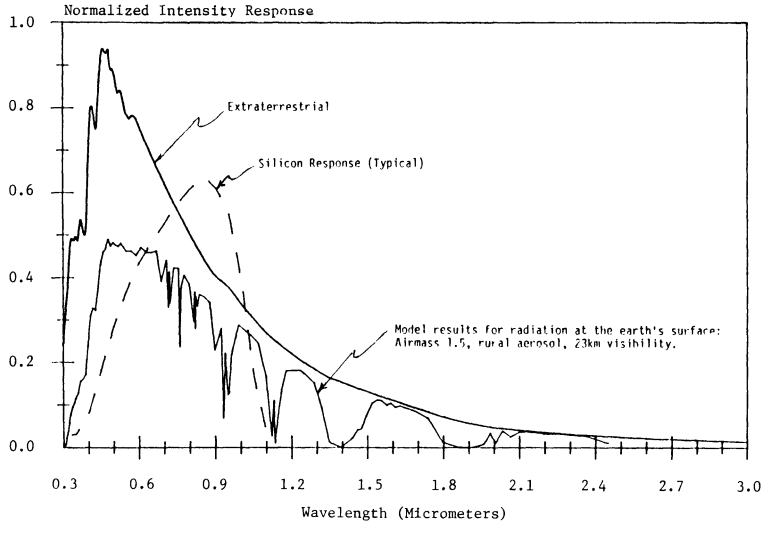


Fig. 1.5: A summary of spectral distributions of extraterrestrial radiation incident at the top of the atmosphere, the results of the SOLTRAN computer model of transmitted radiation, and a typical response curve.

1.3.9 Leveling

The detector surface and the reference surface of the spirit level are generally assumed to be coplanar. Production tolerances must allow for some departure from this ideal condition. The accuracy of typical spirit-level, commercially available instruments is generally better than $\pm 0.3\%$. The level can be adjusted to coincide with the true optical level of the detector by illuminating the pyranometer at some angle from the normal to the sensing element, usually 70° to 80°, rotating the instrument in azimuth, and adjusting the altitude until the output signal is constant with azimuthal position with respect to the light source. Azimuthal dependencies of the sensitivity must also be accounted for in this test.

1.3.10 Spectral Response

The ideal instrument for measuring solar radiation would have a uniform sensitivity, or "flat" response, and not detect radiation outside this spectral region (see Fig. 1.5). In practice, this is not the case with commercially available pyranometers.

Variation among pyranometers is caused by many factors, including:

- variations in the spectral characteristics of the transmission and reflection coefficients of cover glass dome(s), windows, and radiation shields and differences in the absorption characteristics of sensor surfaces;
- variations in the electrical nature of the detection mechanism (particularly in photovoltaic detectors); and

- variations in the methods of heat removal or cooling of the instruments, or in operating temperatures of the instruments (particularly in thermal detectors).

Photovoltaic detectors have distinct spectral response characteristics resulting from the photoelectric effect displayed by silicon (see Fig. 1.5). A number of conclusions are worth drawing at this point:

- If a pyranometer does not have the desired flat spectral response from 0.3 to 3.0 micrometers, its sensitivity will vary with atmospheric conditions which alter the spectral distribution of the solar radiation.
- Under changing atmospheric conditions. two pyranometers with the same spectral response would produce measurement agreement, even if their response was not flat. but they would not agree with a unit that did have a flat response or a different spectral response.
- Two different models of radiation detectors which might agree in sunlight may differ by several percentage points under artificial light (or vice versa), because of the differences between the spectra of the two radiation sources.[11]

1.3.11 Stability

Pyranometer sensitivity changes with time and exposure to radiation. Periodic calibrations are suggested by most manufacturers and are required for accurate measurement capability. Pyranometers in continuous use should be calibrated annually as a minimum policy.

1.3.12 Pyranometer Sensitivity Function

The characterizing of pyranometers is defined as the quantifying of the responses of the instrument to the various parameters mentioned above which produce the sensitivity function:

$$R = f(E_g, \dot{E}_g, \beta, \theta, a, T, \dot{T}, \Delta T_n, \lambda, P, \ldots) \tag{1.6}$$

where R = sensitivity (typically, v/W/m^2); E_g = global irradiance at receiver (effects of linearity); $\dot{D}_g$ = time rate of change of global irradiance (effects of time constant); β = angle between the normal to the instrument and the horizontal (effects of tilt); Θ = angle between the incident beam and the receiver normal (effects of cosine error); a = angle to the incident beam measured about the receiver normal with respect to a reference direction, typically the center line of the connector (effects of azimuthal dependence); T = temperature of the instrument body, usually intended to indicate the thermopile heat sink or cold thermojunction temperature, but often approximated by measuring ambient air temperature surrounding the instrument; $\dot{T}$ = thermal transients or time rate of change in temperature; ΔT_n = gradients and temperature differences between parts within the instrument (for example, glass domes and body, or body and thermopile cold junctions); λ = wavelength of incident radiation (effects of spectral response); and P = pressure (pressure dependence of thermal conductivity of air).

Note that this analysis of the response of the pyranometer is to be contrasted with the classical view of the instrument calibration in which a single value of sensitivity (calibration factor, C_f) is determined by averaging the output signals from a specified test or tests performed in the laboratory or outdoors. Such techniques do not isolate the individual effects described above and limit the application of any detailed characterization information. It has been shown that C_f does not vary measurably with respect to some of the above parameters. For increased accuracy in pyranometry, it is apparent that the documentation (characterization) of the effects of the variables in the sensitivity function is necessary. When these factors are measured, we can construct a transfer function,

$$\hat{f} = \prod_i f_j \tag{1.7}$$

or

$$\hat{f} = \sum_i f_k \tag{1.8}$$

which applies these effects as corrections to a basic sensitivity, R_0, thus yielding more accurate pyranometer measurements.

If a single sensitivity, R_0, can be defined based on proper testing procedure which quantifies the individual characteristics of a pyranometer, then

$$R = R_0 \cdot \hat{f}(E_g, \dot{E}_g, \beta, \theta, a, T, \dot{T}, \Delta T_n, \lambda, P \ldots) \tag{1.9}$$

It may not be possible to separate the effects of some individual variables. This means that it is not possible in every case to produce a set of independent functions which can be combined to form Eqs. (1.7) or (1.8). More explicitly, with

$$\frac{S}{R_0} = E_0 \tag{1.10}$$

where S = instrument output signal; R_0 = basic sensitivity; and E_0 = first estimate of global irradiance, the applications of the transfer function may result in the computation of the corrected irradiance value, E_{corr}, according to some function of the form

$$E_{corr} = E_0 \cdot f_1(E_0, T) \cdot g_1(E_0, \beta) \cdot h_1(E_0, \theta) \ldots \tag{1.11}$$

or

$$E_{corr} = E_0 + f_2(E_0, T) + g_2(E_0, \beta) + h_2(E_0, \theta) \ldots \tag{1.12}$$

or combinations of products and sums of correction functions. The structure of the transfer function will depend upon the order, manner, and form in which the correction functions are derived.

Generalized discussions of the mathematical and engineering implications of the transfer function concept, sensitivity (responsivity), linearity analysis (including nonlinear systems), detector calibration, and sources of uncertainty are covered in detail by Wyatt.[7] Additional insight into this topic may also be gained from

discussions in Wolfe and Zissis[8] and the National Bureau of Standards tutorials on optical radiation,[9] especially chapter 5.

In the final applications of this transfer function to solar collector tests, more detailed measurements of environmental and other parameters influencing the output of the pyranometer will be required to achieve enhanced accuracy over the more common applications of this instrument in meteorology.

1.4 PYRANOMETERS

1.4.1 Clear-Sky Analysis

Instruments may degradate over time or may be slightly damaged so as to affect the data output. One procedure which may be practiced annually is that of clear-day analysis or clear-hour analysis. Clark and Rudzki describe the procedure which may be used to test pyranometers and pyrheliometers on an annual basis.[12] If the data output are similar year to year, serious degradation or damage has not affected the signal output. The following is an excerpt from Clark's report:

> *"Clear-Day" Solar Noon Insolation.* As part of our quality control procedures (and as part of the development of insolation models), we have examined the pattern of measured "clear-day" solar noon insolation. The definition of "clear-day" is problematic. Figures 1.6 through 1.9 illustrate the global and direct normal insolation as classified using four different definitions of clear days.

Figures 1.6 and 1.7 show the measured noon values of global and direct normal insolation in San Antonio, Texas, for all days with clear noon periods between November 1978 and June 1979.

Figures 1.8 and 1.9 show the resulting atmospheric transmittances corresponding to these measured insolation values. Figure 1.8 shows noon transmittance for direct normal insolation, which is defined as

$$\gamma_{DN} = \frac{I_{DN}}{(1367\ \mathrm{W/m^2})\left(\frac{r_o}{r}\right)^2} \tag{1.13}$$

where I_{DN} is the measured noon direct normal insolation, r_o is the average distance to the sun, and r is the distance on the day in question.

Figure 1.9 shows the noon transmittance for global insolation, which is defined as

$$\text{global} = \frac{I_{\text{global}}}{(1367\ \mathrm{W/m^2})\left(\frac{r_o}{r}\right)^2 \cos\,(\text{zenith angle})} \tag{1.14}$$

In all cases, we have required that the chart-recorder trace of global insolation indicate that no opaque clouds passed the disk of the sun within 20 min., centered on solar noon. (Naturally, this still permits clouds to be present.) Another definition

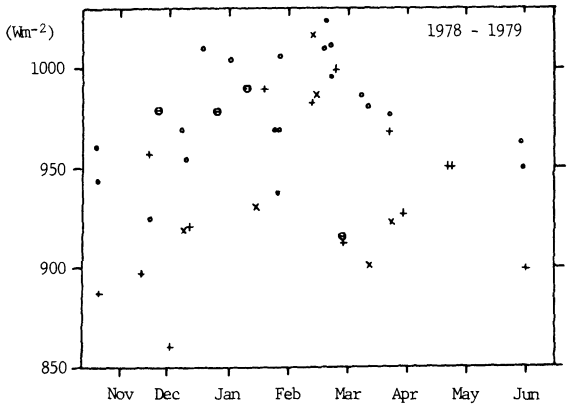


Fig. 1.6: Direct normal insolation of local solar noon: Trinity University, San Antonio, Texas.

° No clouds and no variation in global insolation within 1 hour each side of solar noon.

Θ Clouds within 1 hour of solar noon, but not within 10 minues each side of solar noon. No variation in global insolation during cloud-free period.

\+ No clouds within 1 hour each side of solar noon. Global insolation variable during cloud-free period.

× Clouds within 1 hour of solar noon, but not within 10 minutes each side of solar noon. Global insolation variable during cloud-free period.

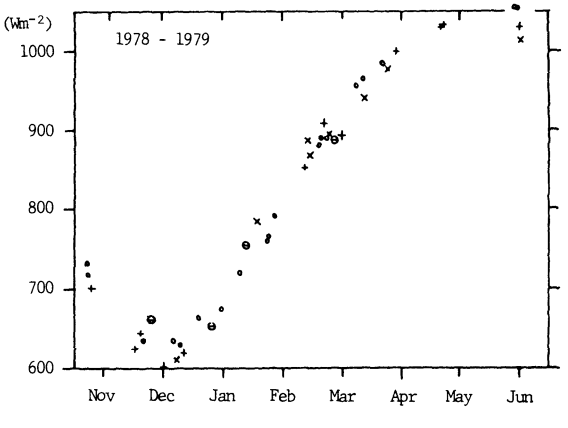


Fig. 1.7: Total insolation on the horizontal: Trinity University, San Antonio, Texas.

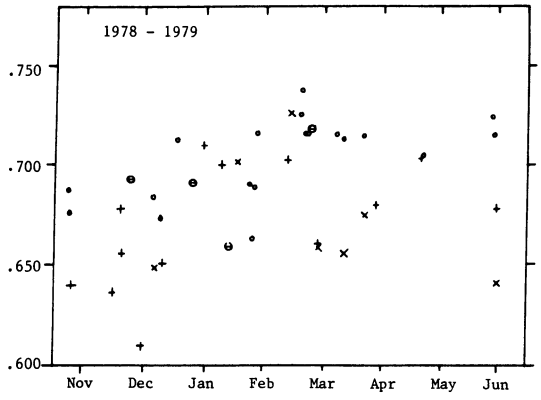


Fig. 1.8: Atmospheric transmittance for direct normal insolation: Trinity University, San Antonio, Texas. Solar constant = 1367 W/m^{2*} (See Fig. 1.6 for definition of symbols).

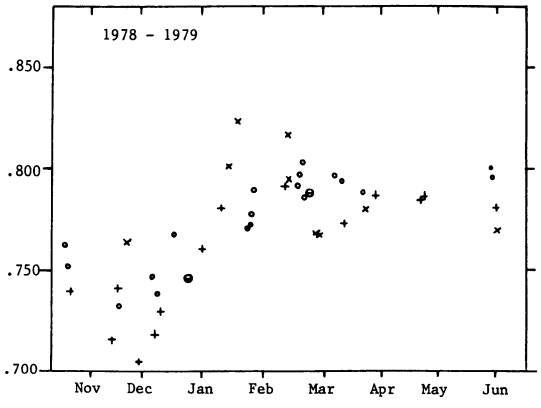


Fig. 1.9: Atmospheric transmittance for global insolation: Trinity University, San Antonio, Texas. Solar Constant: 1367 W/m^{2} * (See Fig. 1.6 for definition of symbols).

requires that the chart trace indicate no clouds in front of the disk during two hours, centers on noon.

The surprising effect appeared when we required that the direct insolation should not exhibit any rapid, small-scale (±5%) variations during this cloud-free period. The highest solar noon, direct normal insolation, occurred on days when there were no such variations exhibited. Days on which these small-scale variations occurred had consistently low direct normal insolation at noon; this remained true even when no opaque clouds passed the solar disk within an hour, each side of noon. Apparently these variations are an indication of the presence of turbidity or

high cirrus clouds. We are attempting to use spectral measurements to distinguish between these two hypotheses.

1.4.2 Installation of the Instrument

On a horizontal mount, many instruments use a bubble-level indicator to determine instrument relation to the ground plane. These bubble-level indicators are useful if the indicator is parallel to the instrument receiver surface, otherwise they are misleading. Care should be taken to check the accuracy of the bubble-level. An instrument which is not level will read too high (low) at solar noon if it is tilted to the south (north), and will read higher in the morning (afternoon) than in the afternoon (morning) if it is tilted to the east (west). A tilted. ESE-facing instrument will "see" sunset earlier. Additionally, the output signal may peak before solar noon, or will at least have a flattened bell-shaped curve before solar noon.

1.4.3 Thermopile Construction

The thermopile of an instrument may or may not be circular, depending upon the manufacturer. While the visible sensor may be circular, the thermopile may be rectangular, causing the various portions of the sensor disk to respond unequally. This effect is especially noted when a portion of the disk is shielded and the position of the instrument changed. Figure 1.10 provides an example of the sensitivity of a thermopile.[12] Under these circumstances, all instruments should be installed in a similar manner to standardize the data output.

1.4.4 Moisture

Occasionally, the review of clear-sky data will reveal a data plot with dual peaks and a depression around solar noon (see Fig. 1.11). The most likely explanation is moisture within the dome. If the pyranometer includes a desiccant, it should be changed immediately, the instrument operated to determine if the data plot is correct, and then the desiccant changed again as it has absorbed at least the amount of visible moisture. Whether the instrument has been damaged or not may be tested partially by performing a clear-day analysis. If there is any question as to performance, the instrument should be checked and recalibrated. Daily checks of the desiccant will prevent this problem.

1.4.5 Deposition

A daily maintenance program will eliminate many of the deposition problems, whether they be frost, dew, birds, or other acts of nature. A dry air delivery system which rings a pyranometer will aid in decreasing frost, dew, and snow deposition. The air must be dried first, and may be warmed slightly. However, care must be taken not to upset the temperature compensation which is built into some instruments. Peterson et al. have described such a ring air blower.[13]

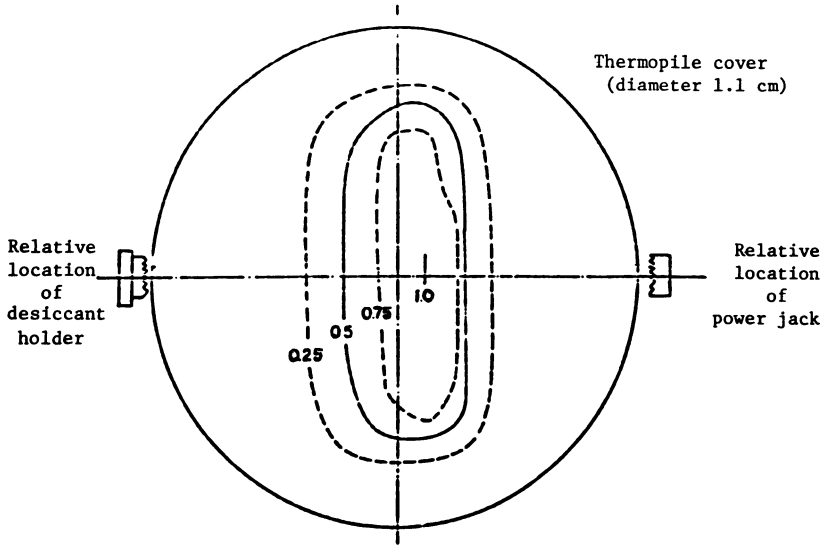


Fig. 1.10: Relative sensitivity contours based on irradiation with a 5 MW laser. Beam diameter = .095 cm. Output voltage measured at square grid points separated by .032 cm.

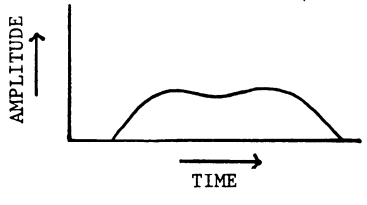


Fig. 1.11: Example of a data plot from a pyranometer with moisture trapped inside the dome.

1.4.6 Negative Values

Thermopile instruments often reradiate at night and may give a small negative signal. These values should be ignored when calculating averages. A quality control system can do this automatically with data recorded after sunset and before sunrise.

1.4.7 Readings Which Exceed the Values for Extraterrestrial Insolation

Occasionally, a reading will exceed the expected ground level insolation by 20 to 30%, or may exceed that reading expected at the top of the atmosphere. If this is not due to instrument error, then the general sky condition at the time of that reading should be checked. Partly cloudy skies may cause an increased amount of solar radiation to be reflected toward the instrument for a short period of time. This will result in a spike in the data (Fig. 1.12), normally followed by a sharp decrease if the cloud passes between the sun and the instrument. If instead the cloud does

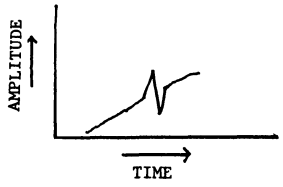


Fig. 1.12: Example of a spike signal caused by the edge of a cloud, then the cloud itself, passing between the instrument and the sun.

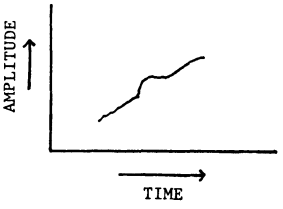


Fig. 1.13: Example of increased insolation impinging on a pyranometer due to reflection from a cloud north of the instrument.

not pass between the sun and the instrument, the readings will return to normal as the reflection off the cloud decreases as the cloud moves (Fig. 1.13). In either case, the instrument is recording the actual insolation received at that moment and at that site. However, computerized quality control systems may print these data out as if they are in error.

If this occurrence takes place on a clear day, then a visual check of the surroundings (or an all-sky camera) may pinpoint a man-made structure which is reflecting insolation toward the instrument. If this occurs only on tilted pyranometers but not on horizontal, then some surface in the field of view of the tilted instrument has increased the albedo above expected limits. A ground shield which acts as an artificial horizon will alleviate this problem.

1.4.8 Diffuse Solar Radiation

The acquisition of accurate diffuse data normally requires daily or weekly attention to the tracking disk or shadowband. Even then, errors will occur due to the construction of the instruments. During early morning or late afternoon there may be reflections off the inner surface of the shadowband. Additionally, there are errors introduced simply by the size of the band as it obscures a portion of the sky which is contributing diffuse radiation. Considerable discussion has been published concerning corrections for the shadowband highlighted by Drummond.[14] LeBaron et al. modeled a correction factor to reduce the early morning/late afternoon reflection input as well as a cloudless sky model, and a variable correction factor for overcast skies. They concluded that 3% of the diffuse input is lost in winter and 8 to 10% in summer.[15] Painter developed shade ring (shadowband) corrections for long-term averages, but noted that an attempt to apply these corrections to individual

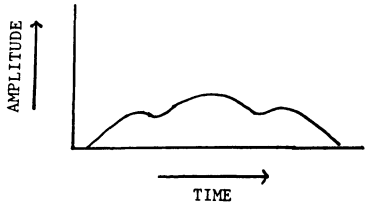


Fig. 1.14: Example of a signal received from a diffuse insolation measurement when the shadowband is misaligned (correctly aligned during midday—misaligned during early morning and late afternoon).

readings could lead to errors of 10%.[16] Spencer developed shadowband corrections based upon minute-by-minute data using the band and the tracking disk. The data were massaged to develop a series of correction factors for each, six sky conditions, and nine different zenith angles.[17] Table 1.1 provides an example of the results as compared to the Drummond factors (clear-sky conditions).

1.4.9 Shadowband Orientation

The orientation of the shadowband, if done properly, will not cause problems. However, a quality control program which checks the data will quickly spot misalignment. Usually this misalignment is obvious during part of a day, but not all day. This is especially true as the installation of the band will take place near midday, but the errors normally occur nearest to day beginning/end. Figure 1.14 provides one example of a misalignment when the band does not have a proper south-facing orientation.

1.5 PYRHELIOMETERS

1.5.1 Calibration

Most of the commonly used pyrheliometers, such as the Eppley NIP, are not self-calibrating instruments. They are generally calibrated at manufacturers' facilities by comparison with standard sensors. These standards are themselves calibrated in reference to primary standards, such as the Angström compensation pyrheliometer (IPS) maintained in Davos, Switzerland, at the World Radiation Center. Originally conceived to represent the absolute scale of units, this instrument has only been used as a relative reference in order to achieve agreement measurements performed throughout the world (see, for example, Thekaekara[18]). Cavity radiometers equipped with a view-limiting aperture, such as the PACRAD[19] or the Eppley H-F,[20] now increasingly are used as primary standards.

It is recommended that field instruments be sent periodically for recalibration to manufacturers' facilities or a national laboratory. Such a procedure prevents slow degradations from modifying the available data. In addition to periodical factory checks, it is good to perform, if possible, interinstrument comparisons in order to monitor the sensor's sensitivity.

1.5.2 Temperature Correction

Thermopiles show variations in intensity with ambient temperature. Manufacturers generally specify the temperature coefficients defined as the variation of the output voltage with ambient temperature, over the instrument operational range.

Pascoe and Forgan investigated the validity of this information for a Linke-Feussner pyrheliometer by comparing its output with that of a Wilson Active Cavity Radiometer for several clear-sky periods, covering a temperature range between 17° and 31.5°C.[21]. They found a discrepancy of 32% between the temperature coefficient provided by the manufacturer and their experimental value, while indicating that a 15% value should be required to better a 0.5% precision in radiation measurements. They recommend that, in order to achieve the precision necessary for turbidity measurements, the determination of both the temperature coefficient and its variation with ambient temperature be found for each individual instrument of this type, for the environment specific to each measuring site.

1.5.3 Spectral Measurements Using Broadband Filters

The International Radiation Commission recommends the use of sharp cut-off filters in order to divide the solar spectrum into well-defined bands. The most widely used Schott filters have varying lower cut-offs and a common upper cut-off of 2800 nm. The cut-off is defined as the wavelength at which 50% transmission is achieved by the filter. Physically, the cut-off typically takes place over a range of 50 nm.

Transmittance through the filters should be accounted for, and in the case of pyrheliometers equipped with a fixed window, an additional correction should be applied to account for multiple reflections between the filter and the window, as pointed out by Sadler.[22] Variations in filter transmittance were observed by Suraqui et al. to be of the order of 1% for an operational period of two years in Jerusalem, Israel.[23]

The lower filter cut-off is known to vary slightly with temperature.[24] Typical values are 1 to 2 nm for 10°C temperature change. Sadler observed the lower filter cut-off of the RG2 filter to vary between 626 – 629 nm for warm summer months to 624 – 627 nm during cooler months with occasional decrease to 621 nm under extreme temperature conditions in Edmonton, Canada. However, instantaneous lower cut-off correction is rarely applied, because of the size of the error generated; rather, it is suggested to adjust the lower cut-off according to the prevailing temperature of the operating area for the period during which measurements are taken.[25]

It is also suggested that the user perform a verification of the lower cut-off values specified by the manufacturer. As observed by Spencer,[26] this value may vary slightly from batch to batch or may be altered after prolonged use. The results of a test performed with an Optical Multichannel Analyzer to determine the actual cut-off value of six sets of filters utilized at the State University of New York is presented in Table 1.3. The values announced by the manufacturer were 525, 630, and 690 nm for the OG1, RG2, and RG8 filter, respectively.

Table 1.3

Cutoff wavelengths in nm for OG1, RG2, and RG8 filters

Type	OG1	RG2	RG8
PSP	530	630	692
NIP	523	630	689
PSP	523	631	692
NIP	523	625	689
NIP	523	625	689
NIP	523	625	689

Incident energy above 2800 nm is small and depends primarily on the amount of precipitable water in the atmosphere (calculated values may be obtained from *Annals of the International Geophysical Year*[27]). Due to the difficulty of monitoring this parameter systematically, averaged corrections, functions of the solar altitude, are generally performed for data analysis.[25]

1.5.4 Time Constant of the Pyrheliometer

The Eppley NIP has an announced one second (1/e time) time constant. Clark and Budski have tested the time constant on specific NIPs and found a time constant of 3.5 on one and over two seconds on two other instruments. Eppley has verified that the time constant can vary 100% between instruments due mainly to the variations in the sensor black coating thickness. It would be wise to establish the individual time constant for each NIP independent of the average time constant quoted in the literature. This is especially so for programs which require a fast response time for filter wheel measurements.

An example (see Fig. 1.15) of the simultaneous test response times of the PSP and NIP are less than 2.5 seconds.[28]

1.5.5 Tracking

Tracking accuracy is necessary to obtain meaningful direct solar radiation data. Accuracy should be within a small fraction of the angular field of view of the pyrheliometer being used. The main characteristics of several current trackers are presented in Table 1.4, containing a summary of the infomation provided by eight manufacturers and five research organizations.[29]

Two methods of achieving continuous alignment with the solar position have been developed. In the first method, tracker alignment is achieved after the sun's position in the sky has been predetermined. A passive tracker is mounted on a fixed position, leveled and aligned according to the geographic coordinates of the

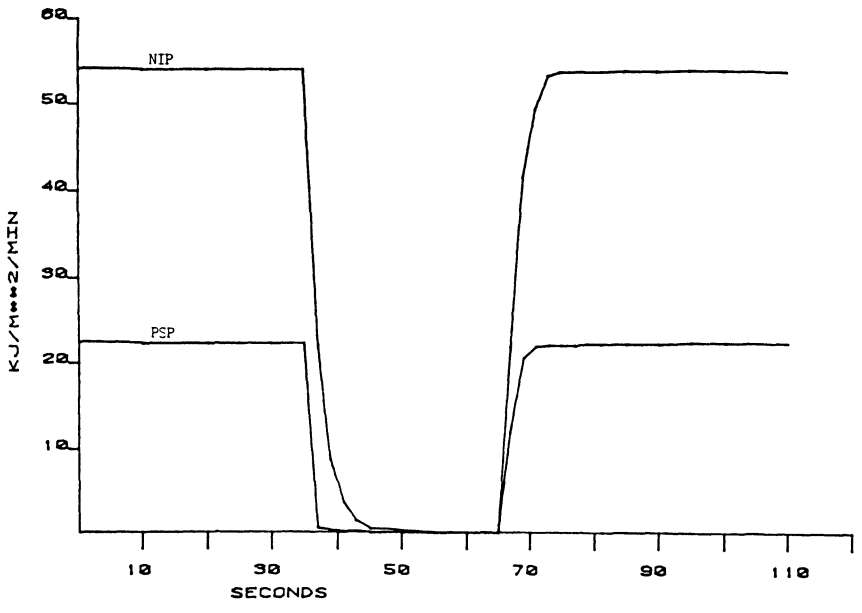


Fig. 1.15: Response test NIP and PSP.

operating area. The positioning of the instrument platform may be accomplished by microprocessor-based computations of the sun's location, for either one or two axis. Then command motors drive the appropriate mechanical devices. Pruett and Evans point out the three limitations of this method of tracking: (1) accuracy with which the tracker polar axis is aligned with the Earth's, (2) the accuracy of the ephemeris data used in the computation of the sun position, and (3) the accuracy of the real time clock used by the control unit.[30]

The second method uses optically controlled electronics to align the instrument platform with the sun (closed-loop feedback). This technique does not require mechanical alignment of the tracker with the true north and is well suited for operation on mobile platforms as long as slew rates are fast enough to maintain constant alignment with the sun. The design of these active trackers should account for occurrences when the sun's position is not the brightest point in the sky. Trackers listed as hybrid in Table 1.4 are of a design which incorporates a passive tracking technique allowing for periodical adjustments by an active feedback mechanism.

The pyrheliometer tracking device, similar in orientation to the tracking disk, should be checked daily. If it is not level, aligned due south and with the proper latitude setting, then the field of view of the instrument will deviate from the perfectly aligned expected field of view. The accompanying figure provides a schematic which aids in correcting a misalignment.

The aiming target on an Eppley Normal Incidence Pyrheliometer (NIP) includes a white circle with a dot in the center (see Fig. 1.16). A shaft of light from a drilled hole in the mounting plate shines onto the circle, its position determines the

Table 1.4

Summary of responses to the questionnaire

Manufacturer or Institution	Mode of Operation	Tracking Accuracy (deg)	Maximum Period of Unattended Operation	Availability	Cost ($)	Comments
PEPCO Electronics	Hybrid	±0.74	100 days	Commercial	12,000	Commercial version of MIT design
Systems and Applied Sciences Corporation	Active	±0.40	[a]	Prototype	5,900	Two research prototypes at SERI
Autonik	Passive	±0.50	1 year	Commercial	11,500	Control unit integral to tracker
SMHI	Passive	±0.50	N/A	Commercial	12,000	Control unit can also as data acquisition system
Eppley Laboratory	Passive	±0.25	daily[b]	Commercial	1,015	Standard unit for U.S. National Weather Service (NOAA Solar Monitoring Network)
LI-COR	Passive	±1.0	1 month[b]	Commercial	2,600	Mechanical design based on NASA patent
SCI-TECH	Passive	±0.2	daily[b]	Commercial	11,000[c]	Commercial version of AES prototype
G. K. Systems	Active	<0.1	N/A	Prototype	N/A	Not included in original survey[d]
NASA	Active	N/A	N/A	Concept	N/A	Not included in original survey[d]
SERI (Gee)	Active	N/A	N/A	Commercial	N/A	Line-focus design overview[d] (see SERI/TP-632-645)
NASA	Hybrid	N/A	N/A	Concept[e]	N/A	
RSI	Hybrid	±0.1 per axis	[a]	Commercial	21,000	For land/sea platforms; able to scan the sky

[a] Active tracker acquires the sun as available with specified accuracy.
[b] Period of operation corresponding to tracking accuracy specification.
[c] $10-12 thousand Canadian.
[d] No questionnaire response available.
[e] See (0) for details of individual designs.

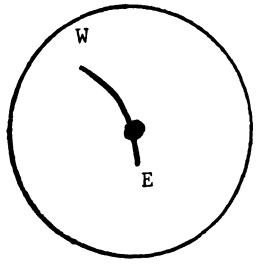


Fig. 1.16: Track of light shaft across aiming circle.

Table 1.5

Data error vs. misalignment angle and interval between maintenance periods (after Rudski et al.[29])

Maintenance Period	Misalignment Angle (deg)	Typical Data Error (%)
daily	0.4	<0.5
2 days	0.8	<0.5
3 days	1.2	0.5
4 days	1.6	1
5 days	2.0	8
6 days	2.4	20
7 days	2.8	25

accuracy of the alignment. The figure indicates a tracker which is aligned slightly west of south. The enlargement of the alignment circle shows the track which the shaft of light would follow during the day. Simple hourly observation of the position of the shaft of light will indicate corrections to be made.

Rudski et al.[31] provided a measurement of the typical data error due to misalignment of the high voltage output Eppley NIP based upon maintenance period (time between realignments) and misalignment due to declination and the equation of time. They also note that the standard NIP would exhibit a 0.5% error at 1°C and approximately 5% at 2° misalignment (see Table 1.5).

1.5.6 Importance of the Solar Aureole

Although pyrheliometers with large field of view are more forgiving of tracking errors than instruments with a small aperture thus requiring less tracking accuracy, the data they record may be questionable for certain sky conditions. Indeed, besides true direct radiation, pyrheliometers also measure this portion of forward

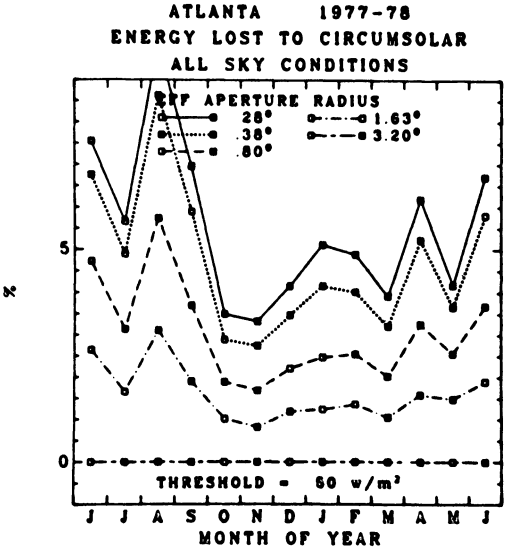


Fig. 1.17: Energy lost to circumsolar for Atlanta, Ga., for the period June 1977 through June 1978.

scattered radiation which is included in their angular acceptance. If data are used for calculations involving point or line focusing concentrators or for precise determination of atmospheric transmission, then this problem should be accounted for and could be limited with minimum tracking errors.

Heys and Vant-Hull, using pyrheliometers of various apertures, observed the intensity of the solar aureole between 1.01° and 2.86° off the center of the sun's position for an eight-day period.[32] This was found to be less than 1.5% for cloudless conditions, but reached as much as 11% when a thin cloud cover was present.

Grether et al. came to a similar observation over an extended period of time (one year) at two recording stations.[33] They used a system described as a circumsolar telescope to obtain systematic detailed intensity of the solar and circumsolar radiation out to 3° from the center of the sun. Their results presented in Fig. 1.17 for Atlanta show that energy contained between 0.38° and 3° accounted for an average of up to 8% of recorded radiation in the month of August.

As a field test to a new variable field of view pyrheliometer, Hickey and Karoli provided a clear day analysis of the solar aureole.[34] Their results, shown (Fig. 1.18) in terms of percent measured irradiance versus aperture angle, indicate that for such clear conditions the classical Eppley pyranometer delivers readings 2.5% superior to the true direct radiation.

1.5.7 Pyrheliometer Tracker Wiring

The pyrheliometer tracker rotates once each day. This twists the umbilical cord and will soon break the wire or cause misalignment, or both. A rewinding of the cord once a day will correct this problem. Occasionally, for a weekend period for

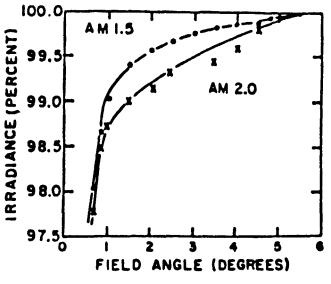


Fig. 1.18: Clear sky analysis of the solar aureole.

instance, the cord may be wound in the opposite direction of the normal twisting. This will allow the cord to unwind during day one and complete the normal twisting on day two. If care is not taken, the cord will be quickly severed.

1.5.8 Quality Control of Data

Although regular inspection of recording instruments is a necessary condition to the quality of measurements, it is strongly recommended that frequent data checks be performed. Wendler and Eaton suggest that global and direct radiation data be plotted daily in order to detect obvious errors, such as those caused by physical damage to an instrument or misalignment of a sensor.[35]

If a systematic quality control is implemented, it should be elaborate enough to account for the lens effect described earlier, and not to reject this data as erroneous. It should also account for the specific environment of the operating site (for example, horizon definition).

A systematic use of the Kastrov formula, giving the ratio of direct to extraterrestrial radiation corrected for Rayleigh scattering and ozone absorption but not for water vapor, provides an upper bound for direct radiation at the earth surface and should detect any excessive direct radiation measurements.[36] For large aperture pyrheliometers, the lens effect may occur on rare occasions within the instrument field of view and generate, for a brief instant, values exceeding Kastrov's. The interrelationship between several instruments operating at a given site provides another means to check data a posteriori.

If diffuse radiation is recorded, the ratio K defined as, K = (NIP $\cos\Theta$ = diffuse)/global and Θ = solar zenith angle, should be checked and its value should be slightly inferior to one in normal condition.*

When filters are used simultaneously on similar instruments, it is useful to establish a systematic control, knowing that energy values recorded with, for instance, an RG8, an RG2, and an OG1 filter should increase respectively and that in most cases ratios between these values should remain fairly constant.

---

* This value depends on the type of instrument, for example, shading disk, shadowband.

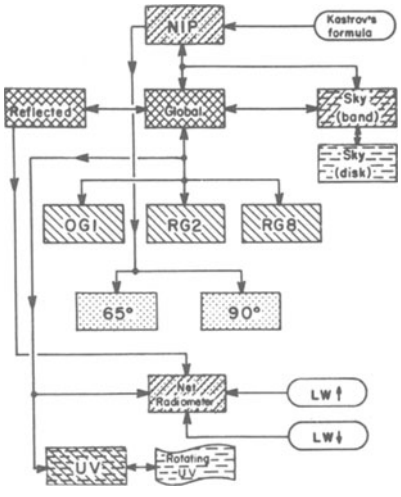


Fig. 1.19: Block diagram showing interrelationships of the radiation measurements in the data control method. Right hashed lines, combination of values from more than one sensor; left hashed lines, ratio of sensor values; dashed and dotted lines, comparison of values from more than one sensor; and dotted lines, calculation of values through the use of empirical formulas.

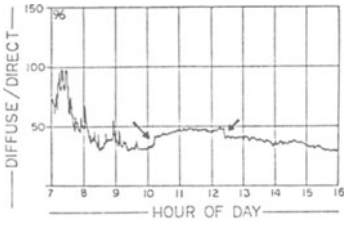


Fig. 1.20: Daily variations of the diffuse to direct ratio in the 520-2800 nm spectral band, September 7, 1982. The arrows point out the results of a misalignment of the pyrheliometer.

The following block diagram (Fig. 1.19) illustrates the data control method implemented at the University of Alaska using the interrelationship between their instruments.[35]

A finer interinstrument analysis may be performed by calculating diffuse radiation from direct and global recorded on the horizontal or on a sloping surface. Since on clear days this value, which is relatively small, is obtained from the difference of two large values, any slight error in the direct or the global measurement becomes apparent when diffuse is plotted for a complete day. The two following

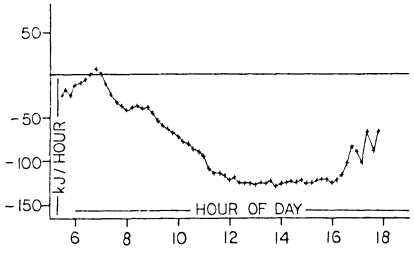


Fig. 1.21: Calculated diffuse radiation impinging on a 43° slope, south-facing pyranometer (based upon the isotropic model) minus measured diffuse radiation (global-direct). The asymmetry of the daily curve indicates that the sloping sensor azimuth is slightly off south-facing.

figures illustrate this last point. Figure 1.20, where the ratio of diffuse obtained from direct and horizontal global to direct, is plotted against time, shows a slight misalignment of the pyrheliometer. On Fig. 1.21, a plot of the difference between diffuse impinging on a sloping pyranometer (global-direct) and this value obtained from the horizontal using an isotropic model, revealed that the orientation of the sloping sensor was slightly off south (less than about 1° west).

Besides a direct feedback of measurements on the maintenance of the instruments, data may also be used to assess the uncertainty with which they are recorded. Hay and Wardle showed that the comparison between equivalent data recorded at two neighboring stations was a good indicator of this uncertainty and its seasonal trends for given time frames.[37]

1.5.9 Solar Radiation Instruments (Kinsell L. Coulson and Yvonne Howell)

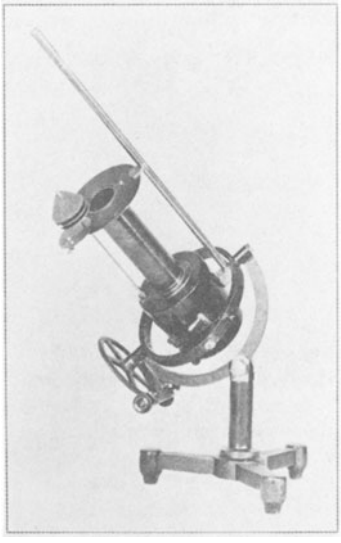
This section has been adapted from *Sunworld* (Solar Radiation Instruments by Kinsell L. Coulson and Yvonne Howell) and is reprinted with the kind permission of the publishers, Pergamon Press, Ltd. Additional information has been added from recent publications.

1.5.9.1 The Pyrheliometers

Pyrheliometers–for measuring the intensity of normal incidence of the direct radiation from the solar disk.

The various types of pyrheliometers are classified by the Commission for Instruments and Methods of Observation of the World Meteorological Organization (Geneva) as *standard, first class*, or *second class* according to the following criteria:

	Standard	1st class	2nd class
Sensitivity (mW cm^{-2})	± 0.2	± 0.4	± 0.5
Stability (% change per year)	± 0.2	± 1	± 2
Temperature (maximum error due to changes of ambient temperature—%)	± 0.2	± 1	± 2
Selectivity (maximum error due to departure from assumed spectral response—%)	± 1	± 1	± 2
Linearity (maximum error due to nonlinearity not accounted for—%)	± 0.5	± 1	± 2
Time constant (maximum)	25 sec	25 sec	1 min



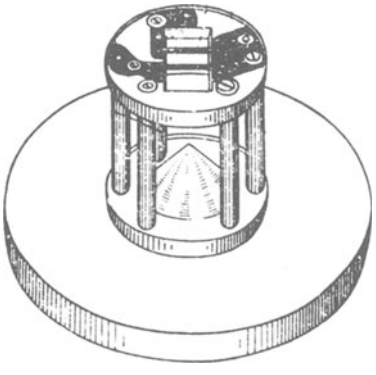
ABBOT SILVER DISK PYRHELIOMETER

Sensor: Calorimeter type, consisting of a blackened silver disk with a thermometer inserted into its edge. Disk is thermally isolated from the environment. A metal shutter alternately shades and exposes the disk to sunlight in a carefully timed sequence. Rate of change of disk temperature determines the energy flux.

Use: No longer extensively used.

Performance: Very stable over time.

WMO Rating: Standard class.



ANGSTROM ELECTRICAL-COMPENSATION PYRHELIOMETER

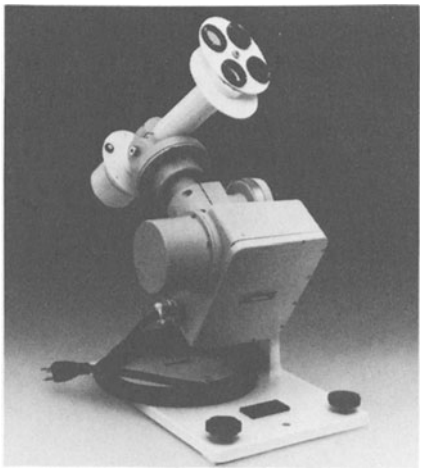
Sensor: Variant of calorimeter type, in which the temperature rise due to absorption of sunlight by one strip of blackened manganin foil is matched in a shaded companion strip by means of a measured amount of electrical energy. In use, each strip is alternately shaded and exposed by a shutter on the collimating tube.

Use: Capable of determining absolute energy although is usually calibrated against a primary standard. Is primary standard in the USSR.

Performance: High degree of stability. One of the most accurate and convenient pyrheliometers.

WMO Rating: Standard class.

EPPLEY NORMAL INCIDENCE PYRHELIOMETER



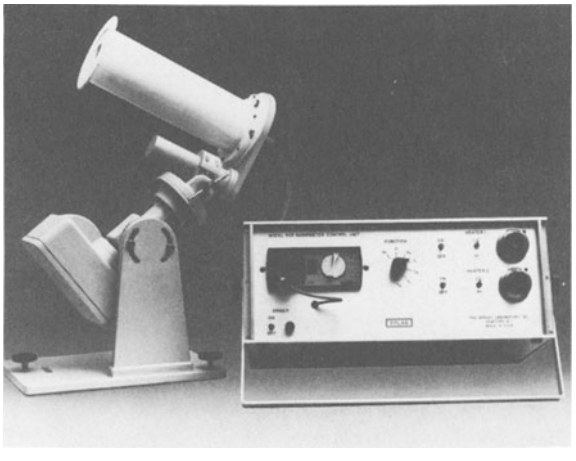
Sensor: Wire-wound copper-constantan plated thermopile with a blackened 9-mm diameter receiver. Solar heat absorbed produces a temperature difference, and hence a voltage, between the thermocouple junctions and a reference junction.

Uses: Sometimes used for continuous operation at remote locations.

Performance: Very stable. By occasional restandardization, can be used as a secondary standard for calibrating other pyrheliometers.

WMO Rating: First class.

EPPLEY H-F SELF-CALIBRATING PYRHELIOMETER



Sensor: Hickey-Frieden (H-F) circular wire-wound thermopile consisting of copper-plated junctions on a continuous constantan wire. Radiation is absorbed on the blackened surfaces of two conical cavities; heat flux is determined by electrical substitution.

Uses: Can be used in several modes: as a thermopile pyrheliometer; as an Angstrom-type, by irradiating one receiver while the other is electrically heated; in an "active" mode, by sequentially equating the signal due to radiation (shutter open) to that due to electrical heating (shutter closed). The various cavity radiometers now made in Europe and the United States are increasingly being used as primary standards.

Performance: Self-calibrating, very stable, very accurate.

LINKE-FUESSNER PYRHELIOMETER

Sensor: Thermopile consisting of 40 manganin-constantan thermocouples arranged in a 1-cm diameter circle. Half are exposed to the sun while the other half are shaded, which produces a voltage related to the solar radiation absorbed.

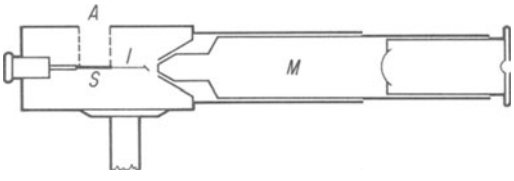
Use: Used with filters to limit its broad spectral range. Can be used to measure diffuse (sky) radiation as well as direct radiation.

Performance: Responses rapidly to temperature changes. Sensitivity somewhat higher than most pyrheliometers.

WMO Rating: First class.

MICHELSON BIMETALLIC PYRHELIOMETER

Sensor: Blackened bimetallic (constantan-invar) lamina. A delicate fiber attached to the lamina deflects when the lamina is exposed to the solar beam. Variant of the calorimeter type of detector.



Use: Portable. Used for daily routine measurements and as a traveling substandard in a radiation-measuring network.

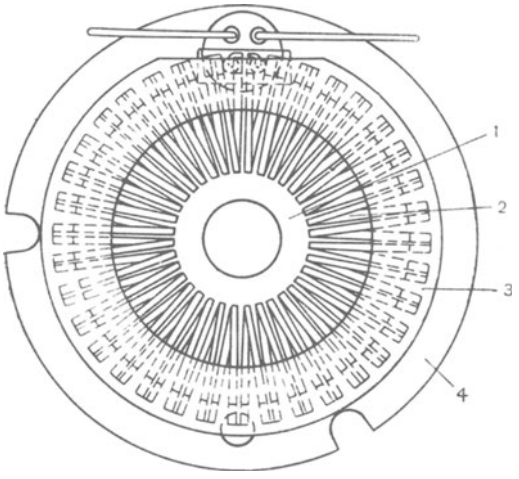
Performance: Fragile. Needs frequent calibration. Reasonably rapid response. No longer widely used.

WMO Rating: First class.

SAVINOV-YANISHEVSKY PYRHELIOMETER (THERMOSTAR)

Sensor: Thin blackened silver disk, around which are arranged 36 pairs of manganin-constantan thermocouples. Sensor assembly is mounted at the bottom of a collimator tube.

Use: Main operational pyrheliometer in USSR. Requires calibration against an absolute standard. Used with recorder and with a clock-driven equatorial mount.



1.5.9.2 The Pyranometers

Pyranometers–*for measuring the global radiation (direct sunlight plus diffuse light from the sky) which falls on a horizontal surface.*

Pyranometers come in many forms, but the best and most widely accepted are those with thermopile-type sensors, because of their long-term stability and because their response is not strongly wavelength dependent. Pyranometer response usually depends on the angle of the incident light; so it is desirable for this dependence to be predictably proportional to the cosine of the incident angle.

The various types of pyranometers are classified by the Commission for Instruments and Methods of Observation of the World Meteorological Organization as *first class, second class,* and *third class,* according to the following criteria:

	1st class	2nd class	3rd class
Sensitivity (mW cm^{-2})	± 0.1	± 0.5	± 1.0
Stability (% change per year)	± 1	± 2	± 5
Temperature (maximum error due to changes of ambient temperature—%)	± 1	± 2	± 5
Selectivity (maximum error due to departure from assumed spectral response—%)	± 1	± 2	± 5
Linearity (maximum error due to nonlinearity not accounted for—%)	± 1	± 2	± 3
Time constant (maximum)	25 sec	1 min	4 min
Cosine response (deviation from that assumed, taken at Sun elevation 10° on clear day—%)	± 3	± 5-7	± 10
Azimuth response (deviation from that assumed, taken on clear day—%)	± 3	± 5-7	± 10

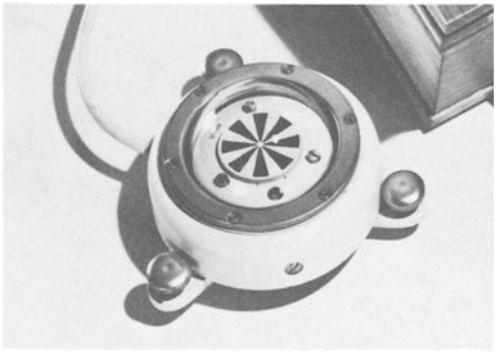
DIRMHIRN-SAUBERER (STAR) PYRANOMETER

Sensor: 32 (or 16) copper plates, half blackened and half painted white, mounted as alternate black and white segments on thermally isolated rings. Copper-constantan or manganin-constantan thermojunctions are soldered to the underside of the copper plates.

Use: Widely used throughout the world. Recommended by WMO for measuring global and sky radiation.

Performance: Very stable with time. With proper calibration, yields reliable and consistent measurements. No azimuthal dependence. Relatively true cosine response.

WMO Rating: Second class.



EPPLEY BLACK AND WHITE PYRANOMETER

Sensor: Wire-wound thermopile made by electroplating copper on constantan. Hot junctions are coated with a black paint and cold junctions with barium sulfate.

Use: Has replaced the older Eppley (180°) Pyranometer.

Performance: Has built-in temperature compensation.

WMO Rating: Second class.

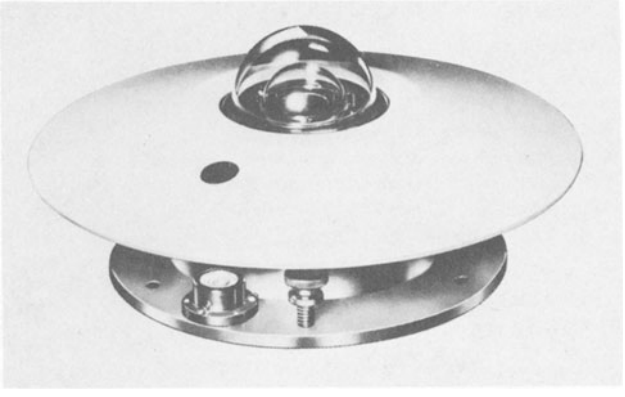
EPPLEY PRECISION SPECTRAL PYRANOMETER

Sensor: Thermopile of copper electroplated on constantan wire, for one-half of each turn of the wire-wound thermopile.

Use: The outer of the two glass hemispheres can be replaced with filters for measurement of solar radiation in selected spectral bands, or with a quartz hemisphere for measurement of ultraviolet.

Performance: Has electrical compensation for ambient temperature changes, and optical compensation for deviation of the response from the cosine law.

WMO Rating: First class.



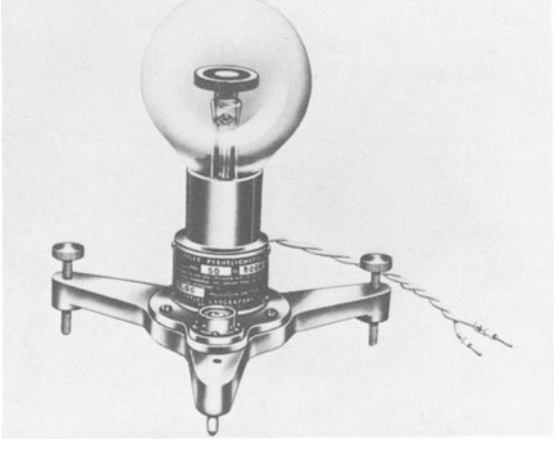
EPPLEY PYRANOMETER *(originally termed 180° Pyrheliometer)*

Sensor: Thermopile with junctions of gold-palladium and platinum-paladium. These are in thermal contact with a receiver disk made of thin silver rings which are coated either black or white. Exposed to sunlight, the black and white rings reach different temperatures, which results in a voltage across the junctions.

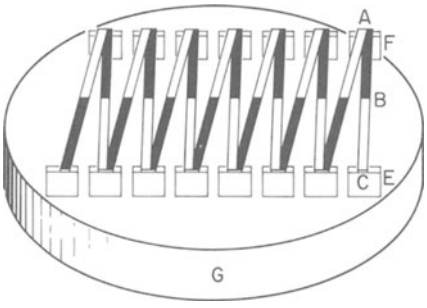
Use: Has been the most widely used instrument in the U.S. No longer in production but many are still in operation.

Performance: Sensitivity depends on temperature unless compensated (by connecting a thermistor in series with one of the thermopile leads).

WMO Rating: Second class.



MOLL-GORCZYNSKI (KIPP) PYRANOMETER



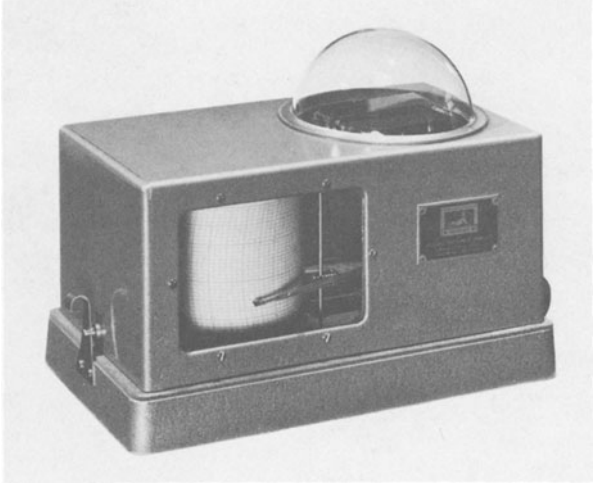
Sensor: Moll thermopile with 14 blackened manganin-constantin thermoelements arranged in a rectangular configuration.

Use: Frequent use in Europe; somewhat less in Africa and Asia.

Performance: Deviation of response form the cosine law and changes in response with azimuth can introduce considerable error.

WHO Rating: Second class.

ROBITZSCH BIMETALLIC PYRANOMETER (ACTINOGRAPH)



Sensor: Blackened metallic strip that acts as a bimetallic thermometer. Absorption of sunshine causes distortion of the strip, thereby moving a recorder pen on a clock-driven chart.

Use: Has been widely used in various parts of the world because of its simplicity and low cost and because it is self-contained and self-recording.

Performance: Very sensitive to ambient temperatures although many ingenious devices have been employed to decrease this sensitivity. Marginal performance. Suitable only for daily totals of radiation in which accuracies of 20% are adequate.

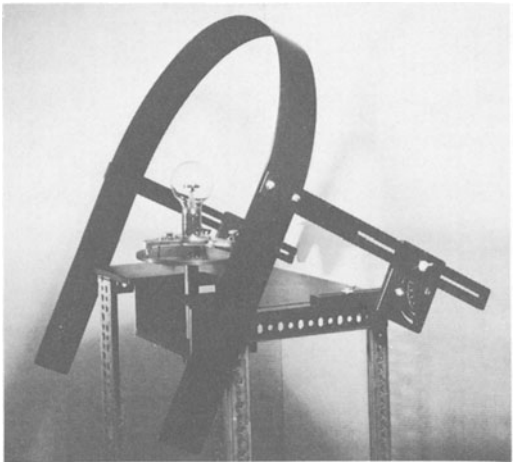
WHO Rating: Third class.

SHADOW RING PYRANOMETERS

Sensor: Two pyranometers, one receiving (as usual) both the direct sunlight and the diffuse sky radiation, and the other shaded from the direct sunlight.

Use: To determine separately the direct and diffuse components of solar radiation.

Performance: Gives useful data if correction is made for the part of the diffuse radiation cut off from the sensor by the shadow ring itself; the correction may be based on either theory or experiment, neither being completely satisfactory because of the variable nature of the radiation intensity and distribution over the sky.



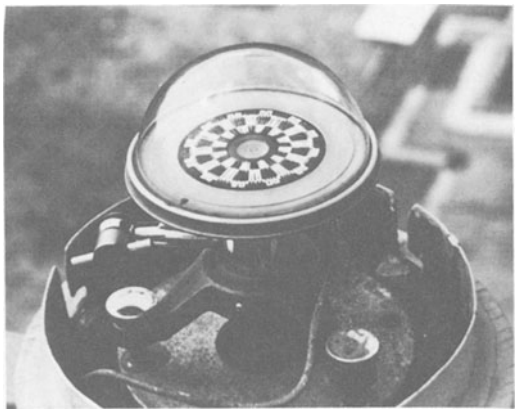
YANISHEVSKY PYRANOMETER

Sensor: Thermocouples of alternate strips of manganin and constantan, arranged in either a checkerboard pattern or alternate black and white segments.

Use: Principal instrument in USSR for measuring global and sky radiation and surface albedo. Used as a relative instrument, calibrated against a primary standard (usually an Angstrom pyrheliometer).

Performance: Requires corrections for deviation from cosine law and for being wavelength dependent.

WMO Rating: Second class.



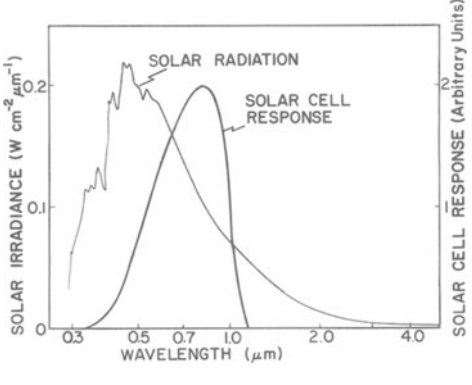
SOLAR CELL PYRANOMETERS

Sensor: Silicon phtovoltaic devices in several different configuations, by various manufacturers.

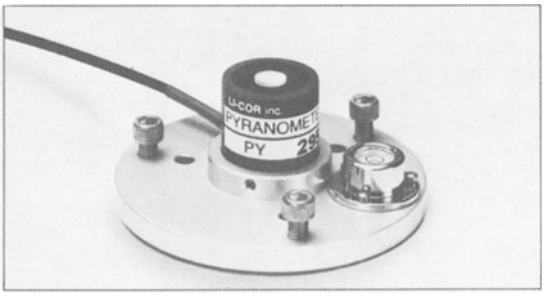
Use: Acceptable for integrating solar radiation over periods of a day or longer. Practical for remote sites.

Performance: Very sensitive, with a fast response, but accuracy not high because does not respond evenly to the full range of the solar spectrum.

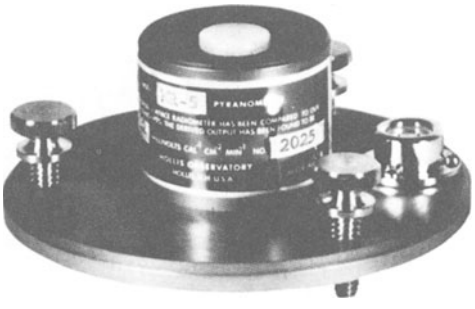
Cost: A fraction of the cost of thermopile-type pyranometers.



Example: LI-COR LI-200S Pyranometer
Silicon detector is mounted in a fully cosine-corrected miniature head. For clear unobstructed daylight conditions, performance compares favorably with first-class thermopile pyranometers.



Example: Hollis MR-5 Pyranometer
Rugged silicon-cell instrument, temperature compensated over a wide range. For routine field use. Best used under open sky conditions.



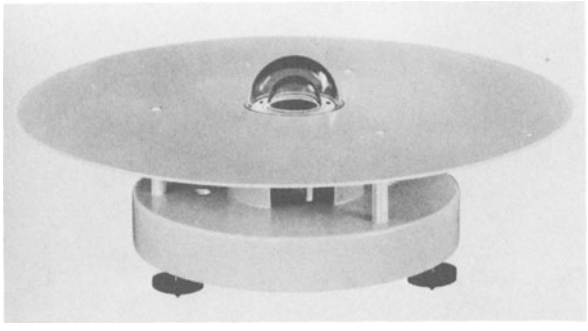
KIPP & ZONEN PYRANOMETER–CM11

Sensor: One hundred thermocouples imprinted on a ceramic substrate using thickfilm techniques.

Use: Comparatively new on the market but suggested uses include meteorology, horticulture, agriculture, solar collector testing, etc.

Performance: Temperature compensated so that sensitivitiy remains constant within $\pm 1\%$ over at least a temperature range of $-10°C$ to $40°C$. Cosine and azimuth error are less than 3% at 10° sun altitude.

WHO Rating: First class.



1.5.9.3 Duration of Sunshine Instruments

DURATION OF SUNSHINE INSTRUMENTS–for measuring the fraction of time during the day that the solar disk is not hidden by clouds.

Sunshine duration—or percent of possible sunshine—is important in characterizing the climate of a given location; its value is also used in roughly deducing the flux of solar radiation on a horizontal surface at locations for which no pyranometric data are available.

The quantity actually measured by the various devices is the amount of time the direct solar radiation is sufficiently intense to activate the instrument (roughly the time during which the Sun casts a visible shadow). But since the activation intensity varies among instruments, the World Meteorological Organization has recommended that all data be reduced to that for a designated standard instrument (Campbell-Stokes Sunshine Recorder) operating under the same conditions.

CAMPBELL-STOKES SUNSHINE RECORDER

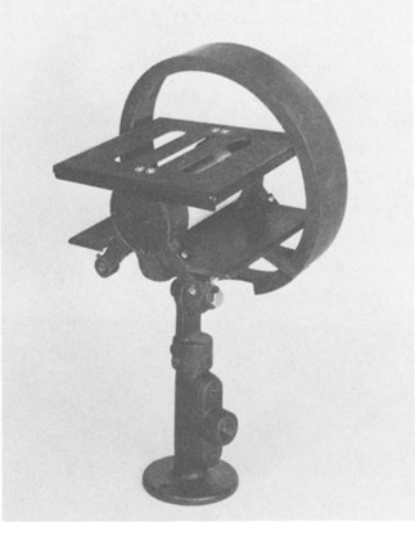
Operation: Direct solar radiation, focused by a glass sphere, burns a trace on a special pasteboard card—except when clouds cover the solar disk. The cards are marked with hour lines so that the periods can be read and tabulated daily.

Performance: Provides useful data on trends of sunny and cloudy periods.

FOSTER SUNSHINE SWITCH

Operation: Pair of selenium photovoltaic cells are mounted so that one is shielded from direct solar radiation by a shading ring while the other is exposed. Diffuse light (which they both detect) results in a zero (balanced) signal; but sunlight produces an (unbalanced) signal that actuates a recorder.

Performance: Is stable and reliable, requiring very little maintenance. Sensitivity is sufficient to give meaningful data throughout the sunrise to sunset period. Used by the U.S. National Weather Service.



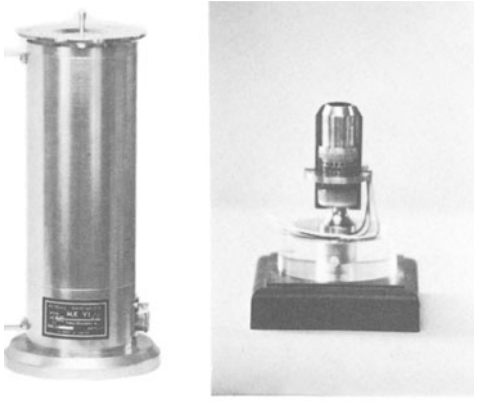
In addition to those instruments presented in "Solar Radiation Instruments," there are specific function instruments which may be of interest to those involved in solar radiation measurement. Three of these are discussed briefly.

PRIMARY ABSOLUTE CAVITY RADIOMETER (PACRAD)

Developed at the Jet Propulsion Laboratory,[6] this instrument was compared with the Angstrom I.P.S. in Davos and found to give readings in the order of 2% higher. It was used in International Pyrheliometric Comparisons (IPC) at Davos in 1970 and 1975. Then it was recommended and accepted as an international reference for the scale of units.

The sensor consists of a windowless cavity mounted in a massive heat sink and has equal sensitivity to ultraviolet visible and infrared radiation. Absorbed incoming radiation is coverted into heat which flows to the heat sink through a metallic thermal resistor producing a temperature difference recorded by a thermopile. A totally enclosed electric heater serves as a source of cavity heating equivalent to radiation heating, providing a built-in means of calibration. The view limiting aperture can vary from 5° to 15°.

The TMI* Mk VI is a further development of the PACRAD by J. Kendall, Sr. Radiometers of this type are used as reference standards in the U.S., England, France, West Germany, Italy, and Argentina.



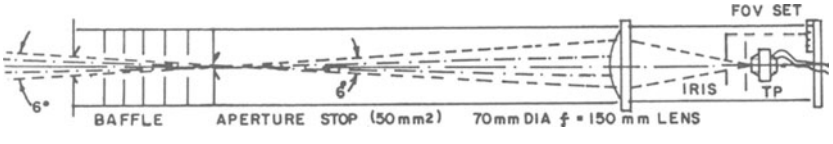
(a) (b)

EPPLEY HICKEY VARIABLE FIELD OF VIEW PYRHELIOMETER

This instrument is designed for measurements of solar and circumsolar radiation within a total subtended angle varying from 40′ to 6°.[23] A baffle assembly limits the field of the incoming radiation to the single lens sensor. An iris diaphragm is placed at the focal plane for selection of the desired field angle.

The detector, a circular wire-wound thermopile with a blackened silver disk receiver, is mounted in a massive copper case.

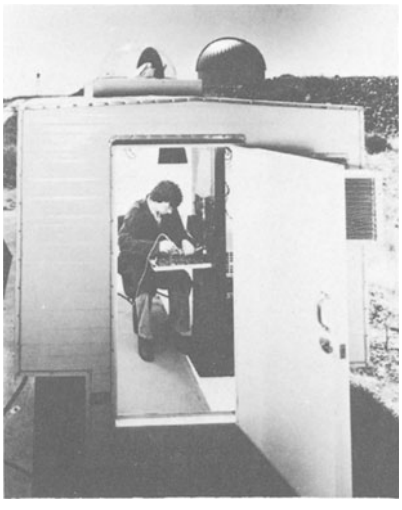
Operational sensitivity is approximately 0.06 mV/mW cm^{-2}, and time constant is of the order of 2 sec.



KLECKNER SCANNING PHOTOMETER FOR MULTISPECTRAL MEASUREMENTS OF DIRECT AND DIFFUSE RADIATION

Designed for unattended operation at remote sites, this computer-controlled instrument measures and records both direct and diffuse radiation in seven spectral bands from 368 to 1025 nm.[27] Its field of view is 1.5°. The sky is scanned in a series of almucantars to determine the distribution of diffuse radiation. A raster scan about the sun is used to measure direct intensity.

The sensor, a silicon PIN photodiode, has a linearity deviation of responsivity of less than 0.5% over six months. By controlling the diode environment temperature and operating this one in a zero-bias configuration. an overall stability bettering 1% is achieved. Photometers of this type are in operation in several sites in the U.S. and Canada.



APPENDIX 1.1 HISTORICAL PERSPECTIVES

1.1.1 Historical Perspectives—From Sundials to Satellites

Some 6,000 years ago, the Chaldeans of Asia Minor constructed sundials, understood the eclipse cycles, and divided the day into hours. Today, the intensity of the sun's energy, not just its position, has become the focal point for data recording and analysis. The following chronological sequence of events is just one way of recognizing the many people and institutions which have provided the basis, the direction, and the spirit which has affected our view of solar radiation today. This historic perspective begins during the heyday of the sundial, at a time when Casella of London was just opening its doors for business.

	Solar History
1666	Visible spectrum first investigated by Newton
1800	Infrared radiation discovered by Herschel
1810	The firm which became Casella of London founded (Tagliabuc and Casella in 1838, later just Casella)
1825	Herschel's pyrheliometer (actinometer) developed: first instrument in which the rate of cooling was introduced into solar radiation measurements
1830	Radiation thermocouple invented by Nobile
1833	Thermopile (extension of thermocouple) introduced by Melloni
1836	Belloni designs the first spherical pyranometer (subsequently modified to the Swiss model, the British model, Henry modification, German model)

1837	Invention of Pouillet's pyrheliometer; first use of the term "pyrheliometer," solar constant measurement by Pouillet yielded 1.76 (cal $cm^{-2}min^{-1}$)
1837	Construction of pyrheliometer on calorimeter principle by Pouillet
1838 to 1840	Invention of first photographic sunshine recorders by T.B. Jordan
1839	Photovoltaic effect observed (sunlight on a voltaic cell) by E. Becquerel
1840	Sir John Herschel obtained the first photograph on a glass plate (a picture of a telescope)
1845	First daguerreotype of the sun was taken by Foucault and Fizeau
1853	Campbell develops the sunshine recorder
1859	Development of Kirchhoff's law
1879	Experimental development of Stefan (later called Stefan-Boltzmann) law. Device introduced by Stokes for use of cards with the Campbell sunshine recorder (later called Campbell-Stokes sunshine recorder)
1881	Observations from top of Mt. Whitney extended known solar spectrum to 1.8 (μm). Value of solar constant estimated to be greater than 3 (cal $cm^{-2}min^{-1}$)
1881	Introduction of the bolometer (sensitive electrical thermometer) by Langley
1883	Becquerel obtained map of solar spectrum out to 1.4 (μm)
1884	Boltzmann's derivation of Stefan-Boltzmann law on basis of thermodynamic theory
1885	Invention of photographic sunshine recorder by J. B. Jordan. Invention of new form of polarimeter and of polestar recorder, both by E. C. Pickering
1885	H. McLeod sunshine recorder
1886	Abney published table of 429 lines of solar spectrum at wavelengths below 1 (μm). Use of first thermoelectric pyrheliograph by Crova. Invention of new type of radiometer by Knut Angstrom
1891	Concept of Maring-Marvin sunshine recorder by D. T. Maring
1893	First construction of electrical compensation pyrheliometer (final form in 1896). Introduction of Angstrom pyrheliometric scale
1894	Law relating wavelength of maximum emission with blackbody temperature by Wein
1897 to 1898	Langley's determination of location and approximate strength of over 700 lines of solar spectrum between 0.4 and 6 (μm)

1898 Invention of Callendar pyranometer; improved version brought out in 1905

1900 Development of Planck's law. Measurements of total solar radiation begun in Washington, D. C., by Smithsonian Institution

1900 Planck's law for wavelength distribution of blackbody emission as a function of temperature (based on his quantum theory)

1902 Monochromatic observations begun by Langley in Washington, D. C. Estimates of solar constant varied between 1.75 to 4 cal ($cm^{-2}min^{-1}$). Work started on a mercury pyrheliometer, which was later developed into the secondary, standard, silver-disk pyrheliometer of C. G. Abbot

1902 Commencement of Smithsonian Institution's "solar constant" observing program by Langley (subsequently continued by Abbot and his collaborators)

1903 Smithsonian Institution started construction of the water-flow pyrheliometer, which was perfected into the primary standard about 1910

1904 First attempt, by Langley, to correlate weather with solar radiation

1905 Recommended universal adoption of Angstrom pyrheliometer as a standard in solar radiation networks and associated Angstrom scale by the International Meteorological Organization

1913 Mean value of solar constant given as 1.933 (cal $cm^{-2}min^{-1}$. First revision of Smithsonian pyrheliometric scale, based on the water-flow pyrheliometer. Abbot constructed hemispherical pyrheliometer and a recording balloon pyrheliometer which was carried to 45,000 ft altitude. First daylight illumination measurements made at Mt. Weather, Virginia, with Sharp-Millar photometer

1913 First revision of Smithsonian pyrheliometric scale, based upon the water-flow pyrheliometer by Abbot

1914 Introduction of U. S. National Bureau of Standards (NBS) reference of total radiation (blackbody) by Coblentz

1914 Balloon flights yielded solar constant of 1.88 (cal $cm^{-2}min^{-1}$)

1915 First vacuum bolometer constructed

1915 Robitzsch pyranometer designed

1916 Pyranometer for measuring global radiation devised. Abbot's "solar cooker," for using direct solar radiation for cooking food, installed

1917 Ultrasensitive vacuum bolometer constructed. Observation station installed on Hump Mountain, North Carolina

1918 Solar observation station installed near Calama, Chile. New instrument constructed for comparing brightness of Sun and sky. Balloon measurement, by Aldrich, of cloud albedo at Arcadia, California, gave

value of 78%. "Short method" of solar constant determination devised and put into use

1919 Honeycomb (melikeron) pyranometer constructed. Marvin pyrheliometer and A. K. Angstrom's electrical compensation pyrgeometer invented

1920 Calama, Chile, station moved to Mt. Montezuma, Chile, where it was operated continuously until 1955. Mt. Wilson station moved to Mt. Harqua Hala, Arizona; abandoned in 1925: mean of 1244 observations at various locations from 1912 to 1920 yielded a solar constant of 1.946 (cal $cm^{-2}min^{-1}$

1922 Dorno's pyrheliograph invented

1923 Invention of Kimball-Hobbs pyranometer (forerunner of first Eppley pyranometers). Invention of Moll thermopile. First weather forecasts based on (indicated) variations of solar constant made by H. H. Clayton

1924 Moll thermopile used by Gorczynski for first Moll-Gorczynski pyranometer, later known as Kipp solarimeter

1925 Solar observation station on Table Mountain, California, established (discontinued in 1958). Strong questioning of Smithsonian Institution's determinations of solar variability begun

1952 Sunshine recorder developed by C. L. Taylor

1926 Solar observation station established at Mt. Brukkaros, South West Africa; abandoned in 1931

1926 Pollak actinometer patented in Czechoslovakia

1926 Henry modifies the Bellani spherical pyranometer

1927 Double water-flow pyrheliometer developed by Shulgin

1929 Method introduced by Kalitin for measuring radiative flux from various zones of the sky. Richardson developed method of measuring reflectivity of natural surfaces from aircraft

1931 Method developed by Kalitin for measuring albedo of natural surfaces from ground

1931 Potsdam absolute pyrheliometer developed

1932 Double water-flow pyrheliometer was adopted as standard by Smithsonian Institution, and first revision of 1913 pyrheliometric scale. (Although the revised scale was reconfirmed in 9134, 1947, and 1952, and showed the 1913 scale to be approximately 2.3% too high, it was never generally adopted.) First standard design of Robitzsch pyranometer developed

1932 Robitzsch actinograph developed

1932	Abbot adopts Shulgin's modification for the pyrheliometer
1932	Smithsonian pyrheliometric scale further revised, according to improved water-flow pyrheliometer (and confirmed in 1934, 1947, and 1952), but such revisions never applied in meteorological practice
1932	Pers sunshine recorder
1933	Solar observation station established on Mt. St. Katherine on Sinai Peninsula; abandoned in 1937
1937	Comparison, indirectly, of NPL, NBS, Angstrom, and Smithsonian scales of radiation. Introduction of absolute drift radiometers to maintain the NPL radiometric reference
1939	Solar observation station established on Burro Mountain, New Mexico; abandoned in 1946
1945	Solar observation station established at Camp Lee, Virginia; abandoned in 1947
1948	Menzel developed sensitive sky photometer
1950	Observation begun at Table Mountain with Menzel sky photometer
1952	The silver-disk pyrheliometer recommended as an instrument for measuring direct solar radiation by Subcommission on Actinometry, World Meteorological Organization, Brussels. Best estimate of solar constant: 1.94 (cal $cm^{-2}min^{-1}$)
1952	Robitzsch actinograph improved by Fowler (Canada)
1953	Invention of sunshine switch by Foster and Foskett of U. S. Weather Bureau
1954	Weather Bureau method of calibrating Eppley pyranometers changed. Old method utilized natural solar radiation; new method uses artificial radiation in an integrating sphere (similar methods later introduced by Canadian Meteorological Service and Eppley Laboratory). Reevaluation of solar constant, by Johnson, through use of new rocket observations in ultraviolet and revised estimates in infrared. Best estimate given as 2.00 (cal $cm^{-2}min^{-1}$) with a probably error of 2%
1956	Introduction of International Pyrheliometric Scale (IPS), based upon Angstrom and Smithsonian references
1956	Statistical study of 30 years of solar radiation data of Mt. Montezuma and Table Mountain by Sterne and Dieter showed root mean square of real changes of the solar constant no greater than 0.0032 (cal $cm^{-2}min^{-1}$). "International Pyrheliometric Scale of 1956" recommended by International Radiation Conference, Davos
1957	"International Pyrheliometric Scale 1956" put into effect; original Angstrom scale increased by 1.5% and 1913 Smithsonian scale decreased by 2.0%

1960	Establishment of NBS reference of spectral radiance
1960	Construction of low-temperature standard of total radiation (blackbody). Introduction and maintenance, independently, of Canadian National Research Council (NRC) radiometric reference
1960 to 1968	Balloon program for solar constant measurements conducted in Soviet Union
1962	Comparison of Guild (drift) and Gilham (Cavity: electrical compensation) absolute radiometers. Verification of NPL radiometric reference
1962	Adoption of the Campbell-Stokes sunshine recorder as an "Interim Reference Sunshine Recorder" by the Commission on Instruments and Methods of Observation, W. M. O. Evaluation of solar radiation pressure on motion of satellite, by Jacchia and Slowey, provides value of solar constant as 2.00 (cal $cm^{-2}min^{-1}$)
1963	Establishment of NBS reference of spectral irradiance
1964–	Cavity radiometer developed for simulated testing
1965	of satellite (Mariner) equipment. This concept can be traced to the work of Abbot and Coblentz
1965	Revision of NBS reference of total radiation
1965	Introduction of Eppley precision pyranometer. Development of automatic control of Angstrom pyrheliometer by Marsh
1966 to 1969	Program using aircraft, balloons, and spacecraft for direct measurement of solar constant and its spectral distribution conducted in the United States
1967	Kerr, Thurtell, and Tanner announce an integrating pyranometer based upon a siliconsensor
1968	Value of solar constant as 1.952 (cal $cm^{-2}min^{-1}$) obtained by Lane and Drummond from rocket aircraft flight to 83 km altitude on 17 October 1967
1968	Development of an NBS high-intensity standard of total and spectral irradiance
1968	Introduction of (NBS) high-accuracy radiance standard
1968	Jet Propulsion Laboratory—Primary cavity radiometer developed to make measurements of irradiance under ambient conditions (Table Mountain test)
1969	Introduction of Eppley black and white (star-type) pyranometer. Previous model, based on Kimball-Hobbs design, discontinued
1970	LI-COR develops LI-2005 pyranometer

1971	Announcement by Thekaekara and Drummond that 1.940 (cal cm^{-2} min^{-1}) adopted as an engineering design value of the solar constant by the U. S. National Aeronautics and Space Administration
1975	Hollis develops MR-5 and MR-5A
1977	Hollis Observatory develops a pyrheliometer— NIP 221 (replaced in 1981 by Autotrak II)
1979	LI-COR develops LI-1776 Solar Monitor
1981	Kipp and Zonen announce the CM 11 pyranometer—a Class 1 instrument
1982	Hollis Geosystems Corporation announces the SPR-10/20 Net Radiometer for salt-gradient ponds

Note: As the Historical Perspective mentions units which are not SI, the following may be of interest:

To convert: 1 cal $cm^{-2}min^{-1}$ = 697.4 watts m^{-2}.

Additionally, the "solar constant" is presented as the authors published it in their own papers. Therefore, there is some variance in the numerical value used throughout this publication.

REFERENCES

1. Michael R. Riches, T. L. Stoffel, and C. V. Wells, eds., *International Energy Agency Conference on Pyranometer Measurements*, SERI-/TR-642-1156, (1981).
2. Kinsell L. Coulson, *Solar and Terrestrial Radiation* (Academic Press, New York, 1975).
3. A. J. Drummond, ed., *Advances in Geophysics, V 14* (Academic Press, New York, 1970).
4. John E. Hay, "Terminology, symbols and units for solar studies," *Proceedings, First Canadian Solar Radiation Data Workshop* Toronto, Ontario: 17–19 April 1978 (1980).
5. Philip Babcock Gove, ed., *Webster's Third New International Dictionary of the English Language, Unabridged* (G. & C. Merriam Co., Springfield, MA, 1976).
6. Kinsell L. Coulson, *Solar and Terrestrial Radiation, Methods and Measurements* (Academic Press, New York, 1975).
7. Clair L. Wyatt, *Radiometric Calibration: Theory and Methods* (Academic Press, New York, 1978).
8. William L. Wolfe and George J. Zissis, *The Infrared Handbook* (Office of Naval Research, Department of the Navy, Washington, D. C., 1978).

9. Fred E. Nicodemus, ed., *Self-Study Manual on Optical Radiation Measurements*, NBS Technical Notes 910-1,2,3 (National Bureau of Standards, Washington, D. C., 1976).

10. J. R. Latimer, "Canadian procedures for monitoring solar radiation," *Proceedings, First Canadian Solar Radiation Data Workshop.*

11. Raymond J. Bahm and John C. Nakos, *The Calibration of Solar Radiation Measuring Instruments* Bureau of Engineering Research, University of New Mexico, Albuquerque, NM, 1979).

12. E. Clark and J. E. Rudzki, *Region IV: In Solar Energy Meteorological Research and Training Site Program Second Annual Report*, SERI/SP-290-1478, (1979).

13. J. T. Peterson, E. C. Flowers, and J. H. Rudasill, "Dew and frost deposition on pyranometers," *J. Appl. Meteor.* **12** (7), 1230–1231 (1973).

14. A. J. Drummond, "On the measurement of sky radiation," *Arch. Meteor, Geophys. & Bioklim.* **7**, 413–426 (1956).

15. Brock A. LeBaron, W. A. Peterson, and I. Birmhirn, "Corrections for diffuse irradiance measured with shadowbands," *Solar Energy* **25**, 1–13 (1980).

16. H. E. Painter, "The shade ring correction for diffuse irradiance measurements," *Solar Energy* **26**, 361–363 (1981).

17. D. W. Spencer, B. S. Oettinger, and R. Stewart, "Diffuse band correction factors for short time intervals," in *Progress in Solar Energy*, Vol. 6 (American Solar Energy Society, Inc., Boulder, CO, 1982) pp. 1253–1257.

18. M. P. Thekaekara, "Solar radiation measurement: Techniques and instrumentation," *Solar Energy* **18**, 309–325 (1976).

19. M. S. Reid, C. M Berdahl, and J. R. Kendall, Sr., "Calibration standards and field instruments for the precision measurement of insolation," *Solar Energy* **20**, 357–358 (1978).

20. J. R. Hickey et al., "The self-calibrating sensor of the eclectic satellite pyrheliometer (ESP) program," *Proceedings of the 1977 Annual Meeting*, edited by Charles Beach and Edward Fordyce (ASES, Boulder, CO, 1977).

21. D. J. Pascoe and B. W. Forgan, "An investigation of the Linke-Feussner pyrheliometer temperature coefficient," *Solar Energy* **25**, 191–192 (1980).

22. G. W. Sadler, "Characteristics of clear sky normal incidence solar insolation measured with the RG2 filter," *Solar Energy* **20**, 139–142 (1978).

23. S. Suraqui, B. Goldberg, and W. H. Klein, "An analysis of the errors found in broadband filter radiometry," *Solar Energy* **19**, 123–127 (1977).

24. A. J. Drummond and J. H. Roche, "Corrections to be applied to measurements made with the Eppley (and other) spectral radiometers, when used with Schott colored glass filters," *J. Appl. Meteor.* **4**, 741–744 (1965).

25. G. E. Sadler, "Measurements of clear sky solar radiation at Edmonton, Alberta, Canada," *Proccedings of International Solar Energy Society*, Atlanta, Georgia (1979).

26. D. W. Spencer, Ph.D. diss., State University of New York at Albany, 1983.

27. *Annals of the International Geophysical Year* (Pergamon Press, Oxford, 1958), Vol. 5, Parts IV, V and VI.

28. J. Healey, D. W. Spencer, and R. Stewart, "Radiometer calibration," SEMRTS Occnl. Rpt., ASRC Rpt. No. 861.

29. *Catalog of Automatic Sun-Following Trackers* edited by Thomas L. Stoffel, SERI-/TR-215-1490, 1982.

30. H. D. Pruett and T. L. Evans, "A computer-driven solar tracking mount," *Proceedings of the 1981 Annual Meeting* (ASES, Boulder, CO, 1981) pp. 1530–1532.

31. John E. Rudski, et al., "The solar-meteorological research program for the south-central U. S.," *Proceedings of the ASME 1980 Winter Annual Meeting*, Chicago, IL (1980).

32. T. H. Heys and L. Vant-Hull, "The contribution of the solar aureole to the measurements of pyrheliometers," *Solar Energy* **18**, 343–348 (1976).

33. D. F. Grether et al., "Application of circumsolar measurements to concentrating collectors," *Proceedings of ISES*, Atlanta, Georgia (1979) pp. 2193–2197.

34. J. R. Hickey and A. R. Karoli, "A variable field of view pyrheliometer," *Proceedings of the 1977 Annual Meeting* (ASES, Boulder, CO, 1977), pp. 15-5.

35. G. Wendler and F. Eaton, "Quality control for solar radiation data," *Solar Energy* **25** 131–138 (1980).

36. V. G. Kastrov, "On the basic actinometric formula," (1928).

37. J. E. Hay and D. I. Wardle, "An assessment of the uncertainty in measurements of solar radiation," *Solar Energy* **29** 271–278 (1982).

CHAPTER 2

ENVIRONMENTAL REQUIREMENTS FOR ANAEROBIC DIGESTION OF BIOMASS

R. E. Speece

2.1 ABSTRACT

The anaerobic digestion process is experiencing renewed interest worldwide. Findings indicate that methanogenic bacteria have catalyzed a revolution in our understanding of procaryote diversity. A host of publications has emerged on this topic. Pure culture techniques have improved, allowing further elucidation of reactions and pathways. Agricultural and environmental engineers have demonstrated renewed interest in the anaerobic digestion process in attempts to address waste treatment management needs and provide supplemental energy supplies, or both. Energy corporations are evaluating anaerobic digestion as a possible source of alternative energy supplies from biomass and other renewable sources. Research in each aspect and overall interest in anaerobic digestion has never been greater.

This review summarizes the literature through 1982 on the microbiology of methanogens, their reactions, characteristics, and pathways. Historical information on nutrient requirements of both the methanogens and fermentation bacteria is compiled.

Particular emphasis is made on the recent discovery of nickel as an essential trace metal requirement for methanogens. Other trace metals such as selenium and tungsten have been discovered. Iron appears to be required in rather high concentrations.

Sulfur requirements for anaerobic digestion are not widely documented, but sulfur appears to be required in concentrations much higher than previously thought. The complex chemistry in anaerobic digestion makes it difficult to clearly identify the requirement for sulfur and trace metals.

Recent investigations of the toxicity response of methanogens has demonstrated they are much less sensitive than previously considered. Acclimation to a wide variety of toxicants has been demonstrated. In addition, toxicity is commonly reversible at the toxicant concentrations commonly observed in the environment.

The economic burden of alkalinity supplementation costs for anaerobic digestion of some types of biomass and industrial wastewaters is so great that a separate section is devoted to this matter. Sources, requirements, and conservation of alkalinity are addressed.

Efforts to model the anaerobic digestion process are reviewed. A number of general and specific models are available. Some models assume volatile acids utilization is rate limiting, while others assume lipids or cellulose degradation to be rate limiting.

The review also considers various design prerequisites. Process configuration, temperature, and type of feedstock are discussed. Some of the industrial wastewaters successfully treated by anaerobic digestion are summarized. Finally, treatability screening procedures and scale-up factors affecting performance are summarized.

2.2 INTRODUCTION

This review addresses the environmental requirements of the anaerobic digestion process. In addition to anaerobic digesters per se, the anaerobic digestion process occurs in the natural ecosystem of marshes, rice patties, and benthal deposits.[1,2] Existing supplies of fossil fuels are reputed to be residues of anaerobic digestion.[3] In more controlled environments, though still in the natural ecosystem, a variation of the anaerobic digestion process occurs in the rumen of animals such as cattle, sheep, and goats. Methane production and anaerobic digestion or both are also major functions in the gastrointestinal tract of many nonruminant animals such as pigs, certain monkeys, iguanas, and so on. It also occurs in some human beings.[2,4] Recently it has been shown that methane production occurs in the gut of termites[5] and in the heartwood of some trees.[6] Methane bacteria have been found in the hot springs of Yellowstone National Park.[7] Methane has also been detected in the deep regions of the oceans.[1,2]

In addition to the natural ecosystem, the anaerobic digestion process has been operated historically with various degrees of control and using a variety of biomass feedstocks. In this review, the term biomass will be used in its broadest sense to include domestic sludges, industrial wastewaters, animal residues, and crops grown specifically for conversion to methane.

In the early practice of anaerobic digestion, the level of control was primitive, being primarily that of simple storage. Consequently, the rates of anaerobic digestion were correspondingly low. With time, various parameters including heating, mixing, alkalinity, nutrients, and trace metals were controlled to increasing degrees. Process configuration was adapted to the feedstock; solids' retention time control evolved to improve methane conversion and operational stability.

The microbial ecology of the process involves rather intricate symbiotic relationships between the microbial species. In addition, the nutrient requirements of the methanogens, one of the classes of organisms involved in anaerobic digestion, appear to be unique. Moreover, the methanogens themselves are phylogenetically distinct from all other bacteria.

The anaerobic digestion process has been employed for a variety of purposes. Historically it was used to stabilize putrescible domestic sludges, often with no attempt to utilize the methane produced. The city of Chicago is only now conducting energy studies to fully use all of the methane generated from its anaerobic digesters.

In some industrial wastewater treatment applications, the major reason for employing anaerobic digestion over aerobic treatment processes was to take advantage of lower energy consumption and greatly reduced excess sludge production characteristic of the anaerobic digestion process. The resultant production of methane may only be a secondary advantage. In a very limited number of applications, the organic waste components may be more degradable by anaerobic treatment than by aerobic treatment.

In some cases, anaerobic digestion has been specifically proposed to convert biomass crops to methane as a supplement to existing natural gas supplies. Corn, sugar cane, and giant kelp are examples of candidate biomass crops. Prototype anaerobic digestion systems have been constructed to convert cattle feedlot manures to saleable methane. Swine manure is anaerobically digested to deodorize it and to produce methane for on-site heating of the farrowing operation. The effluent from animal residue digesters, both solubles and solids, can be an excellent source of fertilizer for crops. Nitrogen and phosphorus are conserved during the digestion process. However, heavy metals in sludge from domestic sludge digesters have severely limited land applications. Finally, one unique application of anaerobic digestion in Finland has been to detoxify chlorinated wood pulping wastewaters so that the effluent can be satisfactorily treated by a subsequent aerobic treatment process. Methane production in this final case is not required, nor for the matter, even desired.[8]

The incentives for anaerobic digestion of industrial wastewaters when compared to aerobic biological treatment lie in the areas of reduced electrical power consumption, reduced biological sludge synthesis, and production of an energy-rich end product, i.e., methane gas (see Table 2.1).

The advantages of anaerobic digestion over aerobic treatment (activated sludge) translate to a primary energy savings of approximately 22×10^6 Btu or \$166.00 per metric tonne of COD destroyed (assuming electricity at \$0.06/kWh, generated at 35% efficiency, sludge disposal costs at \$100.00 per metric tonne of dry solids, and methane valued at \$4.50 per 10^6 Btu). For a 600 metric tonne per day papermill, treating only the condensate wastewater stream, the potential savings would be approximately \$330,000 per year. For a large petrochemical or pharmaceutical plant, the potential savings would be approximately \$500,000 per year. For a typical cheese plant, the potential savings would also be approximately \$500,000 per year.[9] While these costs may be relatively small compared to overall plant cash flow, they nevertheless are worthy of consideration.

Table 2.1

Comparison of aerobic treament and anaerobic digestion[8]

	Activated Sludge	Anaerobic Digestion
Electrical power	1100 kWh	0
Excess biological sludge production	400-600 kg	20-150 kg
Methane production	0	340 m^3

Based on conversion 1000 kg of Chemical Oxygen (COD) equivalent of pollution.

The first requirement for successful methane fermentation of biomass feedstock or industrial wastewaters is that the organic components must be biodegradable by the mixed microbial population contained in the anaerobic digester. This requirement is generally satisfied by a wide range of organics and will be discussed in more detail later in the review.

The second requirement has to do with toxicity inherent in the feed. This is a major consideration with respect to industrial wastewaters since various toxicants are commonly produced or used industrially. For instance, coking wastewater contains phenols and cyanide; petrochemical wastewater contains about 1 percent of the product being made as well as heavy metal catalysts; pharmaceutical wastewater may contain detergents or bacteriocidal agents used in the washdown operation and, in some cases, residual antibiotics. In these cases, the anaerobic digestion treatment system must be designed to operate with stability under transient or chronic toxicity as it occurs in the wastewater stream.

A third requirement is that the system must accommodate the nonsteady-state pollution load and hydraulic flow conditions characteristic of many industrial operations and some biomass to methane schemes. Production may be intermittent or seasonal, e.g., a bakery operating 40 hours per week or a winery operating three months per year. Product lines may change, such as when a petrochemical plant switches production. Some biomass production is also seasonal. The microbial population must adapt adequately to quantitative and qualitative changes in the organic components. In addition, the process design must insure proper retention of biomass during hydraulic surges.

In summary, after the biodegradability of organic components in biomass feedstocks or in industrial wastewater is established, the primary considerations for successful anaerobic digestion treatment must properly insure:

(a) tolerance to transient or chronic toxicity;

(b) nonsteady-state pollution load and/or hydraulic flow conditions; and

(c) qualitative changes in the nature of the organic components.

The key process variable controlling operational performance and stability is the biological solids retention time (SRT). If SRT is properly controlled, the biological definition of an organic component may be changed from toxic to nontoxic, and in some cases to degradable. This is the phenomenon of acclimation, a strong function of SRT.

Traditionally, anaerobic digestion has primarily been applied to the stabilization of municipal wastewater sludges. This class of sludge is composed of a continuum of organics, ranging from the readily biodegradable to the nonbiodegradable. About half of the organics in municipal sludges are functionally nonbiodegradable, resulting in a residual sludge disposal problem. Grease (primarily long chain fatty acids) and waste bacterial cells are the precursors of much of the methane produced. The waste bacterial cells originate from secondary biological treatment. Both of these components degrade relatively slowly. Traditionally, anaerobic digestion works well on these substrates so long as sufficient SRT (8 to 15 days) is provided. The methane product has economic value and is commonly used to generate electricity, drive air blowers, heat buildings, and/or warm the incoming sludge.

Anaerobic treatment of industrial wastes, however, presents a different set of problems. On the one hand, many industrial wastewaters are warm and contain high concentrations of simple, soluble organic components which are rapidly degraded. Anaerobic digestion of such wastes can often be accomplished with removal efficiencies in excess of 90 percent, without additional heat input, in hydraulic detention times of several hours involving little or no net waste sludge production beyond solids loss in the effluent.

The pertinent microbiological characteristics and the critical design criteria will be discussed. The spectrum of degradable organics will be covered. The matter of substrate toxicity and toxicants inherent in the wastewaters will be addressed, with particular attention given to the response of methanogens after exposure to toxic substances.

Optimum temperature for mesophilic methanogens is 35° to 40°C. However, stable methanogenesis has been reported at as low as 10°C.[10] Thermophilic digestion operates at 55° to 60°C. The pH optimum is between 6.8 and 7.4. Here again stable methanogenesis as low as pH 5 and as high as 9 has been observed.[11]

The methanogens require an oxygen-free environment for growth, but oxygen appears to be bacteriostatic rather than bacteriocidal for one common species. The oxidation-reduction potential, E_h, should be below -450 mV. Above -330 mV, the methanogens are inactive.[12] There is some speculation that a low E_h is the result and not the cause of good digestion. Some consider oxygen more important than E_h, although traces of oxygen commonly raise E_h.[13]

Seldom are all environmental conditions optimum for anaerobic digestion, nor do they have to be to provide satisfactory operation and treatment. Generally, an increase in SRT can compensate for adverse environmental conditions, at least up to a certain degree. This principle is demonstrated by the vast scale of methano-

genesis which occurs under less than optimum environmental conditions in swamps, marshes, and benthic deposits.

Anaerobic digestion of organic wastes simultaneously can accomplish pollution abatement and energy conservation. In addition, an order of magnitude less quantity of biological sludge may be produced, and proportionately less nutrients are required when compared with aerobic treatment processes. The anaerobic process offers tremendous potential as an alternative to aerobic treatment for industrial waste treatment. Thus far, industrial applications are limited due to a preconceived notion that anaerobic digestion processes are not able to operate efficiently and reliably in the real industrial wastewater world replete with toxicants, synthetic organics, and hydraulic flow in a dynamic state. This concept does not recognize adequately the key role of SRT and recent technological developments. These provide prolonged SRT by insuring biomass retention with relatively low hydraulic retention times and associated (relatively) small reactor volumes.

However, recent research, pilot and full-scale applications have shown that methane fermentation systems can accommodate dilute organic concentrations, temperatures as low as 25°C (lower in some cases), and both chronic and slug doses of toxicants. Such systems can also degrade a wide variety of petrochemicals and other organics heretofore thought to be resistant to anaerobic degradation at hydraulic detention times equivalent to aerobic treatment processes. Present technology defines solids retention time as the most crucial design and operational parameter. Proper SRT control provides sufficient acclimation time to allow efficient treatment of the nonsteady-state toxicity and synthetic organics common to industrial wastewaters. Application is not predicated on any new scientific breakthrough, but only on pilot treatability studies with the actual industrial wastewater in question. In the future, applications will no doubt become more widespread, and the benefits of treating organic wastewaters using methane fermentation will be more fully realized. When the present surplus of natural gas supplies becomes a limiting factor again, renewed attention will be directed to energy crops grown specifically for conversion to methane by the anaerobic digestion process.

2.3 MICROBIOLOGY OF METHANOGENS

The conversion of complex organic material under anaerobic conditions in the absence of light, nitrate, nitrite, and sulfate is accomplished by chemoheterotrophic, nonmethanogenic bacteria, and methanogenic bacteria.[14] The complex organics are hydrolyzed by the chemoheterotrophic nonmethanogens to free sugars, alcohols, volatile acids, hydrogen, and CO_2. The alcohols and volatile acids longer than two carbons are then oxidized to acetate and H_2 by obligate proton-reducing organisms which must exist in symbiotic relation with H_2-utilizing methanogens.[15] Finally, the methanogens convert acetate and H_2 to methane.[16] This conversion is shown graphically in Fig. 2.1.

The reason for this symbiotic relationship is that for favorable thermodynamic conditions, the H_2 partial pressure must be kept extremely low for the conversion of alcohols, propionate, and longer carbon chain fatty acids to acetate and hydrogen. At higher H_2 levels, this conversion cannot proceed, and therefore higher volatile

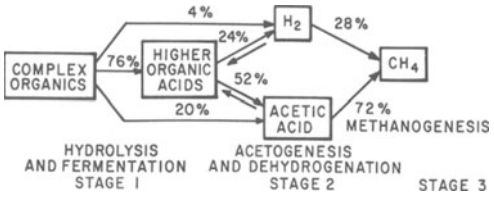


Fig. 2.1: The three stages of methane fermentation and the percentage flow of the energy content of complex organic materials through each stage into methane as represented by chemical oxygen demand.[17]

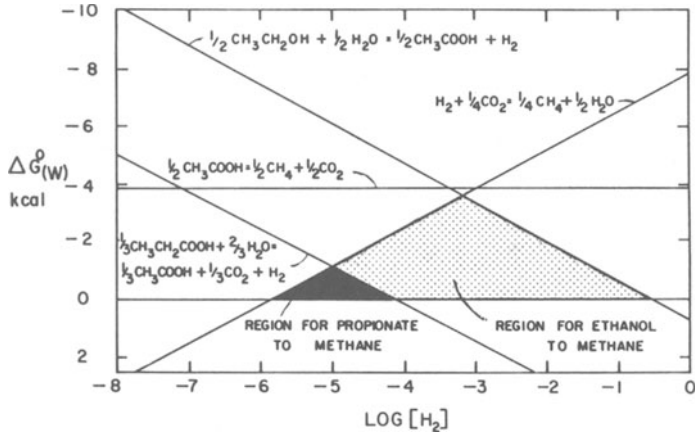


Fig. 2.2: Effect of hydrogen partial pressure on the free energy of conversion of ethanol, propionate, acetate, and hydrogen during methane fermentation.[17]

acids accumulate in the system instead of freely being converted to methane and CO_2. McCarty[17] has graphed this relationship in Fig. 2.2.

This obligate, syntrophic relationship was only recently discovered by Bryant et al. in 1967.[18] These authors demonstrated that *Methanobacillus omelianskii* was actually a two-membered culture. This discovery was a milestone in our understanding of the anaerobic digestion process. Lorowitz and Bryant[19] have recently succeeded in isolating a stearate-utilizing bacterium in coculture with an H_2-utilizing *Desulfovibrio*.

The standard free energies for the conversion of propionate, butyrate, and ethanol are as follows:[20]

$$CH_3CH_2COO^- + 3H_2O \longrightarrow CH_3COO^- + HCO_3^- + H^+ + 3H_2 \quad + 76.1 \text{ kJ}$$

$$CH_3CH_2CH_2COO^- + 2H_2O \longrightarrow 2CH_3COO^- + H^+ + 2H_2 \quad + 48.1 \text{ kJ}$$

Table 2.2

Methanogenic reactions[14]

	$\Delta G°$ (kJ)
1. Acetate	
$^{\circ}CH_3{}^{\Delta}CO^- + H_2O \longrightarrow {}^{\circ}CH_4 + H^{\Delta}CO_3^-$	− 30
2. Methanol	
$4CH_3OH \longrightarrow 3CH_4 + HCO_3^- + H^+ + H_2O$	−314
3. $H_2 - CO_2$	
$4H_2 + H + HCO_3^+ \longrightarrow CH_4 + 3H_2O$	−136
4. Amines	
$4CH_3NH_3 + 3H_2O \longrightarrow 3CH_4 + HCO_3^- + 4NH_4^+ + H^+$	−225
$2(CH_3)_2NH_2^+ + 3H_2O \rightarrow 3CH_4 + HCO_3^- + 2NH_4^+ + H^+$	−220
$4(CH_3)_3NH^+ + 9H_2O \rightarrow 9CH_4 + 3HCO_3^- + 4NH_4^+ + 3H^+$	−669
5. Formate	
$4HCOOH + H_2O \longrightarrow CH_4 + 3HCO_3^- + 3H^+$	−130

$$CH_3CH_2OH + H_2O \longrightarrow CH_3COO^- + H^+ + 2H_2 \quad + 9.6 \text{ kJ}$$

Thus, if the reactions are to be thermodynamically favorable, the partial pressure of H_2 must be maintained at less than 10^{-4}, 10^{-3}, and 10^{-1} atmospheres (less than $4 \cdot 10^{-6}$, $4 \cdot 10^{-5}$, and $4 \cdot 10^{-3}$ moles H_2 per liter of medium) for propionate, butyrate, and ethanol respectively.[21]

2.3.1 Interspecies H_2 Transfer

Thus, a physiological partnership between nonmethanogenic chemoheterotrophs and methanogens must exist if the higher volatile fatty acids and ethanol are to be converted to methane.[16] Hungate[22] postulated that the electrons formed from the oxidation of fermentation substrates were used for the reduction of CO_2 to methane in the rumen. Iannotti et al.[23] were the first to systematically verify this phenomenon. In a well operating anaerobic digester, the complex organics are converted all the way to methane and CO_2 with low concentration of volatile acids and almost nondetectable levels of H_2 in the system. In addition, H_2 added to the system is rapidly converted to methane.[24,25] In malfunctioning digesters, acetate, propionate, butyrate, and H_2 accumulate in the system.

2.3.2 Methanogenic Reactions

Table 2.2 lists the methanogenic reactions and the accompanying free energy changes.

2.3.3 Characteristics of Methanogenic Species in Pure Culture

Table 2.3 lists the methanogenic species, morphology, and growth substrates.

2.3.4 Methanothrix Soehngenii

Huser et al.[26] described a new genus of methanogenic bacteria which is probably the filamentous bacterium earlier referred to as *Methanobacterium soehngenii*,[27,28] "fat rod" or "acetate organism." This organism is often noted to predominate in anaerobic digesters having chronically low volatile acids concentrations.[29-31] *Methanosarcina* is usually not present in great amounts in digested sludge.[26,31-34] *M. soehngenii* can only metabolize acetate to methane (not H_2, methanol, methylamine, or formate); it is one of only three methanogens capable of converting acetate to methane, the other two being *M. barkeri* and *M. mazei*.

Huser et al.[26] reported on the characteristics of this organism. It has a temperature optimum of 37°C and is inactivated at 50°C. It has an optimal pH range of 7.4-7.8. Sulfide and ammonia serve as the sulfur and nitrogen source. Oxygen completely inhibited growth and methane production, but the organism did not lose its viability upon restoration of anaerobic conditions. It has a growth-yield coefficient (dry wt. of cells per wt. of acetate) of 0.019, a half-saturation coefficient of 41 mg/L acetate, and a doubling time of 3.4 days. Comparative analysis of the 16S ribosomal RNA (r RNA) revealed only a distant relationship to other methanogens. The highest degree of relatedness was found to *M. barkeri*. The organism has been reclassified as *Methanothrix soehngenii*.

2.3.5 Biochemical Mechanisms and Pathways

McCarty[17] set forth the pathway for methane fermentation of a complex waste, as shown in Fig. 2.1. It is significant to note that approximately 70% of the methane generated in the digestion of complex wastes comes from acetate.[35-37]

Balch et al.[38] have shown by means of 16S rRNA oligonucleotide catalog comparisons that the methanogens are phylogenetically distinct from typical procaryotes. The cell wall of *Methanosarcina barkeri* consists of a heteropolysaccharide, but does not contain peptidoglycan, the rigid cell wall of all other procaryotes except *Halococcus morrhaiae*.[39] They have been classified as members of the archaebacteria,[40] a proposed phylogenetically distinct biological grouping. All archaeobacteria have different cell walls (no muramic acid or true peptidoglycan), different membrane lipids, and so on, as compared to *Eubacteria*.[41] Therefore, it may not be unreasonable to expect to discover an array of metabolic features such as coenzymes which are confined to the methanogens.[42]

Cofactors generally have been found to be ubiquitous. However, a recently discovered cofactor, 2-mercaptoethanesulfonic acid (HS-*CoM* or just plain *CoM*), has been found in all methanogens available in pure culture except *M. ruminantium* which requires it for growth. However, none of this coenzyme has been found in a wide range of nonmethanogenic, eucaryotic tissues and procaryotic organisms.[43]

Table 2.3

Characteristics of methanogenic species in pure culture[14]

Species	Morphology	Growth Substrates
Methanobacterium		
formicicum	Long rod to filament	$H_2 - CO_2$, formate
bryantii	Long rod	$H_2 - CO_2$
thermoautotrophicum	Long rod to filament	$H_2 - CO_2$
Methanobrevibacter		
ruminantium	Lancet-shaped cocci	$H_2 - CO_2$, formate
smithii	Lancet-shaped cocci	$H_2 - CO_2$
arboriphilus	Short rod	$H_2 - CO_2$
Methanomicrobium		
mobile	Short rods	$H_2 - CO_2$, formate
Methanogenium		
cariaci	Irregular small cocci	$H_2 - CO_2$, formate
marisnigri	Irregular small cocci	$H_2 - CO_2$, formate
Methanospirillum		
hungatei	Short to long wavy spirillum	$H_2 - CO_2$, formate
Methanosarcina		
barkeri	Pseudosarcina	$H_2 - CO_2$, methanol methylamines, acetate
*mazei***	Pseudosarcina to irregular coccoid elements	$H_2 - CO_2$, methanol methylamines, acetate
Methanococcus		
vannielii	Irregular small cocci	$H_2 - CO_2$, formate
voltae	Irregular small cocci	$H_2 - CO_2$, formate
Methanothrix		
*soehngenii**	Long filaments	acetate

* Reclassified by Huser.[26]

** Reclassified by Mah and Kuhn.[27]

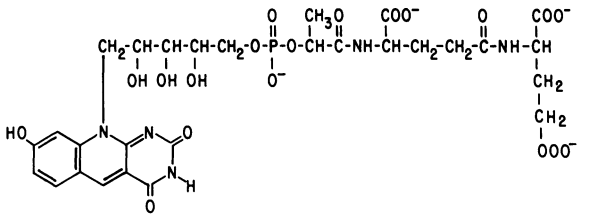


Fig. 2.3: Structure of F_{420}.

Coenzyme M (*CoM*), as a required growth factor for rumen methanogens, has further elucidated the pathway of methane formation. *CoM* was discovered by McBride and Wolfe[44] as a required cofactor for the transfer of methyl groups from methylcobalamin to methane. Taylor and Wolfe[45] developed an assay for *CoM* presence. *CoM* was shown to be one of the growth factors present in rumen fluid required for the growth of *M. ruminantium*.[46]

HS-*CoM* is involved in the terminal reduction step as the *X* factor in the pathway proposed by Barker [43-45,47,48]

$$CH_3 - S - CoM \xrightarrow[\text{Methyl Reductase}]{H_2 Mg^{++} ATP} CH_4 + HS - CoM \qquad (2.1)$$

where CH_3-S-*CoM* is 2- (methylthio) ethanesulfonic acid and HS-*CoM* is 2-mercaptoethanesulfonic acid.

Another organic compound apparently specific to the methanogens is factor 420 (F_{420}) first discovered by Cheeseman et al.[49] F_{420} exhibits a marked absorption peak at 420 nm when oxidized and no fluorescence or absorption at 420 nm when it is reduced. Tzeng et al.[50,51] determined the role of F_{420} as an electron transfer coenzyme. They detected an *NADP*-linked F_{420} dependent hydrogenase system in *M. smithii* and *M. bryantii*. They also discovered a formate-dehydrogenase-F_{420}-dependent-*NADP*-linked system in *M. smithii*.[51] F_{420} has tentatively been identified as shown in Fig. 2.3.[52] It has a "riboflavin-like" component which all methanogens contain. Fluorescence may be used in identifying methanogenic bacteria in ecosystems.[53,54]

Recently still another coenzyme, F_{430}, was discovered in extracts from *Methanobacterium thermoautotrophicum*. F_{430} was found to be of low molecular weight and is possibly another novel compound characteristic of methanogens.[55] Whitman and Wolfe showed that F_{430} contains substantial amounts of nickel and lacks other metals commonly associated with molecules of biological origin. Diekert et al.[56] found that nickel was an essential component of factor F_{430} and that iron, cobalt, and molybdenum are not involved. Whitman and Wolfe[55] also state that F_{430} is the first nickel-containing, biological compound of low molecular weight to be reported. Ellefson and Wolfe[57] recently obtained evidence that factor F_{430} may be the prosthetic group of methyl coenzyme M (*CoM*) reductase, and there is evidence that F_{430} has a nickel tetrapyrrol structure.[58] Diekert et al.[59] also found that nickel

was incorporated into a protein fraction. There are also some indications that the hydrogenase of methanogenic bacteria could be a nickel protein.[60]

Most pure culture studies have used H_2, CO_2, and formate as the preferred substrates for methanogenesis.[34] Many methanogens are fully autotrophic, requiring only some inorganic salts and an atmosphere of H_2 and CO_2.[61] Some require the addition of organic substrates and *M. smithii* needs acetate to supply about 60% of its cell carbon.[62,63] Until Mah et al.[32] adapted *M. barkeri* to acetate in the absence of H_2, it was believed by Zeikus et al. that H_2 was a requirement for growth in the atmospheres of a pure methanogenic culture, even though in sewage sludges and lake sediments the concentration of hydrogen gas is very low or undetectable. Mixed methanogenic cultures thrive without having been given H_2 gas. Zeikus et al.[64] suggested that H_2 could be provided by the growth of heterotrophic contaminants on lysed cells.

Much work has been done on the use of acetate in methanogenesis because of its importance in the natural habitat and in anaerobic digesters. The cleavage of acetate to CH_4 and CO_2 results in the release of some free energy ($\Delta G^\circ = -28$ kJ/mole), but Zeikus et al.[64] contended that it was insufficient to sustain bacterial growth. The fact that enriched methanogens have been maintained for years on acetate alone, with no added H_2, is not in agreement with their hypothesis. Recent work by Mah et al.[32] and Smith and Mah[65] has documented with pure cultures that acetate serves as the sole carbon and energy source. Approaching the problem of methane production from the precursors has reached a stalemate until the intermediate products and enzymes of the first steps of methanogenesis have been isolated. Approaching the problem from the product end has uncovered the final step of methanogenesis, involving coenzyme M, or 2 mercaptoethane-sulfonic acid.[44,45,63] Methane is formed from its methyl derivative, CH_3-CoM, by the addition of H_2. This reaction has been described by Gunsalus et al.[47] Daniels and Zeikus[63] found CH_3-CoM and other labeled compounds in cell extracts fed labeled CO_2.

Methane must be a by-product of some more useful cell function, because from the cell's point of view, methane is useless. Because methanogenesis parallels bacterial growth,[66] two possibilities have been considered: the production of a form of energy useful to the cell and the synthesis of the cell's organic molecules. It is understandably difficult to determine which of these processes is responsible for methane production, since both of them parallel cell growth and each is dependent on the other.

Before the discovery of CoM, Bryant et al.,[67] Blaylock and Stadtman[68] and Wolin et al.,[69] obtained cell-free extracts which produced methane upon the additions of (1) methyl-cobalamin and 5-methyltetra-hydrofolate, which are common coenzymes for methyl group transfer; (2) serine and pyruvate, which are common early intermediates in cell organic molecule synthesis; and (3) CO_2. Zeikus and Wolfe[66] also examined cell-free extracts of *M. thermoautotrophicum* for methanogenic activity. They found the highest activity using 0.05 M *TES* buffer, with somewhat less activity using a K_3PO_4 buffer and no activity using a Tris-HCl buffer.

Doddema et al.[70] attempted to link *ATP* synthesis with methanogenesis. They also cited other workers who demonstrated a rise in *ATP* concentration in whole

cells during methane production.[71,72] Thauer et al.[20] suggest that ATP production in methanogens is coupled to electron transport, resulting in reduction of CO_2 to methane. However, the mechanism coupling methane production and ATP synthesis remains a mystery.[73]

Methanogenesis, then is probably an energy-releasing process for the production of ATP, involving the stepped reduction of CO_2 to methane. This process is probably independent of macromolecule synthesis and involves several derivatives of the coenzyme M and compound F_{420} which are unique to methanogens. Delafontaine et al.[74] have demonstrated a good correlation between methane production rate and F_{420}.

2.3.6 Predominance of Species

A competition for available substrate occurs between species. The predominance of a species is dependent on complex ecological conditions. It has been proposed that this competition can be most readily explained on the basis of the kinetic coefficients which describe the growth of the organism in a given complex ecosystem.

Sulfate reducing organisms compete with methanogens for hydrogen under anaerobic conditions. The half saturation concentration (K_S) for hydrogen metabolism by methanogens is 6.6 μM but only 1.3 μM for sulfate reducers.[75] Thus, sulfate reducers have a competitive advantage over methanogens for hydrogen in high sulfate, hydrogen-limiting environments. Also, the apparent K_S values for acetate were determined to be 0.2 μM and 3 μM for sulfate reducers and methanogens, respectively. It has been often observed that when domestic wastewater containing several hundred mg/L of sulfate is treated anaerobically, methane production is nil, and hydrogen sulfide production accounts for the major reduction in liquid oxygen demand.[16] Therefore, sulfate concentration in the carriage water of biomass to methane systems is an important parameter that may affect anaerobic digestion of sea-water-grown biomass which has high salinity and high sulfate concentrations.

Competition between methanogenic species also determines which species predominates. Kinetic coefficients for acetate catabolism of *M. barkeri* have been reported to have a specific utilization rate, k, of 6 to 10 days^{-1} and a half velocity constant, K_S, of approximately 400 mg/L.[24] For *M. soehngenii* k = 2 day^{-1} and K_S = 20 to 41 mg/L.[76] Figure 2.4 shows a plot of comparative specific activity for these two species as a function of acetate concentration. At low acetate concentrations below about 70 mg/L, *M. soehngenii* would have a competitive advantage. At acetate concentrations greater than about 70 mg/L, *M. barkeri* would have the competitive advantage. This is confirmed by photomicrographs of anaerobic digesters with characteristically low volatile acids levels, which show that highly enriched *M. soehngenii* predominates.[77] In pH Stat systems operated in our laboratory which automatically maintain the acetate level between 2000 to 3000 mg/L, we observed that *M. barkeri* (or *M. mazei*) predominate over *M. soehngenii* after about one month.[29]

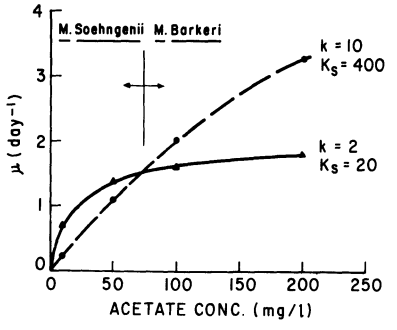


Fig. 2.4: Specific activity of *M. soehngenii* and *M. barkeri* versus acetate concentration.

2.3.7 Fermentative Bacteria

Much of our understanding of the anaerobic digestion process as it occurs in digesters is derived from studies of other anaerobic ecosystems, especially the rumen.[78-81] In addition, much of the work on the acid phase of the domestic anaerobic digestion processes has been done with partially purified wastes, one segment of the population, or with minimally described cultures.[82-85] The acetogenic bacteria, except for a few cases, have not been isolated and characterized.[79,82,86] However, the methanogens have been studied in greatest detail.[1,7,80,82,87]

The nutritional requirements of fermentative bacteria are generally not well defined. Initially, rumen fluid and digester fluid were added to the medium to supply nutritional requirements. Caldwell and Bryant[88] then developed a more defined medium in which rumen fluid was replaced by trypticase, yeast extract, hemin, and volatile fatty acids. The viable count per total microscopic count with the above media was about 10 to 30%.[89] The best recovery was obtained from human feces, in which the bacteria are apparently less fastidious than those of other ecosystems.[90] Mah and Sussman[84] recovered approximately 1 to 10% of the bacteria from a domestic digester with a medium containing glucose, digester fluid, minerals, a reducing agent, and buffer. Kirsch[91] recovered 8.6% of the bacteria using Caldwell and Bryant's medium to which 16% digester fluid was added.

Iannotti et al.[89] developed a habitat-simulating medium for enumeration and isolation of bacteria from a swine waste digester. A roll tube medium with growth factors for strict anaerobes from previously studied anaerobic ecosystems was tried to evaluate the effects of deletion, addition, or level of digester fluid, digester fluid treated with acid or base, rumen fluid, fecal extract, anaerobic pit extract, tissue extract, carbohydrates, peptones, short-chain fatty acids, minerals, vitamins, nitrogen and phosphorous sources, reducing and solidifying agents, buffers, and gases on colony counts. With a medium containing digester fluid, peptones, minerals, cysteine, sodium carbonate, and agar, colony counts were 60% of the microscopic count. These yields represent 2.5 to 20 times greater improvements over previous media compositions.[89]

The level of growth of some of the isolates is still low. The data are indicative of complex interactions within the microbial population. Organisms that initiate the hydrolysis of complex molecules, such as cellulose, supply partial breakdown products to bacteria with less hydrolytic capabilities. The latter organisms supply growth factors to the former specialists. Growth factors are also supplied as the plant material is degraded.[78]

Iannotti et al.[78] have isolated and characterized 130 strains of bacteria from a swine digester. The isolates were divided into 11 groups that included organisms identified as *Peptostreptococcus*, *Peptococcus*, *Eubacterium*, *Lactobacillus*, *Bacteriodes*, and unidentified genera plus miscellaneous facultative and strict anaerobes. The organisms were shown to require mixtures of known factors for growth plus unknown factors in crude extracts such as from digester fluid, swine manure extracts, and rumen fluid. They developed a medium that permitted at least minimum growth for 80% of the isolates. This medium is tabulated in Table 2.4.

Deletion of vitamins, volatile acids, hemin, and tract metals from the medium in Table 2.4 reduces the growth of most strains, indicating that digester fluid does not contain sufficient levels of these factors for maximum growth. When crude extract of swine manure was substituted for digester fluid, the growth of all groups of bacteria was greatly increased. Even though maximum absorbance did not differ significantly, the time required for maximum absorbance was much shorter.[82] The requirement for both known and unknown factors indicates that fermentative bacteria have complex requirements and this limits the ability to define optimum conditions.

McInerney and Bryant[21] reported that a complex mixture—mostly anaerobic but also facultative bacteria—uses the major carbohydrates present in the substrate material as energy sources for growth, while these plus other bacteria utilize the simpler carbohydrate hydrolytic products or end-products of the metabolism of other bacteria, such as lactate or glycerol. They also report that ammonia usually serves as the main nitrogen source and is essential for many bacterial species. Amino acids are required by only a few bacterial species. However, methionine is highly stimulatory or essential and cysteine may also be essential for many strains of *Bacteroides ruminicola*.[92]

Most strains of fermentative anaerobes require one or more B-vitamins for growth.[93] Sulfide often serves as the main sulfur source, although methionine and cysteine are sometimes required, and at least a few species carry out assimilatory or dissimilatory sulfate reductions.[21,94-96] Nutritional interactions among bacterial species are very important in these ecosystems.[21]

2.4 NUTRIENT REQUIREMENTS

Our lack of knowledge of the nutritional requirements of methanogens, particularly nickel, selenite, and molybdate, has hindered the development of the anaerobic digestion process for decades. Even though iron and cobalt requirements were reported on by Speece and McCarty,[97] there is evidence that these elements are still limiting in some studies.[98,99] The first substrates used for the anaerobic digestion

Table 2.4

Medium for defined growth of swine manure digester bacteria[83]

Component	Concentration (mg/L or v/v)
Glucose	300 mg/L
Cellobiose	300 mg/L
Soluble starch	300 mg/L
Pyruvic acid	250 mg/L
Phytone	200 mg/L
Casamino acids	100 mg/L
Yeast extract	100 mg/L
Resazurin	0.1 mg/L
Cysteine - HCl	50 mg/L
Sodium carbonate	800 mg/L
Digester fluid	45 %
Minerals (vol/vol)[a]	3.75 %
Volatile fatty acids - (vol/vol)[b]	0.45 %
$Na_2S.9H_2O$	50 mg/L
Porphyrin compounds (vol/vol)[c]	1 %
Vitamins (vol/vol)[d]	1 %
Trace metals (vol/vol)[e]	1 %

[a] Minerals: The final mineral concentrations in the medium were: K_2HPO_4, 1.3×10^{-3} M; KH_2PO_4, 1.7×10^{-3} M; NaCl, 7.7×10^{-3} M; $(NH_4)_2\ SO_4$, 1.7×10^{-3} M; $MgSO_4$, H_2O, 3.7 × 10−4 M; $CaCl_2.2H_2O$, 2.7 × 10−4 M.

[b] Volatile fatty acids were added to result in a final concentration in the media of: acetic, 1.3×10^{-2}; propionic, 4.0×10^{-3} M; butyric, 2.0×10^{-3} M; isobutyric, 4.1×10^{-4} M; n-valeric, isovaleric and DL - 2 - Methylbutyric, 4.1×10^{-4} M.

[c] Hemin, Vitamin K, and Vitamin K_3 had a final concentration in the media of 1 mg/L.

[d] The concentrations of the vitamins in this stock solution were: thiamin - HCl; calcium pantothenate; nicotinic acid; and riboflavin, all at 2000 μg/L; pyridoxal, 200 μg/L; p-amino benzoic acid, 100 μg/L; biotin 50 μg/L; folic acid, 50 μg/L; and cyanocobalamin 20 μg/L.

[e] Final concentrations of trace metals were: $FeSO_4.7H_2O$, 100 μg/L; $MnCl_2.4H_2O$, 30 μg/L; H_3BO_3, 300 μg/L; $CoCl_2.6H_2O$, 10 μg/L; $NiCl_2.6H_2O$, 20 μg/L; and $Na_2MoO_4.H_2O$, 30 μg/L.

process were mainly domestic wastewater sludges. Such sludges have generally been assumed to contain all of the proper kinds of nutrients and in adequate concentration so as to impose no nutrient limitations on the anaerobic digestion process. However, results in our laboratory indicate there is some preliminary evidence that iron availability may be limiting methanogenic activity even with domestic wastewater sludge as a feedstock.

As the anaerobic digestion process is applied to other feedstocks, the nutritional requirements will demand increasing attention. Some candidate feedstocks for biomass to methane conversion include corn, sugarcane, water hyacinth, and giant

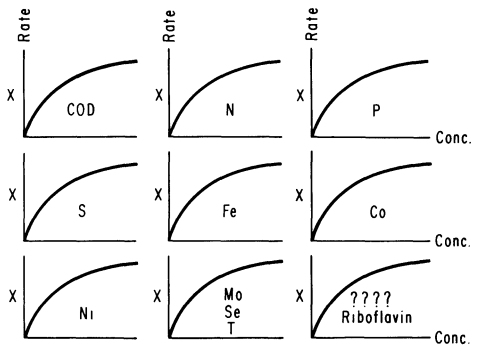


Fig. 2.5: Rate of methanogenesis. The graphs emphasize the multiplicative function of all the nutrients.

kelp. The inherent adequacy of these feedstocks to provide the nutrient requirements for anaerobic digestion is yet to be demonstrated. One trend seems apparent: the diversity of components in these feedstocks is more limited than those in domestic wastewater sludges. It is logical to extrapolate that the required nutrients would also be more limited than in domestic wastewater sludges. This progressive limitation of inherent nutrients is further manifested with many industrial wastewaters that are candidate feedstocks for anaerobic digestion. It presently appears that the ultimate concern for inherent nutrient adequacy must be demonstrated with evaporative condensate streams. These are candidate feedstocks for anaerobic digestion because of their concentrated nature and lack of background chemical elements required for microbial growth.

The microbial regeneration time is a function of the nutrients present. Ideal nutrient concentrations are not essential, but process compensation must be made in either lower loading rates or lower resulting treatment efficiency as a trade-off for nonideality of chemical nutrients present. In addition, it is difficult to say that one nutrient is more important than another because all of the required nutrients are important and must be supplied. This is graphically depicted in Fig. 2.5 which emphasizes the multiplicative nature of all the nutrients. Nutrient limitations at best prolong the minimum generation time of methanogens and at worst can cause generation to eventually cease altogether. The rates of substrate metabolism are also limited by nutrient limitations. Specific substrate utilization rates can be increased severalfold when all required nutrients are in excess. There is a possibility that some of the published kinetic coefficients may reflect nutrient-limiting conditions under which the data were taken, e.g., nickel and iron deficiency. In addition, toxicity response is compounded by nutrient limitations.

Not only do the microbes in the anaerobic digestion process require nutrients, but certain nutrients must be present before other required nutrients will be effective. In the author's experience[100] on nutritional stimulation of methanogens, the nutrients in decreasing order of importance are: nitrogen, sulfur, phosphorous, iron, cobalt, nickel, molybdenum, selenium, riboflavin, and vitamin B_{12}. Methane fermentation will occur without all of the nutrients present but the rates will be

reduced. All of the nutrients of greater importance must be present before a lesser required nutrient will demonstrate stimulation. Without adequate nitrogen, sulfur will not show stimulation and without adequate sulfur, iron will not show stimulation. Ammonium appears to serve as the nitrogen source for all methanogens. Most methanogens utilize sulfide as a sulfur source, but some can utilize cysteine.[12]

Economics will dictate which nutrients can best be added to achieve a stimulation in activity. In the conversion of acetate to methane, approximately 3 to 6 kg of nitrogen is required per 1000 kg of acetate utilized, depending on the SRT.[97] For a completely nitrogen-deficient fatty acid wastewater, if supplemental nitrogen costs \$0.40 per kg (as N) and methane has a value of \$4.50 per 10^6 Btu, the nitrogen supplementation costs would be approximately 2% of the value of the methane produced. The phosphorous requirement is approximately 0.5 kg per 1000 kg of acetate utilized.[97]

For carbohydrate wastewaters, the requirement for nitrogen, phosphorous, and sulfur may be as much as six times greater than for a fatty acid wastewater, due to the increased synthesis of fermentative bacteria. This has significant impact on some nitrogen-deficient biomass feedstocks.

The heavy metals are relatively inexpensive to supplement. Due to the presence (and requirement) of sulfide, most of the heavy metals are precipitated from solution, and it is difficult to determine their actual requirement. Generally for methanogens, iron is added at approximately 10 mg/L; cobalt at 5 mg/L; and nickel, molybdenum, and selenium at 0.1 mg/L. The purified form of riboflavin and vitamin B_{12} would not be supplemented, due to the high cost and because both vitamins are synthesized by many bacteria present in most mixed systems. However, addition of excess sludge heat treatment liquor could conceivably supply such vitamins in restricted cases. Presently, supplementation of these two vitamins is only of academic interest to achieve maximum stimulation rates. Recent studies in our laboratory[29] have demonstrated that rates of acetate fermentation to methane can be stimulated tenfold by proper inclusion of trace metals and yeast extract. The controls (a completely stirred tank reactor—CSTR at 20 days SRT) operate at about 3300 mg $L^{-1}day^{-1}$ of acetate conversion to methane. Rates of 80,000 mg $L^{-1}day^{-1}$ have been achieved[29] at an SRT of 20 days when proper trace elements are included.

If a nutrient limitation exists, it will result in a decreased rate of growth. The search for possible stimulants is therefore very dependent upon the nutrient media that the researcher is using and the situation is further complicated by the ecological interactions which must be taken into account. A compound which is stimulatory to a pure species of bacteria may have no effect on a mixed culture such as found in an anaerobic digester. Mah et al.[101] found the rate of methane production differed in pure culture versus an enriched culture and state:

> The interactions demonstrated between methanogenic and nonmethanogenic species show that methane production in the mineral acetate enrichment is not the function solely of the organism catalyzing the split of acetate to methane and CO_2; it is a function of a community of organisms, each contributing nutrients to the common environment,

> and withdrawing others.... Decomposition of acetate to methane occurs far more rapidly in the community of mixed species than with the sarcina alone. This increase in rate through the interaction of many microbes is the most significant feature of these studies insofar as practical problems of methanogenesis are concerned. Complex organic growth factors are released into the environment common to all the cells, hastening their growth and the rate of substrate disappearance. If provision of increased complex nutrients is the explanation for the increased methanogenesis by the consortium, provision of the appropriate complex nutrient substrates to the enrichment ecosystem should increase the rate of methane formation. The greater rate of methanogenesis by the sarcina in the complex medium as compared to acetate alone is consistent with this view.

One further problem exists with the nutritional study of methanogens. Among the known methanogens, the ability to convert acetate to methane is only observed for *M. barkeri* strains, *M. soehngenii* and *M. mazei*. Some researchers have suggested separating the *methanosarcina* species into a separate family.[1] The implications as far as differing metabolic pathways and different required nutrients are not yet properly understood.

The organic growth factors, which in some cases are required for the methanogens' activity, include: coenzyme M, Factor 420, acetate, 2-methyl butyric acid, vitamins, N-acetyl glucosamine, riboflavin, B_{12}, and possibly some as yet unidentified compounds. Tanner[102] developed a defined medium for *Methanomicrobium mobile*. This organism was found to require acetate, isobutyrate, 2-methylbutyrate, isovalerate, tryptophan, para-aminobenzoic acid, pyridoxine, thiamine, biotin, and vitamin B_{12} for its growth. ATP is also required during the methane formation process.[72,103,104] Several of these compounds are only required by particular species. CoM and F_{420} have been discussed above. Trace amounts of nickel are stimulatory to an enriched culture of acetate utilizing methanogens. Diekert and others[105] found nickel to be a component of F_{430}.

2.4.1 Nitrogen

Generally, nitrogen is required as the major nutrient other than energy source for microbial systems. Speece and McCarty[97] determined the nitrogen requirements for the anaerobic digestion of fatty acids, carbohydrates, and proteins. The following growth equations were derived:

$$\text{amino and fatty acids:} \quad A = 0.054F - 0.03M \tag{2.2}$$

$$\text{glucose and starch:} \quad A = 0.46F - 0.088M \tag{2.3}$$

$$\text{nutrient broth:} \quad A = 0.076F - 0.014M \tag{2.4}$$

The empirical formula for the anaerobic microbial cells was found to be $C_5H_7O_3N$ which has an N/cell ratio of 11% (w/w). At a 5% net cell synthesis ratio, the nitrogen requirement would be 6 kg N/1000 kg of COD or 1 kg of N/60 m^3 of

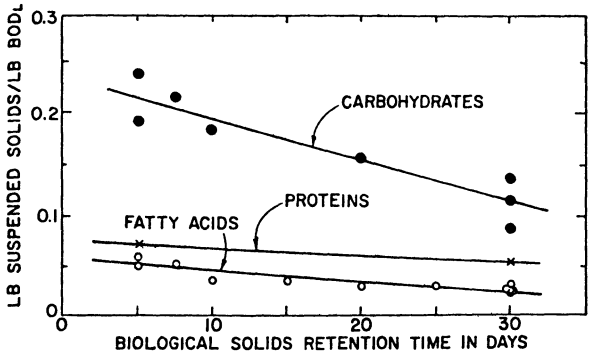


Fig. 2.6: Biological solids production resulting from methane fermentation.

methane produced. The empirical formulation of anaerobic microbial cells was reported to be $C_5H_8O_2N$ by Symons and McKinney[106] while Hoover and Porges[107] reported $C_5H_7O_2N$.

It is noted that microbial synthesis—and thus the nitrogen requirement of carbohydrate digestion—is about six times greater than for proteins and fatty acids, respectively. This is shown in Fig. 2.6. Therefore, the anaerobic digestion of high carbohydrate feedstocks to methane gas deserves special consideration because of the relatively high microbial synthesis to substrate ratio.

The half velocity constant of ammonia nitrogen for methanogens is not reported in the literature, but our preliminary studies indicate it may be as low as 1 mg/L.

2.4.2 Phosphorus

The microbial uptake of phosphorus in anaerobic digestion has been reported to be approximately 1/7 of that for nitrogen.[97] This would yield an empirical microbial formulation of $C_5H_7O_2NP_{.06}$. No data on the half-saturation constant for phosphorus are reported in the literature, but it is assumed to be on the order of 1 mg/L as P or less.

2.4.3 Sulfur

The sulfur requirement for methanogens is not extensively documented. The limited information in the literature makes reference to only two of the many methanogenic-pure cultures and one mixed culture, which had cellulose as the substrate.[108–114] The sulfur requirement is part of a complex picture. On the one hand, the presence of sulfate may inhibit methanogenesis because the sulfate reducers have a lower half-saturation coefficient than the methanogens for substrates such as hydrogen and acetate. On the other hand, the methanogens are dependent on the production of sulfide for growth. It is known that methanogenic bacteria require fully reduced sulfur as a sulfur source.[34,62,115]

Ronnow and Gunnarsson[109] reported that a thermophilic methanogenic bacterium has a specific sulfide requirement for methane production and growth. They indicated that the consumed sulfide was used for producing sulfide compounds taking part in methane production and as a general source of compounds required for growth, e.g., sulfur containing amino acids and proteins.

Scherer and Sahm[116] reported that optimal growth of *M. barkeri* occurred on a defined medium containing methanol when 2.5 to 4.0 mM sodium sulfide was added. This resulted in a measured total soluble sulfide concentration of 0.04 to 0.06 mM. They also reported that iron sulfide, zinc sulfide, or L-methionine could also act as sulfur sources. However, addition of sodium sulfate to a sulfide-depleted media failed to restore growth.

Mountfort and Asher[108] found that at least part of the sulfide requirement for growth is used as a precursor of *CoM*, which is 40% sulfide by weight. Ronnow and Gunnarsson[109] state that 2.6% of the cell mass of *M. thermoautotrophicum* is sulfur. According to their data this would give an empirical cell composition of

$$C_5H_7O_2NP_{.06}S_{0.1}$$

At sulfide levels below 0.1 mM, growth of *M. thermoautotrophicum* was poor and the methane production rate decreased. However, prolonged sulfide starvation did not affect the growth rate of the organism after readdition of sulfide. They noted a linear relation between growth, methane production, and sulfide concentration. They also noted that sulfate or thiosulfate could not replace the sulfide requirement. Only frequent injection of sulfide could sustain rapid growth and continuous methane production. After injection of sulfide, a large increase in methane production was noted within 30 minutes.

Ronnow and Gunnarsson[110] later observed that when sulfide was added in increments of 20 mg/L, the methane production rate increased until sulfide was completely depleted in the medium (as determined by sulfide analysis) and that the methane production rate then decreased. Growth continued to increase during sulfide starvation. However, upon ammonium starvation, growth stopped while the methane production rate was only slightly affected, sulfide still being available. This observation provides further indication of a sulfur compound requirement for high rates of methane formation. They further stated that although methanogenic bacteria contain large amounts of coenzyme M, between 0.3 to 16 μmole/g dry weight,[42] coenzyme M cannot account for a large part of the sulfur absorbed. From a cell yield of 1200 g dry weight/mole of sulfur,[109] one can calculate that the sulfur content is 830 μmole/g dry weight (2.6%). Thus, *CoM* can only account for approximately 4% or less of the total sulfur content of the cells. A large pool of sulfur compounds, possibly low molecular weight compounds, still remains to be discovered.[110] Methionine and cysteine obviously account for a large part of the sulfur.

Gunnarsson and Ronnow also reported that the methane production rate increased until the sulfide was completely depleted. Subsequently, the *ATP* pool increased. In the ammonium-limited experiments, on the other hand, the *ATP* level

increased immediately after ammonium was added to a nitrogen-starved culture. The *ATP* level remained constant until ammonium was fully depleted and then rapidly decreased to almost one-third of the constant level; however, the methane production rate did not fluctuate in this way. They found no clear correlation between *ATP* pool level and the methane production rate in *M. thermoautotrophicum* as was observed for *M. bryantii*.[72] It should be emphasized that the rate of energy production is not related directly to the *ATP* pool level but rather to the turnover of the *ATP* pool. Gunnarsson and Ronnow[110] reported recently that few *ATP* measurements in methanogenic bacteria have been made, making further comparisons with methanogens unfeasible. They found no simple correlation among the methane production rate, the *ATP* concentration in the cell, and growth.

Minimal growth of *M. barkeri* occurred when sulfide in the media was replaced by sulfate or *CoM*.[108] These authors suggested that coupling between growth and energy for methanol fermentation occurred at high sulfide concentrations.

Zehnder[117] noted the effects of various sulfide concentrations on growth and specific methane production rate of *M. arboriphilus* at pH 7.0. Optimal growth and specific rate of methane production required the presence of between 10^{-6} and 10^{-3} M sulfide.

A sulfur source of about 0.85 mM was found to be essential for degradation of cellulose to methane.[113] At 9 mM, all inorganic sulfur compounds other than sulfate inhibited both cellulose degradation and methane formation. The inhibition increased in the following order: $thiosulfate < sulfite < sulfide < H_2S$.

Sulfide was found to act as a sulfur source rather than a reducing agent.[114] The optimum sulfide concentration, 10^{-4}, coincides with the sulfide concentration in the rumen. In the absence of sulfide, cysteine was the only compound which stimulated methanogenesis.

The sulfide requirement can be easily overlooked in anaerobic digestion because there are many sinks for sulfide:

(1) H_2S in the off-gas;

(2) total sulfides in the effluent;

(3) microbial synthesis of sulfide (150% of P synthesis); and

(4) precipitation of sulfides by heavy metals.

One could reasonably hypothesize from these references that about 11.5 mg/L (corresponds to 0.5% H_2S in gas) of un-ionized H_2S is required for optimum growth of methanogens. If it is also assumed for convenience that the digester were operated at a pH of 6.85, equal to the pK of H_2S, then the total sulfide requirement for anaerobic digestion can be calculated as shown in Fig. 2.7.

To maintain 11.5 mg/L of un-ionized H_2S in solution at a digester $pH = pK$ of H_2S would result in a total dissolved sulfide concentraiton of 23 mg/L which would be lost in the effluent of a CSTR. A concentration of 11.5 mg/L of un-ionized H_2S in solution would be in equilibrium with about 0.5% H_2S in the gas phase. The mass of H_2S stripped into the gas phase is related to the volume of

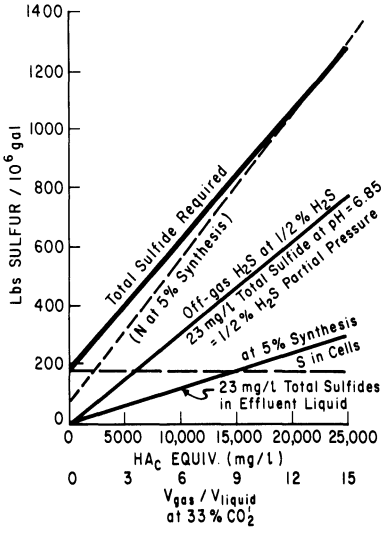


Fig. 2.7: Sulfur requirement for anaerobic digestion. At Y = 0.05, the pH = pK = 6.85. Total sulfides in the digester, 23 mg/L.

digester gas produced per volume of liquid feed to the digester and is shown in Fig. 2.8. The net result of this hypothesis is that the total sulfide requirement for optimum anaerobic digestion essentially is equivalent to the nitrogen requirement if a CSTR configuration of digester is used. This is quite surprising since it is generally considered that the nitrogen requirement for anaerobic digestion is about an order of magnitude higher than any other nutrient other than the energy source. The second surprising observation is that the sulfur concentration of methanogenic cells is 150% greater than that of phosphorous (phosphorous is normally assumed to be second only to nitrogen).

If it is assumed that both the required nutrient form and also the toxic form of sulfide is the un-ionized sulfide form, then Fig. 2.8 can be constructed and the ionized sulfide concentrations can be ignored in our range of interest. The toxic level of total dissolved sulfide in anaerobic digestion has been reported to be 200 to 300 mg/L.[118] In such a case, if the digester pH is assumed to be 6.85, the pK for H_2S, then the toxic concentration of un-ionized H_2S would be 100 to 150 mg/L. Other investigators have reported that approximately 6% H_2S in the head gas is the maximum limit for methanogenesis.[119] This would correspond to 138 mg/L of un-ionized H_2S in solution, which reasonably confirms the results of Lawrence and McCarty.[118]

Consequently, it would be concluded that approximately 0.5% H_2S is required in the head gas for optimum anaerobic digestion—11.5 mg/L un-ionized H_2S in solution. Unfortunately, this value (0.5% or 0.115 kg H_2S/100 m^3) amounts to six times the H_2S allowable by internal combustion engine warranties. If 6% H_2S is the maximum toxic level for methanogenesis, then this illustrates the relatively

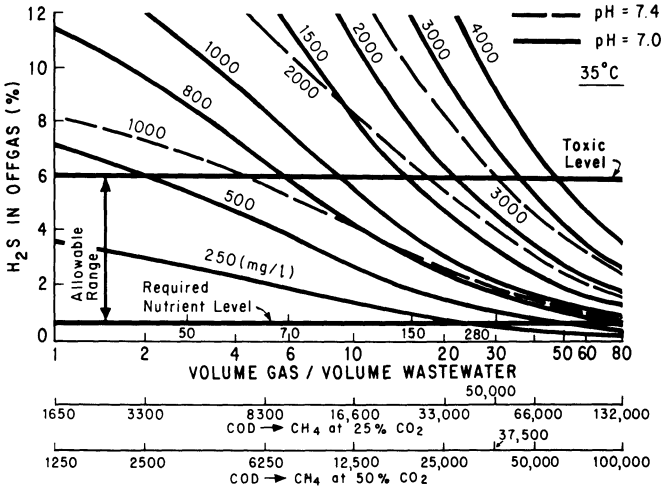


Fig. 2.8: H_2S in offgas versus V_g/V_w (or COD→ CH_4). Numbers indicate sulfide sources in the wastewater (mg/L).

tight ecological niche occupied by methanogens—at minimum desirable levels of approximately 0.5% and a maximum toxic level of 6% H_2S. Few inorganics go from ideal to toxic within such a small range as the twelve-fold concentration hypothesized for sulfide.

2.4.4 Trace Metals

Nickel is generally not essential for growth of bacteria.[58] However, in the last few years it has become evident that the methanogenic bacteria are clearly a unique group in this respect. Nickel is an essential component of bacterial urease and in bacterial enzymes which convert $H_2 - CO_2$ to acetate.

In view of the apparently universal occurrence of F_{430} in methanogens and the inherent nickel requirement of F_{430}, the question arises as to the source of nickel, since it has not generally been included in defined media. The medium described by Balch et al.[1] for growth of most methanogenic bacteria contains 0.2% yeast extract and 0.2% trypticase, but it is not supplemented with nickel salts. Diekert et al.[59] found that the methanogenic bacteria grew readily on this media. However, if the concentration of yeast extract was significantly lowered, good growth was only observed upon addition of nickel. They found that no other transition metal could substitute for nickel.

Yeast extract contains considerable amounts of nickel and significant nickel contamination occurs as a result of contact with stainless steel needles and fittings. Diekert et al.[59] conclude that the high contamination level of nickel in the medium is the reason why the nickel requirement for growth of methanogenic bacteria has long been overlooked.

Mah et al.[32] showed that the ash of yeast extract was approximately as stimulatory as yeast extract itself, indicating that trace minerals, limiting in the defined medium, were supplied by yeast extract. They ruled out the possibility that the trace metal could be cobalt.

Diekert et al.[59] also report that yeast extract contains considerable amounts of nickel. In addition, the mineral salts and trace elements used to make the medium are contaminated with low amounts of nickel.[120] The nickel concentration in the medium prepared by the method of Balch et al.[1] was found to contain between 0.1 and 0.5 nM, even though nickel was not added. Diekert et al.[59] estimated the maximal amount of nickel required for growth of different methanogens to be 250 to 1100 nmoles of nickel per gram (dry weight) of cells. Scherer and Sahm[99] found nickel was required for optimal growth of *M. barkeri*.

Other trace metal requirements have been demonstrated with methanogenic bacteria. Iron, cobalt, molybdenum, selenium, and tungsten have been shown to be stimulatory to methanogens.[97,99,120-122] However, it appears that nickel is essential and that iron, cobalt, or molybdenum cannot replace the nickel requirement for F_{430}.[56] Thus, there is a clear basis for the essential role of nickel as a nutrient requirement in anaerobic digestion.

McCarty and Vath[123] have reported that very high acetate utilization rates were observed when dried supernatant solids from a municipal digester were supplemented in digesters receiving a defined medium of acetate and mineral salts. Conventionally, high-rate digestion is considered to be 3.3 g $L^{-1}day^{-1}$, yet McCarty and Vath observed acetate utilization rates as high as 21.9 g $L^{-1}day^{-1}$ when the digesters were supplemented with dried supernatant solids. McCarty and Vath also stated that previous studies on the digestion of pure volatile acids alone have indicated that only relatively low rates of fermentation were possible. They believed the low rates were not due to an inability of the methane organisms to carry out much higher rates, but rather to improper environmental conditions. They concluded that there appears to be no practical limit to the possible rate of volatile acid fermentation under proper environmental conditions and process control.[123]

Murray and van den Berg[124] reported that the conversion of acetic acid to methane and carbon dioxide by a mixed methanogenic population from an anaerobic fixed-film digester was stimulated by the addition of nickel (100 nM), cobalt (50 nM), and especially by the addition of these elements in combination. Molybdenum addition (50 nM) was only slightly stimulatory when added in combination with both nickel and cobalt. Addition of these trace metals greatly enhanced anaerobic digestion of food processing wastes. Methane production was increased 42% and reactor residence time was shortened.

Speece and McCarty[97] attempted to define the stimulatory components of the dried digester supernatant solids and identified the requirement of the methanogens for iron and cobalt. In addition, some organic compounds such as proline, benzimidazole, and thiamine were also shown to be stimulatory, resulting in acetate utilization rates in excess of 20 g $L^{-1}day^{-1}$. By including iron and cobalt in the medium, it was possible to maintain acetate utilization rates of approximately 2 to 3 g $L^{-1}day^{-1}$.

Hoban and van den Berg[98] reported that addition of iron to a methanogenic culture utilizing acetic acid markedly increased conversion of acetate to methane. The optimum soluble iron concentration was between 0.2 and 2 mM. Most of the iron precipitated mainly as carbonate, to a lesser extent as phosphate, and to an even lesser extent as sulfide salts within days after addition. A large part of the iron was precipitated immediately. Precipitation of iron during the tests was related to the amount of CO_2 produced by the microorganisms. Conversion of acetic acid to methane in liquid from municipal sewage digesters and from laboratory food processing waste digesters was also increased markedly by addition of iron. They concluded that optimization of the conversion of acetic acid to methane in methanogenic fermentations requires soluble iron concentrations which are many times higher than those often required for maximum growth and activity in microbial cultures.[112,125,126] Also they stated that it is conceivable that in practice some of the difficulties often encountered in controlling the digestion process are caused by variations in soluble iron content.

Microscopic observation of trace-metal-deficient, low-acetate concentration cultures maintained in the author's laboratory consistently revealed that the initial inoculum was predominantly similar to the "fat rod" described by Zehnder et al.[117]; whereas, in some cases where the acetate-utililzation rate was stimulated, the culture was predominantly composed of sarcina, typical of *M. barkeri.*

In the absence of nickel, specific acetate-utilization rates were in the ranges of 2 to 4 day^{-1} and did not exceed 4.6 day^{-1}. In the presence of nickel (with no yeast extract), specific acetate-utilization rates of 10 day^{-1} were observed. When yeast extract was supplemented along with nickel, specific acetate-utilization rates as high as 12 to 15 day^{-1} were observed.[29]

Daily phosphate addition appeared to be required to prevent an abrupt "crash" in high acetate-utilization rates and volatile suspended solids (VSS) levels. However, omission of daily phosphate supplementation did not prevent temporary achievement of relatively high acetate-utilization rates. In fact, the highest acetate-utilization rates observed in this study, 51 g $L^{-1}day^{-1}$, were achieved without daily supplementation of phosphate. Phosphate precipitation of iron and/or other trace metals must be considered because, in our laboratory, stimulation of acetate conversion to methane is correlated with low phosphate concentrations.

The methanogens require seemingly contradictory growth conditions. For good growth, a sulfide concentration of 10 to 27 mg/L is required.[108,109] But the sulfide may remove metal ions essential for growth, such as nickel, iron, and cobalt.[122,127] Likewise high rates of gas production tend to strip the essential sulfide from solution.[118] In our experiments the sulfur source was sulfate, which apparently was continually reduced to sulfide. This provided a continual source of sulfide to replenish that stripped by the high gas production rates. In addition, the trace metals nickel, iron, and cobalt were fed in a 50% acetic acid feed with a low pH, keeping the metals in solution. The pH Stat automatically injected 1 to 2 ml of this acidified feed containing the soluble trace metals once per 15 or 30 min when the acetate-utilization rate was 25 g $L^{-1}day^{-1}$.[29]

We believe that even though the sulfide would precipitate the nickel, iron, and cobalt thus leaving very low concentrations of the metals at equilibrium, the equilibrium was frequently disturbed by the acidified feed pulses, making the trace metals momentarily available for uptake by the methanogens. Thus, nutrient limitation in the methanogens could be minimized, and the acetate energy source could be assayed as the limiting nutrient. It is not known if all nutrients were in fact unlimiting, but exceptionally high acetate-utilization rates were observed (in the order of 15 times greater than that which is conventionally considered to be high rate digestion. i.e., 3.3 g $L^{-1}day^{-1}$). This is quite significant since enrichment cultures generally demonstrate lower acetate-utilization rates than normal sludge digesters.

It also appears that nickel, iron, and/or cobalt must be supplemented to achieve high VSS concentrations.[29] With yeast extract supplementation only, the highest VSS observed was 1800 mg/L. When nickel, iron, and cobalt only were supplemented, the highest VSS observed was 3000 mg/L. When nickel and yeast extract both were supplemented, the highest VSS observed was 7000 mg/L.

Very high growth and decay of biomass were observed. On occasion, the biomass doubled in 24 hours, as also was reported by Mah et al.[32] When phosphate was not supplemented daily, biomass was observed to decay just as rapidly. To our knowledge, biomass decay rates of this magnitude have not been reported.

The inherent capability of achieving such high acetate-utilization rates (51 g $L^{-1}day^{-1}$, 20 vol. CH_4· vol. of culture^{-1}· day^{-1}) and associated high specific acetate-utilization rates (15 day^{-1}) appears to be another bit of evidence indicating that the methanogens are indeed a unique biological group. With proper process control and adequate nutrient supplementation, there does not appear to be a practical limit to the possible rates of acetate fermentation. It could be that propionate degradation, however, limits some systems.

2.5 TOXICITY RESPONSE

A common misconception about anaerobic digestion is that the process cannot tolerate toxic materials and that the biota "dies" when exposed to toxicants. This concept arose from municipal sludge digestion experiences in that when methane bacteria were inhibited by a toxicant, the result was that volatile acid production continued, exceeded the buffer capacity of the system, depressed the pH, and led to a "sour" digester. This is a rather complex phenomena in which the original toxicant may play only a minor triggering role. The Chicago Metropolitan Sanitary District has extensive anaerobic digestion facilities and receives a host of toxic materials from the broad spectrum of industrial wastewaters received, but due to careful operational control it has never had a sour digester.[128]

It is true that methanogens and the other organisms in anaerobic digestion can be inhibited by toxicants. Due to the relatively low fraction of substrate synthesized into cells by the methanogens and the associated prolonged generation times, the recovery period can be considerably extended if the toxicant is indeed bacteriocidal. Design safeguards against such a situation are discussed elsewhere. However, studies

Table 2.5

Relative concentration of petrochemicals assayed with acetate enriched methanogenic cultures resulting in a 50% decrease in gas production.[31]

Compound	Concentration Resulting in 50% Acticity (mM)	Compound	Concentration Resulting in 50% Acticity (mM)
1-Cl-Propene	0.1	2-Cl-Propionic acid	8
Nitrobenzene	0.1	Vinyl acetate	8
Acrolein	0.2	Acetaldehyde	10
1-Cl-propane	1.9	Ethyl acetate	11
Formaldehyde	2.4	Acrylic acetate	12
Lauric acid	2.6	Catechol	24
Ethyl benzene	3.2	Phenol	25
Acrylonitrile	4	Aniline	26
3-Cl-1,2-propandiol	6	Resorcinol	29
Crotonaldehyde	6.5	Propanal	90

on toxicity recovery of only the methanogens in our laboratory indicate that most toxicants exhibit a bacteriostatic or reversible effect on the methanogens at the lower concentration ranges normally encountered in industrial wastewaters.[129]

The effect of molecular structure of 52 petrochemicals was assayed with unacclimated, acetate-enriched methanogens.[130] The following structural chacteristics influenced the selection of petrochemicals which were assayed: double bonds, position of functional group, length of carbon chain, branching, oxidation state, type of functional groups, number of identical functional groups, and configuration. The results showed a definite correlation between molecular structure and toxicity to unacclimated methanogens. Chloro substitution, adlehydes, double bonds, and benzene rings exhibited toxicity. The addition of hydroxyl groups and increased carbon chain length decreased toxicity. Table 2.5 indicates the relative toxicity of the petrochemicals which inhibited unacclimated methanogens.

Suspended growth systems of acetate-enriched methanogens which have received a wide variety of toxicants all show a characteristic recovery pattern when a neutral pH is maintained and the acetate concentration is restored daily to the range of 1000 to 2000 mg/L. Figures 2.9 through 2.12 show the response of suspended growth systems to chloroform, cyanide, ammonium, and oxygen. Gas production rate is inhibited for a time proportional to the toxicant concentration. Even after prolonged periods of nil gas production, eventual recovery was very rapid and complete. This indicated that biomass viability was maintained and that the toxicity was bacteriostatic not bacteriocidal.[129]

The reversibility of toxic inhibition was confirmed in two ways.[129] First, the feed to plug flow anaerobic filters was spiked with various toxicants. The plug flow through these filters resulted in almost complete displacement of the feed and thus of

the toxicant spikes in less than two days. Figure 2.13 indicates that when 2000 mg/L of sulfide was added to the feed of a filter for a day, gas production almost ceased. However, after two days gas production was back to normal. The second method involved experiments with suspended growth systems. Very high concentrations of formaldehyde (300, 600, 1800 mg/L) were injected into duplicate systems. One of the duplicates was centrifuged after exposure for one hour. The supernatant was then discarded and replaced with an unadulterated nutrient salt solution. The other duplicate served as a control with no removal of the soluble fraction. Figure 2.14 shows that gas production in the systems exposed for one hour to 300 to 600 mg/L of formaldehyde exhibited only a slight, temporary reduction in gas production. The system exposed to 1800 mg/L of formaldehyde for one hour took about 10 days to recover to 75% of the pretoxicant gas production rate. The gas production ceased in the controls.

2.5.1 Acclimation and Metabolism

There is no correlatioan between the toxicity characteristics of an organic compound to an unacclimated, acetate-enriched methanogenic culture and the eventual metabolism of that same organic by the acclimated culture.[131] Toxicity was interpreted as meaning no metabolism and acclimation was interpreted as meaning the substance was actually metabolized. For example, aldehydes show toxic inhibition to an unacclimated, acetate-enriched methanogenic culture, whereas ketones exhibited no toxicity. However, upon acclimation aldehydes actually degrade more readily than ketones.

The time required for an organic compound to be degraded via methane fermentation depends on the substrate structure and the source of the inoculum.[132] These authors also studied the cross acclimation of potential aromatics formed during the heat treatment of lignin. They suggested that a microbial population acclimated to a particular aromatic substrate can be simultaneously acclimated to other aromatic substrates which have similar arrangements of side group on the aromatic ring. Weng and Jeris[133] reported that glutamic acid-acclimated cultures could utililze acetic, propionic, mesaconic, α-ketoglutaric, and pyruvic acid, converting them to methane and carbon dioxide. Chou et al.[131] assayed the biodegradability of a wide spectrum of petrochemicals by methanogenic cultures enriched on acetate. (Of course these enriched cultures contained some nonmethanogens also.) Table 2.6 shows the petrochemicals which were eventually metabolized by the acetate-enriched, methane-producing cultures.

Figure 2.15 shows the cross-acclimation characteristics of phenol to an acetate-enriched, methane-producing culture. The control received only acetate. After approximately 30 days, the culture started to metabolize the phenol, even though at high concentrations phenol has been utilized as a disinfectant. The same is true with formaldehyde. However, at dilute concentrations both are readily metabolized. In our laboratory, we fed an anaerobic filter a wastewater which contains 1000 mg/L phenol and 5 mg/L cyanide, and the system converted the phenol to methane at a process efficiency in excess of 90% at a hydraulic retention time of 12 hours.

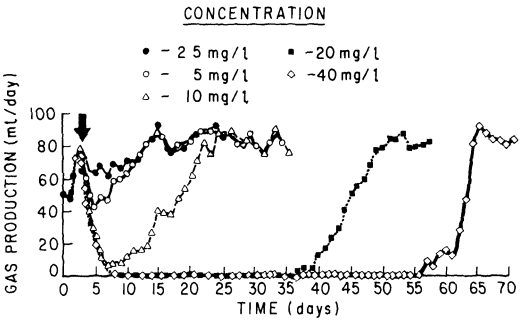


Fig. 2.9: S.R.T. response to toxicant—chloroform 50 days S.R.T. 35°C.

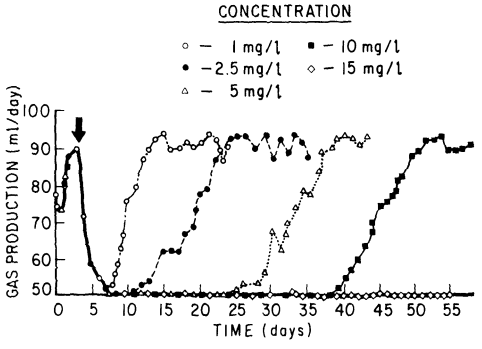


Fig. 2.10: S.R.T. response to toxicant—cyanide 50 days S.R.T. 35°C.

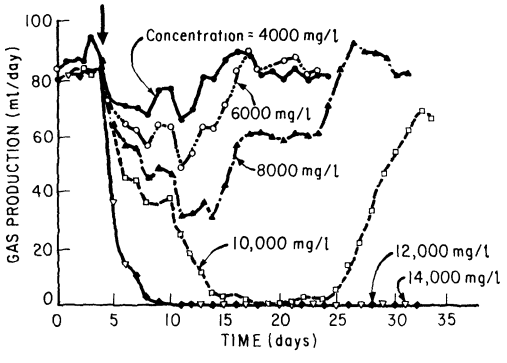


Fig. 2.11: S.R.T. response to toxicant—NH_4^+ 50 days S.R.T. 35°C.

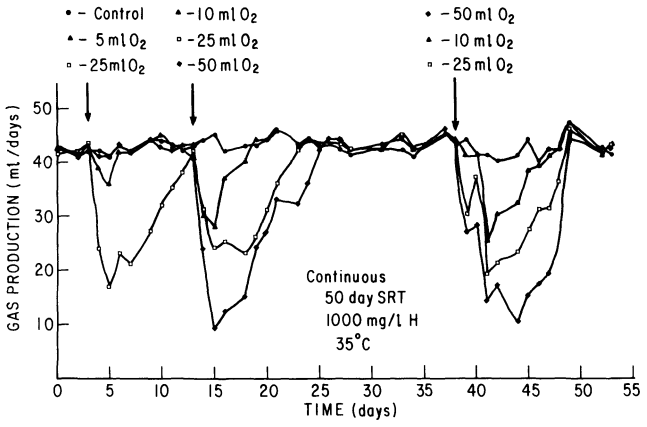


Fig. 2.12: The effect of oxygen on methane fermentation.

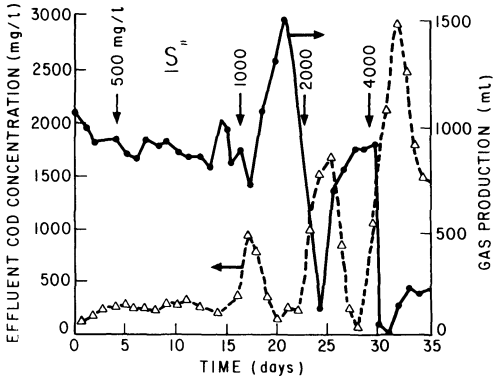


Fig. 2.13: Response of anaerobic filter to toxicant—$S^{=}$.

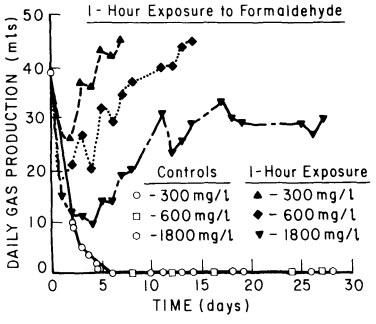


Fig. 2.14: Effect of exposure time on toxicity response: formaldehyde.

Table 2.6

Petrochemicals metabolized by enriched methane cultures[131]

Acetaldehyde	Formaldehyde	Phthalic acid
Acetone	Formic acid	Propanal
Adipic acid	Fumaric acid	Propanol
1-amino-2-propanol	Glutaric acid	2-propanol
4-amino-butyric acid	Glycerol	Propionic acid
Benzoic acid	Hexanoic acid	Propylene glycol
Butanol	Hydroquinone	Resorcinol
Butyraldehyde	Isobutyric acid	Sec-butanol
Butyric acid	Maleic acid	Sec-butanol
Catechol	Methanol	Sorbic acid
Crotonaldehyde	Methyl acetate	Succinic acid
Crotonic acid	Methyl ethyl ketone	Tert-butanol
Ethyl acetate	Nitrobenzene	Valeric acid
Ethyl acrylate	Phenol	Vinyl acetate

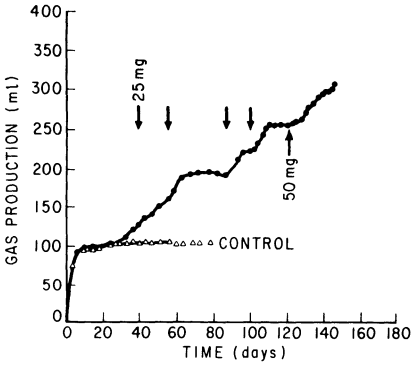


Fig. 2.15: Cross acclimation of acetate culture to phenol.

The structural characteristics of compounds which affect cross acclimation and metabolism were recognized as follows:[131]

(1) Nontoxic aliphatic compounds containing carboxyl, ester, or hydroxyl groups were all readily metabolized by acetate-enriched cultures. Butyric acid, valeric acid, propanol, propylene glycol, ethyl acetate, and vinyl acetate were all metabolized within four days.

(2) Compounds with carbonyl functional groups or double bonds were metabolized within a short period of time (butyraldehyde and sorbic acid required

7 to 10 days to acclimate although these compounds exhibited toxicity to unacclimated, acetate-enriched cultures).

(3) Compounds with amino functional groups showed difficulty in being metabolized. Aniline, 4 amino-butyric acid, sec-butylamine, and 1-amino-2-propanol degraded very slowly.

(4) Dicarboxylic acids required a longer period of time to metabolize than monocarboxylic acids of the same carbon chain length.

(5) The position of the functional group affected the lag period for metabolism, e.g., primary butanol required four days for metabolism, secondary butyl alcohol required fourteen days, and tertiary butyl alcohol was not metabolized in suspended growth systems (but eventually was metabolized in an anaerobic filter).

2.5.2 Microbial Kinetics and Toxicity

Industrial wastewaters commonly contain chronic or transient levels of toxicity. Consequently, a wastewater treatment system receiving industrial wastes would rarely operate at true steady state. For instance, under ideal conditions the microbial regeneration time of methanogens metabolizing acetate would be approximately 2 to 4.2 days.[26,134] However, when a toxicant is present in the wastewater, the specific utilization rate, k, may be reduced (noncompetitive inhibition) or the half-velocity constant, K_S, may be increased (competitive inhibition). The minimum biological SRT, θ_c^{min}, is defined as

$$\theta_c^{min} = \frac{1}{Yk - b} \tag{2.5}$$

where b is the endogenous decay coefficient in unit of day^{-1}.[5] Therefore, anything which decreases the product of the bacterial yield, Y, and k will increase the generation time. Since a toxicant inhibits substrate utilization by definition, it increases the microbial generation time.

The "dynamic" nature of θ_c^{min} with toxicant exposure is demonstrated in Fig. 2.16A. After addition of the toxicant, θ_c^{min} increases with concomitant decrease in methane production and process efficiency as shown in Fig. 2.16B. Although drawn as steady-state response for each condition, the "elastic" nature of θ_c^{min} is clear. Actual treatment systems are operated at a reasonably fixed SRT (θ_c), and thus, temporarily at least, the system may be operating under washout conditions after exposure to toxicants. It is also apparent that failure is a function of time, and thus also whether a substance is considered toxic or nonbiodegradable. A sufficient biological safety factor (defined as θ_c/θ_c^{min}) in the form of a proper design θ_c guards against the above-described failure and allows for acclimation. For example, Chou et al.[131] could not achieve metabolism of acrylic acid in a suspended-growth, acetate-enriched culture with a 25-day hydraulic and solids retention time, because the biomass washed out before acclimation occurred. However, acrylic acid was readily metabolized in an anaerobic filter having a solids retention time in excess of 100 days. Thus the generation time of the culture was apparently in excess of 25 days during

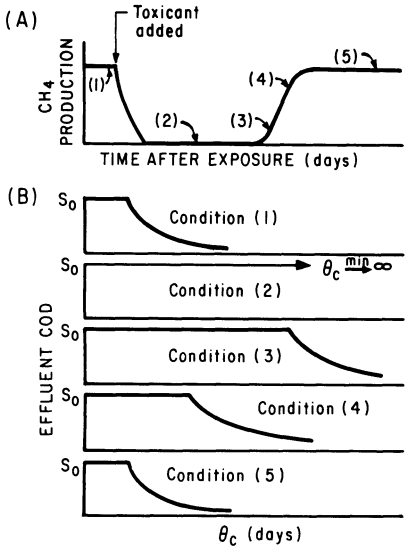


Fig. 2.16: Effect of transient nature of toxicity response on process efficiency.

the acrylic acid acclimation phase. The need for prolonged solids retention times in methane fermentation processes treating industrial wastewaters is not sufficiently recognized.

2.5.3 Toxicity in Anaerobic Digestion Studies

Kugelman and Chin[135] made the following observations:

A review of the literature on anaerobic waste treatment indicates considerable variation in the toxicity reported for most substances. In some cases, the range is over more than one order of magnitude. The major reason for these variations is the complexity of the phenomenon of toxicity. Unfortunately, in most studies of toxicity in anaerobic waste treatment, little or no attention is paid to the effect of antagonism, synergism, acclimation, and complexing reactions. Indeed, in very few studies are the chemical characteristics of the control defined adequately.

The main reasons for these experimental deficiencies are:

(1) Much published toxicity data for anaerobic waste treatment are the product of studies in which toxicity was not to be studied! Rather, toxicity developed during a study of some other phenomenon, and the toxicity data were reported, sometimes as an afterthought, with the other data.

(2) Another main source of toxicity data is from reports on failure of and attempts to revive field digestion systems. These reports were primarily filed by treatment plant operators who were not equipped to run the sophisticated chemical analyses required to categorize adequately the environment in the digester. Neither, for that matter, did they have the time, since coping with a field scale sour digester is a formidable task.

(3) Most studies have been conducted in full strength or diluted sewage sludge, which is an incredibly complex substance whose characteristics vary with geographical location. This procedure has been necessary because it is difficult to develop a chemically defined medium in which methane bacteria will flourish. These organisms will thrive in sewage sludge.

Although there are good reasons why much of the data on toxicity in anaerobic systems are suspect, such data have not and cannot be interpreted properly. At best, most of these data are useless; at worst, they are misleading.

2.6 ALKALINITY

The digestion of complex substrates involves the production of intermediate, volatile acids. It is important to maintain a measure of alkalinity in the system to buffer the pH in the event there is an accumulation of volatile acids. This alkalinity may be generated by the digestion process, as in the case with municipal sludges, or it may need to be purchased in chemical form and supplemented in the feed.

With municipal sludges there is a significant protein fraction. Raw sludge is proteinaceous and waste-activated sludge is of microbial origin with a high protein fraction. Microbial digestion of the protein breaks down the organic nitrogen to the end product, ammonia. This ammonia then combines with carbonic acid in solution to provide an ammonium bicarbonate buffer

$$NH_4OH + H_2CO_3 \longrightarrow NH_4HCO_3 + H_2O \tag{2.6}$$

Thus, there is an increase in alkalinity as raw domestic sludge is digested. The concentraiton of alkalinity formed in a municipal digester is mainly related to the protein concentration in the feed sludge. As an example, a municipal sludge of 4% solids resulted in an alkalinity of 2300 mg/L in the digester, while an 8% sludge resulted in an alkalinity of 4300 mg/L. The respective ammonia nitrogen concentrations were 500 and 980 mg/L.

Early attempts to digest blood anaerobically from packing plants met with failure due to the build-up of excessive ammonium bicarbonate alkalinity levels, which increased the pH. The combination of elevated pH and high ammonia nitrogen resulted in digestion failure, due to excessive free ammonia levels. This is one of the few cases where HCl (not H_2SO_4) should be added to a digester to remedy process

failure. It should be noted that neutral salts, such as NH_4Cl, added to a digester, exhibit little or no buffer capacity and will not supplement the alkalinity.

Other sources of alkalinity in a municipal digester are soaps and other salts of organic acids, especially weak acids such as fatty acids. (Here again, microbial digestion of the soap or organic salt yields an end product cation which combines with carbonic acid.) The reaction for sodium acetate would be as follows:

$$CH_3COONa + H_2O \longrightarrow CH_4 + NaHCO_3 \tag{2.7}$$

Thus, the digestion of sodium acetate would produce alkalinity as an end product and no CO_2, whereas digestion of acetic acid would not produce alkalinity but would produce CO_2. Under some conditions, excessive alkalinity and cation toxicity may result from digestion of excessively high concentrations of salts of organic acids.

Carbohydrates, alcohols, aldehydes, and ketones have no end product cation produced after degradation to methane and CO_2. Consequently, no alkalinity is produced in the digestion process. This is another major consideration in the anaerobic digestion of higher carbohydrate biomass feedstocks. The relatively high microbial synthesis rate during the catabolism of carbohydrates in anaerobic digestion could result in a depletion of alkalinity, if ammonium bicarbonate were present, and if the microbes consumed it as their nitrogen source for biological synthesis.

2.6.1 Alkalinity Concentration Required

The concentration of alkalinity required in an anaerobic digestion process is dependent on many factors, including:

(1) the alkalinity generated from the waste feed;

(2) the necessity for purchase of alkalinity;

(3) the organic concentration in the feed;

(4) the degree of process control;

(5) the availability of effluent recycle; and

(6) the rate of conversion of substrate to volatile acids.

Domestic sludge digesters receive a feed sludge of about 2 to 8% solids which are normally 75% volatile. The chemical oxygen demand (COD) of these feed sludges ranges from about 25,000 to 100,000 mg/L. The alkalinity range for domestic sludge digesters is approximately 1500 to 6000 mg/L; this is an alkalinity/COD ratio of about 1/16. The higher the alkalinity, the higher the volatile acids concentration which can be buffered without a serious pH depression. Therefore, from an operational viewpoint, it is desirable to maintain higher alkalinity levels when the alkalinity is generated from the waste sludge feed.

If the alkalinity must be purchased as a bulk chemical for treatment of industrial wastewater or biomass feedstock, then economics dictate that its use be minimized. The cost of purchased alkalinity can be significant and conceivably

may offset the expected credit for methane. Costs of alkalinity purchased in various forms are discussed later. It would be wasteful to supplement excess alkalinity, and the process does not require it for stability. In addition, high alkalinity results in undesirably high total dissolved inorganic salts in the effluent.

However, the pertinent question for biomass feedstocks is to ascertain the minimum acceptable alkalinity requirement. This alkalinity requirement is related to the biomass feedstock strength, which determines the potential for volatile acids generation. With reasonable process control, the alkalinity requirement can be a small fraction of the influent COD. As the degree of process control is relaxed, increased alkalinity is required to buffer possible higher excursions of volatile acids. A well-operating domestic sludge digester may have volatile acid concentrations of 100 to 300 mg/L, which would require very little alkalinity to maintain a neutral pH.

A well-balanced digester converts the volatile acids to methane as rapidly as they are formed, minimizing the need for alkalinity. In the long run, for biomass feedstocks, it may be less costly to provide more careful process control than to pay the added chemical cost for maintaining excess alkalinity levels to buffer occasional volatile acids upsets. Automatic feed diversion should be one process control option. If odor generation is not excessive and the feed can be temporarily diverted to a holding pond or parallel digester, this very effective means of process control can be maintained. In any case, it is not necessary to supplement continuously high levels of alkalinity when a process control loop could automatically add it at the level that is actually required. Alternatively, the process control loop automatically could stop feed input. Removal of CO_2 from the off-gas and recycling the scrubbed gas back into the digester could also reduce alkalinity requirements and is economically attractive.

The highest concentration of alkalinity is in the effluent, where it is least needed. Effluent recycle can be used to reduce or possibly even eliminate the need to provide supplemental alkalinity or inorganic nutrients. This is particularly true of waste streams high in organic acids. There is no need to neutralize such a waste stream prior to its introduction into an anaerobic filter if the effluent is recycled and if it has sufficient alkalinity generation potential.

2.6.2 Choice of Purchased Alkalinity

There are four main chemical sources of purchased alkalinity—lime, ammonia, sodium bicarbonate, and sodium carbonate. Each has advantages and disadvantages, and the final selection may be quite site specific. Urea is also a possible source of alkalinity, but it may give rise to ammonia toxicity.

Lime is usually the cheapest source of alkalinity and costs approximately $60 per ton in bulk lots. This price applies to both slaked lime $[Ca(OH)_2]$ or quick lime (CaO). Slaked lime costs about one-third more on an equivalent weight basis, but it avoids the need for an on-site slaking machine operation. An operational problem with lime is that both $Ca(OH)_2$ and $Ca(HCO_3)_2$ are rather insoluble and their use can be inefficient. Scaling of pipes and heat exchanger surfaces may be a problem.

Carbon dioxide is a reactant in converting slaked lime to calcium bicarbonate. This reaction can generate a vacuum in a closed digester. On the other hand, if insufficient carbon dioxide is available to react fully with the lime, the pH easily can overshoot neutrality and end up being too high. An excessively high pH level can be just as detrimental as an excessively low pH.

With lime assumed to cost \$60 per ton of CaO and methane assumed to be valued at \$4.50 per 10^6 Btu, the cost of lime supplemental addition alone to produce 1500 mg/L of alkalinity as $CaCO_3$ would just about equal the value of the methane produced from the conversion of 1000 mg/L of COD.

Sodium bicarbonate is a very convenient source of alkalinity. It is easy to handle, quite soluble, and unlike lime, requires no CO_2. In addition, it does not drive the pH excessively high if overdosed. Sodium bicarbonate is relatively more expensive than lime, costing approximately \$300 per ton in bulk. Using the above assumptions, addition of 180 mg/L of $NaHCO_3$ (112 mg/L as $CaCO_3$) for each 1000 mg/L of COD converted to methane represents the break-even point when the value of the methane produced equals the cost of the sodium bicarbonate supplementation.

Sodium carbonate (soda ash) costs about \$180 per ton. The break-even point where supplemental soda ash addition costs exactly offset the value of methane produced is about 300 mg/L of soda ash (as $CaCO_3$) per 1000 mg/L COD converted to methane.

Ammonia as an economic alkalinity source is very site specific. In addition, caution must be exercised to prevent ammonia toxicity.

Alkalinity supplementation is a major economic consideration in the treatment of some biomass feedstocks and industrial wastewaters which generate very little alkalinity during fermentation, e.g., paper mill condensate.[136] Process control feedback with equalization is, therefore, the most cost effective means to minimize the economic burden of alkalinity supplementation.

2.7 MODELING OF THE ANAEROBIC DIGESTION PROCESS

2.7.1 Cellular Synthesis of Organic Substrate

Cellular synthesis is the means whereby the microbial mass is generated and maintained for biological wastewater treatment. Typical cell yields for aerobic processes are 0.4 kg VSS/kg COD removed.[137] This high value results in the generation of considerable quantities of excess sludge which are a waste to the system and must be disposed of.

In anaerobic digestion, cell synthesis is not directly related to COD removed, but is related to the class of organics being digested.[97] In the breakdown of organics, the chemoheterotrophs (Stages I and II in Fig. 2.1) convert the more complex organics mainly the acetate, CO_2, and H_2, which are then converted to methane by the methanogens. The carbohydrate class of organics provides relatively more energy to these organisms than proteins, a fact which is reflected in significantly high synthesis rates. The synthesis rates are as follows:[138]

volatile fatty acids – 0.054 $g \cdot g^{-1}$ dry wt. substrate

proteins – 0.076 $g \cdot g^{-1}$ dry wt. substrate

carbohydrates – 0.24 $g \cdot g^{-1}$ dry wt. substrate

Consequently, specific provision must be made for the disposal of considerably higher excess sludge production volumes in the anaerobic treatment of high carbohydrate-type biomass feedstocks and industrial wastewaters than for protein or simpler volatile fatty acid substrates.

2.7.2 Rate-Limiting Step in Anaerobic Digestion

2.7.2.1 Lipids and Acetate. The rate-limiting step in anaerobic digestion depends on the loading rate. the characteristics of the organic component being treated, and the predominant bacterial populations in the system. In the anaerobic digestion of municipal wastewater and refuse sludges, hydrolysis of the complex volatile solids is rate-limiting at conventional loading rates.[12,139] The lipid fraction of domestic sludges was found by O'Rourke[140] to degrade at the slowest rate. As the loading rate is increased and the SRT is decreased, or both, eventually the conversion of acetate to methane becomes rate-limiting. This is evidenced by a build-up in the volatile acids concentration.

With simpler organics, such as the short-chain compounds found in petrochemical wastewaters, little or no hydrolysis step is required; only the fermentation stage is needed. However, these organic acids, alcohols, aldehydes, ketones, etc., must be converted to acetate, H_2, and CO_2, and this fermentation step may be rate-limiting. Butyrate, propionate, longer-chained fatty acids, and benzoate require specific organisms to convert them to acetate, H_2, and CO_2.[21,141] Increased concentrations of propionic acid have been noted in digesters which were operating poorly; there is speculation that propionic acid is responsible for the poor digestion. The obligate H_2-producing (proton-reducing) acetogens will not be active unless H_2-using methanogens are quite active. It has also been suggested that un-ionized acetic acid is toxic to the methanogens above certain concentrations.[142]

Acetate conversion to methane may be rate-limiting for a number of reasons. We have postulated that the temperature dependence of methanogenesis is stronger than the hydrolysis of acetogenic stages.[143] Thus, a decrease in temperature would not reduce the rate of acetate production as severely as it would reduce the rate of acetate conversion to methane. This would lead to a rise in volatile acids concentration, indicating that acetate conversion to methane had become rate-limiting. Reduced temperature is a common cause of elevated volatile acids concentrations.

Another factor which may cause acetate conversion to methane to become rate-limiting may be the absence of required nutrients, particularly a lack of iron. This was discussed in more detail above.

The presence of toxicity will generally decrease the rate of activity of the various microbial species. However, if the toxicity decreases the rate of activity of the methanogens relatively more than it affects the fermentative and acetogenic organisms. then a build-up in volatile acids will occur. Conversely, if the toxicity

depresses the activity of the fermentative and acetogenic organisms relatively more than the methanogens, then a decrease in volatile acids will be noted. There is general speculation that the methanogens are more sensitive than the fermentative organisms, but exceptions are noted in the literature.[144]

2.7.2.2 Cellulose. Cellulose, in its various natural forms, represents the largest potential biomass feedstock for conversion to methane. Cellulose composes a large fraction of crop residues, e.g., corn stover, straws, and wood. It also comprises a large proportion of cattle manures and urban solid wastes. The complex structure of natural cellulose and the lignin sheath encapsulating it make these biomass feedstocks difficult to degrade microbially. Consequently, cellulose hydrolysis to simpler molecules is often the overall rate-limiting step in anaerobic digestion.

Hobson, Bousfield, and Summers[145] have recently authored an authoritative overview of the subject of anaerobic digestion of cellulose in their book, *Methane Production from Agricultural and Domestic Wastes.* The following is a summary of their coverage of this topic. They point out that polymeric carbohydrates vary in physical structure and solubility as well as in the simple sugars which compose them. As a result, a basic problem arises in the hydrolytic action needed to degrade the polymers into simple sugars in a digester for absorption, utilization, and conversion to methane.

Monomers, the linkages between them, and the structural form of polysaccharides are of great importance in the hydrolysis step of digestion, for the latter may be the rate-limiting problem in carbohydrate fermentation. Bacteria can assimilate and metabolilze glucose in mono- and disaccharide forms relatively rapidly, enabling the bacteria to double in size and divide in 20 to 30 minutes. But polymer glucose is available to the bacteria only at a rate making division possible once every 24 or even 48 hours.

Different enzymes are required for hydrolysis depending upon the polysaccharide monomers and the way they are linked together. The reason that few bacteria can utilize glucose polymer cellulose is that its physical form is characterized by long, straight, β-linked chains lying close together in bundles. Acid treatment, however, can break up the regular structure of cellulose fibers by breaking interglucose linkages. This allows the enzymes more access to the glucose chains and forms more chain ends fo enzymatic hydrolysis.[145]

They conclude that the rate of hydrolysis of the main feedstock carbohydrate in a digester influences the rate of growth and metabolism of a large part of the digester flora. They state that if carbohydrate is the main source of fermentation acids and hydrogen, then the rate of growth of the methanogenic bacteria can in turn be governed by the rate of metabolism of the carbohydrate.

Cellulase, the enzyme which hydrolyzes cellulose, is found in varying amounts in bacteria, imparting to the organism varying ability to attack different forms of cellulose. "Cellulolytic" bacteria present in the rumen of cattle are probably also present in digesters, but this aspect of cellulose fermentation remains to be studied in depth. However, according to Hobson et al.[145] electron micrographs have shown that many different kinds of bacteria are present in, on, and around plant fibers

in the rumen. They also state that the bacteria involved have differing hydrolytic properties and, thus, will attack different cellulosic structures.

The rate at which cellulose in a digester can be degraded depends upon its physical form. Farm animal manures present a more difficult fermentation problem than human feces. Domestic sewage cellulose has already been partially broken down through cooking and other preparation which the food has received. In addition, the processes which convert wood pulp into toilet paper found in domestic wastewater enhance the biodegradability of the cellulose. Conversely, crop residues and animal manures contain virtually untreated cellulose incorporating waxes, lignin, and silica.[146]

Lignin is a very complex compound and its structure and defense against anaerobic digestion is still not understood. Since the amount increases as plants grow older, lignin acts as an increasingly significant barrier to bacterial attack on plant cellulose. Thus, well-dried grass found in the rumen of cattle is not easily degradable and remains so even in the excreta. Roof-thatching demonstrates the permanence of straw, the hardest lignified plant material. Its use as barn bedding is also common. In the rumen of cattle, straw is only half digested and very little further degradation of the manure form is possible.[145]

Ghosh and Klass[147] found that municipal sewage sludge could be partially hydrolyzed and fermented to acids in 0.5 to 1.0 day and that the conversion of acids to methane required a retention time of 6.5 days. A retention time of 1.7 days was suggested for cellulose fermentation; however, the type referred to was purified cellulose capable of being nearly 100% digested. Hog waste cellulose was found by Hobson[148] to be only 41% digested during a retention time of 10 days. Barley straw added to the hog excreta was 35% fermented in 20 days.[146,149] Pfeffer[150] also reported on wheat straw in a thermophilic digester at 59°C, stating that about 38% degradation occurred in 13.7 days retention time. He believed that even with infinite retention time, the straw would not degrade beyond 50%.

Conversion of acids to methane can be rate-limiting in the digestion of sewage sludge and purified cellulose, while the rate-limiting step in straw digestion is the degradation of the cellulosic substrate. Robbins et al.[151] tested the influence of lignin on cellulose degradation and methane gas production. They found that delignified straw was almost entirely digestible whereas untreated straw was only about 32% degradable. Healy and Young[132] established that a number of compounds which make up the lignin polymer could be degraded to methane and carbon dioxide by adding domestic sludge digester bacteria. Hungate[152] found no cellulolytic bacteria in domestic sewage sludge but did find 0.8 to $2.0 \cdot 10^3$ cellulolytic bacteria per ml of digesting sludge. Maki[153] later found $1.6 \cdot 10^4$ to $9.7 \cdot 10^5$ per ml in sewage digesters. Hobson and Shaw[154] measured the cellulolytic level in hog waste and in sludge from a mesophilic hog waste digester, finding none in the former but $4 \cdot 10^3$ per ml present in the latter.

Hobson and Shaw[154,155] predicate that hemicellulose does not require the complex enzyme system of the celluloses for hydrolysis. They demonstrated through anaerobic studies of hog waste that starch is much more easily digested than cellulose. Torien[156] also supported the suggestion that starch-degrading bacteria

are fairly common in domestic sludge digesters even though analyses indicate virtual absence of starch. Hobson et al.[157] also found bacteria fermenting glycerol, formed by lipid hydrolysis, in hog waste digesters.

Thus, the natural cellulosics (excepting the alginates of kelp) appear to be less than 50% degradable in the solids retention times normally used in anaerobic digestion. Aging of the plant materials and predigestion in the rumen decreases the anaerobic biodegradability even more. Chemical, enzymatic, or physical hydrolysis enhances biodegradability of natural cellulosics. Since natural cellulosics comprise such a large, potential biomass feedstock and since these complex, polymeric, carbohydrates often control the overall rate of anaerobic digestion, much additional effort needs to be expended to realize their vast potential for biomass to methane conversion.

2.7.3 Kinetics

Lawrence and McCarty[134] were among the first to attempt to model the methane fermentation phase of anaerobic digestion. They reasoned that in a complex multistage process such as anaerobic treatment, the kinetics of the slowest step would govern the overall kinetics of waste stabilization. O'Rourke[140] concluded that the degradation of long-chain and volatile fatty acids was in fact the rate-controlling step in the anaerobic treatment of complex wastes. Thus, they modeled the methane fermentation of mainly acetate, propionate, and butyrate. They based their model on the Monod expression

$$\frac{dF}{dt} = \frac{kXS}{K_S + S} \tag{2.8}$$

where F = substrate concentration (mg/L); S = volatile acids concentration in the reactor (mg/L); k = maximum rate of volatile acids utilization per unit biomass ($mg \cdot mg^{-1} day^{-1}$; K_S = half velocity constant equal to volatile acids concentration when dF/dt equals $1/2\ k$ (mg/L); and X = organism concentration (mg/L).

They also reported that the average values of substrate utilization coefficients for acetic, propionic, and butyric acid in a CSTR system at 35°C were:

	k ($mg \cdot mg^{-1} \cdot day^{-1}$)	K_S (mg/L)
Acetic	8.1	154
Propionic	9.6	32
Butyric	15.6	5

O'Rourke[140] investigated the digestion of complex sludges and he summarized the kinetic coefficients developed for lipid material as follows:

Temperature °0	K_S (mg/L as COD)	k (day^{-1})
20	4620	3.85
25	3720	4.65
30	2000	6.67

He calculated the combined Michaelis-Menten coefficient, K_c, for prediction of effluent quality as the sum of K_S for acetate, propionate, and lipids (long-chain fatty acids). At 35°C, this was $164 + 71 + 2000 = 2235$ mg/L.

K_c is temperature-dependent and varies as follows:

$$(K_c)_T = (K_c)_{35°C} \quad 10^{0.046(35-T)} \tag{2.9}$$

He also proposed that the kinetics of waste treatment should be based only on the degradable fraction of the various waste components. The total effluent COD, w, from a system treating a complex waste would then be the sum of the nondegradable and degradable components in the effluent as follows:

$$w = \frac{K_c[1 + b(SRT)]}{SRT(Yk - b) - 1} + R \tag{2.10}$$

where b = microbial decay coefficient (day^{-1}), e.g., a typical value is 0.03 day^{-1}; Y = microbial yield coefficient, e.g., a typical value is 0.04; k = maximum substrate utilization (day^{-1}), e.g., a typical value is 6.67 day^{-1}; and R = influent nondegradable COD (mg/L).

The temperature dependence of the maximum utilization rate for complex wastes was found by O'Rourke to be

$$(k)_T = 6.67 \cdot 10^{(-0.015)(35-T)} \qquad (day^{-1}) \tag{2.11}$$

With a complex waste which contained 50% lipids, he calculated the following effluent degradable COD versus SRTs at 35°C:

SRT days	Effluent Degradable COD (mg/L)
3.75	washout
5.0	10.1
7.5	2.78
10.0	1.69
20.0	0.72
60.0	0.30

From this he concluded that significant reduction in the effluent COD beyond 7.5 days SRT are minor and that the operating safety factor to be applied would be site specific.

Gossett and Belser[158] evaluated and modeled the effect of SRT in the activated sludge process, SRT^{AS}, on the anaerobic digestibility of the waste-activated sludges produced. They found that the mass of waste-activated sludge per day increases as the activated sludge SRT decreases, but this does not allow prediction of the net effect on sludge treatment costs. In order to predict the dependence of anaerobic digester performance upon activated sludge SRT, they developed two separate models. The first model related the ultimate anaerobic biodegradability of activated sludge to its SRT^{AS}. The second was a kinetic model for the anaerobic digestion of activated sludge.

They found that the fractional biodegradability of waste-activated sludge, f_d^{WAS}, could be expressed as

$$f_d^{WAS} = \frac{f_d}{1 + (1 - f_d) b^{AS} SRT^{AS}} \tag{2.12}$$

where f_d = net biodegradable fraction of active biomass volatile suspended solids (VSS); b^{AS} = specific decay rate of active biomass VSS (day^{-1}); and SRT^{AS} = activated sludge SRT (day).

Consequently, the fractional biodegradability of waste-activated sludge depends only on the fractional anaerobic biodegradability of active biomass, its specific decay rate, and the SRT of the activated sludge process. It is also noted that f_d represents the biodegradability of active biomass VSS from the point of view of the activated sludge organisms themselves, i.e., aerobic endogenous respiration. Gossett and Belser's model assumed that the same fraction which was aerobically biodegradable would also be degradable anaerobically, and this assumption was confirmed by their experimental results. Their results also showed that $f_d = 0.683$ and b^{AS}=0.0956 day^{-1} fit the data well and demonstrated that the above equation adequately predicted ultimate anaerobic digestibility of waste activated sludge. Gossett and Belser summarized their data in Table 2.7. These results were compared with O'Rourke's model developed for raw sludge which assumed methane fermentation of volatile acids was rate-limiting, regardless of the nature of complex sludge originally fed. O'Rourke's model significantly overestimated the performance of anaerobic digesters fed waste-activated sludge, unless K_c was increased to 4.48 g/L. Gossett and Belser felt there was no fundamental reason for altering this parameter.

Instead, Gossett and Belser[158] postulated that anaerobic digestion of waste-activated sludge was rate-limited by enclosure within viable cells and required that the potentially degradable portion of viable activated sludge organism must first be converted to available substrate by autolysis or endogenous fermentation decay or hydrolysis, or all three. They then proposed a first order "availability rate," dX_a^{DIG}/dt, with respect to the digester concentration of active organisms, X_a, as follows:

$$\frac{dX_a^{DIG}}{dt} = -\beta X_a^{DIG} \tag{2.13}$$

Table 2.7

Anaerobic digestion of waste activated sludge–summary of results[158]

	SRT^{AS} (days)					
	5	10	15	20	25	30
Influent						
COD (g/L)	9.91	13.7	13.8	15.9	15.2	13.9
VSS (g/L)	9.36	11.02	10.95	12.67	12.17	11.19
Effluent						
COD (g/L)	—	10.2	10.9	11.6	12.1	11.6
VSS (g/L)	6.49	8.08	8.81	9.47	9.90	9.61
Destruction						
COD (g/L)	—	25.5	21.0	27.0	20.4	16.5
VSS (g/L)	30.6	26.7	19.5	25.3	18.6	14.1
ml CH_4/g COD fed	137	107	88	102	82	79

where β = specific rate of availability (day^{-1}) = (0.22 day^{-1}).

The parameter, β, would be the sum of a variety of coefficients, including the applicable specific lysis rate and the specific decay rate via endogenous fermentation. The effect on the pool of available digester substrate formed would be given by

$$\left(\frac{dS^{DIG}}{dt}\right)_{\text{formed}} = 1.42 f_d \beta X_a^{DIG} \tag{2.14}$$

Finally these authors proposed a general model which would predict the anaerobic digestion performance of combined primary and waste activated sludges. For their special case where no primary sludge was fed and no VSS were in the activated sludge feed, their generalized equation is as follows:

$$\Delta BOD_{\text{processed}} = COD_{in} \frac{f_d}{1 + (1 - f_d) b^{AS} SRT^{AS}} \frac{\beta \cdot SRT^{DIG}}{1 + \beta \cdot SRT^{DIG}} - \frac{K_c(1 + b^{DIG} \cdot SRT^{DIG})}{SRT^{DIG}(Y_a k_a - b^{DIG}) - 1} \tag{2.15}$$

where $SRT^{DIG} = SRT$ in digester (days); f_d = net biodegradale fraction of active (viable) biomass VSS; b^{AS} = specific decay rate of activated sludge (day^{-1});

b^{DIG} = specific decay rate of methane bacteria (day^{-1}); Y_a = yield coefficient for acetate-using methane bacteria ($gVSS \cdot gBOD^{-1}$); and k_a = maximum specific substrate utilization rate for acetate-using methane bacteria ($g\ BOD \cdot VSS^{-1} \cdot day^{-1}$). The same authors concluded their study with an economic analysis of the relative effect of SRT^{AS} on digester costs and methane production. They concluded that the additional methane production credit at low SRT^{AS} offsets the costs for additional required digester volume. Consequently, total anaerobic digester costs were independent of SRT^{AS}. They also observed that SRT^{AS} caused little change in quantity of digested sludge solids for ultimate disposal.

Chen and Hashimoto[159] worked with complex sludges and developed a model which incorporated the provision for toxicity. Their model is as follows:

$$\gamma_v = \frac{B_o S_o}{\theta}\left(1 - \frac{K}{\theta\mu_m - 1 + K}\right) \tag{2.16}$$

where γ_v = volumetric methane production rate ($m^3\ CH_4 \cdot m^{-3}$ reactor$\cdot day^{-1}$); B_o = ultimate methane yield ($m^3\ CH_4 \cdot kg\ VS^{-1}$); θ = solids retention time (day); S_o = influent VS ($kg \cdot m^{-3}$); μ_m = maximum specific growth rate (day^{-1}); K = kinetic parameter (dimensionless); and where μ_m increases with temperature and K increases with toxicity.

Andrews and Graef[142] developed a dynamic model to simulate the anaerobic digestion process and included an inhibition function to relate the volatile acids concentration and specific growth rate of methanogens. They assumed the un-ionized acid as the growth-limiting substrate and the inhibiting agent. They incorporate the interactions which occur in and between the liquid, gas, and biological phases of the digester and also assumed valid Monod kinetics.

Carr and O'Donnell[160] evaluated Andrews and Graef's model with experimental results from a laboratory digester which they subjected to sudden increases in loading. By modifying the values assumed for some of the coefficients, they obtained good agreement with Andrews and Graef's model.

Olesczkiewicz[161] used the following equation for organic removal in anaerobic treatment:

$$\frac{S_e}{S_o} = \exp^{-K/L} \tag{2.17}$$

where S_e = effluent concentration ($kg\ COD \cdot m^{-3}$); S_o = influent concentration ($kg\ COD \cdot m^{-3}$); L = loading rate ($kg\ COD \cdot m^{-3} \cdot day^{-1}$); and K = empirically derived coefficient (not the same K as in Eq. (2.15)).

Preliminary modeling of the response of methane fermentation to toxicity has begun.[135,162-165] A wide spectrum of toxicants has been assayed for effect on methane fermentation, e.g., cations, bisulfite, phenol, nickel, ammonium, sodium, cyanide, chloroform, gasoline, and formaldehyde. The general response pattern has been similar to the spoon-shaped, dissolved oxygen sag curve in a stream, regardless of the specific toxicant.[165]

2.7.4 Stoichiometry

The chemical composition of the feedstock determines the resulting off-gas composition. Organic matter destruction is directly related to CH_4 production. Buswell and Mueller[166] developed the following equation to predict off-gas composition:

$$C_nH_aO_b + \left(n - \frac{a}{b} - \frac{b}{2}\right)H_2O \longrightarrow \left(\frac{n}{2} - \frac{a}{8} + \frac{b}{4}\right)CO_2 + \left(\frac{n}{2} + \frac{a}{8} - \frac{b}{4}\right)CH_4 \quad (2.18)$$

However, McCarty[167,168] found an ingenious system for predicting the stoichiometry of biological reactions which is much more fundamental and useful. He developed a generalized reaction for the overall methane fermentation of a waste (with an empirical chemical formulation of $C_nH_aO_bN_c$) to methane, CO_2, and bacterial cells ($C_5H_7O_2N$) as follows:

$$\begin{aligned} C_nH_aO_bN_c + &\left(2n + c - b - \frac{9sd}{20} - \frac{ed}{4}\right)H_2O \\ &\longrightarrow \frac{de}{8}CH_4 + \left(n - c - \frac{sd}{5} - \frac{de}{8}\right)CO_2 \\ &+ \frac{sd}{20}C_5H_7O_2N + \left(c - \frac{sd}{20}\right)NH_4^+ + \left(c - \frac{sd}{20}\right)HCO_3^- \end{aligned} \quad (2.19)$$

where $d = 4n + a - 2b - 3c$.

McCarty went on to state that s represents the fraction of waste COD synthesized or converted to cells and e represents the portion converted to methane where $s + e = 1$.

The value of s varies with the waste composition, the average solids retention time in the system (θ_c), and the cell decay rate (b) as follows:

$$s = a_e\frac{1 + 0.2b\theta_c}{1 + b\theta_c} \quad (2.20)$$

The value 0.2 represents the refractory portion of bacterial cells formed during cell decay. Values for a_e, the maximum value for s obtained when θ_c equals zero, are listed in Table 2.8 for various components of wastes. Carbohydrates give the maximum and fatty acids the minimum value of a_e.

McCarty then gave an example for $C_{10}H_{19}O_3N$, which is an average empirical formula for sewage sludge, for which a_e equals 0.11. If θ_c is assumed to equal 20 days and b equals 0.03 day^{-1}, then $s = 0.11[1 + 0.2(0.03)(20)]/[1 + 0.03(20)]$ or 0.076 and e equals 0.924. He then gives the overall reaction

$$\begin{aligned} C_{10}H_{19}O_3N + 4.74H_2O \rightarrow &\ 5.78CH_4 + 2.47CO_2 + 0.19C_5H_7O_2N \\ &+ 0.81NH_4^+ + 0.81HCO_3^- \end{aligned} \quad (2.21)$$

Table 2.8

Values for a_e and Y for methane fermentation of various wastes components[168]

Waste Component	Chemical Formula	a_e	Y ($\frac{\text{g cells}}{\text{g COD consumed}}$)
Carbohydrate	$C_6H_{10}O_5$	0.28	0.20
Protein	$C_{16}H_{24}O_5N_4$	0.08	0.056
Fatty acids	$C_{16}H_{32}O_2$	0.06	0.042
Domestic sludge	$C_{10}H_{19}O_3N$	0.11	0.077
Ethanol	C_2H_6O	0.11	0.077
Methanol	CH_4O	0.15	0.11
Benzoic acid	$C_7H_6O_2$	0.11	0.077

McCarty concluded by stating that from this equation, the decomposition of one empirical mole or 201 grams of sewage sludge at a retention time of 20 days results in the production of 5.78 moles or 129 liters of methane, 0.19 moles or 21.5 grams of bacteria, and 0.81 moles or 40.5 grams of bicarbonate alkalinity. Similar calculations can be made for any biomass feedstock.

2.7.5 Calculations of Methane Production

BOD can be satisfied in two ways. The mode used in aerobic or anoxic processes is the addition of an oxidizing agent such as dissolved oxygen, nitrates, or sulfates, but strong chemical oxidants such as chlorine and ozone will also satisfy BOD. In anaerobic processes, a reduced substance is removed (primarily methane and secondarily hydrogen sulfide). Therefore, the ultimate BOD (BOD_{ult}) or COD reduction can be calculated on the basis of methane produced as follows:

$$CH_4 + 2O_2 \longrightarrow CO_2 + 2H_2O \tag{2.22}$$

Since one gram mole (22.4 liters at STP) of methane has a COD of 64 g, then (22.4/64) 0.35 L of CH_4 at 0°C and 760 mm Hg is equivalent to 1 g COD removed. Correcting for a temperature of 35°C yields 0.395 L CH_4/g COD removed.

There are 33 Btu/L of methane at 35°C. Therefore, the methane produced from the removal of 1000 mg COD L^{-1} would be sufficient to raise the sludge feed temperature 3.3°C at 100% heat transfer efficiency.

In the anaerobic digestion process, there is no net COD removal in the acid formation phase, since the organics are merely being converted from one form to another. There may be an increase in five day biochemical oxygen demand (BOD_5) in the acid-formation phase, since the organics may be more easily degradable as the volatile acids when compared to the more complex precursor materials. Due to the

relatively high synthesis, which occurs in the acid formation from carbohydrates, there may be a significant reduction in soluble COD, but not total COD.

Since COD reduction in anaerobic digestion is stoichiometrically related to methane production (assuming hydrogen sulfide production is relatively minor), the anaerobic digester performance can be easily calculated with respect to overall treatment plant performance. In municipal wastewater treatment plants, approximately 60% of all COD destruction occurs in the anaerobic digestion process.

2.8 DESIGN PREREQUISITES

One of the major advantages of anaerobic digestion is the relatively low yield of cells or sludge production (when compared with the aerobic, activated sludge process). Under normal operating conditions, low cell yield is an advantage. However, it becomes a disadvantage in process start up. Likewise, if a toxicant is bacteriocidal to the biomass, low cell yield would prolong the recovery. (This would not be true if the toxicant was reversible in its inhibitory action.) The role of SRT, therefore, requires considerably more attention in anaerobic digestion than with aerobic processes. This fact has not been adequately appreciated by design and operational personnel. Since anaerobic digestion has been observed under extremely varied environmental conditions and research has demonstrated the ability of enriched methane producing cultures to acclimate and/or recover from a wide spectrum of toxicants (and eventually to metabolize a considerable variety of organic compounds), it is apparent that ideal conditions are not required to achieve effective anaerobic digestion of most biomass feedstocks and industrial wastewaters. However, it is paramount to recognize that a proportionate increase in SRT must be provided in the design and operation to compensate for adverse conditions. Under ideal conditions, an SRT of 10 days may provide a biological safety factor of three; but under adverse conditions, experience in treating some industrial wastewaters suggests that an SRT in excess of 100 days may have to be provided to insure a biological safety factor of three.

Young and McCarty[169] and Coulter et al.[170] were among the first to recognize this "SRT effect" when they introduced the concept of the anaerobic filter for treatment of soluble or colloidal wastewaters, or both. The packed column is inherently efficient in retaining biomass at SRTs in excess of 100 days, with hydraulic retention times of less than a day. Pretorius[171] extended this principle by providing a layer of sand on top of a column packed with stones to enhance biomass retention to an even greater extent for the effective treatment of relatively dilute domestic wastewaters. Lettings et al.[172] demonstrated that effective and efficient biomass retention could be realized without the need of packing, by proper development and control of a sludge blanket. Jeris[173] and Switzenbaum and Jewell[10] have developed methanogenic cultures attached to fluidized particles such as sand or plastic. Suidan and Cross[174] grew methanogens attached to activated carbon. Friedman[175] developed methanogens on a rotating biological contactor.

All the above-described reactors provide inherently large values of SRT (in excess of 100 days) at low liquid retention times. These high values of SRT insure

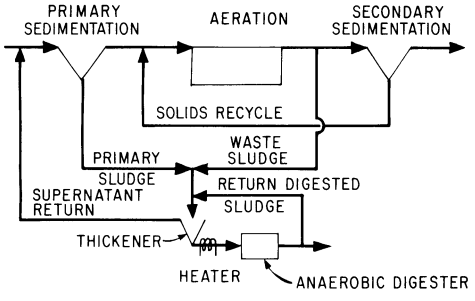


Fig. 2.17: Anaerobic digestion of municipal sludges.

maximum process stability by providing increased acclimation time and allowing cool, dilute wastewaters to be treated.

2.8.1 Solids Concentration in Feed Sludge

Solids concentration control in the feed sludge is crucial to the anaerobic digestion process. A strong case can be presented to demonstrate that surplus water is generally detrimental to process performance. (An exception to this is that surplus water may be advantageous for dilution of toxicity.) Thickening of sludges prior to feeding has been practiced for improved process control, increased methane production, and reduced volumes for ultimate sludge disposal. Torpey and Melbinger[176] reduced the required digestion volume by a factor of three, utilizing thickening (Fig. 2.17). Pfeffer[139] indicated that solids destruction varied linearly with the log of SRT and was independent of solids concentration.

The detrimental effects of surplus water in the feed are noted as follows:

- Energy is wasted to heat the surplus water.
- Alkalinity concentration in the digester is reduced.
- SRT, and thus solids, destruction is reduced.
- Required digester volumes are increased.

2.8.2 Batch versus Continuous Feeding

A particular practice of convenience has evolved in the operation of many municipal anaerobic digesters: pumping sludge to the digester once per shift, thus feeding the digester three times per day. This arrangement is convenient and has some rationale in work scheduling. From the point of view of the biological process, however, it is undesirable to slug feed the digester rather than provide continuous feeding. A biological system can process more substrate when fed continuously rather than batchwise, since substrate becomes alternately excessive and limiting in the latter case. Additional alkalinity must also be maintained to buffer the surges in volatile acids concentrations.

However, in an attempt to feed continuously, caution must be exercised to avoid pumping dilute sludge into the digester. Automatic feedback control of the sludge pump utilizing a sludge concentration meter is desirable in order to keep excessive water out of the digester. With soluble industrial wastewaters, concentration of the feed sludge is not an issue.

2.8.3 Plug Flow versus CSTR

The anaerobic filter concept can be operated as a quasi-"plug-flow" system. It is relatively easy to convert it into a quasi-continuously stirred tank reactor (CSTR) by high rate recycle of the effluent back to the inlet. Both the plug flow and CSTR modes have advantages. If a slug of toxicant enters a plug flow reactor, there is no initial dilution and gas production may be more severely retarded. However, the plug flow nature will flush the toxicant out of the system much more rapidly than would occur in a CSTR reactor. Thus, there is less adverse impact on overall treatment efficiency in the case of the plug flow reactor. This has been observed when 100 mg/L of cyanide was administered in the daily feed to a plug flow filter.

There are also situations in which benefits accrue with the CSTR operational mode. We have found it difficult to acclimate methane fermentation systems to metabolize formaldehyde concentrations in excess of 400 mg/L and phenol concentrations in excess of 1500 mg/L due to substrate toxicity. In order to treat higher concentrations effectively, dilution must be provided and the most suitable source of dilution water is the treated effluent. Thus, if the formaldehyde, phenol, or other organic components which exhibit substrate toxicity are metabolized in the process, the effluent concentration is low and suitable for recycle as dilution water. Similar economy can be realized for the treatment of low pH wastewaters which release alkalinity-producing cations during treatment, as is the case with proteins, salts of volatile acids, and soaps. Recycle of the effluent, which contains increased alkalinity, to the influent can reduce or eliminate the need to supplement the wastewater with commercial forms of alkalinity.

It must be realized that especially with a sludge blanket, methane fermentation system, recycling of the effluent will effectively increase the hydraulic throughput. Ultimately this can hydraulically disturb the sludge blanket to the point of causing biomass loss in the effluent. Thus the limits and trade-offs of an effluent recycle need careful consideration.

2.8.4 Temperature

Buhr and Andrews[7] reviewed the thermophilic digestion literature. They reported the following advantages of thermophilic digestion over mesophilic digestion:

- Increased reaction rates for destruction of organic solids.
- Increased efficiency for destruction of organic solids.
- Improved solids-liquid separation.
- Increased destruction of pathogenic organisms.

They also cited the following possible disadvantages of the thermophilic process over mesophilic processes:

- Higher energy requirements for heating.
- Poorer supernatant quality.
- Poorer process stability.

In summarizing the literature, Buhr and Andrews[177] concluded that the preponderance of the results presented for both bench and plant scale studies showed that the rate of destruction of organic solids was higher at thermophilic temperatures than at mesophilic temperatures. Consequently at equal residence times, a larger fraction of organic solids is destroyed, although the difference may not be noticeable at longer residence times. In addition, the literature indicated that in all cases thermophilic digested sludge had a higher solids concentration than mesophilic digested sludge. Higher filtration rates with lower coagulant demands were observed,[178,179] but a poorer supernatant quality was noted for the thermophilic process.[179]

Plant scale operation of thermophilic digestion is rather limited.[179-181] Garber[179] reported on the extensive experience with thermophilic digestion at the Hyperion Treatment Plant in Los Angeles. He investigated digestion in the range of 29° to 52°C and found (1) that sludge digested at the higher temperatures filtered much more easily and (2) that under Hyperion conditions, 49°C was the maximum temperature that could be easily controlled. Future plans call for conversion of all digesters from mesophilic (32° to 35°C) to thermophilic (49°C) because of the improved dewatering characteristics at the higher temperature. Garber reports that operation in the thermophilic range increased process sensitivity, necessitating closer process control. Operation at 38°C could tolerate a temperature differential of ±2.8°C, while at 49°C this differential was reduced to approximately ±0.8°C. It was also observed that a properly digesting thermophilic sludge always had about 2 to 2.5 times the volatile acids content of the mesophilic product and the volatile acids to alkalinity ratio was about twice that of mesophilic sludge.

Garber[179] reported on the inactivation of pathogens at higher temperatures. Present evidence is that digestion at approximately 54°C will substantially inactivate pathogenic bacteria, virus, or parasites. At 49°C there was almost complete destruction of *Salmonella sp.* and *Pseudomonas aer.*, whereas fecal coliform and streptococci were not substantially inactivated below 51°C. Virus inactivation data were similar, with substantially complete virus inactivation above 50°C for Hyperion sludges. Ascaris worm parasites also seemed to be inactivated at temperatures above 51°C in the Hyperion thermophilic digestion process.

The Metropolitan Sanitary District of Greater Chicago plans to convert its mesophilic digesters to thermophilic operation in order to double the plant's sludge processing capability and also to achieve energy self-sufficiency.[180] When the procedure used by Garber[179] at Hyperion to obtain a thermophilic culture (i.e., raising the temperature of the digester at a rate of 0.6° C day^{-1}) was followed, the digester "soured." To overcome this difficulty, the temperature of the digester was increased at a much lower rate. When the change to thermophilic digestion was accomplished, the following observations were made:

- The quality of the thermophilic sludge was improved.
- The concentration of volatile acids was higher in the thermophilic than in the mesophilic digested sludge (270 versus 40 mg/L respectively).
- Volatile solids destruction was 9% greater in the thermophilic digester operated at 51°C and 7-day retention time as compared to mesophilic digestion at 35°C and 14-day retention time.
- The ammonia levels were higher in the thermophilic sludge, but they did not result in ammonia toxicity. A greater breakdown of proteinaceous material had occurred in thermophilic digestion.
- COD reductions were similar in both thermophilic and mesophilic processes.
- The thermophilic sludge had an increased odor.
- The thermophilic sludge was grainy in appearance and did not possess the gelatinous structure of the mesophilic sludge.
- Gas production per kilogram of volatile solids feed was consistently higher for thermophilic digestion with a yearly average of 25% more gas per kilogram of volatile solids for thermophilic digestion compared to mesophilic digestion.

Pfeffer[182] discussed the significant temperature effect on the pH-alkalinity-gas composition. The temperature effect manifests itself in two ways—change in the equilibrium constants for the bicarbonate system and change in the vapor pressure of water. Table 2.9 shows the change in the values of Henry's Law Coefficient (K'_H), the bicarbonate ionization coefficient (K_1), and the water ionization coefficient with temperature. The solubility of CO_2 in water is only about half as much at 60°C as at 30°C this markedly reduces the alkalinity requirement at thermophilic temperatures. Pfeffer illustrates that a digester operating at 40°C and producing a gas containing 40% CO_2 would require an alkalinity of about 2250 mg/L to maintain a pH of 7.0. The same digester operating at 60°C would require only 1300 mg/L of alkalinity to maintain a pH of 7.0. For bioconversion of high carbohydrate feedstocks requiring almost total supplementation of alkalinity, thermophilic digestion would be a strong economic incentive for reducing the relatively high costs of supplemental purchased alkalinity.

The net energy production efficiency from anaerobic digestion at various temperatures was analyzed by Ghosh.[183] He considered various substrate qualities, biodegradabilities, and methane yields in the context of temperature variations. He found that sludge heating or digester heat losses accounted for the largest fraction of external energy input, depending on the feed slurry concentration and the loading rate. External energy inputs to dilute-slurry and low-loading-rate digesters can exceed 250% of those required for typical high-rate digestion conditions receiving a concentrated feed. Sludge heating energy requirements for thermophilic digestion (55°C) are about double that for mesophilic digestion (35°C) for all practical feed-solids concentrations. Net energy production becomes negative for mesophilic digestion below 2% total solids, and for thermophilic digestion below 4% total solids.

Table 2.9

Equilibrium constants for CO_2, bicarbonate, and water vapor pressure for various temperatures[184]

°C	$K'_H(CO_2)$ (mm Hg. mole fraction^{-1})	K_1	K_w	Water Vapor Pressure (mm Hg)
30	1.39×10^6	4.7×10^7	1.47×10^7	31.8
40	1.73×10^6	5.04×10^7	2.92×10^7	55.3
50	2.17×10^6	5.13×10^7	5.47×10^7	92.5
60	2.58×10^6	5.19×10^7	9.62×10^7	149.4

Heat input to compensate for mesophilic digester heat losses could range from 20% to less than 6% of product methane yields at loading rates of 1.6 to 8.0 kg volatile solids (VS) $m^{-3}day^{-1}$, respectively. For thermophilic digestion, heat losses may increase to 50% of product methane at similar loading rates.

O'Rourke[140] studied the effect of temperature on the anaerobic digestion of municipal raw sludge. He observed that lipid decomposition controlled the minimum SRT which could be operated satisfactorily. He observed that the minimum SRT for lipid decomposition was about 4, 10, 12, and greater than 60 days, at temperatures of 35, 25, 20, and 15°C, respectively.

Varel et al.[185] studied the effect of temperature and retention time on methane production from beef cattle waste. A temperature range of 30° to 65°C was studied in 5°C increments. The fermentors were fed once per day with a 6% volatile solids feed, and retention times varied from 18 down to 2.5 days. The methane production rates ($vol \cdot vol^{-1}day^{-1}$) and methane yields ($L \cdot g^{-1}$ of VS) are shown in Table 2.10. Volatile solids degradation at these retention times and temperatures was between 46 and 54%. The concentration of volatile acids in the 30° to 55°C fermentors was generally below 2000 mg/L, with the exception of the three-day retention time. The 60° and 65°C fermentors had volatile acids levels usually higher than 2000 mg/L for all retention times.

Kennedy and van den Berg[186] evaluated the effects of temperature and overloading on the performance of anaerobic fixed film reactors digesting bean blanching wastes. In mature fixed film reactors, high loading rates without substantial changes in COD removal efficiency were observed at temperatures between 10° and 35°C. At 35°C, a maximum steady-state loading rate of 10.4 kg COD $m^{-3}day^{-1}$ could be maintained for several months without reactor failures. The COD removal rate was 88 ±2% and the volumetric methane production was 5.4 $vol \cdot vol^{-1}day^{-1}$. Decreasing the temperature to 25°C decreased the maximum steady-state loading rate by 37% while at 10°C the loading rate was reduced by 75% (compared to 35°C). However, the COD removal was independent of temperature. The feeding rate was

Table 2.10

Highest methane production rates and methane yield for beef cattle waste[185]

Temperature (°C)	Methane Production Rates ($L \cdot L^{-1} day^{-1}$)	Methane Yield ($L \cdot g^{-1}$ VS)	Retention Time (days)
30	1.27	0.19	9
35	1.60	0.16	6
40	2.28	0.23	6
45	2.42	0.24	6
50	2.83	0.14	3
55	2.75	0.14	3
60	3.18	0.14	2.5
65	1.69	0.17	6

regulated so that volatile acids always remained below 500 mg/L in the effluent at all temperatures.

Van den Berg et al.[31] studied methane formation from acetic acid with enriched methanogenic cultures and found that temperature affected the rate of acetate conversion markedly. The optimum temperature was 40° to 45°C, not 35°C. Methane formation was still measurable at 10°C. At both 50° and 55°C, acetic acid conversion stopped after one to two days. Speece and Kem[143] studied the effects of short-term temperature variations on methane production using bench scale mesophilic (35°C) digesters. Their retention time was 20 days and the digesters were fed raw sludge once per day. Methane production was particularly sensitive to decreases in temperature and practically ceased when the temperature was dropped to 20°C. Gas production rapidly resumed with no lag when the temperature was restored to 35°C. An increase in temperature to 45°C rapidly reduced a supplemental spike of acetic acid, in addition to reducing the volatile acids. All were converted from raw sludge at a higher rate at the higher temperature. They suggested that an increase in digestion temperature would be a control method for reducing a temporary increase in volatile acids.

Cooney and Wise[187] investigated thermophilic anaerobic digestion of municipal solid waste for fuel gas production. The overall conversion of solid waste to fuel gas proceeded faster under thermophilic conditions than under mesophilic conditions. In addition, the rate of acetic acid consumption was 3.2 times faster under thermophilic conditions.

2.8.5 Process Configuration

It was reported by van den Berg et al.[188] that methane production rates of anaerobic fixed film fermenters exceeded those of anaerobic contact and fully mixed

continuous fermenters, both with simulated sewage sludge and bean blanching waste as substrate. With bean blanching waste, the methane production rates were 5.7, 2.0, and 0.3 $m^{-3}CH_4 \cdot m^3$ of digester day^{-1} for anaerobic fixed film, anaerobic contact process, and fully mixed continuous fermenter, respectively.

2.8.6 Soluble versus Sludge Feedstocks

A significant distinction has to be made between the treatment of soluble and sludge feedstocks. Soluble feedstocks (and also colloidal feedstocks such as starch which can be completely liquified) do not result in accumulation of a significant fraction of nonbiological, refractory suspended solids such as is the case in digestion of domestic sludges. In the latter case, considerable solids wasting is required to move the nonbiological, refractory solids out of the system. The viable acetogens and methanogens cannot be separated from this waste sludge and unfortunately are also lost from the system, reducing the SRT/hydraulic retention time (HRT) ratio, process efficiency, and stability.

With soluble feedstock wastewaters, much greater SRT/HRT ratios can be maintained. This decreases the relative capital cost and increases the process efficiency and stability. Carbohydrate wastewaters tend to produce much higher mass ratios of acid formers/methanogens and likewise tend to reduce the mass of methanogens which can be maintained within a given system.

2.8.7 Industrial Wastewaters Treated by Methane Fermentation

A wide variety of industrial wastewaters has successfully been treated by methane fermentation. Table 2.11 lists the types of wastewaters that have been successfully treated at full- or pilot-scale installations.

The liquor produced from the heat treatment of waste-activated sludge has been successfully treated by suspended-growth methane fermentation at a 150 metric tonne per day sludge processing plant on the Emscher River.[189] Haug et al.[190] also demonstrated the efficiency of anaerobic filter treatment of waste-activated sludge heat treatment liquor. McCarty et al.[191] have demonstrated enhanced methane production realized by various heat treatment schemes of previously nonbiodegradable organics such as lignins, lignite, and wood fibers. Research has also been conducted on the potential catabolism of the components of coal conversion wastewater via methane fermentation. Phenol, resorcinol, catechol, o-cresol, p-cresol, and 4-methyl catechol were biodegraded to methane in our laboratory. Based upon the estimate of Singer et al.[192] of the composition and concentration of the components of coal conversion wastewater, I found that at least 70% of the total COD would be biodegradable to methane by anaerobic digestion. Associated research in our laboratory on methane fermentation of phenol weak liquor produced from the coking operation indicates that if the cyanide level can be reduced to 5 mg/L by precipitation as Prussian blue, or other means, the phenol weak liquor can be biodegraded to methane at an efficiency of 80 to 90% by anaerobic digestion.

Buswell found that a wide range of organic compounds could be degraded to methane by anaerobic digestion.[193–196] Hovious et al.[197] reported the partial

Table 2.11

Industrial wastewaters treated by methane fermentation

Type	References
Distillery	Sen and Bhaskaran[198]
Distillery	Stander and others[199]
Distillery	Painter[200]
Sugar cane	Guzman[201]
Wool and textile	Buswell and others[202]
Wool and textile	Grishina[203]
Wool and textile	Tanaka and others[204]
Wool and textile	Nevzorov[205]
Potato processing	Hindin and Dunstan[206]
Potato processing	Olson and others[207]
Tannery	Ivanov[208]
Tannery	Gates and Line[209]
Slaughter house & packing plant	Schroepfer and Ziemke[210]
Slaughter house & packing plant	Silvester[211]
Slaughter house & packing plant	Steffen and Bedker[212]
Slaughter house & packing plant	Henzen[213]
Slaughter house & packing plant	Bradney and others[214]
Rum distillery	Hiatt and others[215]
Starch	Taylor and Burm[216]
Yeast	Koehler[217]
Methanolic	Lettinga and others[30]
Pharmaceutical	Jennet[218]
Sugar beet	Lettinga and others[172]
Cheese whey	Switzenbaum[219]
Pulp condensate	Ferguson[220]

removal of the following organic chemicals: methanol, ethanol, n-propanol, isopropanol, n-butanol, isobutanol, n-pentanol, isopentanol, hexanol, acetaldehyde, n-butyraldehyde, isobutyraldehyde, acetone, methyl ethyl ketone, ethylene glycol, crotonaldehyde, and ethyl acrylate. Jeris et al.[221] reported degradation by anaerobic digestion of succinic, lactic, glutamic, oleic, and benzoic acids. Gossett et al.[222] reported evidence that ferulic, dimethoxybenzoic, para-hydroxy benzoic, benzoic, syringic, vanillic, protocatechuic acids, catechol, phenol, syringaldehyde, and vanillin could be degraded by methane fermentation. Chmielowski and Wasilewski[223] degraded phenol via methane fermentation. Witt et al.[224] demonstrated the biodegradation of a variety of petrochemical wastewater components.

2.8.8 Screening Studies

There are a wide variety of biomass feedstocks and industrial wastewaters which are potential candidates for treatment by anaerobic digestion. With some exceptions, it is generally considered that if an organic is biodegradable aerobically, it will also biodegrade anaerobically.

McCarty and his co-workers have made extensive studies on anaerobic digestion of organic components. In particular, they have investigated various heat treatment schemes for enhancement of anaerobic digestion of waste-activated sludges and lignin-containing waste materials. They have proposed a very useful treatability assay procedure called the Biochemical Methane Potential (BMP) assay. The BMP is the anaerobic counterpart of the aerobic ultimate BOD assay.[225] The BMP assay is simple and replicable. A sample of the organic wastewater is injected into a serum bottle containing an appropriate inoculum of anaerobic digestion sludge (well digested sludge from a municipal wastewater sludge digester generally contains a wide spectrum of organisms). The assay is then incubated at 35°C and the gas production measured with time. Methane production in excess of the control—containing only the anaerobic sludge inoculum—is that contributed by degradation of the organics injected. Incubation occurs for a period of at least 30 days in most cases and longer for oganics which necessitate considerable acclimation by the biomass. Nutrient deficiencies are not elucidated by the BMP assay and therefore must be determined by additional studies.

The BMP is quite useful because it gives an accurate measure of the potential methane which can be biochemically generated from the organic components in question. The extended incubation period may overpower some inherent toxicity in the wastewater, but toxicity could significantly reduce the BMP. Requirements for microbial acclimation may need further definition. The BMP does not give direct information on the kinetics of decomposition. Such kinetics would have to be derived from pilot plant studies. The BMP simply gives a measure of the maximum amount of methane which can be produced from the organics in question. The treatment process provided will determine how much of this methane potential is actually realized.

In associated work, McCarty and his co-workers developed an Anaerobic Toxicity Assay (ATA) to evaluate the effect of any inherent toxicity contained in an organic wastewater sample.[225] Toxicity is measured with reference to the predominant methanogens and acetogens which convert acetate and propionate to methane. Thus the inoculum sludge in the serum bottle is "spiked" with an excess of acetate and propionate, and gas production is measured with time. A reduced slope of gas production versus time, when compared with the "spiked" control, is a measure of any inherent toxicity. Successive additional "spikes" of acetate and propionate allow for the possibility of acclimation of the methanogens and propionate-degrading acetogens to the toxicity. The methanogens have demonstrated significant potential to acclimate to a wide variety of toxicants. (Toxicity acclimation is discussed in more detail in a previous section.)

Considerable care needs to be exercised in performing screening and treatability studies to ensure that valid conclusions are reached. In the BMP assay, there is no wasting of biomass and, thus, no washout. In most pilot plant treatability studies, wasting of biomass is required when each new batch of wastewater is fed. Under these conditions, excessive solids washout occurs before biodegradation is established.

2.8.9 Scale-Up Factors Affecting Performance

Pilot plant treatability studies of industrial wastewaters are necessarily performed in small-scale systems. Flow distribution through the horizontal cross section of an anaerobic filter is not difficult to achieve in a small-scale system with introduction of the feed at only one point. However, in a full-scale system, flow distribution becomes a major design consideration.

In suspended growth systems, short-circuiting of the influent to the effluent can easily be avoided with adequate mixing. However, in prototype systems, short-circuiting must be specifically addressed by adequate design provision, e.g., baffles or mixing.

Many treatability studies are performed with batch feeding because of either convenience or lack of suitable low-volume pumps. The biological effect of batch feeding to a CSTR is to simulate plug flow through a reactor. Some wastes demonstrate significantly improved treatment efficiency in the plug flow mode. From biological kinetics, it can be shown that wastewaters exhibiting a relatively high half velocity constant, K_S, are more efficiently treated in the plug flow mode than in the CSTR mode. If the prototype treatment process is a continuously fed CSTR, the treatment kinetics will respond accordingly, and effluent quality will be inferior to that predicted from the "plug flow" treatability studies.

2.9 FUTURE DIRECTIONS

Anaerobic digestion occurs widely in natural ecosystems and has been commonly applied for treatment of domestic wastewater sludges. Considerable interest has been expressed in the application of anaerobic digestion for treatment of industrial wastewater, but the number of systems actually constructed and operating is quite limited. Likewise, anaerobic digestion is a viable route for bioconversion of a wide variety of renewable biomass feedstocks to methane as an alternative energy source. Once again, the number of systems actually constructed and operated for this specific purpose are very few.

The relatively low cost of methane gas has been a major deterrent in application of anaerobic digestion. This process can rarely be justified economically when based solely on the methane gas produced. Additional credits associated with the process are necessary for economic justification, such as waste stabilization, reduced excess biological solids production, fertilizer value of residue, pyrolysis value of solids, potential acclimation to toxicity in feed, aesthetic considerations, etc. The current glut of natural gas supplies has also noticeably eroded interest in construction of biomass to methane projects.

Aside from the methane economic considerations, several aspects of anaerobic digestion require additional clarification to encourage more widespread application of the process. Attachment and associated start-up phenomena in fixed film versions of anaerobic digestion are not well understood. It has been proposed to develop the initial bacterial film on virgin media with temporary nitrate supplementation to speed up the subsequent methanogenic film development. Attached film start-up times are still prolonged and there is a definite need to reduce this interval.

Nutrient requirements have been clarified to some extent, but it is still not possible to sustain high acetate-utilization rates indefinitely. Preliminary evidence in our laboratory indicates that high phosphate levels prevent high acetate utilization rates from being realized. Phosphate precipitation of essential trace metals such as iron, cobalt, nickel, and/or molybdenum may be involved but this has not been proven. The role of natural or synthetic chelators in stimulation of methane fermentation by making essential trace metals available to the microbiota is hypothesized but, as yet, not demonstrated.

Propionate metabolism appears to be inhibited in some systems. It is not clear whether this is due to elevated hydrogen concentrations or to nutritional deficiencies. It deserves much attention because increased propionate levels are commonly reported in malfunctioning field installations. Hydrogen levels observed in anaerobic digesters are rarely reported in the literature. Perhaps with recent developments in analytical instrumentation for detection of very low levels of hydrogen, the hydrogen concentrations can be correlated with overall digestion parameters.

The nutritional requirements for the acid fermentation process need definition. Frequently, this limits the overall rate of anaerobic digestion and therefore efforts to reduce the required digestion time must address this common rate-limiting step. There is general speculation that symbiotic relationships exist between the classes of microorganisms, but documentation is lacking.

Remedial treatment procedures for "stuck" digesters or for digesters operating with chronically high volatile acids concentration need to be developed. Most process applications cannot tolerate lapses in effective treatment. There is preliminary evidence that iron additions enhance recovery of malfunctioning digesters. Reducing agents have also been shown to enhance digestion recovery after exposure to cyanide toxicity. There is much research to be done in the area of remedial treatment which would probably lead to increased process stability.

Finally, the choice of engineering process configuration is still open to debate. Some side-by-side comparisons of different process configurations have been made for a given wastewater.[226] However, there is little information in the literature to support any general selection guidelines for a wide spectrum of wastewaters or biomass feedstocks. The phenomena controlling the development of the "granular" floc in the upflow anaerobic sludge blanket reactor is not clear. Neither is the phenomena leading to abrupt disintegration of the same granular sludge well understood. The choice of engineering process configuration must avoid undue loss of biological solids on the one hand and prevention of plugging of the support media on the other.

In summary, the anaerobic digestion process has been developed to the point of reliable process application for many potential biomass feedstocks. With further developments, process rate, efficiency, and stability can be further enhanced, making the economics more attractive. Finally, even though short term natural gas supplies are temporarily in surplus, the predicted depletion of these natural gas supplies within 20 to 30 years is indicative of a long-term, growing interest in the anaerobic digestion process.

REFERENCES

1. W. E. Balch et al., "Methanogens: Re-evaluation of a unique biological group," *Microbiol. Rev.*, **43** 260–296 (1979).
2. R. A. Mah and M. R. Smith, "The methanogenic bacteria," Chap. 76 in *The Prokaryotes* edited by Starr et al. Springer-Verlag, 948–977 (1981).
3. M. P. Bryant, "Microbial methane production–theoretical aspects," *J. Animal Science*, **48**, 193–201 (1979).
4. R. A. Prins, "Methanogenesis in the gastrointestinal tracts of ruminants and man," *Antonie van Leeuwenhoek*, **45**, 339–345 (1979).
5. P.R. Zimmerman et al., "Termites, a potentially large source of atmospheric methane, carbone dioxide, and molecular hydrogen," *Science*, **218**, 563–565 (1982).
6. J.G. Zeikus and J. G. Ward, "Methane formation in living trees: A microbial origin," *Science* **184**, 1181–1183 (1974).
7. J.G. Zeikus, "The biology of methanogenic bacteria," *Bact. Rev.* **41**, 514–541 (1977).
8. M. Salonen-Salkinoja et al., "Biodegradation of recalcitrant organochlorine compounds in fixed film reactors," IAWPR specialized seminar—Anaerobic treatment of wastewater in fixed film reactors. Copenhagen, Denmark, 149–162 (1982).
9. R.E. Speece, "Fundamentals of the anaerobic digestion of municipal sludge and industrial wastewaters," Seminar on Anaerobic Wastewater Treatment and Energy Recovery by Duncan, Lagnese and Assoc. Inc., Pittsburgh, PA (1981).
10. M.S. Switzenbaum and W. J. Jewell, "Anaerobic attached film expanded bed reactor treatment," *J. Wat. Poll. Control Fed.*, **52**, 1953–1965 (1980).
11. R.H. Clark and R. E. Speece, "The pH tolerance of anaerobic digestion," *Proc. 5th Int'l. Conf. on Water Poll. Res.*, **17**, (1970).
12. A. J. B. Zehnder, "Ecology of methane formation," Chap. 13 in *Microbiology of Water Pollution* by Mitchell, Wiley Interscience Pub., 349–376 (1978).
13. J.G. Morris, "Oxygen and the obligate anaerobic," *J. Appl. Bacteriol.*, **40**, 229 (1976).
14. Robert A. Mah, "The methanogenic bacteria, their ecology and physiology," in *Trends in the Biology of Fermentations for Fuels and Chemicals*, edited by Hollaender, et al., Plenum Pub., 357–373 (1981).

15. M. J. McInerney, M. P. Bryant, and N. Phennig, "Anaerobic bacterium that degrades fatty acids in syntrophic assocaition with methanogens," *Arch. Microbiol.*, **122**, 129–135 (1979).

16. R.A. Mah, "Methanogenesis and methanogenic partnerships," *Phil. Trans. R. Soc. London B* **297**, 599–616 (1982).

17. P.L. McCarty, "History and overview of anaerobic digestion," *Second Int'l. Symp. on An. Dig.*, Travemunde, W. Germany (1981).

18. M.P. Bryant et al., "*Methanobacillus omelianski*, a symbiotic association of two species of bacteria," *Arch. Microbiol.*, **59**, 20–32 (1967).

19. W.H. Lorowitz and M. P. Bryant, "Co-culture of stearate-utilizing bacterium in co-culture with an H_2-utilizing *Desulfovibris*," Personal communication (1982).

20. R.K. Thauer, K. Jungerman, and K. Decker, "Energy conservation in chemotrophic anaerobic bacteria," *Bacteriol. Rev.*, **41**, 514–541, 100-180 (1977).

21. M.J. McInerney and M. P. Bryant, "Review of methane fermentation fundamentals," *CRC Press* (1981).

22. R.E.Hungate, *The Rumen and its Microbes* (Academic Press, 1966).

23. E.L. Iannotti et al., "Glucose fermentation products of *Ruminococcus albus* grown in continuous culture with *Vibrio succinogenes*, changes caused by interspecies transfer of H_2," *J. Bact.*, **114**, 1231–1240 (1973).

24. P.H. Smith, "Studies of methanogenic bacteria in sludge," EPA—600/ 2–80–093 Research Report (1980).

25. P.H. Smith and P. S. Shuba, "Terminal anaerobic dissimilation of organic molecules," *Proc. Bioconversion Energy Res. Conf.*, 8–14 (1973).

26. H.A. Barker, "Studies upon the methane producing bacteria," *Arch. Microbiol.*, **7**, 420–438 (1936).

27. R.E. Speece, G. F. Parkin, and D. Gallagher, "Nickel stimulation of anaerobic digestion," *Wat. Res.* **17** 667-683 (1983).

28. G. Lettinga et al., "Anaerobic treatment of methanolic wastes," *Water Research*, **13**, 725 (1979).

29. L. van den Berg, et al., "Factors affecting rate of methane fermentation from acetic acid by enriched methanogenic cultures," *Can. Soc. Microbiol.*, **22**, 1313–1319 (1976).

30. R.A. Mah, M. R. Smith, and L. Baresi, "Studies on an acetate-fermenting strain of *Methanosarcina*," *Appl. and Environ. Microbiol.*, **35**, 1174–1184 (1978).

31. M.A. Pretorius, "The effect of formate on the growth of acetate utilizing methanogenic bacteria," *Wat. Res.*, **6**, 1213–1217 (1972).

32. R.S. Wolfe, "Microbial formation of methane," *Adv. Microbial Physiol.*, **6**, edited by A. H. Rose, J. F. Wilkinson (Academic Press, New York-London, 1971), pp. 107–146.

33. J.S. Jeris and P. L. McCarty, "The biochemistry of methane fermentation using C^{14} tracers," *J. Water Poll. Control Fed.*, **37**, 178–192 (1965).

34. P.H. Smith and R. A. Mah, "Kinetics of acetate metabolism during sludge digestion," *Appl. Microbiol.*, **14**, 368–372 (1966).

35. H. Kaspar and K. Wuhrmann, "Kinetic parameters and relative turnovers of some important catabolic reactions in digesting sludge," *Appl. Envir. Microbiol.*, **36**, 1–7 (1978).

36. W.E. Balch, et al., "An ancient divergence among the bacteria," *J. Mol. Evol.*, **9**, 305–311 (1977).

37. O. Kandler and H. Konig, "Chemical composition of peptidoglycan-free cell walls of methanogenic bacteria," *Arch. Microb.*, **118(2)**, 141–152 (1978).

38. C. R. Woese and G. E. Fox, "Phylogenetic structure of the prokaryotic domain: the primary kingdoms" *Proc. Nat'l. Acad. Sci.*, **74**, 5088–5090 (1977).

39. W.E. Balch, "Methanogens: their impact on our concept of procaryote diversity," *Zbl. Bakt.Hyg. I. Abt. Orig.*, Archaebacteria: Proceedings of the First International Workshop on Archaebacteria, **3**, 295–303 (1982).

40. W.E. Balch and R. S. Wolfe, "Specificity and biological distribution of Coenzyme M (2–mercaptoethanesulfonic acid)," *J. Bacteriol.*, **137**, 256–263 (1979).

41. B.A. Huser, K. Wurhmann, and A. J. B. Zehnder, "*Methanothrix soehngenii* non-hydrogen oxidizing methanobacterium," *Arch. Microbiol.*, **132**, 1–9 (1982).

42. R.R. Mah and Kuhn, *Intern. J. Syst. Bacteriol.* (In press).

43. W.E. Balch and R. S. Wolfe, "Transport of Coenzyme M (2–mercaptoethanesulfonic acid) in *Methanobacterium ruminantium*," *J. Bacteriol.*, **137**, 264–273 (1979).

44. B.C. McBride and R. S. Wolfe, "A new coenzyme of methyl transfer, coenzyme M," *Biochemistry*, **10**, 2317–2324 (1971).

45. C.D.Taylor and R. S. Wolfe, "A simplified assay for coenzyme M ($HSCH_2CH_2SO_3$)," *J. Biological Chem.*, **249**, 4886–4890 (1974).

46. C.D. Taylor, et al., "Coenzyme M, essential for growth of a rumen strain of *Methanobacterium ruminantium*," *J. Bacteriol.*, **120**, 974–975 (1974).

47. R.N. Gunsalus, et al., "Methyl transfer and methane formation," Proceedings of the *Symposium on Microbial Production and Utilization of Gases* (H_2, CH_4, CO), E. Goltze, Gottingen, 191–197 (1977).

48. S. Shapiro and R. S. Wolfe, "Methyl *CoM*, an intermediate in methanogenic dissimilation of C_1 compounds by *Methanosarcina barkeri*," *J. Bacteriol.*, **141**, 728–734 (1980).

49. P. Cheeseman, A. Toms-Wood, and R. S. Wolfe, "Isolation and properties of a fluorescent compound, $factor_{420}$, from *Methanobacterium strain M.o.H.*," *J. Bacteriology*, **112**, 527–531 (1972).

50. S.F.Tzeng, R. S. Wolfe, and M. P. Bryant, "Factor 420-dependent pyridine nucleotide—linked hydrogenase system of *Methanobacterium ruminantium*," *J. Bacteriology*, **121**, 154–191 (1975).

51. S.F. Tzeng, M. P. Bryant, and R. S Wolfe, "Factor 420-dependent pyridine nucleotide—linked formate metabolism of *Methanobacterium ruminantium*," *J. Bacteriology*, **121**, 192–196 (1975).

52. L.D. Eirich, G. D. Vogels, and R. S. Wolfe, "Proposed structure of coenzyme F_{420} from methanobacterium," *Biochemistry*, **17**, 4583–4593 (1978).

53. H. Doddema and G. Vogels, "Improved identification of methanogenic bacteria by fluorescence microscopy," *Appl. Envir. Microbiol.*, **36**, 752–754 (1978).

54. T. Edwards and B. McBride, "New method for the isolation and identification of methanogenic bacteria," *Appl. Microbiol.*, **29**, 540–545 (1975).

55. W.B.Whitman and R. S. Wolfe, "Presence of nickel in factor F_{430} from *Methanobacterium bryantii*," *Biochem. and Biophys. Res. Comm.*, **92**, 1196–1201 (1980).

56. G. Diekert, B. Klee, and R. Thauer, "Nickel, a component of factor F_{430}," *Arch. Microbiol.*, **124**, 103–106 (1980).

57. W.L. Ellefson and R. S. Wolfe, "Component C of the methylreductase system of methanobacterium," *J. Biol. Chem.*, **256**, 4259–4262 (1981).

58. G. Diekert, R. Jaenchen, and R. K. Thauer, "Biosynthetic evidence for nickel tetrapyrrole structure of factor F_{430} from *Methanobacterium thermoautotrophicum*," *FEBS Lett.*, **119**, 118–120 (1980).

59. G. Diekert, et al., "Nickel requirement and factor F_{430} content of methanogenic bacteria," *J. Bact.*, **148**, 459–464 (1981).

60. H.J. Doddema, "Anaerobic biochemical techniques applied in the purification of the hydrogenase of *Methanobacterium thermoautotrophicum*," *Antonie van Leeuwenhoek, J. Microbiol. Serol.*, **46**, 7 (1980).

61. G. Fuchs, E. Stupperich, and R. K. Thauer, "Acetate assimilation and synthesis of alanine, aspartate and glutamate in *Methanobacterium thermoautotrophicum*," *Arch. Microb.*, **117(1)**, 61–66 (1978).

62. M.P. Bryant, et al., "Nutrient requirements of methanogenic bacteria," *Adv. Chem. Ser.*, **105**, 23–40 (1971).

63. L. Daniels and J. G. Zeikus, "One-carbon metabolism in methanogenic bacteria—Analysis of short-term fixation products of $C^{14}O_2$ and $C^{14}H_3OH$ incorporated into whole cells," *J. Bacteriol.*, **136**, 75–84 (1978).

64. J.G.Zeikus, et al., "Bacterial methanogenesis: acetate as a methane precursor in pure culture," *Arch. Microb.*, **104**, 129–134 (1975).

65. M. R. Smith and R. A. Mah, "Acetate as sole carbon and energy source for growth of *Methanosarcina* strain 227," *Appl. Envir. Microbiol.* **39**, 993–999 (1980).

66. J. G. Zeikus and R. S. Wolfe, "*Methanobacterium thermoautotrophicum* sp. n., an aerobic, autotrophic extreme thermophile," *J. Bacteriol.* **109**, 707–713 (1980).

67. M. P. Bryant, B. C. McBride, and R. S. Wolfe, "Hydrogen oxidizing methane bacteria. I. Cultivation and methanogenesis," *J. Bacteriol.* **95(1)**, 1118–1123 (1968).

68. B. A. Blaylock and T. C. Stadtman, "Methane biosynthesis by *Methanosarcina barkeri*. Properties of soluble enzyme system," *Arch. of Biochem. and Biophysics* (1966).

69. E. A. Wolin, M. J. Wolin, and R. S. Wolfe, "Formation of methane by bacterial extracts," *J. Biol. Chem.* **238**, 2882–2886 (1963).

70. H. J. Doddema et al., "ATP Hydrolysis and synthesis by membrane-bound ATP—synthetase complex of *Methanobacterium thermoautotrophicum*, *J. Bacteriol.*, **136(1)**, 19–23 (1978).

71. E. S. Pantskhava and V. N. Bukin, "Effect of ATP, ADP, and AMP on production of methane from cobalamin by cell-free extracts of *Methanobacillus kuzneceovii*," *Dokl. Akad. Nauk SSSR* **206**, 494–496 (1972).

72. A. M. Roberton and R. S. Wolfe, "Adenosine triphosphate pools in *Methanobacterium*," *J. Bacteriol.* **102**, 43–51 (1970).

73. J. G. Zeikus et al., "Oxidoreductases involved in cell carbon synthesis of *Methanobacterium thermoautotrophicum*," *J. Bacteriol.* **132**, 604–613 (1977).

74. M. S. Delafontaine, H. P. Naveau, and E. J. Nyns, "Fluorimetric monitoring of methanogenesis in anaerobic digesters," *Biotech. Let.* **1**, 71–73 (1980).

75. J. K. Kristjansson, P. Schonheit, and R. K. Thauer, "Different K_S values for hydrogen of methanogenic bacteria and sulfate reducing bacteria: An explanation for the apparent inhibition of methanogenesis by sulfate," *Arch. Microbiol.*, **131**, 278–282 (1982).

76. B. A. Huser, K. Wuhrman, and A. J. B. Zehnder, "*Methanothrix soehngenii* non-hydrogen oxidizing methanobacterium," *Arch. Microbiol.* **132**, 1–9 (1982).

77. G. Lettinga and N. N. Vinken, "Feasibility of the VASB process for the treatment of low strength wastes," *Proc. Purdue Ind. Wastes Conf.* (May 1980).

78. E. L. Iannotti et al., "The effect of digester fluids, swine manure extract and rumen fluid on the growth of bacteria from an anaerobic swine manure digester," *Dev. Ind. Microbiol.* **22**, 565–576 (1981).

79. M. P. Bryant, "Microbial methane production—theoretical aspects," *J. Anim. Sci.* **48**, 193–201 (1979).

80. R. A. Mah et al., "Biogenesis of methane", *Ann. Rev.Microbiol.* **31**, 309–341 (1977).

81. E. J. Kirsch and R. M. Sykes, "Anaerobic digestion in biological waste treatment," *Progr. Ind. Microbiol.* **9**, 155–239 (1971).

82. E. L. Iannotti, J. R. Fischer, and D. M. Sievers, "Medium for enhanced growth of bacteria from a swine manure digester," *Appl. Environ. Microbiol.* **43**, 247–249 (1982).

83. D. P. Chynoweth and R. A. Mah, "Bacterial populations and end products during anaerobic sludge fermentation of glucose," *J. Wat. Poll. Control Fed.* **49**, 405–412 (1977).

84. R. A. Mah and C. Sussman, "Microbiology of anaerobic sludge fermentation. I. Enumeration of the nonmethanogenic anaerobic bacteria," *Appl. Microbiol.* **16**, 358–361 (1968).

85. D. F. Toerien, "Population description of the nonmethanogenic phase of anaerobic digestion. I. Isolation, characterization and identification of numerically important bacteria," *Water Res.* **4**, 129–148 (1970).

86. D. R. Boone and M. P. Bryant, "Propionate-degrading bacterium, *Syntrophobacter wolinii* sp. nov. gen. nov. from methanogenic ecosystems," *Appl. Environ. Microbiol.* **40**, 626–632 (1980).

87. R. S. Wolfe, "Microbial biochemistry of methane—a study of contrasts," *Intl. Review Biochem. Microbiol Biochemistry* **21**, edited by Quayle (1970).

88. D. R.Caldwell and M. P. Bryant, "Medium without rumen fluid for non-selective enumeration and isolation of rumen bacteria," *Appl. Microbiol.* **14**, 794–801 (1966).

89. E. L. Iannotti, J. R. Fischer, and D. M. Sievers, "Medium for the enumeration and isolation of bacteria from a swine waste digester," *Appl. Environ. Microbiol.* **36**, 555–566 (1978).

90. C. Eller, M. R. Crabill, and M. P. Bryant, "Anaerobic roll tube media for nonselective enumeration and isolation of bacteria in human feces," *Appl. Microbiol.* **22**, 522–529 (1971).

91. E. J. Kirsch, "Studies on the enumeration and isolation of obligate bacteria from digesting sewage sludge," *Dev. Ind. Microbiol.* **10**, 170–176 (1969).

92. K. A. Pittman, K. Lakshmanan, and M. P. Bryant, "Oligopeptide uptake by *Bacteroides ruminicola*," *J. Bacteriol.* **93**, 1499 (1967).

93. M. P. Bryant and I. M Robinson, "Some nutritional characteristics of predominate culturable ruminal bacteria," *J. Bacteriol.* **84**, 605 (1962).

94. M. P. Bryant, "Nutritional requirements of the predominant rumen cellulolytic bacteria," *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **32**, 1809 (1973).

95. M. P. Bryant, "Nutritional features and ecology of predominant anaerobic bacteria of the intestinal tract," *Am. J. Clin. Nutr.* **22**, 1313 (1974).

96. R. S. Emery, C. K. Smith, and L. Fai To, "Utilization of inorganic sulfate by rumen microorganisms. II. The ability of single strains of rumen bacteria to utilize inorganic sulfate," *Appl. Microbiol.* **5**, 363 (1957).

97. R. E. Speece and Perry L. McCarty, "Nutrient requirements and biological solids accumulation in anaerobic digestion," *Adv. in Wat. Poll. Res.* **2**, 305–322 (1964).

98. D. J. Hoban and L. van den Berg, "Effect of iron on conversion of acetic acid to methane during methanogenic fermentations," *J. Appl. Bacteriol.* **47**, 153–159 (1979).

99. P. Scherer and H. Sahm, "Effect of trace elements and vitamins on the growth of *Methanosarcina barkeri*," *Acta Biotechnologica 1*, 57–65 (1981).

100. R. E. Speece, E. P. A. Seminar on Anaerobic Digestion, St. Louis.

101. R. Mah, R. Hungate, and K. Ohwaki, "Acetate, a key intermediate in methanogenesis," *Seminar on Microbial Energy Conversion*, edited by H. G. Schlegel (E. Goltz KG, Gottingen, Germany, 1976) 97–106.

102. R. S. Tanner, "Novel compounds from methanogens: Characterization of component B of the methylreductase system and mobile factor," Ph. D. Diss., Univ. of Illinois–Urbana, 1982.

103. A. M. Roberton and R. S. Wolfe, "ATP requirement for methanogenesis in cell extracts of *Methanobacterium strain M. o. H.*," *Biochem. Biophys. Acta*, 420–429 (1969).

104. D. O. Mountfort, "Effect of adenosine 5′-monophosphate on adenosine 5′-triphosphate activation of methyl coenzyme M methylreductase in cell extracts of *Methanosarcina barkeri*," *J. Bacteriology* **143**, 1039–1041 (1980).

105. G. Diekert, B. Weber, and R. K. Thauer, "Nickel dependence of factor F_{430} content in *Methanobacterium thermoautotrophicum*," *Arch. Microbiol.* **227**, 273–278 (1980).

106. J. M. Symons and McKinney, "The biochemistry of nitrogen in the synthesis of activated sludge," *Sew. and Ind. Wastes* **30**, 874 (1961).

107. S. R. Hoover and N. Porges, "Assimilation of dairy wastes by activated sludge II. The equation of synthesis and oxygen utilization," *Sew. and Ind. Wastes* **24**, 306 (1952).

108. D. O. Mountfort and R. A. Asher, "Effect of inorganic sulfide on the growth and metabolism of *M. barkeri*, strain DM," *Appl. Envir. Microbiol.* **37**, 670–675 (1979).

109. P. H. Ronnow and L. A. H. Gunnarsson, "Sulfide dependent methane production and growth of thermophilic methanogenic bacterium," *Appl. Envir. Microbiol.* **42**, 580–584 (1981).

110. P. H. Ronnow and L. A. H. Gunnarsson, "Response of growth and methane production to limiting amounts of sulfide and ammonia in two thermophilic methanogenic bacteria," *FEMS Microbiol. Lett.* **14**, 311–315 (1982).

111. L. A. H. Gunnarsson and P. H. Ronnow, "Variation of the ATP pool in thermophilic methanogenic bacteria during nitrogen or sulfur starvation," *FEMS Microbiol. Lett.* **14**, 317–320 (1982).

112. G. B. Patel, A. W. Khan, and L. A. Roth, "Optimum levels of sulfate and iron for the cultivation of pure cultures of methanogens in synthetic media," *J. Appl. Bacteriol.* **45**, 347–356 (1978).

113. A. W. Khan and T. M. Trotter, "Effects of sulfur-containing compounds on anaerobic degradation of cellulose to methane by mixed cultures obtained from sewage sludge," *Appl. Environ. Microbiol.* **35**, 1027–1034 (1978).

114. A. Wellinger and K. Wuhrman, "Influence of sulfide compounds in metabolism of methanobacterium strain AZ," *Arch. Microbiol.* **115**, 13–17 (1977).

115. A. J. B. Zehnder and K. Wuhrman, "Physiology of a *Methanobacterium* strain AZ," *Arch. Microbiol.* **111**, 199–205 (1977).

116. P. Scherer and H. Sahm, "Influence of sulfur containing compounds on the growth of *Methanosarcina barkeri* in a defined medium," *European J. Applied Microbiol. Biotechnol.* **12**, 28–35 (1981).

117. A. J. B. Zehnder et al., "Characterization of an acetate-decarboxylating, non-hydrogen-oxidizing methane bacterium" *Arch. Microbiol.* **124**, 1–11 (1980).

118. A. W. Khan and T. M. Trotter, "Effects of sulfur-containing compounds on anaerobic degradation of cellulose to methane by mixed cultures obtained from sewage sludge," *Appl. Environ. Microbiol.* **35**, 1027–1034 (1978).

119. A. W. Lawrence and P. L. McCarty, "The effects of sulfides on anaerobic treatment," *Int. J. Air and Wat. Poll.* **10** 207–221 (1966).

120. D. Wise, personal communication, 1982.

121. P. Schonheit, J. Moll, and R. K. Thauer, "Nickel, cobalt and molybdenum requirement for growth of *Methanobacterium thermoautotrophicum*," *Arch. Microbiol.* **123**, 105–107 (1979).

122. J. B. Jones and T. C. Stadtman, "*Methanococcus vannielii*: culture and effect of selenium and tungsten on growth," *J. Bact.* **130**, 1404–1406 (1977).

123. G. T. Taylor and S. J. Pirt, "Nutrition and factors limiting the growth of a methanogenic bacterium. (*Methanobacterium thermoautotrophicum*)," *Arch. Microbiol.* **113**, 17–22 (1977).

124. P. L. McCarty and C. A. Vath, "Volatile acid digestion at high loading rates," *Int. J. of Air and Wat. Poll.* **6**, 65–73 (1962).

125. W. D. Murray and L. van den Berg, "Effects of nickel, cobalt, and molybdenum on performance of methanogenic fixed-film reactors," *Appl. Environ. Microbiol.* **42**, 502–505 (1981).

126. M. E. Jack, G. J. Farquhar, and G. M. Cornwall, "Anaerobic digestion of primary sludge containing iron phosphate," *Water Poll. Res. in Canada* **8**, 91–109 (1976).

127. C. E. Lankford, "Bacterial assimilation of iron," *CRC Critical Reviews of Microbiol.*, 273–331 (1973).

128. M. R. Winfrey and J. G. Zeikus, "Effect of sulfate on carbon and electron flow during microbial methanogenesis in freshwater sediments," *Appl. and Environ. Microbiol.* **33**, 275–281 (1977).

129. Cecil Lue-Hing, Chicago Metropolitan Sanitary District, personal communcation, 1974.

130. J. Yang et al., "The response of methane fermentation to cyanide and chloroform," *Prog. Wat. Tech.* **12**, 977–989 (1980).

131. W. Lin Chou et al., "The effect of petrochemical structure on methane fermentation toxicity," *Prog. Wat. Tech.* **10**, 545–550 (1978).

132. W. Lin Chou, R. E. Speece, and R. H. Siddiqi, "Acclimation and degradation of petrochemical wastewater components by methane fermentation," *Biotech. and Bioeng. Symposium No. 8*, 391–414 (1978).

133. J. B. Healy, Jr., and L. Y. Young, "Anaerobic biodegradation of eleven aromatic compounds to methane," *Appl. Envir. Microbiol.* **38**, 84–89 (1979).

134. C. Weng and John S. Jeris, "Biochemical mechanisms in the methane fermentation of glutamic and oleic acids," *Wat. Res.* **10**, 9–18 (1976).

135. A. W. Lawrence and P. L. McCarty, "Kinetics of methane fermentation in anaerobic treatment," *J. Wat. Poll. Control* **41**, R1–R17 (1969).

136. I. J. Kugelman and K. K. Chin, "Toxicity, synergism, and antagonism in anaerobic waste treatment processes in *Anaerobic Biological Treatment Processes.* Advances in Chemistry Series," **105**, 55–90 (1971).

137. J. F. Ferguson, B. J. Eis, and M. M. Benjamin, "Neutralization in anaerobic treatment of an acidic waste," I. A. W. P. R. Specialized Seminar–Anaerobic Treatment of Wastewater in Fixed Film Reactors, Copenhagen, Denmark, pp. 339–355 (1982).

138. Metcalf and Eddy, *Wastewater Engineering: Treatment Disposal Reuse*, (New York, McGraw-Hill, 1979).

139. P. L. McCarty, "Anaerobic waste treatment fundamentals," *Public Works*, Sept. Oct. Nov. Dec., 1964.

140. J. T. Pfeffer, "Increased loadings on digesters with recycle of digested solids," *J. Wat. Poll. Control Fed.* **40**, 1920–1933 (1968).

141. J. T. O'Rourke, "Kinetics of anaerobic treatment at reduced temperature," Ph. D. Diss., Stanford Univ., 1968.

142. D. O. Mountfort and M. P. Bryant, "Isolation and characterization of an anaerobic, syntrophic bacterium degrading benzoate from sewage sludge," *Arch. Microbiol.* (in press).

143. J. E. Andrews and S. P. Graef, "Dynamic modeling and simulation of the anaerobic digestion process," *Anaerobic Biological Treatment Processes*, Adv. in Chemistry Series, **105**, 126 (1971).

144. R. E. Speece and J. Kem, "The effect of short-term temperature variations on methane production," *J. Wat. Poll. Control Fed.* **42**, 1990–1997 (1970).

145. A. W. Lawrence and P. L. McCarty, "The role of sulfide in preventing heavy metal toxicity in anaerobic treatment," *J. Wat. Poll. Control Fed.* **37**, 392–409 (1965).

146. P. N. Hobson, S. Bousfield, and R. Summers, *Methane Production from Agricultural and Domestic Wastes*, Halsted Press Div. of John Wiley & Sons, 1981.

147. P. N. Hobson, *Microbial Polysaccharides and Polysaccharases*, edited by R. C. W. Berkeley, G. W. Gooday, and D. C. Ellwood, (Academic Press, London and New York, 1979) p. 377.

148. S. Ghosh and D. L. Klass, *Proc. Biochem.* **13**, 15 (1978).

149. P. N. Hobson, Solar Energy in Agriculture Joint Conf. (C9), Univ. of Reading and UK-ISES, UK Section of ISES, London, pp. 12–14.

150. R. Summers, *Report on Straw Utilization Conf.*, MAFF, p. 84 (1978).

151. J. T. Pfeffer, *Proc. Biochem.* **13**, 6 (1978).

152. J. E. Robbins, M. T. Arnold, and S. L. Lacher, *Appl. Environ. Microbiol.* **38**, 175 (1979).

153. R. E. Hungate, *Bact. Rev.* **14**, 1 (1950).

154. L. R. Maki, "Experiments on the microbiology of cellulose decomposition in a municipal sewage plant," Antonie von Leewenhoek, **20**, 185–200 (1954).

155. P. N. Hobson, B. G. Shaw, *Water Res.* **8**, 507 (1974).

156. P. N. Hobson and B. G. Shaw, *Microbial Aspects of Pollution*, edited by G. Sykes and F. A. Skinner (Academic Press, London and New York, 1971) p. 103.

157. D. F. Torien, *Water Res.* **1**, 147 (1967).

158. P. N. Hobson, S. Bousfield, and R. Summers, "Anaerobic digestion of organic matter," *CRC Critic. Rev. Environ. Control* **4**, 131–191 (1974).

159. J. W. Gossett and R. L. Belser, "Anaerobic digestion of waste activated sludge," *J. Env. Eng. Div. ASCE.* **108**, 1101–1120 (1982).

160. Y. R. Chen and A. G. Hashimoto, "Kinetics of methane fermentation," *Biotech. and Bioeng.*, Symp. No. **8**, 269–282 (1978).

161. A. D. Carr and R. C. O'Donnell, "The dynamic behavior of an anaerobic digester," *Prog. Wat. Tech.* **9**, 727–738 (1977).

162. J. Olesczkiewicz, Seminar on Anaerobic Wastewater Treatment and Energy Recovery by Duncan, Lagnese Assoc. Inc., Pittsburgh, Pa., 1981.

163. F. Pearson, C. Shium-Chung, and Maryse Gauteer, "Toxic inhibition of anaerobic digestion," *J. Wat. Poll. Control Fed.* **52**, 472–482 (1980).

164. Brian Eis et al., "The fate and effect of bisulfite in anaerobic treatment," 1981 Annual *Wat. Poll. Control Fed. Mtg. Detroit.* Preprint.

165. R. D. Neufeld, J. D. Mack, and J. P. Strakey, "Anaerobic phenol biokinetics," *J. Wat. Poll. Control Fed.* **52**, 2367–2377 (1980).

166. G. F. Parkin and R. E. Speece, "Modeling toxicity in methane fermentation systems," Am. Soc. Civil Eng., EE3, June 1982, p. 515.

167. A. M. Buswell and H. F. Mueller, "Mechanisms of methane fermentation," *Ind. Eng. Chem.* **44**, 550 (1952).

168. P. L. McCarty, "Stoichiometry of biological reactions," Intl. Conf. "Toward a Unified Concept of Biological Waste Treatment Design," Atlanta, Ga. (1972).

169. P. L. McCarty, "Anaerobic processes," Birmingham Short Course on Design Aspects of Biological Treatment, I. A. W. P. R.

170. J. C. Young and Perry L. McCarty, "Anaerobic filter for waste treatment," *J. Wat. Poll. Control* **41**, 5 (1969).

171. J. B. Coulter, S. Soneda, and M. B. Ettinger, "Anaerobic contact process for sewage disposal," *Sew. and Ind. Wastes*, **29**, 648 (1957).

172. W. A. Pretorius, "Anaerobic digestion of raw sewage," *Wat. Res.* **5**, 681–687 (1971).

173. G. Lettinga et al., "Use of upflow sludge blanket (USB) reactor concept for biological wastewater treatment, especially for anaerobic treatment," *Biotech. and Bioeng.*, **22**, 699–734 (1980).

174. J. S. Jeris, DOE workshop, "Anaerobic Filter. An Energy Plus for Wastewater Treatment," Howey-in-the-Hills, Fla., ANL-CNSV-TM-50, 1980, pp. 171–178.

175. M. T. Suidan and W. H. Cross, Ibid., pp. 59–76.

176. A. A. Friedman, Ibid., pp. 95–114.

177. W. N. Torpey and N. R. Melbinger, "Reduction of digested sludge volume by controlled recirculation," *J. Wat. Poll. Control Fed.* **39**, 1464–1474 (1967).

178. H. O. Buhr and J. F. Andrews, "The thermophilic anaerobic digestion process," *Wat. Res.* **11**, 129–143 (1977).

179. H. O. Buhr and J. F. Andrews, "The thermophilic anaerobic digestion process," *Wat. Res.* **11**, 129–143 (1977).

180. C. G. Golueke, "Temperature effects an anaerobic digestion of raw sewage sludge," *Sew. and Ind. Wastes J.* **30**, 1225 (1958).

181. W. F. Garber, "Operating experience with thermophilic anaerobic digestion," *J. Wat. Poll. Control Fed.* **54**, 1170–1175 (1982).

182. R. R. Rimkus, J. M. Ryan, and E. J. Cook, "Full-scale thermophilic digestion at the West-Southwest Sewage Treatment Works, Chicago, Illinois," *J. Wat. Poll. Control Fed.* **54**, 1447–1457 (1980).

183. H. M. Popova and O. T. Bolotina, "The present state of purification of town sewage and the trend in research work in the City of Moscow," *Int. J. Air and Wat. Pol.* **7**, 145 (1963).

184. J. T. Pfeffer, "Methane fermentation engineering aspects of reactor design," 34th Int'l. Conf. of Cebedeau, Belgium, 1981.

185. *Handbook of Chemistry and Physics*, edited by R. C. Weast (CRC Press, 1979).

186. S. Ghosh, "Energy production efficiency in anaerobic digestion—a basis for advanced process development," presented at annual meeting, Wat. Poll. Control Fed., 1981.

187. V. H. Varel, A. G. Hashimoto, and Y. R. Chen, "Effect of temperature and retention time on methane production from beef cattle waste," *Appl. Environ. Microbiol.* **40**, 217–222 (1980).

188. K. J. Kennedy and L. van den Berg, "Effects of temperature and overloading on the performance of anaerobic fixed film reactors," Purdue Ind. Wastes Conf., 1981.

189. C. L. Cooney and D. L. Wise, "Thermophilic anaerobic digestion of solid waste for fuel gas production," *Biotech. and Bioeng.* **17**, 1119–1135 (1975).

190. L. van den Berg, C. P. Lentz, and D. W. Armstrong, "Methane production rates of anaerobic fixed film fermenters as compared to those of anaerobic contact and fully mixed continuous fermenters," *Advances in Biotechnology* **II**, 251–256 (1981).

191. Emschergenossenschaft, Bothrop Central Sludge Treatment Plant—Status: June 1978.

192. R. T. Haug et al., "Effect of thermal pretreatment on digestibility and dewaterability of organic sludges," *J. Wat. Poll. Control* **50**, 73–85 (1978).

193. P. L. McCarty et al., "Thermochemical treatment of lignocellulosic and nitrogenous residuals for increasing anaerobic biodegradability," Second Annual Symposium on Fuels from Biomass, Renssalaer Polytechnical Inst., June 1978.

194. P. C. Singer et al., "Assessment of coal conversion wastewaters: Characterization and preliminary biotreatability." EPA-600/7-78-181, Sept. 1978.

195. A. M. Buswell and S. L. Neave, *Laboratory studies of sludge digestion, III.* State Water Survey, Bulletin No. 30 (1930).

196. G. E. Symons and A. M. Buswell. "The methane fermentation of carbohydrates," *J. Am. Chem. Soc.* **55** (1933).

197. D. Tarvin and A. M. Buswell, "The methane fermentation of organic acids and carbohydrates," *J. Am. Chem. Soc.* **56** (1934).

198. A. M. Buswell and W. F. Hatfield, *Anaerobic Fermentations*, Illinois State Water Survey, Bulletin No. 32 (1936).

199. J. C. Hovious, R. A. Conway, and C. W. Ganze, "Anaerobic lagoon pretreatment of petrochemical wastes," *J. Wat. Poll. Control* **45**, 71 (1973).

200. J. S. Jeris et al., "The biochemistry of anaerobic digestion," Final Report: Research Grant 17070 DFK, Fed. Wat. Poll. Control Admin., August 1970.

201. J. M. Gossett et al., "Heat treatment of refuse for increasing anaerobic biodegradability." *Civil Engineering Technical Report No. 205*, Stanford Univ., June 1976.

202. J. Chmielowski and W. Wasilewski, "An investigation of the dynamics of anaerobic decomposition of some phenols in the methane fermentation," *Zeszyty Naukowe Politechniki Slaskij, Inzy ieria Sanitaria* **9**, 121 (1966).

203. E. R. Witt, W. S. Humphrey, and T. E. Roberts, "Full scale anaerobic filter treats high strength wastes," *Proc. Purdue Ind. Waste Conf.*, 1979.

204. W. F. Owen et al., "Bioassay for monitoring biochemical methane potential and anaerobic toxicity," *Water Research* **13**, 485–492 (1979).

205. B. P. Sen and T. R. Bhaskaran, "Anaerobic digestion of liquid molasses distilling wastes," *J. Wat. Poll. Control.*, **34**, 1015 (1962).

206. G. J. Stander et al., "Treatment of wine distillery wastes by anaerobic digestion," *Proc. 22nd Purdue Ind. Wastes Conf.* (May 1967).

207. H. A. Painter, "Treatment of malt whiskey distillery wastes by anaerobic fermentation," *Brewers' Guide* (Brit.) (August 1960).

208. R. N. Guzman, "Control of caning sugar wastes in Puerto Rico," *J. Wat. Poll. Control* **34**, 1213 (1962).

209. A. M. Buswell, P. E. Gaffney, and R. S. Ingols, "Anaerobic digestion treats cotton mill desize waste," *Wastes Eng.* **35** 402 (1962).

210. E. E. Grishna, "The purification of wastewaters from wool washing by means of anaerobic fermentation," *Chem. Abs.*, **64** (1964).

211. M. Tanaka, M. Dazai, and H. Ono, "Biological treatment of wool spinning wastes," *Chem. Abs.* **64**, 19181 (1966).

212. M. J. Nevzorov, "Clarification of industrial wastes by anaerobic fermentation," *Chem. Abs.* **64**, 12355 (1964).

213. E. Hindin and G. H. Dunstan, "Anaerobic digestion of potato processing wastes," *J. Wat. Poll. Control* **35**, 486 (1963).

214. O. O. Olson, J. W. Vennes, and W. von Huevelen, "Microbiological treatment of protein water waste from potato starch manufacturing." Official Bull., N. Dakota Wat. and Sew. Wks., **34**, 11 (1966).

215. C. I. Ivanov, "Anaerobic purification of tannery waste waters," *J. Am. Leather Chem. Assoc.* **58**, 389 (1963).

216. W. E. Gates and S. D. Lin, "Pilot plant studies of anaerobic treatment of tannery effluents," *J. Am. Leather Chem. Assoc.* **61**, 516 (1966).

217. G. S. Schroepfer and N. R. Ziemke, "Development of the anaerobic contact process I. Pilot plant investigations and economics," *Sew. and Ind. Wastes* **31**, 164 (1959).

218. D. K. Silvester, "The treatment of slaughterhouse waste by anaerobic digestion," *J. Inst. of Public Health Eng.* **61**, 266 (1962).

219. A. J. Steffen and M. Bedker, "Operation of full scale anaerobic contact treatment plant for meat packing wastes," *Proc. 16th Purdue Ind. Wastes Conf.*, 423 (1961).

220. M. R. Henzen, Report on Overseas Visit, National Institute for Water Research, P. O. Box 395, Pretoria, South Africa, 1966.

221. L. Bradney, W. Nelson, and R. E. Bragstad, "Treatment of meat packing wastes," *Sew. and Ind. Wastes* **22**, 807 (1950).

222. W. C. Hiatt, A. D. Carr, and J. F. Andrews, "Anaerobic digestion of rum distillery wastes," *Proc. 28th Purdue Ind. Wastes Conf.*, 1973.

223. D. W. Taylor and R. J. Burm, "Full scale anaerobic filter treatment of wheat starch plant wastes," 71st Am. Inst. Chem. Eng. Mtg., 1970.

224. R. Koehler,"Anaerobic abban von Jefekabrikabwasser," *Wasser Luft, Betrieb* **17**, 342 (1973).

225. C. Jennett. See Jeris[174], pp. 77–94.

226. M. S. Switzenbaum. See Jeris[174], pp. 115–128.

227. John Ferguson, University of Washington, personal communication, 1978.

228. E. R. Hall, "Biomass retention and mixing characteristics in fixed-film and suspended growth anaerobic reactors," I. A. W. P. R. Specialized Seminar—Anaerobic Treatment of Wastewater in Fixed Film Reactors, Copenhagen, Denmark, 1981, pp. 371–396.

CHAPTER 3

PRINCIPLES AND TECHNOLOGY OF BIOMASS GASIFICATION

Thomas B. Reed

3.1 ABSTRACT

After a dormant period following World War II, biomass gasification entered a period of rapid development. The major types of gasifiers and their various modes of operation are discussed. The role of thermodynamic equilibrium and kinetics in gasifier operation is presented. The mechanisms and kinetics of pyrolysis and char gasification are reviewed and the current state of gasifier modeling is surveyed.

3.2 INTRODUCTION

3.2.1 History of Biomass Gasification

Biomass has been the principal chemical fuel of man's activities for 100,000 years. It can supply the U. S. with at least 10 quads of energy (10^{19} J or 10^{16} Btu) per year and the world as a whole with far more on a renewable basis.[1] Unfortunately, biomass is often disperse, wet, and difficult to use. However, gasification can form the bridge between this raw energy resource and the liquid or gaseous fuels upon which our modern society currently depends. Figure 3.1 shows the various products that gasification and combustion can yield.

The process of gasification was discovered independently in France and England in 1798. By 1850, the technology had advanced to the point that much of London was lit by manufactured gas or "town-gas."[2] Manufactured gas soon came to the U. S. and by 1920 most towns and cities had a "gasworks" to supply town-gas for cooking and lighting. In 1930, the first natural gas pipeline was constructed from the oil fields of Texas to Denver. As the pipelines crisscrossed the country, making very low-cost natural gas widely available, manufactured gas disappeared and now

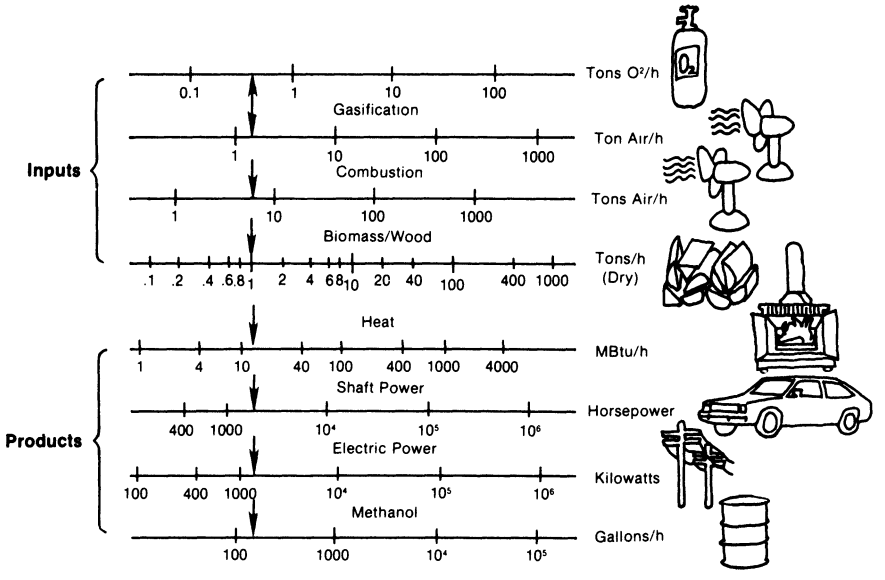


Fig. 3.1: Biomass/wood conversion equivalents. Arrows indicate input requirements and output values for 1 ton of air dry wood. (Requires 0.4 tons oxygen or 1.5 tons of air for gasification; requires 6.4 tons air for complete combustion. Contains 16 MBtu; yields 12 MBtu when burned at 75% efficiency; yields 1406 kWh or 1886 hph at 30% conversion efficiency; yields 160 gal methanol at 60% conversion efficiency.) Shown in common U.S. units. See Appendix A for English/SI conversion factors.

this extensive industry has been forgotten. Town-gas was used in England into the 1970s, but with the advent of North Sea oil, these plants were all dismantled.

Town-gas producers generally used coal for large-scale production. Starting about the time of World War I, small gasifiers were developed using charcoal and biomass for operating vehicles, boats, trains, and small electric generators. Between the wars, development on this scale was done largely by amateur tinkerers because gasoline was relatively inexpensive and simpler to use than biomass. However, as World War II became imminent, many countries speeded development of small biomass gasifiers, and by 1943, 820,000 were in operation worldwide.[3] When cheap gasoline returned, the gasifiers, of course, were abandoned. The historical, technical, and scientific record of this experience has been well documented by the Swedish Academy of Engineering in the book *Gen-Gas*, which was translated into English recently.[4]

Gasification will probably fill an important niche in biomass conversion because it can produce the wide spectrum of high value products shown in the equivalence chart of Fig. 3.1. A number of factors, such as the rising cost of oil, improved biomass production methods, new conversion technologies, and increased government incentives for renewable energy will all hasten the development of biomass gasification. The bibliography below including a number of major works shows that we have already entered a period of rapid development.[4-16]

Table 3.1

Bibliographic citations by year in National Academy of Sciences (NAS) Producer Gas Bibliography[9]

Years	Citations	Years	Citations
1901–1910	4	1941–1945	212
1911–1920	3	1946–1950	3
1921–1930	14	1951–1060	0
1931–1937	22	1961–1970	0
1931–1940	134	1971–1975	3
		1976–1981	23

In this paper, I will first acquaint the reader with the principal reference works and research organizations in this field. The technical and scientific basis of gasification and the latest experimental and theoretical work in the field will then be discussed.

3.2.2 Major References on Biomass Gasification

While present activity in gasification is high, it is small compared to the activity between 1850 and 1950. The reader is referred especially to Rambush[17] for updraft and coal gasifier experience, to Generator Gas[4] for downdraft gasifiers, and to Gumz[18] for the thermodynamics and kinetics of coal and wood gasification.

One indication of the recent growth of interest in biomass gasification is the number of publications written over the years. In a recent bibliography published by the National Academy of Sciences on vehicle gasifiers,[9] the number of papers per year roughly indicates the degree of interest in each period, and this distribution is shown in Table 3.1.

In the last five years, seven volumes of over 200 pages each have been published on biomass gasification in addition to hundreds of more specialized papers that have not yet been collected in the above bibliography. The above figures, while showing the explosive interest in small vehicle gasifiers during World War II, only suggest the reemergence that is now underway. Other bibliographies have appeared recently and some are computerized for ready access.[8]

3.2.3 Gasification Research Centers

As is evident from Table 3.1, there was very little interest in gasification as soon as low-cost gasoline returned following World War II. Only Sweden continued to work in this area, maintaining a program at the National Machinery Testing Institute in Uppsala.[19]

The recent interest in gasifiers occurs in many fields—government, university, industry, and so on—and the number of publications in each does not necessarily reflect the level of interest or funding.

The Environmental Protection Agency (EPA) initiated a number of gasification demonstrations for municipal waste disposal in the early 1970s and supported relevant research at universities. This support was discontinued in 1978, leaving a residue of personnel with experience in this area and a few partially successful processes.[12]

A moderate level of funding ($2 to 20 million per year) has been maintained since 1975 by the U. S. Department of Energy (DOE) for what may be called "advanced concept" gasification and pyrolysis. Most of the work is aimed at large industrial processes and is supported in government laboratories, industrial firms, and universities. The work going on in various university departments around the world, however, is difficult to trace if it is unfunded or on a small scale. The work of Goss and his students at the University of California at Davis deserves special mention because of its long continuity and because it includes both experimental and theoretical studies.[8] Twente University in Holland has had a large program in gasification for several years.

Another source of gasifier information is that provided by companies developing gasifier systems for sale. These groups write advertising brochures as often as scientific articles, and it is sometimes difficult to separate actual from projected performance.

Excellent work in gasification is proceeding in Canada, Europe, Brazil, New Zealand, South Africa, and other parts of the world, primarily at the university level.[20] Commercial gasification of municipal wastes is particularly advanced in Japan and Europe.[12]

Finally, there are a large number of individuals carrying out programs leading to small gasifiers which are relatively inexpensive to build and test. There are many articles in magazines for general distribution, often quite informative, but sometimes lacking perspective and engineering data.[21,22]

3.2.4 Gasification Meetings

The work done in gasification is reported at a wide variety of meetings and recorded in their symposia, although very few of these are devoted to biomass alone. The Institute for Gas Technology has sponsored seven yearly meetings in Florida in which a number of papers have appeared on gasification.[16] The Department of Energy holds regular meetings for the thermochemical contractors they support; proceedings of these meetings are available form the NTIS.[11] (The DOE supported work is also often presented at other symposia.[6,10,12,16]

The Bioenergy Council has sponsored several large meetings on biomass and, in addition, has published yearly catalogues of research activities and publications.[23] The Forest Products Research Society holds a yearly meeting devoted to wood energy and publishes the proceedings.[15]

In Canada, the Biomass Energy Institute has held yearly meetings with symposia.[24] The Commission of the European Communities, in cooperation with the Department of Energy of the UK, has sponsored several symposia on biomass activities in Europe.[20,25]

The above meetings and symposia deal more generally with energy from biomass, but typically will contain a few to a dozen papers on gasification. Very few meetings have been devoted exclusively to gasification. Retrofit '79 was a workshop devoted to the use of gasifiers to replace gas and oil in heating applications.[5] While most of these meetings deal generally with the larger subject of biomass, the International Conference "Fundamentals of Thermochemical Biomass Conversion" in October 1982 focused very specifically on the thermal conversion of biomass and the papers are referenced subsequently in this review, even though the proceedings have not yet appeared.[14] The Beijer Institute of Sweden held a week long meeting on gasification in Sri Lanka in November 1982.[26]

3.3 GASIFIER TECHNOLOGY

3.3.1 Types of Gasifiers

3.3.1.1 Charcoal Gasifiers. Technically, charcoal is the easiest biomass-derived fuel to gasify. Air and steam are admitted to the hot charcoal on a grate (Fig. 3.2) and produce CO and H_2 (producer gas) which can be burned in vehicles without cleanup.[4] However, less than 40% of the energy of the biomass can be converted to charcoal, so that unlesss the charcoal manufacture recovers the balance of the energy in wood chemicals, charcoal gasifiers are very wasteful of energy. In addition, charcoal forms a great deal of dust and easily absorbs large quantities of water. For these reasons, charcoal gasifiers were common at the beginning of World War II, but as soon as wood gasifiers were developed the charcoal gasifiers disappeared.

An interesting variation on charcoal gasification is the Cargas System now being mass produced in Brazil. A 4 MBtu/h gasifier-burner uses a special bagged charcoal charred at very low temperatures so that most of the total energy content of the wood is still preserved while the energy density increases from 8000 to 14,000 Btu/lb. The fuel will be, in part, fast-growing eucalyptus now grown in Brazil for reforestation.[27]

3.3.1.2 Updraft Gasifiers. The earliest gasifiers used were patterned on the updraft design shown in Fig. 3.2. Air (or oxygen) is admitted to a grate on which charcoal burns. The resulting CO_2 is immediately reduced to CO by the excess carbon present. Due to the very high heat of reaction, it is usually necessary to add steam, move the grate, or in other ways moderate the temperature. The hot gases so formed then become a source of heat to pyrolyze and dry the incoming biomass (or coal), producing a heavy tarry pyrolysis oil which can contain up to 30% of the input energy. This vapor is easily burned if the temperature is kept high enough to prevent condensation. The Lurgi gasifier is an example of a commercial updraft gasifier for coal. The APCO gasifier is an example of a commercial updraft gasifier for biomass.[12,16] It uses up to 36 dry tons of hogged (ground up) wood per day to provide heat for the state hospital at Rome, Georgia.

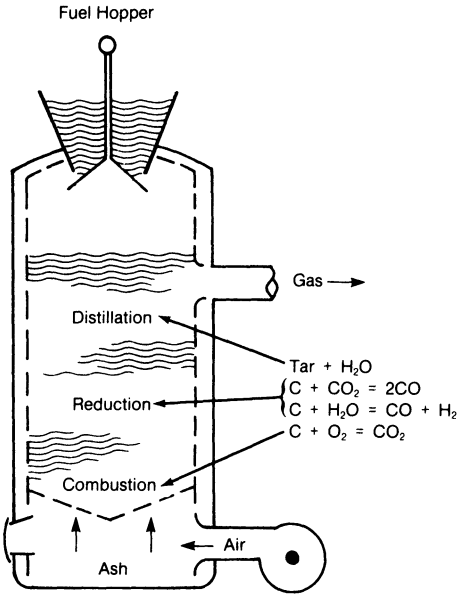


Fig. 3.2: Schematic diagram of updraft gasifier showing reactions occurring in each zone.

The advantage of updraft gasifiers is that they can use a wide variety of high-moisture fuels. The disadvantages are that the tar vapors are almost impossible to remove and are highly corrosive, and so must be burned. Thus, updraft biomass gasifiers are used primarily for retrofitting boilers. If high value uses can be found for biomass tars, updraft gasifiers may become more widely used since updraft gasification is an inexpensive way to generate a liquid fuel from biomass. (All fixed-bed coal gasifiers are updraft because in coal gasification the primary problem is to convert the relatively unreactive coke, or char, to gas and because the relatively small amount of volatiles have a high value as coal liquids.)

While the high grate temperatures in updraft gasifiers may generally be considered a disadvantage, they become an advantage in the processing of municipal solid waste (MSW) which can contain up to 35% ash (dry basis). The Andco Torax process uses air preheated in checkers to gasify MSW and produces a dense slag which is much easier to dispose of than partially melted bottles and cans.[28] The Purox process uses oxygen in an updraft gasifier to give a medium Btu gas (300 Btu/scf) and also produces sufficient heat to melt the inorganic materials in MSW. A 40 ton/day unit has been operated in Japan for several years.[29]

3.3.1.3 Downdraft Gasifiers. In the downdraft gasifier (Fig. 3.3), the air or oxygen is usually admitted to the biomass through nozzles (tuyerres) causing pyrolysis to charcoal and volatile tars that partially burn as they are produced. The gaseous products of this "flaming pyrolytic combustion" (see below) then consume the charcoal produced during this pyrolysis and are reduced to fuel gas. In this way, tar vapors are typically lowered to 0.1% of the total feed. This gas is much more

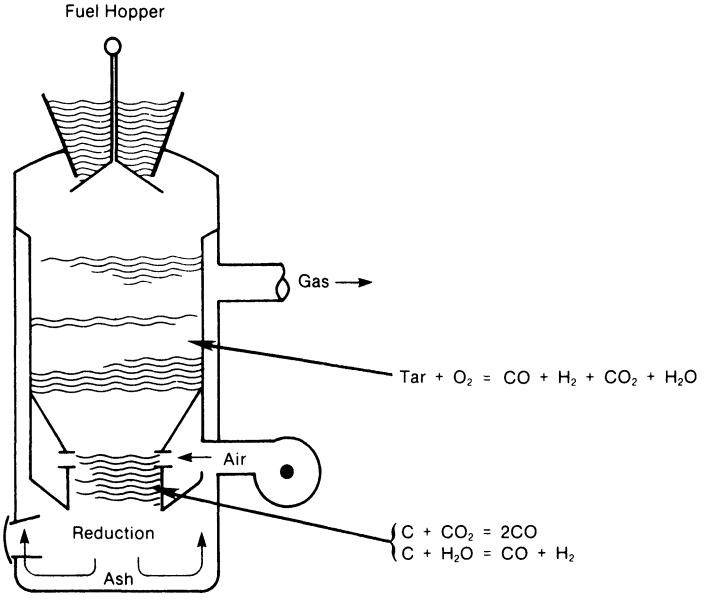


Fig. 3.3: Schematic diagram of downdraft gasifier showing reactions occurring in each zone.

suitable for operation of burners, engines, turbines, or transport in pipelines. Another advantage of downdraft gasification is that the initial reaction of air or oxygen with biomass produces lower temperatures than reaction of air/oxygen with charcoal, due to the endothermic pyrolysis reactions. Thus, the downdraft gasifier does not require steam or other cooling at the grate and produces the heat for pyrolysis by burning the tars in situ.

However, there are disadvantages in downdraft gasification. The gas leaves the charcoal bed at temperatures of 700°to 800°C where the char/gas reactions become slow. To achieve the highest efficiencies, it is necessary to recover the sensible heat which represents about 10% of the energy in the gas by using heat exchangers. The fuels for downdraft gasifiers must be more uniform in size and lower in moisture content than fuels for updraft gasifiers.

Downdraft gasifiers are more difficult to scale up than are updraft gasifiers because the air or oxygen enters at the edge of the reactor and has difficulty penetrating a wide bed. One solution has been to use a central lateral injector as well as the edge nozzles,[30] or to introduce air through a central moving agitator/inlet.[31]

A test was made at Environmental Energy Engineering of West Virginia using oxygen in a World War II gasifier. The temperature in the char zone increased only about 100°C.[32] The thermodynamic equilibrium reaction temperature for biomass gasification lies in the 650°to 850°C range, though local temperatures can, of course, be much higher. This unexpected result suggests that the downdraft gasifier might

be modified to operate with oxygen and still operate at relatively low temperature provided biomass were fed continually to the reaction zone.

An oxygen gasifier for methanol production has been developed at SERI based on modifications of the downdraft gasifier (Fig. 3.4). Biomass and oxygen are fed onto the top of a bed of hot charcoal where the biomass pyrolyzes very rapidly, and the resulting tars are burned in the oxygen. The resulting partial combustion gases then move down through a deep bed of charcoal where they reach equilibrium.[33] Provision is also made to inject oxygen at the edges of the bed to prevent too much heat from being generated in the top combustion seciton. This 1 ton/day gasifier has been operated extensively at both atmospheric pressure and 10 atmospheres. It is now ready for scaleup for commercial application.[34]

During operation of the oxygen gasifier it was found that injection of oxygen to the charcoal bed was not necessary and that all of the oxygen and biomass could be fed together at the top of the bed. In this mode, the flow of oxygen and biomass is uniform across the gasifier cross section; and the downward flow of gas, pyrolyzing biomass, and then charcoal is "stratified" in layers as opposed to earlier gasifiers in which air/oxygen was admitted through nozzles at discrete points. A new "stratified downdraft gasifier" is now being developed at SERI which uses this principle (see modeling section).

A commercial gasifier built on these principles is now being marketed by the Buck Rogers Co. in Olathe, Kansas, and has been evaluated at Kansas State University. The gasifier feeds chips or other fuels into an open vertical stainless steel cylindrical tube. Air is distributed at an intermediate level through rotating stirring arms, and the gas is taken off from a moving grate, as shown in Fig. 3.5. Average operating conditions measured at Kansas State University on a nominal 1.5 MBtu/h gasifier are shown in Table 3.2.[35]

3.3.1.4 Fluidized Bed Gasifiers. Recently, a great deal of work has been done on fluidized beds (Fig. 3.6). In these units a hot, inert solid may recirculate with the biomass to supply high heat transfer for pyrolysis tar cracking and char reduction. Fluidized beds can handle small and mixed particles and wet fuels more easily than fixed beds. Many of the potential advantages of fluidized beds are difficult to realize with biomass because the low density of charcoal causes the charcoal to float on top of the bed rather than recirculate. This results in a "once through" bed and diminishes contact time so that very tall units are required for complete cracking.

The catalytic steam gasification of wood is being investigated at the Pacific Northwest (Battelle) Laboratory initially at atmospheric and now at 10 atmosphere pressure. Wood chips are continuously fed submerged into a bed of catalyst and char, fluidized with superheated steam. Various catalysts are used to alter the gas composition. Work is focused on developing long lived catalysts.[36]

The Omnifuel gasifier is an example of a successful fluidized bed gasifier.[37] A 130 (bone dry) Omnifuel tons/day fluidized bed was installed at the Levesque Plywood Ltd. mill in Hearst, Ontario, and operated between June 1981 and mid-1983. Typical gas analyses for this gasifier are shown in Table 3.3. This gasifier is

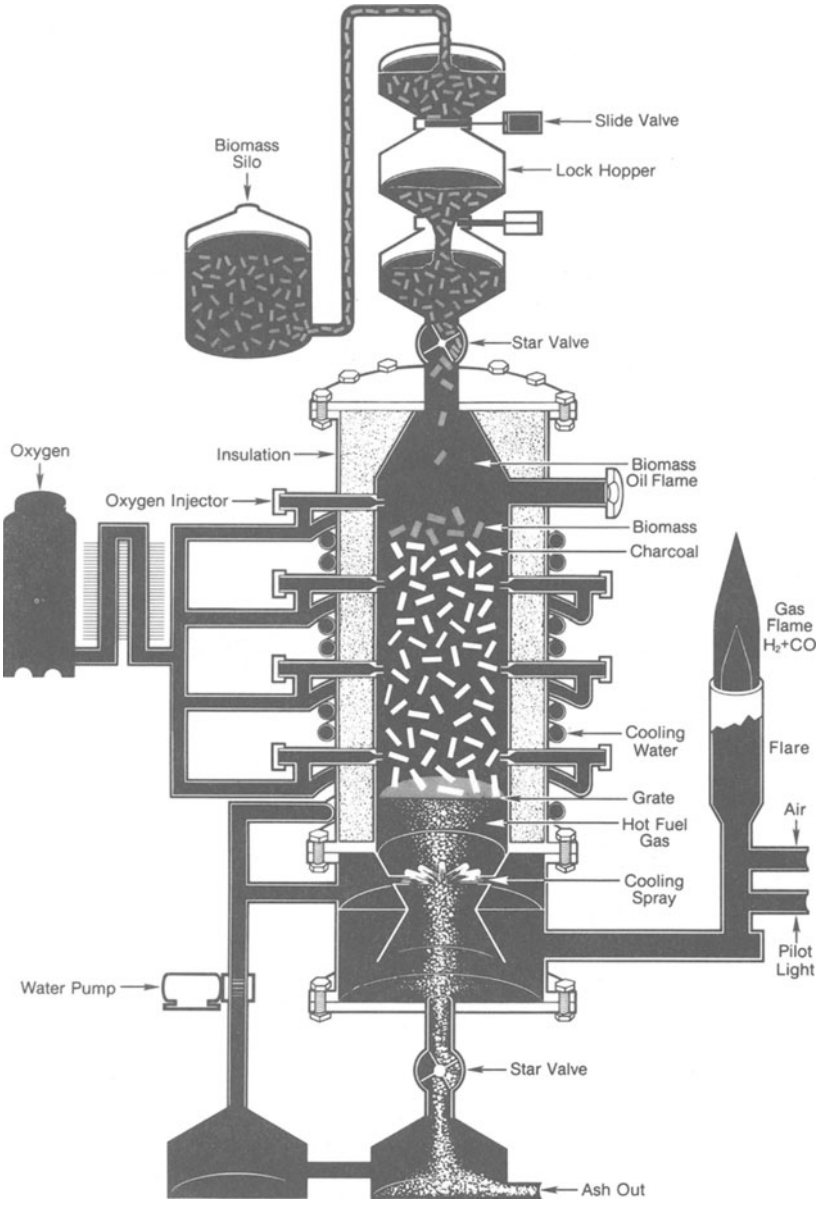


Fig. 3.4: SERI 1 ton/day high pressure oxygen downdraft biomass biomass gasifier.

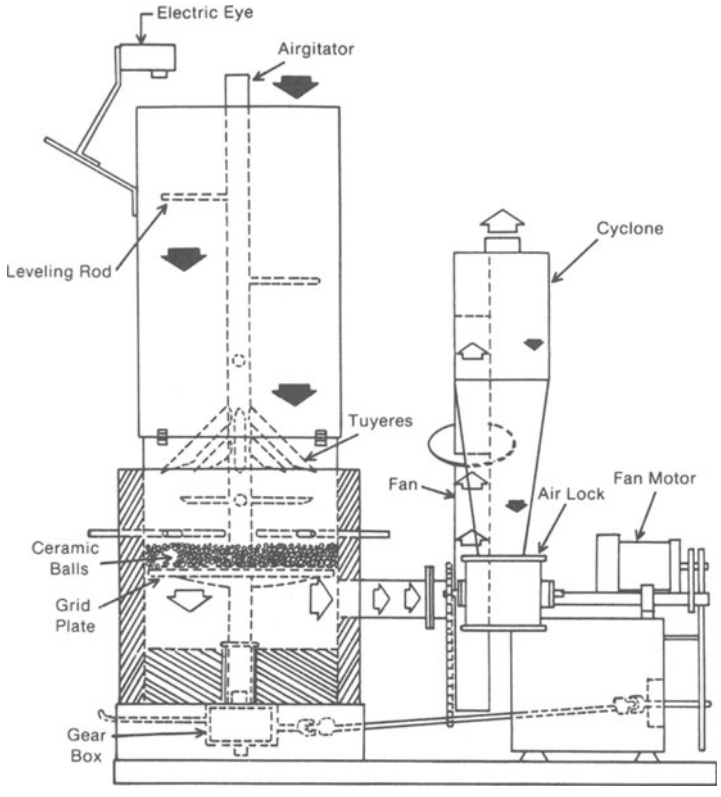


Fig. 3.5: Commercial "stratified downdraft gasifier."

now being considered as a prototype for a fluidized bed high-pressure oxygen gasifier for production of methanol from wood.[38]

3.3.1.5 Suspension Gasifiers. Finely divided coal is burned in large power plants in suspension, and the Texaco Development Corporation has developed coal gasification of suspended particles. While most biomass requires a great deal of energy to reduce it to small size, sawdust and sander dust are suitable in size for suspension gasification. A successful cyclonic suspension gasifier has been tested (but not produced) by the Morebark Co.[39]

The Evergreen Energy Corporation is now working with the Texaco Development Corporation to develop an entrained flow gasification system to make methanol. Ultimately, they plan to build a plant converting 3500 tons of green wood chips/day into 100 tons of methanol/day. Overall energy conversion efficiency is projected to be 55%.[40]

3.3.1.6 Other Gasifiers. The gasifiers discussed so far involve direct heat transfer to the biomass and have in most cases been commercialized. A number of other gasifiers have been built on an experimental basis and embody other approaches.

Table 3.2

Typical operating results for stratified downdraft gasifier[36]

Gas Analysis	Volume %
CO	19.8
CO_2	15.0
H_2	16.5
CH_4	2.6
N_2	46.5
C_2	0.07
C_3	0.47
High Heating Value (HHV)	148 Btu/SCFH
Tar lb/lb	0.0034
Char lb/lb	0.04
Cold gas efficiency	78%

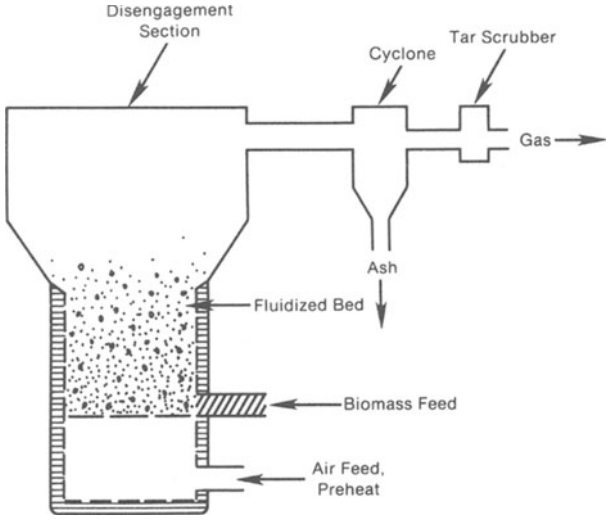


Fig. 3.6: Schematic diagram of fluidized bed gasifier.

When biomass is gasified directly with air, the nitrogen in the air cuts the fuel value of the gas in half, making it unsuitable for pipeline use or chemical synthesis. This can be solved directly by using oxygen for gasification. Since oxygen is produced on a very wide scale, it is relatively inexpensive[41] and has been widely used for coal gasification. Nevertheless, if oxygen use could be avoided, it might reduce the cost

Table 3.3

Typical Gas Analyses for Fluidized Bed Air Gasifier[39]

Component	Wood Moisture Content 20 (Mole %)	44 (Mole %)
H_2	5.3	5.6
CO	15.9	7.7
CH_4	7.0	3.3
C_2H_4	2.3	1.6
CO_2, N_2, H_2O	69.2	81.3
Heating value (Btu/scf)	175.0	99.0

of gasification and possibly improve the gas energy content. Simply stated, the aim would be to burn char with air to generate the pyrolysis energy and to introduce this energy to the pyrolysis zone through heat exchangers.

One scheme for accomplishing this was developed by Bailie.[42] Biomass is pyrolyzed by hot sand in one fluidized bed (Fig. 3.7), producing medium Btu gas and char (operating in a manner somewhat similar to two bed cat crackers used in refineries). The char from the first reactor is separated and burned in a second fluidized bed using air. The heat is transferred to hot sand which is circulated back to the first bed. A 40 ton/day unit has been operated over 7000 hours by the Ibaru Corporation, producing 4700 kcal/Nm^3 (568 Btu/scf) gas.

A similar scheme using a Multi-Solid Fluidized Bed has been under development at Battelle-Columbus for several years. Two sizes of solids are used which increases the residence time for entrained solids.[43]

A process is being developed by the Wright Malta Corporation to convert wet wood into medium Btu (388 Btu/scf) fuel gas without air or oxygen. Biomass moves through a 10 to 20 atmosphere pressurized kiln at temperatures of 1100° to 1150° F.[44] Initial work in "minikilns" established the process conditions. A 100-foot-long process development unit was tested in 1983 but experienced mechanical difficulties.

3.3.2 Gasifier Systems

It is important to realize in developing or selecting a gasifier that there are many choices which must be made consciously or unconsciously in setting up a gasifer system. One must consider the fuel type (wood, agricultural, municipal solid waste, etc.), the size of the fuel available, the method of feeding (batch or continuous), the energy supply (direct oxidation or heat exchange), the state of the

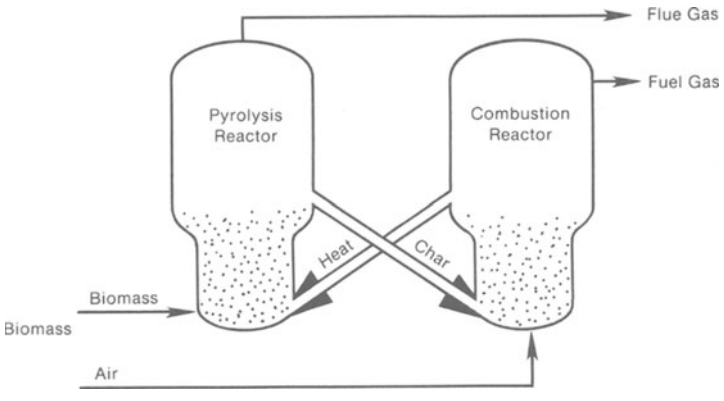


Fig. 3.7: Fluidized bed gasifier-combustor. Biomass is injected into the left hand vessel where it is pyrolyzed by hot sand, yielding medium Btu gas. Char from the left reactor is burned in air in the right hand reactor, reheating sand flowing from the left reactor.

ash (dry or molten), the pressure range required (atmospheric or high pressure), and the products desired (gas only, or combined with char or oil). For instance, several processes now being developed produce charcoal and gas simultaneously.[45,46] All of these choices must be taken into account when selecting a gasifier for a particular purpose.

3.3.3 Gasifier System Specifications

Gasification being more art than science, much of the literature on the subject is qualitative in nature. As gasification develops further, it will be necessary to quantitatively evaluate various operating parameters so that progress can be judged and gasifiers can be compared. These specifications can be divided into fuel specifications, gasifier specifications, and gas specifications.

Although gasifier enthusiasts will sometimes claim that the gasifier will "burn anything," gasifiers often have very specific fuel requirements. Specifications on any gasifier system should indicate the range of fuel properties giving acceptable operation. When purchasing fuel, one should have information on: size, size distribution, shape, shape distribution, density, energy content, water content, and ash content.

Obviously, the gasifier minimum and maximum throughput is a most important parameter. This can be given in terms of energy input (derived from fuel consumption) or energy output (derived from gas flow and gas energy content). The smallest gasifiers generate a few kW and the largest generate up to 10 MW.

A very important gasifier operating parameter is the turndown ratio, the ratio of the highest to the lowest sustainable throughput and this ratio is claimed to vary between about 3 and 20 for existing gasifiers. The turndown must be specified in

terms of the gas quality required. Too low a throughput generates very dirty gas; a too-high throughput yields low-energy gas.

Equally important is the efficiency of the gasifier, usually quoted in terms of energy in the cold clean gas divided by the energy in the biomass feed. Efficiencies range from 70% to over 90%. The balance of the energy appears as sensible heat of the gas, skin losses of the gasifier, char carryover loss, and tar loss.

There are figures of merit that can be used to compare the conversion rates of gasifiers. The first is the energy/area conversion rate which can run from 0.2 to 3 MBtu $ft^{-3}hr^{-1}$ for commercial and experimental gasifiers. The energy/volume conversion rate runs from 0.1 to 0.5 MBtu $ft^{-3}hr^{-1}$. These figures of merit are convenient for comparing gasifiers of different sizes and estimating gasifier size requirements.

Several factors are used to describe gas quality. The energy content (Btu/scf or kJ/m^3) is important if the gas is to be transported and also affects the ease of combustion. Tar and particulate level of the raw and processed gas are needed to estimate cleanup problems and use suitability in burners, engines, and so on (see below). The temperature of the raw gas dictates the degree of cooling required for use in engines or the sensible heat that may be recovered. The moisture content of the gas determines, in part, the cooling load.

While the specifications are generally not available, if the gasification industry is to reach the degree of development of the petroleum industry today, they need to be developed.

3.3.4 Producer Gas Use and Gas Conditioning

The use of producer gas for heat does not require high gas quantity and removal of particulates and provision for proper combustion in the burner is sufficient for most applications. If the gas is to be distributed to remote locations in a pipeline, the gas must either be kept at temperatures between 200° and 400° C or the tars must be removed prior to distribution (to prevent the tars from condensing and blocking the flow).

If the gas is to be used to operate a spark, diesel, or direct-combustion turbine engine, the gas must be very clean. Australian tests on vehicle gasifiers during World War II showed that the gas should have less than 10 mg/m^3 (approximately 10 ppm) of tar and particulates if the engine wear was to be less than that measured with gasoline.[47] Gas turbine blades are particularly sensitive to corrosion by the components of wood ash, and, thus, they may require similar or better conditioning of the gas. The relationship between engine wear and gas cleanliness is an area of prime concern deserving a good measure of research effort.

Depending on the type of gasifier (see above), the raw producer gas can contain up to 5% char and ash, and 20% condensible tars, all of which must be removed to various degrees, depending on the use to be made of the gas. Typically, the solids are removed with 1-2 cyclones or a hot bag house filter. The gas can then be cooled and the tars, soot, and ash removed by condensation or in filters, scrubbers, impingers, and/or demisters. The cost of the cleanup train may well

exceed the cost of the gasifier, and difficulties with gasification systems often lie primarily in gas cleanup. Thus the choice of gasifier may greatly depend on the degree of gas cleanup required for a specific application. The subject of gas cleanup is beyond the scope of this review, but the reader is referred to a number of references on the subject.[47-51]

3.4 PRINCIPLES OF GASIFICATION

Combustion is a very old art. During this century it has also become a science developed by an extensive body of research. Emphasis has been placed on understanding combustion in terms of the thermodynamics, kinetics, and chemistry of the processes in flames. The Combustion Institute meets annually, has its own journal,[52] and has produced an impressive number of volumes supporting this understanding. Because gasification was well developed before this period, but has been less important for the past 40 years, there is no such comparable body of knowledge, though there is a substantial body of scattered information as listed above. It is to be hoped that as the commercial application of gasification progresses, research will be undertaken to develop a similar understanding of basic processes. In this section, I summarize the present state of knowledge and suggest areas for future work.

3.4.1 Gasifier Fuels

Solid fuels are more difficult to work with than liquids and gases, in part because they have a wider range of variability. For purposes of discussion, some approximations must be made. Fortunately, the range of chemical variation and energy content for most biomass is quite small, though there is a wide range of physical forms. (On the other hand, coal has a much narrower range of physical properties, but the chemical and energy variability are much wider.)

The properties of biomass that are important for gasification are summarized by two standard ASTM analyses.[53] The proximate analysis follows a standard heating sequence and gives the weight percent of volatile material, fixed carbon, and ash, and high heating value. Proximate analyses for various feedstocks are given in Table 3.4. The ultimate analysis gives the weight percent of C, H, O, and other elements, and ash; ultimate analyses of various fuels are listed in Table 3.5.

In making calculations on gasification reactions, it is necessary to convert the weight fractions in the above tables to molar quantities or ratios. A typical composition for (generic) wood biomass is [54]

$$CH_{1.4}O_{0.6} \tag{3.1}$$

and this is sufficiently accurate for most calculations. Cellulose ($C_6H_{10}O_5$) has the formula

$$CH_{1.66}O_{0.83} \tag{3.2}$$

Table 3.4

Proximate analysis data for selected solid fuels and biomass materials[54]
(Dry Basis, Weight Percent)

	Volatile Matter (VM*)	Fixed Carbon (FC*)	Ash*
Coals			
Pittsburgh seam coal	33.9	55.8	10.3
Wyoming Elkol coal	44.4	51.4	4.2
Lignite	43.0	46.6	10.4
Oven Dry Woods			
Western hemlock	84.8	15.0	0.2
Douglas fir	86.2	13.7	0.1
White Fir	84.4	15.1	0.5
Ponderosa pine	87.0	12.8	0.2
Redwood	83.5	16.1	0.4
Cedar	77.0	21.0	2.0
Oven Dry Barks			
Western hemlock	74.3	24.0	1.7
Douglas fir	70.6	27.2	2.2
White Fir	73.4	24.0	2.6
Ponderosa pine	73.4	25.9	0.7
Redwood	71.3	27.9	0.8
Cedar	86.7	13.1	0.2
Mill Woodwaste Samples			
−4 Mesh redwood shavings	76.2	23.5	0.3
−4 Mesh Alabama oakchips	74.7	21.9	3.3
Municipal Refuse and Major Components			
National average waste	65.9	9.1	25.0
Newspaper (9.5% of average waste)	86.3	11.5	2.2
Paper boxes (23.4%)	81.7	12.9	1.5
Magazine paper (6.8%)	69.2	7.3	23.4
Brown paper (5.6%)	89.1	9.8	1.1
Pyrolysis chars			
Redwood (790 F to 1020 F)	30.0	67.7	2.3
Redwood (800 F to 1725 F)	23.9	72.0	4.1
Oak (820 F to 1185 F)	25.8	59.3	14.9
Oak (1060 F)	27.1	55.6	17.3

Table 3.5

Ultimate analysis data for selected solid fuels and biomass materials[55]
(Dry Basis, Weight Percent)

Material	C	H	N	S	O	Ash	Higher Heating Value (Btu/lb)
Pittsburgh seam coal	75.5	5.0	1.2	3.1	4.9	10.3	13,650
West Kentucky No. 11 coal	74.4	5.1	1.5	3.8	7.9	7.3	13,640
Utah coal	77.9	6.0	1.5	0.6	9.9	4.1	14,170
Wyoming Elkol coal	71.5	5.3	1.2	0.9	16.9	4.2	12,710
Lignite	64.0	4.2	0.9	1.3	19.2	10.4	10,712
Charcoal	80.3	3.1	0.2	0.0	11.3	3.4	13,370
Douglas fir	52.3	6.3	0.1	0.0	40.5	0.8	9,050
Douglas fir bark	56.2	5.9	0.0	0.0	36.7	1.2	9,500
Pine bark	52.3	5.8	0.2	0.0	38.8	2.9	8,780
Western hemlock	50.4	5.8	0.1	0.1	41.4	2.2	8,620
Redwood	53.5	5.9	0.1	0.0	40.3	0.2	9,040
Beech	51.6	6.3	0.0	0.0	41.5	0.6	8,760
Hickory	49.7	6.5	0.0	0.0	43.1	0.7	8,760
Maple	50.6	6.0	0.3	0.00	41.7	1.4	8,580
Poplar	51.6	6.3	0.0	0.0	41.5	0.6	8,920
Rice hulls	38.5	5.7	0.5	0.0	39.8	15.5	6,610
Rice straw	39.2	5.1	0.6	0.1	35.8	19.2	6,540
Sawdust pellets	47.2	6.5	0.0	0.0	45.4	1.0	8,814
Paper	43.4	5.8	0.3	0.2	44.3	6.0	7,572
Redwood wastewood	53.4	6.0	0.1	39.9	0.1	0.6	9,163
Alabama oak wastewood	49.5	5.7	0.2	0.0	41.3	3.3	8,266
Animal waste	42.7	5.5	2.4	0.3	31.3	17.8	7,380
Municipal solid waste	47.6	6.0	1.2	0.3	32.9	12.0	8,546

while lignin (average for hard and softwood) is typically

$$CH_{1.23}O_{0.38}$$

These composition points are plotted in the fuel diagram of Fig. 3.8 discussed below. (One should be careful when working with nontypical biomass to use actual analyses of the fuel at hand, where possible.)

The high heat of combustion, ΔH_c, of the various forms of biomass varies surprisingly little. It can be estimated with considerable accuracy from the IGT

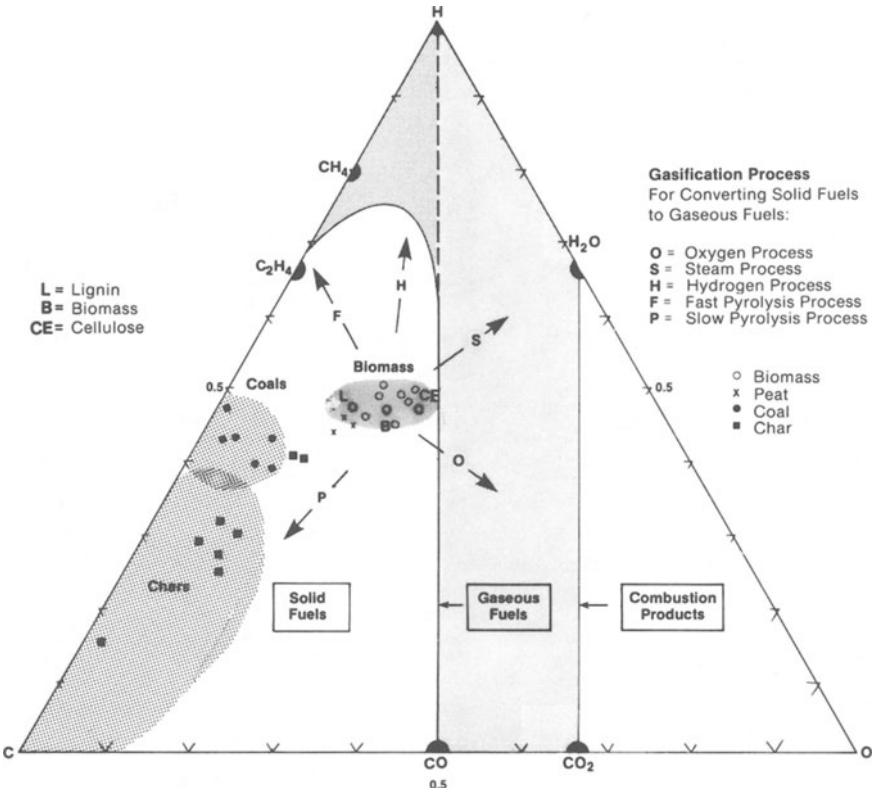


Fig. 3.8: Ternary diagram showing mole fraction of C, H, and O in various solids and gases. Position of the solid fuels coal, char, biomass, and peat indicated at left; the "synthesis gas" line $H - CO$ at center; combustion products $H_2O - CO_2$ at right.[57]

formula [55]

$$\Delta H_c = HHV\,(\text{Btu/lb}) = 146.58\,C + 568.78\,H - 51.53\,O - 6.58\,A + 29.45 \qquad (3.3)$$

where C, H, O, and A are the weight percent of the carbon, hydrogen, oxygen, and ash in the dry fuel, respectively. The variation between this formula and measured values for a wide range of solid fuels is typically less than 2%.[55] The formula is accurate within 6% for other gaseous and liquid fuels, except carbon monoxide which is 22% lower than the accepted value.

3.4.2 Biomass versus Coal Gasification

It has been a common assumption that the gasification of biomass and coal are closely related and that, in fact, a good coal gasifier can be easily modified to use biomass. While there are many similarities, successful biomass gasifiers are typically quite different from coal gasifiers because they must process different fuels. Note in Table 3.4 that only a third of a typical coal is volatile, while over 80% of typical

biomass is volatile. Less than 20% charcoal must be converted in the latter case. Furthermore, the reactivity of the charcoal is 25 times greater than that of coke, due to lower density, and increased internal surface area, low ash content, and the amorphous nature of the carbon.[56,57] Thus, the major focus of biomass gasifiers must be elimination of tars, while charcoal conversion is relatively easy. For coal, the major focus must be gasification of the coke remaining after volatilization.

3.4.3 Chemistry of Biomass Gasification

The principal chemical elements contained in all fuels—carbon, hydrogen, and oxygen—are shown as the vertices of a triangle in Fig. 3.8.[57] The elemental compositions of some biomass, coal, and char samples are plotted in the figure where the hatched areas define the practical range of variation of these solid fuels. The figure also shows the wide variation of char compositions, which overlap the composition (but not the physical structure) of coals. Chars formed below 800° C have a surprisingly high H and O content. The composition of three peats have been included, and they are very close in composition to that of lignin.

In this diagram, solid fuels lie to the left of the line joining CO and H. At high temperatures, only CO and H_2 are stable. The producer gas fuel range is thus confined to the right of the $H - CO$ line. However, at lower temperatures, CH_4 becomes stable and CO becomes unstable, so there is no exact position for the line separating gas fuels from solid fuels unless thermodynamic and kinetic conditions are specified. Finally, the products of complete combustion lie on the line joining CO_2 and H_2O, so that this line defines the low energy limit of gaseous fuels. Compositions to the right of this line represent combustion with excess air or oxygen.

Thus, the problem of gasification becomes the problem of shifting the composition of the solid fuels on the left side of Fig. 3.8 along one or more of the arrows pointing to a gaseous composition. The arrows of Fig. 3.8 show the various methods of accomplishing this. Pyrolysis is the disproportionation of biomass to yield gases, char, and oil as shown by the arrow P. Oxygen/air gasification is mechanically the simplest method of producing gas because the overall reaction is exothermic (arrow O). By far the largest number of gasifiers use this method. Steam can be used alone for biomass gasification, producing a gas high in methane, but the temperature of operation must be kept relatively low and long dwell times and catalysts are required.

Hydrogen has been used in the past for the liquefaction and gasification of coal. It can be seen from the arrow H in Fig. 3.8 that this would shift the composition of solid fuels toward high-methane and other high-energy content fuels. However, reaction with hydrogen requires high pressures, high temperatures, and a source of hydrogen—a valuable chemical feedstock and fuel in its own right. Furthermore, at the low temperatures at which biomass volatilizes (200° to 500° C), there is no primary reaction between the biomass and hydrogen. However, after tar vapors leave the biomass, secondary cracking in other gases may affect the character of the products. Interesting experiments were carried out at Brookhaven National Laboratories on pyrolysis of biomass in hydrogen, helium, and methane. Yields of ethylene and benzene are significantly enhanced with hydrogen and methane.[58]

A new area of biomass gasification involves the production of ethylene and higher olefins such as propylene and butylene.[59] These molecules are quite unstable compared to methane or CO at high temperatures. However, their decomposition is slow, below 900° C, so that they are formed in high yields by the fast pyrolysis of hydrocarbon feedstocks. Fast pyrolysis of biomass to ethylene is shown diagrammatically by the arrow F on Fig. 3.8.

3.4.4 Global Thermodynamics of Biomass Gasification

Thermodynamics cannot predict the course and products of all reactions, but it does delimit possible pathways and products and prevents us from attempting the impossible. Thus, it is useful for understanding old and new processes, and in some cases (e.g., downdraft gasification) comes quite close to predicting the final gas composition. Fortunately, with modern computers and the excellent thermodynamic data base we now have, the calculation of thermodynamic equilibria between the principal species produced during gasification (CO, CO_2, H_2, H_2O, and CH_4) are not too difficult.[6,18,60,61]

The principal reactions of charcoal and producer gas at high temperatures are listed with numerical values for the relevant thermodynamic data in Table 3.6. This data has been taken from the JANAF Thermochemical Tables[62] and fitted to a linear equation for ΔG over the region 800 to 1600 K so that the formulas for ΔG use the effective values of ΔH and ΔS listed over this range. This simplifies calculations for these reactions. Greater precision is not justified by the accuracy of the data.[63,64] The pressure equilibrium constant, K_p, for these reactions is shown in Fig. 3.9 where it is seen that all the reactions cross the line $K_p = 1$ between 550 and 800 K. The values shown here are based on the most recent thermodynamic data available, but differ only slightly from that published over 30 years ago.[4,18]

The adiabatic reaction temperature, T^*, is a very important concept in gasification. It is the temperature at which the heat inputs minus the heat losses are equal to the energy content of the gas and it is derived by solving the energy balance

$$\Delta H_R - \Sigma h_l = \Sigma n_i \int_{T_o}^{T^*} c_i dT + \Sigma n_i \Delta H_{Pi} \tag{3.4}$$

for T^*, where ΔH_R is the heat of combustion of the biomass, ΔH_{Pi} is the heat of combustion of the fuel gas components, $\Sigma n_i \int c_i dT$ is the heat content of the fuel gas components, and Σh_l is the heat losses from the reactor. (n_i is the number of moles of the products and c_i is the heat capacity of each component.) Unfortunately, T^* must be determined by iteration, so that evaluation of T^* would be a tedious process without a computer. Since the heat loss is specific to each reactor situation, it is generally omitted to the examples below.

The calculation of gas compositions using these nonlinear equations is not straightforward and successive approximation methods are discussed by Gumz.[18] The equilibrium reaction temperatures and gas compositions for wood of varying moisture content were calculated by Gumz and are shown in Table 3.7. These compositions agree quite well with the observed gas compositions from downdraft gasifiers.

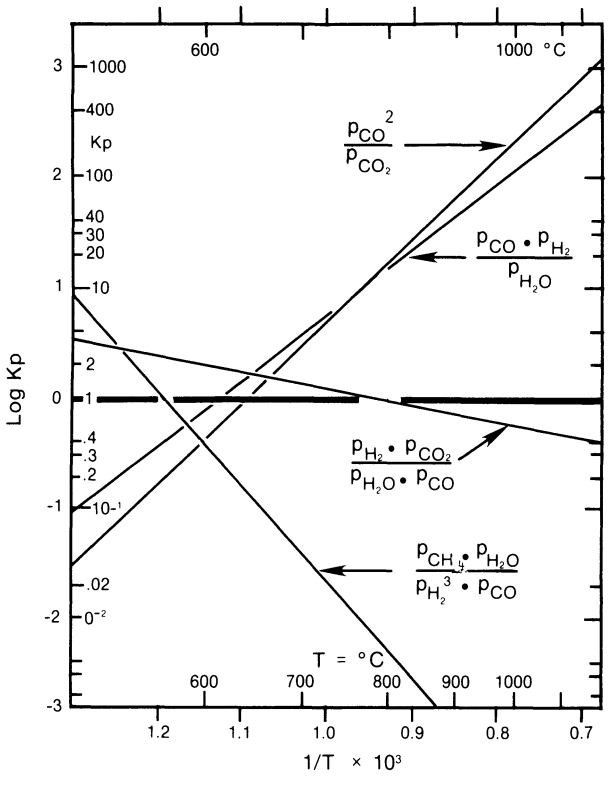


Fig. 3.9: Gasification equilibrium constants.
(Data from Stull[62] and Table 3.6.)

The most significant deviation from equilibrium observed in actual gases is that of methane and an adjusted equilibrium constant is sometimes used to account for this deviation.[18] One possible source of deviation may be the use of zero free energy for graphite in the reactions of Table 3.6 where, in fact, the freshly formed char may differ significantly from zero.[20]

Several computer programs have been written to explore more fully the equilibrium energy relations of the equations in Table 3.6. Desrosiers has presented data for a wide variety of conditions using oxygen, air, and steam.[5] The adiabatic reaction temperature of biomass with oxygen or air is shown in Fig. 3.10. The abscissa of Fig. 3.10 is the equivalence ratio, ϕ, the ratio of oxygen used to that required for complete combustion. (Complete combustion requires a mass ratio of 1.47 for O_2 or 6.36 for air.) The equilibrium compositions associated with these temperatures are shown in Fig. 3.11 and the energy content of the gas and char are plotted in Fig. 3.12.

The distinction between pyrolysis, gasification, and combustion is easily seen in Fig. 3.10. Pyrolysis is the destructive breakdown of biomass in the absence of oxygen, represented by $\phi = 0$. Note in Fig. 3.11 that if pyrolysis can bring biomass to equilibrium, a gas rich in methane results. However, as shown in Fig. 3.8, the

Table 3.6

Thermodynamics of Gasification Reactions[62]

Reaction	$C + CO_2 \rightarrow 2CO$	$C + H_2O \rightarrow CO + H_2$	$CO + 3H_2 \rightarrow CO + H_4 + H_2O$	$CO + H_2 \rightarrow CO_2 + H_2$
Reaction Name	Boudouard	Water Gas	Steam Reforming	Water Gas Shift[b]
ΔG°_R (kJ)[a]	$= 170.72 - 0.175T$	$= 135.9 - 0.143T$	$= -225.6 + 0.252T$	$= -34.83 + 0.0328T$
Temperature, K	$K_p = \frac{p^2_{CO}}{p^2_{CO_2} \cdot a_c}$	$K_p = \frac{p_{CO} \cdot p_{H_2}}{p_{H_2O} \cdot a_c}$	$K_p = \frac{p_{CH_4} \cdot p_{H_2O}}{p^3_{H_2} \cdot p_{CO}}$	$K_p = \frac{p_{CO} \cdot p_{H_2O}}{p_{CO_2} \cdot p_{H_2}}$
800	0.0102	0.0420	34.416	4.11
900	0.1768	0.406	0.795	2.299
1000	1.731	2.499	0.039	1.44
1100	11.200	11.043	0.0033	0.986
1200	53.06	38.094	$4.2E^{-4}$	0.71801
1300	198	108.6	$7.4E^{-5}$	0.549
1400	611	266.6	$1.7E^{-5}$	0.43
1500	162.6	580	$4.6E^{-6}$	0.357
1600	382.6	1147	$1.5E^{-6}$	0.300

a) Linearized fit of data in region 800 < T 1600 K to the equation $\Delta G^\circ_R = \Delta H^\circ_R - T\Delta S^\circ_R$. Constants are effective values of ΔH and ΔS in this range. a_c is the activity of carbon, generally assumed to be 1.

b) This data can be derived from the first three reactions and is not mathematically independent. It is included for convenience.

Table 3.7

Downdraft gasification of wood with different moisture content[18]

Moisture %	0	10	20	30
Net Heating Value, kcal/kg	4430	3927	3424	2921
Reaction Temp., °C	661	631	600	555
Gas Analysis (Wet)				
% CO	21.04	16.50	12.02	6.83
% CO_2	9.99	12.68	15.24	17.88
% H_2	18.18	18.73	18.94	17.45
% H_2O	4.65	6.68	9.41	13.65
% CH_4	0.69	1.07	1.66	2.72
% N_2	45.45	44.34	42.73	41.47
	100.00	100.00	100.00	100.00
Gas Analysis (Dry)				
% CO	22.07	17.68	13.27	7.91
% CO_2	10.48	13.59	16.82	20.71
% H_2	19.07	20.07	20.19	20.21
% CH_4	0.72	1.15	1.83	3.15
% N_2	47.66	47.51	47.17	48.02
	100.00	100.00	100.00	100.00
Gas Net Heating Value, kcal/SCM (Standard Cubic Meters)	1161.6	1071.1	991.7	887.3
Gas yield, SCM/kg	2.967	2.800	2.604	2.401
Gasifying agent, SCM/SCM	0.5858	0.5715	0.5507	0.5344
Fuel consumption, kg/SCM	0.3371	0.3571	0.3841	0.4164
Gas yield × heating value, kcal/kg	3346.5	2999.1	2582.4	2130.4
Gasification efficiency, %	77.8	76.4	75.4	72.9
Equivalent blast saturation temperature, °C*	68.0	69.6	74.3	85.0

*Equivalent blast saturation temperature is the saturation temperature of the blast, gasifying pure carbon, which gives the same reaction temperature.

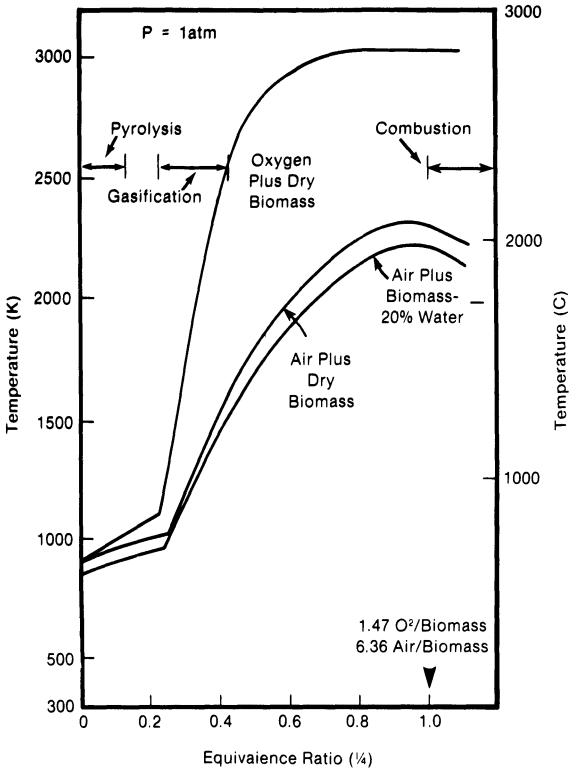


Fig. 3.10: Biomass adiabatic reaction temperatures showing pyrolysis, gasification, and combustion regions for air and oxygen reactions.[54]

composition of biomass does not contain sufficient hydrogen or oxygen to produce all gas, so pyrolysis alone also produces a sizable fraction of charcoal as shown in Fig. 3.12.

The addition of oxygen or air to biomass can be ideally represented for ideal gasification

$$CH_{1.4}O_{0.6} + 0.2O_2 \rightarrow CO + 0.7H_2 \tag{3.5}$$

Unfortunately, this reaction is endothermic, so that external heat would have to be added to the system to accomplish this ideal gasification. In adiabatic gasification, the biomass is converted to fuel by internal partial combustion using an excess of oxygen, e.g.,

$$CH_{1.4}O_{0.6} + 0.45O_2 \rightarrow 0.7CO + 0.3CO_2 + 0.55H_2 + 0.2H_2O \tag{3.6}$$

Note that in this case, a third of the CO and H_2 have been burned to supply the heat for the reaction, but the reaction now requires no external heat transfer. (The exact amount of oxygen used depends on heat loss, water content, etc.) Variations of this reaction form the basis of all oxygen/air gasifiers and provide for high efficiency

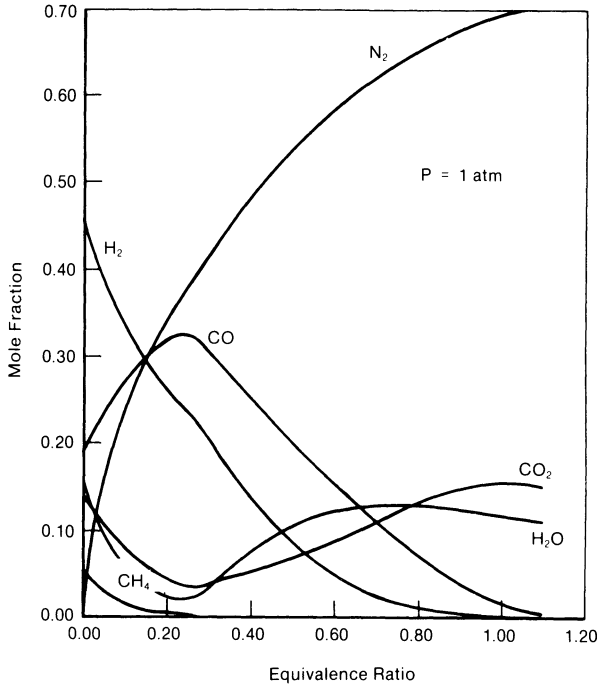


Fig. 3.11: Equilibrium composition for adiabatic air/biomass equilibrium.[54]

and mechanical simplicity, since no external heat is required. However, the energy content of the gas is lowered by the presence of CO_2 and H_2O. If external heat can be supplied, less biomass is consumed. A gasification process using an electrically heated pyrolysis zone is being developed in Brazil.[65]

Finally, complete combustion is represented by

$$CH_{1.4}O_{0.6} + 1.05O_2 \rightarrow CO_2 + 0.7H_2O \tag{3.7}$$

The above reactions represent the overall "global" reaction of pyrolysis, gasification, or combustion, but in practice the reactions occur in several steps. If the gas passes through charcoal just before leaving the gasification zone, the final equilibrium can be described by the four equations in Table 3.6, and the equilibrium constants for these reactions are shown in Fig. 3.9. These equations, plus a material balance, permit calculation of equilibrium gas compositions at any temperature. Use of an energy balance then permits calculation of the adiabatic reaction temperature.

The results of calculations of thermodynamic equilibrium on the downdraft gasification of wood are shown in Table 3.7.[18] The values calculated are typical of those measured in operating downdraft gasifiers, except that the equilibrium methane content is lower by up to a factor of two than that observed in practice (see above).

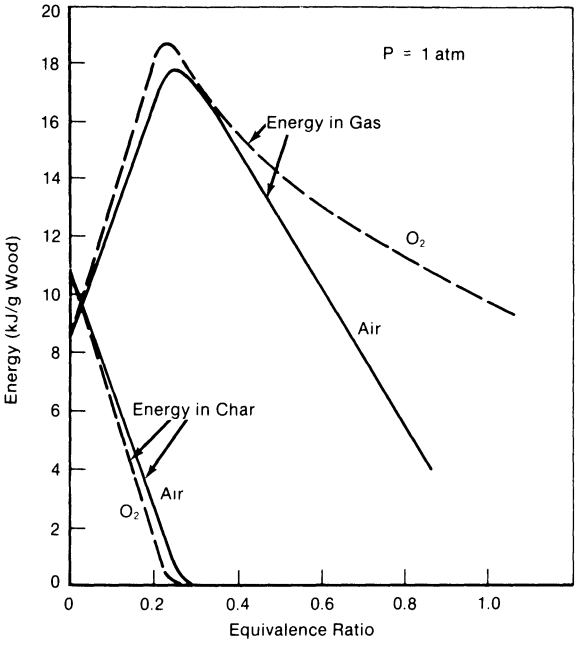


Fig. 3.12: Gas and char energy content for oxygen and air/biomass equilibria.[54]

Reed et al. have examined the variations of this reaction relative to oxygen gasification, exploring the effect of steam addition, heat loss, and pressure on gas composition for methanol production. The results are shown in Table 3.8.[33] The calculated equilibrium is compared to that observed in the SERI Oxygen Gasifier in Table 3.9.

Overend has examined these reactions for air and oxygen using the Gumz model over a wide range of conditions, and his results are shown in Fig. 3.13 for downdraft gasification.[60] The work on equilibrium calculations is by no means complete and it would be desirable to have a universal computer program able to explore all the relevant variables at high speed in an "if, ...then what?" mode.

The results of the above calculations suggest that downdraft gasification can come very close to the equilibrium values predicted by thermodynamics. Downdraft gasifiers are constrained to operate very close to the discontinuity of slope at $\phi = 0.25$ of the curve in Fig. 3.10. To the left of this point, charcoal is produced and the bed increases in depth. To the right of this point, excess carbon is consumed and the bed depth decreases. (The position of this discontinuity varies slightly with char carry over or oil production.) Fluidized and suspended beds are not constrained to operate at the discontinuity of Fig. 3.10, but gas compositions can approach equilibrium, given sufficient time and temperature.

Table 3.8

Calculated equilibrium gas temperature and composition for 3/16 in. (5-mm) pellets[88]

Case	P (atm)	Heat Loss ($\%H_c$)*	Steam Rate ($\frac{\text{mass steam}}{\text{mass biomass}}$)	Moisture (%)	O_2 Consumption ($\frac{\text{mass oxygen}}{\text{mass biomass}}$)		Temperature		Composition (vol %)				
					wet	dry	°C	°F	CO	CO_2	H_2	H_2O	CH_4
1	1	0	0.0	3.6	0.339	0.352	731	1347	46.2	9.9	37.2	5.5	1.1
2	1	5	0.0	3.6	0.391	0.409	709	1307	42.7	13.3	35.5	7.1	1.3
3	1	0	0.0	20.0	0.285	0.356	660	1224	29.8	18.2	37.5	11.9	2.5
4	1	10	0.2	3.6	0.420	0.436	642	1187	25.9	22.3	34.8	14.1	2.5
5	1	0	0.2	3.6	0.315	0.327	672	1242	31.8	16.4	38.5	10.8	2.3
6	10	0	0.0	3.6	0.345	0.358	878	1613	46.2	10.0	32.5	8.3	2.9

*Heat loss in % of heat of combustion of fuel.

Table 3.9

Comparison of calculated and experimental gas compositions[83]

	R^a	Temperature		Composition (vol %)			
		°C	°F	CO	CO_2	H_2	CH_4
Calculated 0% heat loss	0.339	731	1347	48.9	10.5	39.4	1.2
10% heat loss	0.445	690	1273	42.8	18.4	36.9	1.6
Experimental run 14A (N_2-free)	0.43	800	1470[b]	42.8	19.2	33.3	4.1

a) R = oxygen/fuel ratio.

b) Off-gas temperature measured at grate. Probably high due to excess of methane over equilibrium value.

Updraft gasifiers approach closely to a charcoal/air equilibrium near the point of air/oxygen entry, but the pyrolysis step produces tar vapors at low temperatures which do not come to equilibrium.

3.4.5 The Gasification Reactions

Complete gasification of biomass involves several sequential and parallel reactions. Most of these reactions are endothermic and must eventually be balanced by partial combustion of gas or an external heat source as shown in Fig. 3.14.[57] In the following sections, we will take these up one at a time. In the section on gasifier modeling, we will then discuss various attempts to combine them for predicting actual gasifier performance.

The gasification of a biomass particle occurs in two sequential stages. As the particle is heated, it initially pryolyzes to form charcoal plus gases and vapors:

$$\text{biomass} + \text{heat} \longrightarrow \text{charcoal} + \text{volatiles} + \text{gases} \tag{3.8}$$

After pyrolysis is completed, the charcoal can react with oxygen and steam (updraft gasifiers) or the products of pyrolysis (downdraft gasifiers, fluidized bed, and suspended particle gasifiers) according to

$$\text{charcoal} + \text{gases} \longrightarrow \text{reduced gases} \tag{3.9}$$

In addition, the vapors formed initially from the solid may undergo cracking reactions to form secondary products, either gases or other condensable species.

The pyrolysis reaction consumes over 80% of the biomass and produces a volume of gas and vapors approximately 1000 times greater than that of the solid (at

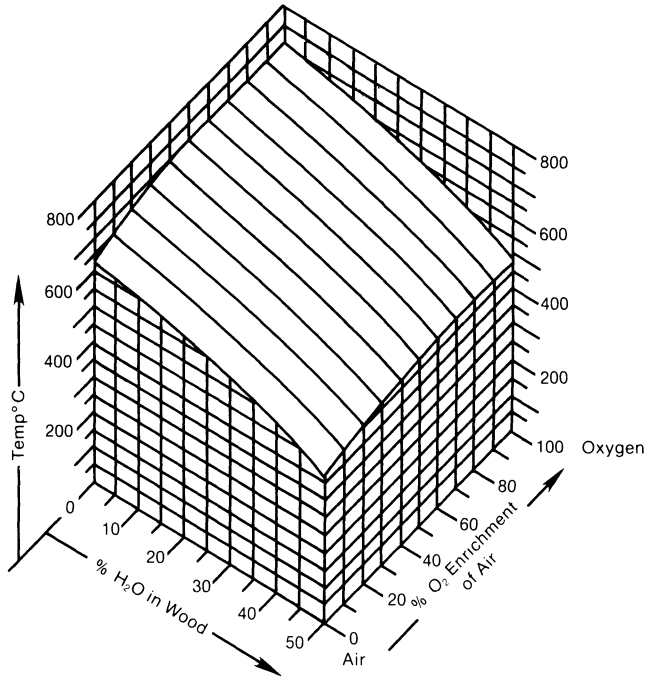


Fig. 3.13: Adiabatic reactor temperature. This figure notes the performance envelope with regard to the adiabatic reactor temperature. The drastic effect of water on the reactor temperature can be seen. From a practical viewpoint, temperatures of <550°C generally destroy the equilibrium nature of the cocurrent gasifier by freezing the "equilibrium;" the composition does not change, since the reactions are kinetically controlled and are very slow at low temperatures.[60]

atmospheric pressure) due to the lower gas density and higher temperature. Gasification and pyrolysis are cleanly separated in time because of the vapor evolution which prevents the contact of external gases with the charcoal until pyrolysis is complete. (This can be clearly seen in a burning kitchen match where the volatile combustion region is clearly separated from the char region.) Therefore, it is possible to discuss the pyrolysis regime and the charcoal gasification regime separately below.

3.4.5.1 Solid and Gas Pyrolysis. Pyrolysis is the breaking down of substances by heat in the absence of air or oxygen. Biomass pyrolysis is a whole field in itself and has been recently reviewed by Williams,[66] Milne,[67] by Antal in the following chapter of this book,[68] by Shafizadeh,[69] and by Soltes.[70] Pyrolysis is the first reaction that biomass undergoes in gasification (and combustion). A representative mass and energy balance for slow wood pyrolysis is given in Table 3.10.[60] However, depending on the conditions, these products may be greatly altered before they leave the biomass particle and after they leave as suggested in Fig. 3.14.[71]

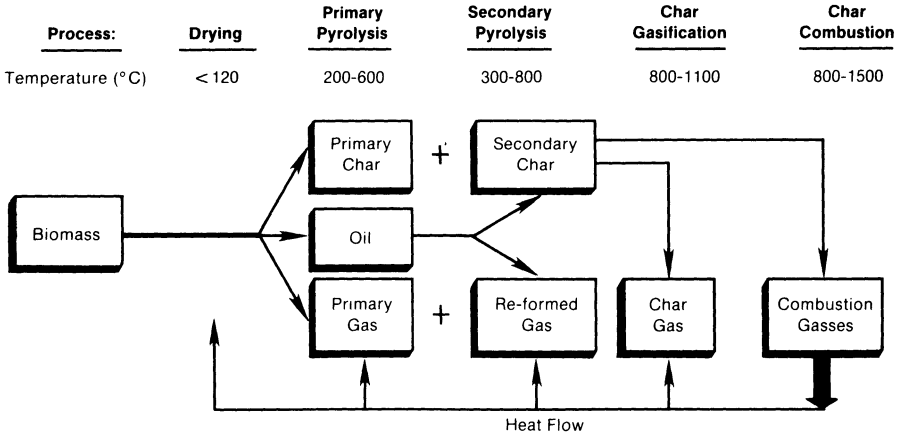


Fig. 3.14: Heat and mass flows in pyrolysis and gasification processes.[57]

Table 3.10

Generalized mass and energy balance for wood pyrolysis[60]

Component	Yield (%)	Energy (MJ/100g)	Energy (%)
Wood	100	18.0	100
Char	30	9.4	52.2
Pryolysis liquid	23.5	5.65	31.4
Gas	20.0	1.07	5.9
Water	26.5	-	-
Process energy	-	1.9	10.5

The char forming reactions in pyrolysis are relatively slow, so that the fraction of tar formed depends inversely on the heating rate as shown in Fig. 3.15.[72] Charcoal manufacture extends for days or weeks and char yields can exceed 30%. Fast pyrolysis converts typically >90% of the biomass to tar vapor or gas by using intense external heat transfer and it thus reduces the difficulty of gasifying the charcoal.[73] In updraft gasification, the tar vapors are carried immediately into the cooler parts of the gasifier where some of them condense, recycle, and are cracked, but where a large fraction pass through unchanged so that tar yields can exceed 15%. In downdraft gasification, the tar vapors escaping from the biomass surface immediately encounter higher temperatures and excess oxygen so that they are burned in situ to provide heat for the pyrolysis process. Typically less than 0.5% of the tars survive in a

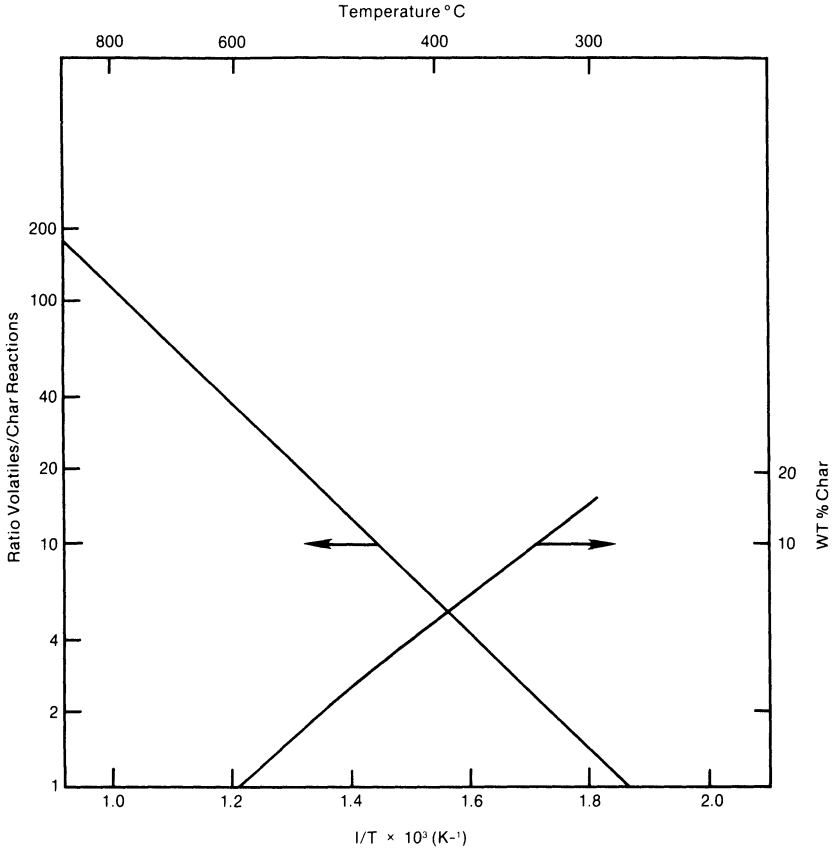


Fig. 3.15: Variation of ratio of volatiles to char with temperature in cellulose pyrolysis.[72]

well-designed downdraft gasifier. Tars are also destroyed in fluidized and suspended beds if they are given sufficient time and temperature.

The rate of pyrolysis is measured by determining the weight loss of a particle as a function of time at constant temperature or a constant heating rate. This process is called thermogravimetric analysis (TGA), and excellent automatic equipment is available from several manufacturers to perform this measurement at fixed temperature or controlled heating rates. A typical TGA curve for wood is shown in Fig. 3.16 where it is seen that the drying, devolatalization, and char devolatalization zones are clearly visible.[67]

3.4.5.2 Kinetics of Solid Pyrolysis Reactions. The sharp, well-defined thermal weight loss, especially for cellulose, suggests that a relatively simple reaction controls the devolatalization. A great deal of effort has gone into fitting classical kinetic theory to TGA data in general and for biomass components in particular.[68] The thermal decomposition curves can be fitted using an Arhenius equation of the

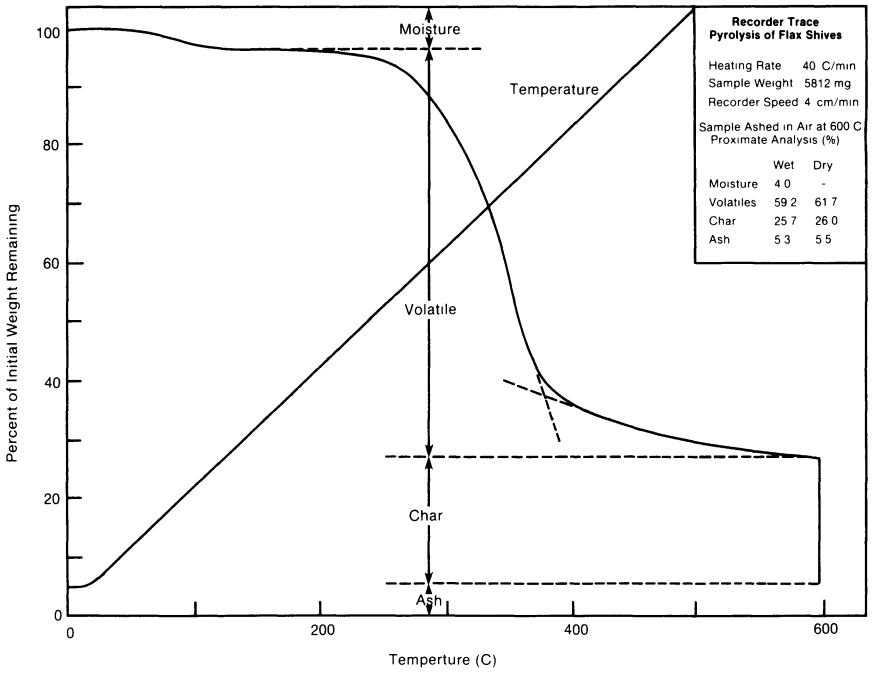


Fig. 3.16: Typical thermogravimetric data for biomass showing loss of moisture, volatiles, and char. Remaining weight is ash.[67]

form [67]

$$\frac{dV}{dt} = kV^n \tag{3.10}$$

where

$$k = A \exp - \frac{E}{RT} \tag{3.11}$$

and V is the fraction of total volatiles remaining at time, t (A is the preexponential frequency factor, E is the activation energy, R is the gas constant, and T is the temperature). The reaction rate k can then be evaluated by measuring the mass as a function of time at constant temperature. However, it is difficult to bring the biomass to temperature without early weight loss. If the sample is heated at a constant rate, $h = dT/dt$, then the first equation becomes

$$\frac{dV}{dT} = k\frac{V^n}{h} \tag{3.12}$$

The rate, k, can also be evaluated from the constant heating rate data and this is more rapid and convenient. A large number of measurements have been made on biomass in both modes to determine the rate constant for pyrolysis, an important constant for understanding gasification and combustion processes.

Table 3.11

Experimental activation energies and frequency factors for first order biomass pyrolysis $[dV/dT = kV$, where $k = A\exp(-E/RT)]$[74,75]

Data Source	Feedstock	Activation Energy (kcal/mol)	Frequency Factory (min^{-1})	Temperature Range (°C)
Stamm	Douglas Fir Sawdust	25.0	1.1×10^{11}	110–220
	α–Cellulose	26.0	2.9×10^{11}	110–220
	Hemicellulose	26.7	2.2×10^{12}	110–220
	Lignin	23.0	8.4×10^{11}	110–220
	Coniferous Wood	29.0	3.1×10^{13}	94–250
Barooah and Long	Wood	4.3	3.2	180–330
	Wood	20.1	1.4×10^{6}	330–400
	Cellulose	17.0	2.4×10^{5}	330–400
Kujirai and Akahira	Manilla Paper	31.7	1.7×10^{14}	145–265
	Linen	31.0	2.7×10^{13}	145-265
	Filter Paper	31.0	4.5×10^{13}	145-265
	Cotton	32.1	7.8×10^{13}	145-265
Tang and Eickner	Wood	54.0	3.9×10^{18}	325–350
	α–Cellulose	56.0	1.0×10^{19}	308–360
Murty and Roberts	Wood	30.0	4.2×10^{9}	250–450
	α–Cellulose	15.0	1.0×10^{6}	25–700
Akita and Kase	α–Cellulose	53.5	6.3×10^{18}	
	Modified Cellulose	32.0	1.0×10^{18}	
Leu	Spruce	51.9	1.6×10^{20}	

Unfortunately, a wide variation of E, A, and the order, n, can give satisfactory fits to the data. Many investigators arbitrarily choose $n = 1$. Selected values for E and A are listed in Table 3.11.[74,75]

A plot of $\log A$ versus E is shown in Fig. 3.17. Here it is seen that the preexponential factor may be a function of the activation energy within the limits of error of the measurements. This suggests that the kinetic constants depend strongly on the method of measurement and, in fact, are probably artifacts of the interpretation rather than true rate kinetic constants. Shafizadah warns "...it should be pointed out that evaporation of levoglucosan and the volatile pyrolysis products is highly endothermic. Thus the increased oven temperature could raise the rate of heat transfer but not necessarily the temperature of the ablating substrate which is cooled by the heat of evaporation ...."[76]

Another more pragmatic approach to the kinetics of gasification is more relevant to real world gasification where bulky fuels are processed at high rates. The global rate of pyrolysis of large particles is primarily determined by heat transfer to the particle, by heat flow through the particle, and by the flow of water vapor and pyrolysis gases from the particle. Due to the interactions between these three fluxes,

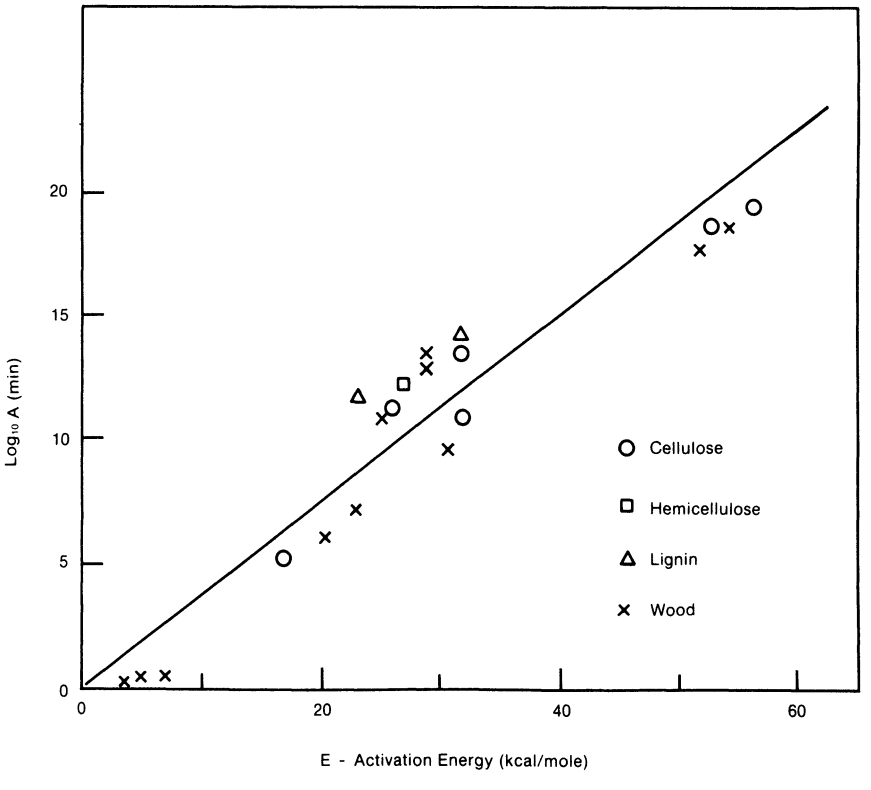


Fig. 3.17: Activation energy for biomass pyrolysis vs. pre-exponential factors.[74,75]

this is a complex problem that can only be solved by numerical techniques for specific cases. The rate of heat transfer and pyrolysis in wood has long been of interest to those doing research in wood combustion, fire prevention, and radiation effects; and there is a voluminous literature on this subject from many viewpoints.[66] Buekens et al. tabulate a number of kinetic and heat flow models of particle pyrolysis.[77] Herman made measurements of the thermal history of densified sawdust cylinders and obtained excellent fit to the experimental data.[74]

The heat source for pyrolysis is important in determining pyrolysis rates. Most studies are done in high temperature furnaces in inert atmospheres. The principal heat transfer mechanism is radiation from the furnace walls, and since the pyrolysis products are largely transparent to this radiation, they do not interfere with the radiant heat transfer.

Another regime important in downdraft gasification can be termed "flaming pyrolysis," intermediate between flaming combustion and externally heated pyrolysis. In flaming combustion, a major source of heat is the combustion or reaction of the pyrolysis gases. This is most easily witnessed in the burning match where the burning pyrolysis vapors supply the heat for continuing pyrolysis until only charcoal remains. In fireplaces, stoves, and furnaces, flaming combustion is aided by radiation from surrounding particles.

In downdraft gasification, air and oxygen are drawn down over freshly pyrolyzing biomass, and initially the combustion supplies the heat needed for pyrolysis. As the air and oxygen are consumed, the flame becomes richer in fuel leaving the combustion products of CO and H_2.

While no complete theory of flaming pyrolysis has been worked out, the data obtained in flaming combustion are relevant to the heat transfer and pyrolysis rates.[63] Recent data taken by Huff correlate the time of flaming combustion with particle size and shape, furnace wall temperature, moisture content, and density in a single equation.[78] Other data related to flaming combustion are presented by Simmons and Ragland and by Reuther and Karsner.[79,80]

3.4.5.3 Kinetics of Biomass Vapor Pyrolysis Reactions. The product of primary biomass pyrolysis is a very complex mixture of organic molecules. The principal initial species from the cellulose is levoglucosan. The principal species from lignin pyrolysis are various complex phenols and polyphenols. However, hundreds of compounds have been identified in the pyrolysis gas. The products that actually survive and are collected depend to a great extent on the type of gasifier used. In an updraft gasifier, the volatile species are immediately swept into a cooler zone where they can condense or pass out of the gasifier and be collected as tar and pyroligneous acids. In a downdraft gasifier, the emerging volatile species are largely burned, but some escape and pass through the high temperature zone where they are cracked to varying degress in secondary pyrolysis.

Antal has constructed a two-furnace system to separate the primary pyrolysis step from the cracking step. A carrier gas is passed over a furnace containing the solid biomass specimen at a temperature typically of 500° C. The gases emerging from this furnace then pass into the second furnace maintained at a temperature between 500° and 1000° C. In this way, it is possible to measure the kinetics of the decomposition of the primary tar species into secondary species.[81]

Antal found that, with residence times up to 8 sec at temperatures from 500° to 750° C, the gases CO, CO_2, H_2, CH_4, and ethylene increase monotonically with time and temperature.[81] Ethane and propylene, however, went through maxima and then decreased in this time range, suggesting further cracking.

3.4.5.4 Mechanisms of Charcoal Gasification. The global thermodynamics of biomass gasification was discussed above and dealt primarily with the relation of the final gas composition to the input fuel, disregarding the mechanism of gas production. In this section, we discuss the narrower subject of char gasification, but the equations of Table 3.4 still apply.

Carbon reacts with either carbon dioxide or water to produce carbon monoxide and hydrogen, according to the Boudouard and Water Gas reactions of Table 3.4. Carbon can also, in principle, react directly with oxygen according to

$$C + O_2 \rightarrow CO_2 \tag{3.13}$$

$$C + CO_2 \rightarrow 2CO \tag{3.14}$$

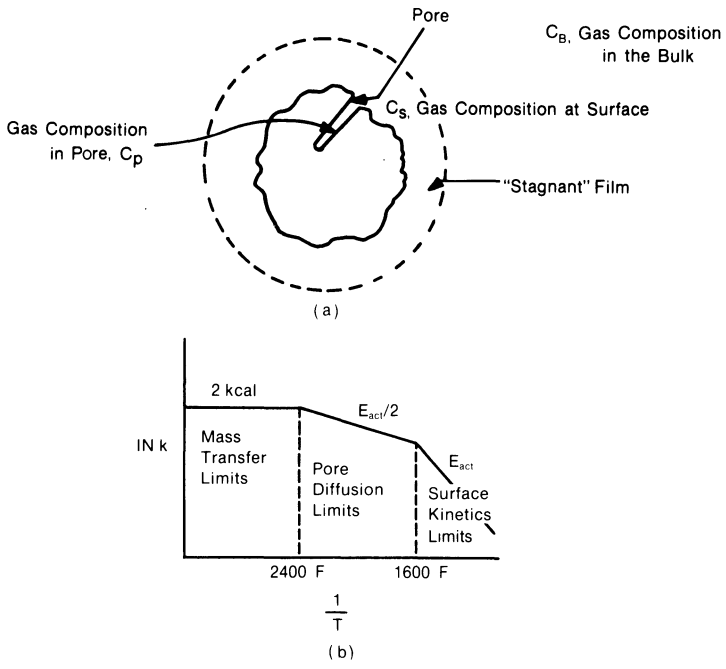


Fig. 3.18: **(a)** Shrinking core model of char gasification and **(b)** three regimes of char gasification.[56]

A great deal of scientific controversy before 1940 revolved around the question of whether carbon reacted with oxygen sequentially to produce CO_2 and then reduced the CO_2 to C) (Eq. (3.13) followed by the Boudouard reaction), or whether it reacted directly according to Eq. (3.14).[82] Current carbon combustion and gasification theory is based on a continuously varying boundary layer, rich in CO near the carbon surface and rich in CO_2, H_2O, or O_2 at the outer edge of the diffusion layer. Heat is conducted to the particle through this layer to provide the endothermic heat of reaction of the carbon with CO_2 or H_2O.

Another reaction of importance in char gasification is that of char and hydrogen to produce methane:

$$C + 2H_2 \rightarrow CH_4 \quad \Delta H = -17 \text{ kcal/mole} \tag{3.15}$$

Freshly devolatalized char is highly reactive, but generally a catalyst is required to promote this reaction at lower temperatures where it is favored.

3.4.5.5 Kinetics of Charcoal Gasification. In the gasification of coal, a relatively unreactive coke is produced. A vast amount of research has looked at the kinetics of coke gasification, and most of the charcoal kinetics work takes this as a starting point.[18,56]

Char is converted to gas by the water gas and Boudouard reactions of Table 3.4 (see Fig. 3.18(a)). At sufficiently high temperatures, these reactions are mass transport limited, and the reaction rate is determined by the rate at which the gases can diffuse to and from the surface of the particle (see Figs. 3.18(a) and (b)). The particles then gradually shrink until they disappear, and this is called the "shrinking core" model.[56,77] At intermediate temperature pore diffusion limits the rate of reaction. Finally, at the lowest temperatures, the reaction rate is limited by the availability of internal surface area for reaction, and the char particle size stays constant while the density decreases. These three regimes are shown in the Arrhenius plot in Fig. 3.18(b).

Char gasification rates are typically measured by recording the weight of char remaining at time, t, and temperature, T, in a flowing gas of CO_2, H_2O, O_2, or mixtures of these. The rate is then defined by

$$\frac{d \ln W}{dt} = r \tag{3.16}$$

where W is the remaining weight of char at time, t. As in pyrolysis studies (see Eqs. (3.10) and (3.11)), the results are often fitted to an Arrhenius equation of the form

$$r = A \exp - \frac{E}{RT} \tag{3.17}$$

Where gas mixtures are involved, it is important to evaluate kinetic data using a Langmuir-Hinshelwood rate law of the type

$$r(T, Pi) = \frac{k_1 p_{H_2O}}{1 + k_2 p_{H_2O} + k_3 p_{H_2}} \tag{3.18}$$

or

$$r(T, Pi) = \frac{k_4 p_{CO_2}}{1 + k_5 p_{CO_2} + k_6 p_{CO}} \tag{3.19}$$

where partial pressures of the various species involve separate rate constants.[56,77]

While formally similar to coke gasification, examination of reported rate data on coke and charcoal suggests that the rates for charcoal are about 25 times faster than those for coke.[61,83] This is probably due to the very low density of the charcoal (due to the high volatile content of biomass, typically 0.05 to 0.1 g/cm^3), the much larger surface area of the charcoal (typically 10 to 100 m^2/g), the low ash content of the charcoal, and the nongraphitic nature of freshly formed charcoal. In addition, char reactivities are even larger in freshly prepared chars than in those stabilized after preparation.[83,84] One should be cautious in comparing data on charcoals taken under different conditions. In practice, examination of actual gasification beds may show a combination of the above mechanisms, with simultaneous shrinking of the charcoal and decrease in density.

For coal chars, the rate of steam gasification is on the order of 0.01 to 0.05 g· g^{-1} min^{-1} at 900°C. Edrich and Graboski measured gasification rates of CO_2 on freshly prepared chars which are on the order of 0.1 g· g^{-1} min^{-1} with a true

activation energy of 54 kcal/mole.[83] However, if the biomass was pretreated at temperatures higher than the gasification temperature, the activation energy was lowered by about 20 kcal/mole due to annealing effect. Thus, it is very important in evaluating char gasification rates to work with chars prepared at a temperature used in gasification if the data is to be relevant. They concluded from data on various size particles that macropore structure did not affect reaction rate to the highest temperature studied (1135°C). Measurements on char from densified (pelletized) biomass showed similar activation energies.

3.4.6 Modeling of Gasifier Operation

The final test of our understanding of gasifier operation is our ability to model or put into mathematical/engineering formulation the operation of a specific gasifier and to predict output and internal conditions from input conditions. Ideally, the perfect model:[77]

- predicts the response of gasifier output to various feedstocks, moisture contents, oxygen concentrations, pressure, etc.;
- predicts the design dimensions of laboratory and industrial gasifiers;
- permits simulation and prediction of gasifier operation under difficult to attain experimental conditions (e.g., high pressure);
- permits extraction of rate data from the operating gasifier; and
- advances the art of reactor modeling and the solution of the complex equations involved.

We are far from having such a model on either coal or biomass for any current gasifier. (Possibly private firms such as Lurgi or Koppers have a model approaching this level, but it is not published in the open literature. It is more likely that they have internal knowledge suggesting the above responses of the gasifier to input variations, but not in mathematical form.)

It is possible that we will never be able to formulate a complete model. Nevertheless, a partial model would be helpful in designing and scaling new gasifiers, and the effort to develop such a model will bring awareness of the true operational parameters to the investigators' attention.

An important step in establishing a realistic model is to have a complete measurement on temperature and gas composition through the gasifier for a number of inlet conditions. However, even these data are not available on most commercial gasifiers, and one must infer internal operation from outlet gas. This is due to the difficulty of inserting temperature and gas probes into a very high temperature, often corrosive atmosphere. In addition, the solid and gaseous species are often not at equilibrium, and one should measure simultaneously radiation, gas, and particle temperature at a point. Temperatures and gas compositions in a 3 m inside diameter commercial coke gasifier are shown in Fig. 3.19.[18]

Unfortunately, the simplifications involved in the measurements and interpretations of measurement of pyrolysis and char gasification rates described in previous

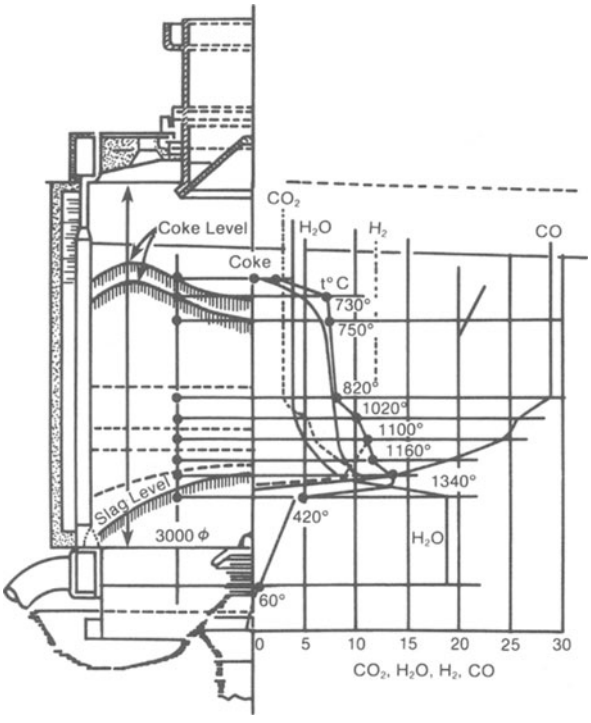


Fig. 3.19: Gas composition and temperatures in a commercial (3 m inside diameter) coke gasifier.[18]

sections do not uniquely permit *a priori* modeling of gasifiers. Many papers on gasifier modeling present some principles but do not predict actual behavior. For lack of experimental data, modelers sometimes compare their predictions to predictions from other models to test the success of the latest model! For this reason, we limit this discussion to modeling activities associated with measurements on real gasifiers. It is hoped that more relevant quantitative data on real gasifiers will be taken in the coming years to guide modelers in their thinking.

Gasifier models can be classified in order of increasing complexity as:[77]

- equilibrium models in which the gas composition is calculated assuming a fixed reaction temperature with heterogeneous- or homogeneous-reaction equilibrium as appropriate;
- kinetics-free models in which the reactor is subdivided into different zones, but the gas composition in the gasification zone is calculated from equilibrium data. The reaction temperature is calculated in each zone by a separate heat balance;
- steady-state models using time independent mass, energy, and heat balances;

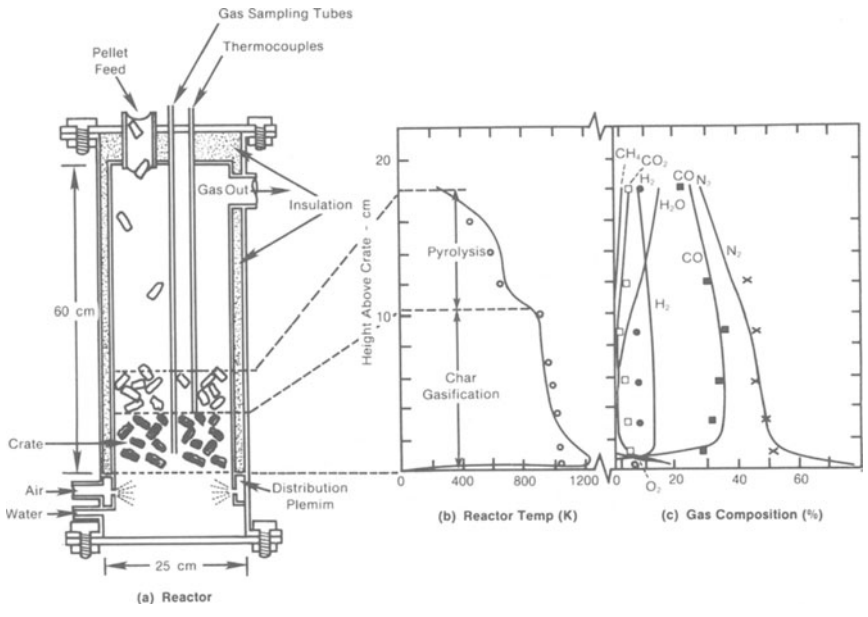


Fig. 3.20: (**a**) Laboratory updraft gasifier, (**b**) temperature profile, and (**c**) gas composition for conditions given in Table 3.12.[61]

- semitransient models in which transients are calculated using the pseudo-steady-state assumption from steady-state models; and
- transient models involving the (computer) solution of the time-dependent, differential-mass energy and heat equations.

3.4.6.1 Updraft Gasifier Modeling. The updraft gasifier is composed of a charcoal-gasification section below and a pyrolysis section above. These can be modeled separately, and the gas contribution of each summed. This kinetic-free model requires at least one set of operating data and can then be readily used to predict variations around this point.[77]

A program was initiated at SERI in 1979 to obtain temperature and gas composition data on the small updraft gasifier shown in Fig. 3.20(a). The temperature distribution and gas composition distribution are shown in Figs. 3.20(b) and 3.20(c) for the conditions listed in Table 3.12. Work is continuing on this gasifier at Texas Tech University.

The solid lines in Figs. 3.20(b) and 3.20(c) are calculated from a dynamic model of gasification based on global rate expressions for pyrolysis and char gasification rates. Pyrolysis is treated as a single reaction (fixed-product slate) whose rate is heat-transfer-limited.

Four heterogeneous char-gas reactions were considered and shift equilibrium was assumed. The model used was the simplest mathematical development that can reasonably account for the phenomena observed in updraft-gasifier operation, but

Table 3.12

Steady-state results: air/steam gasification of densified biomass[61]

Feed(1/4-in. pellets) (wt%*)		
Fixed carbon	14.24	
Volatile matter	77.89	
Moisture	7.15	
Ash	0.72	
C	46.31	
H	5.58	
O	40.24	
LHV [MJ/kg (Btu/lb)]	17.07	(7337)
Rates [kg/h (lb/h)]		
Rate	11.85	(26.13)
Air	7.51	(16.56)
Steam	1.0	(2.205)
Total input	20.36	(40.89)
Offgas	11.83	(26.09)
Oil	3.29	(7.25)
Water make	5.07	(11.18)
Offgas Analysis (mol %[a])		
H_2	14.32	
CO	32.87	
CO_2	5.18	
CH_4	1.67	
C_2H_4	1.67	
C_2H_6	0.09	
C_3H_6	0.09	
O_2 + Ar	0.88	
N_2	44.83	
MW	24.96	
LVH [MJ/m^3 (Btu/ft^3)]	6.469	(174)
Bed height (cm)	18.0	
Pressure (atm)	1	
Grate temperature (°C)	1040	
Offgas temperature (°C)	300	
Duration of run (h)	4.50	

*Except when otherwise indicated.

Table 3.13

Rate constants used in updraft gasifier modelling

Reaction	$A\,(g \cdot g^{-1} \cdot min^{-1})$	Activation Energy, E (MJ/kmol)	(kcal/mole)
$C + C_2 \rightarrow 2CO$	6.0×10^9	27	
$C + H_2O \rightarrow CO + H_2$	$4 - 37 \times 100^{10}$	217	51.8
$C + CO_2 \rightarrow 2CO$	$2.4 - 22.2 \times 10^{10}$	217	51.8

gave quite satisfactory results. The rate expressions used for the carbon, oxygen, steam, and CO_2 reactions are listed in Table 3.13. The C/H_2O and C/CO_2 ratio was calculated using a rate expression containing the sum of the CO_2 and H_2O concentrations raised to the 0.7 power after Groenevelt, with the Boudouard reaction rate taken as 0.6 times the water-gas reaction rate.

While pyrolysis was evaluated as heat-transfer-limited, the gasification reactions were clearly chemical-rate controlled. In addition, the effectiveness factors were close to unity, indicating that more of the internal surface of the particles contributed to the rate.

3.4.6.2 Downdraft Gasifier Modeling. The global operation of the downdraft gasifier lends itself to equilibrium models because the pyrolysis products have a chance to burn and then reach equilibrium in the charcoal zone just before leaving the gasifier. The data cited previously (Table 3.9) show reasonable correspondence between actual operating temperatures and compositions and those predicted from thermodynamics.

The earlier downdraft gasifiers did not lend themselves to detailed modeling in the pyrolysis zone because the high-velocity blast through the air/oxygen nozzles created a nonuniform distribution of pyrolysis regions. Groeneveldt made cold flow studies of the flow around nozzles and other injectors and found that the recirculation patterns around the air blast are very important.[30] The calculated degree of char conversion, temperature, and particle diameters for a kinetic char-conversion model is compared to experimental data taken in a downdraft gasifier with a central air distribution pipe in Figs. 3.21(a), (b), and (c).[84]

We have recently developed the "stratified downdraft gasifier" shown in Fig. 3.22 in which the fuel, air, and gas flow uniformly down through the gasifier.[85] Since there is no seal required at the top of the gasifier, it is possible to insert temperature and gas probes. Since it is a downdraft gasifier, most of the tars and oils are destroyed at the particle surface, and the temperatures are moderated and seldom exceed 1000°C.

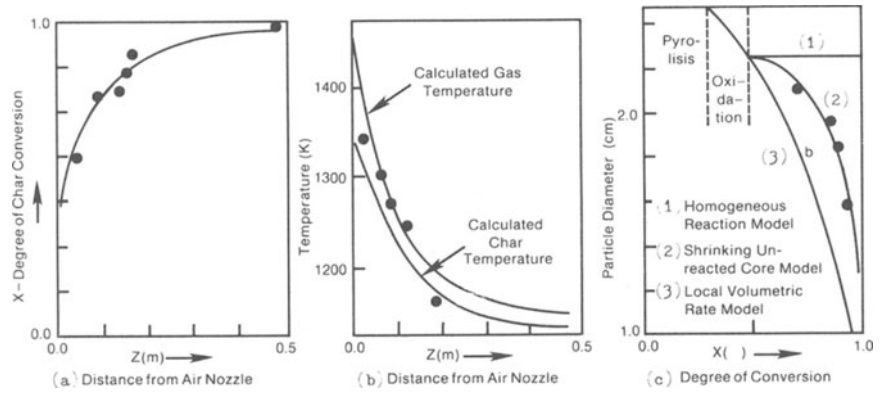


Fig. 3.21: (a) Measured and calculated (b) temperatures, and (c) particle diameters in a downdraft gasifier.

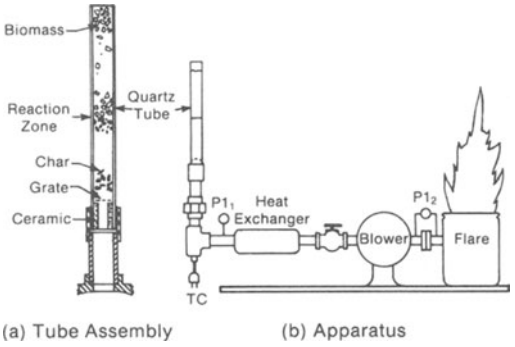


Fig. 3.22: Transparent stratified downdraft gasifier constructed from 5 cm quartz.

The position of the pyrolysis and gasification zones are not inherently stable between the grate and the top of the bed. The rates of motion of pyrolysis and gasification have been measured under various conditions. A number of factors that determine the position of the reacting interface are discussed.[85] We are currently collecting extensive data on this gasifier for use in modeling.

3.4.6.3 Fluidized Bed Gasifier Modeling. In a fluidized bed, distinct flow regimes are possible, depending on the diameter of the biomass and the bed particles. Depending on the relative gas velocity, the bed may be in a homogeneous fluidization region, a fast bubble region, a turbulent fluidization region, or a fast fluidization region. One of the main characteristics of many fluidized-bed reactors is good mixing properties.[77,86,87]

Bacon et al. have developed a computer simulation model for a sand fluidized-bed wood gasifier. The results from the model show that the composition of product gas is almost entirely dependent on the assigned degree of approach to equilibrium

and that the air-to-dry-wood ratio determines the calorific value of the product gas.[86]

In addition, van den Aarsen et al.[87] are developing a fluidized-bed gasifier model based on data taken in a 3-cm diameter fluidized bed operated at a temperature ranging from 700° to 900° C. About 70% of the wood energy is recovered in the gas phase. The particle heating rate becomes the rate-limiting factor at particle diameters between 5 and 10 mm. It was found that gasification kinetics are surface-reaction-limited with an activation energy of 166.2 kJ/mol and a frequency factor of $2.34 \cdot 10^4$ s^{-1}. More experimental data is needed to help evaluate these modeling effects.

3.4.6.4 Fast Pyrolysis Modeling. Fast (or flash) pyrolysis refers to the pyrolysis of biomass at sufficiently high heat-transfer rates so that little or no char is produced. Modeling of fast pyrolysis then involves primary studies of the heat-transfer rate and its effect on char yields. Diebold has developed a very high flux cyclone pyrolysis device that can achieve linear-surface pyrolysis rates of 1 mm/sec without forming the usual char layer. The pyrolysis rate is modeled in terms of reactor temperature and particle force produced by the cyclone.

3.5 CONCLUSIONS

There is a great deal of information available on biomass gasification but much of it is qualitative. While considerable progress has been made in recent years in applying the tools of science to biomass gasification, a great deal more work must be done before we will have a "science of gasification."

3.6 ACKNOWLEDGMENTS

The author would like to acknowledge the invaluable assistance of his co-workers at SERI. He would also like to apologize to the scientists and authors whose work he may have overlooked or not mentioned because of space limitations or ignorance.

We at SERI would like to acknowledge the continuing financial support of the Office of Alcohol Fuels and the Fuels and Chemicals from Biomass offices of the U. S. Department of Energy.

APPENDIX 3.1 UNITS

Although it was our intention to present the collected data within a consistent framework of acceptable metric units, this goal could only be partially achieved. The still widespread use of English units and more convenient practical units did not in all cases allow transfer of the reported data to metric units. The internationally established gram (g), meter (m), seconds (s), and joule (J) system (SI) is therefore occasionally replaced by more convenient units which are more familiar to the reader. (Adapted from *Generator Gas: The Swedish Experience from 1939-1945*[4] and Kaupp[8].)

Length:	1 cm = 0.3937 in.
	1 ft = 30.48 cm
Volume:	1 L = 0.0353 ft^3 = 0.2642 gal = 1000 cm^3
	1 m^3 = 35.3147 ft^3 = 1000 L
Mass:	1 lbm = 453.59237 g
Pressure:	1 atm = 1.0132 bar (10^5 N/m^2) = 760 mm Hg = 103 m H_2O
	1 atm = 14.70 lbf/in^2 = 29.94 in. Hg = 405.7 in. H_2O
Energy:	1 kJ = 0.948 Btu = 238.9 cal
Power:	1 W = 1 J/sec = 3.413 Btu/h
	1 hp = 745.6 W = 2545 Btu/h
Temperature:	Degree Rankin (°R) = 1.8 × Degree Kelvin (°K)
	Degree Celsius (°C) = Degree Kelvin - 273.16
	Degree Fahrenheit (°F) = Degree Rankin - 459.69
	Degree Fahrenheit = 1.8 × Degree Celsius + 32
Miscellaneous:	1 Nm^3 (one norm cubic meter) = one cubic meter of gas at 0° C and 1 atm
	1 scf (one standard cubic foot) = one cubic foot of gas at 77° F and 1 atm (previously at 68° F and 1 atm)
	cfm (cubic feet per minute)
	1 kg-mole of producer gas = 22.4 Nm^3 of producer gas treated as ideal gas
	1 kJ/Nm^3 = 0.239 $kcal/Nm^3$ = 0.0289 $Btu/sft^3(scf)$

REFERENCES

1. Office of Technology Assessment, *Energy from Biological Processes*, OTA- E- 124 (1980).
2. C. J. Singer, *History of Technology*. Vol. 4 (Oxford, 1958) p. 252.
3. G. Egloff and P. Van Arsdell, "Motor Vehicles Propelled by Producer Gas," *Petroleum Engineer* **15**, 645 (1943).
4. *Generator Gas: The Swedish Experience from 1939-1945*, SERI/TP-49-183, edited by T. B. Reed (Solar Energy Research Institute, Golden, Colo., 1979) Republished with index by TIPI Workshop, Allenspark, Colo., Nov. 1982.
5. *Retrofit '79: Proceedings of a Workshop on Air Gasification*, Seattle, Wash.; 2 Feb. 1979, edited by T. B. Reed and D. Jantzen, SERI/TP-49-183, (Solar Energy Research Institute, Golden, Colo., 1979).

6. *A Survey of Biomass Gasification, Vol. I-III*, SERI/TR-33-239, edited by T. B. Reed, (Solar Energy Research Institute, Golden, Colo., 1979). Published in hard cover as T. Reed, editor, *Biomass Gasification: Principles and Technology*, Energy Technology Review No. 67 (Noyes Data Corp., Park Ridge, N. J., 1981).

7. R. H. Hodam, R. O. Williams, and M. C. Lesser, *Engineering & Economic Characteristics of Commercial Wood Gasifiers in North America*, SERI/TR-231-1459. (Solar Energy Research Institute, Golden, Colo., 1982).

8. A. Kaupp and J. R. Goss, *State of the Art for Small (2-50 kW) Gas Producer-Engine Systems*, final report to USDA, Forest Service, Contract No. 53-39R-0-141, (March 1981). A bibliography of the 800 papers used by authors has been published as "Small-scale gas producer-engine systems," by German–Appropriate Technology.

9. *Producer Gas: Another Fuel for Motor Transport*, National Academy of Sciences (National Academy Press, Washington, D. C., 1983). A companion bibliography with 421 references is also available.

10. Solar Energy Research Institute, *Biomass-to-Methanol Specialists Workshop*, Tamarron, Colo.; 3-5 March 1982, edited by T. B. Reed and M. Graboski, SERI/CP-234-1590 (Solar Energy Research Institute, Golden, Colo., 1982).

11. *Proceedings of the 16th Biomass Thermochemical Conversion Contractor's Meeting*, Portland, Ore,; 8-9 May 1984, PNL-SA-12403, edited by Pacific Northwest Laboratory, Richland, Wash., 99352. This and previous proceedings describing work supported by the U. S. Department of Energy are available from National Technical Information Service (NTIS), U. S. Department of Commerce, 5285 Port Royal Rd., Springfield, Va. 22161.

12. J. L. Jones and S. B. Radding, *Thermal Conversion of Solid Biomass Wastes*, ACS Symposium Series No. 130 (Am. Chem. Soc., Washington, D.C., 1980).

13. *Proceedings of the Specialists' Workshop on the Fast Pyrolysis of Biomass*, Copper Mountain, Colo.; 19-22 Oct. 1980, edited by J. Diebold, SERI/CP-622-1096 (Solar Energy Research Institute, Golden, Colo., 1981).

14. Solar Energy Research Institute, *Fundamentals of Thermochemical Biomass Conversion: An International Conference*, Estes Park, Colo.; 18-22 Oct. 1982, edited by R. P. Overend, T. A. Milne, and L. Mudge. (Elsevier Applied Science Publications, Ltd., 1985).

15. Forest Products Research Society (FPRS), *Proceedings of the 6th Industrial Wood Energy Forum, Forum '82*, 1983. See also previous FPRS symposia (available from FPRS, Madison, Wis.).

16. D. Klass, ed., *Energy from Biomass and Wastes Symposium VIII*, Lake Buena Vista, Fla.; Jan. 1983. Proceedings will be published in the continuing series with similar titles numbered 1–7, held approximately yearly in Orlando, (1–3) or Lake Buena Vista, (4–7) Fla., starting in Jan. 1977. Published by the Institute of Gas Technology (IGT), Chicago, Ill.

17. N. E. Rambush, *Modern Gas Producers* (Benn Bros. Lts., London, 1923). Microfiche copy at Energy Branch Library, CISTI, National Research Council of Canada, Ottawa, Ontario, K1A 0R6.

18. W. Gumz, *Gas Producers and Blast Furnaces* (J. Wiley and Sons, New York, 1950).

19. E. Johansson, "Vehicle Gasifiers," in *Retrofit '79: Proceedings of a Workshop on Air Gasification*[5], p. 5-5-1.

20. Commission of the European Communities, *Energy from Biomass*, Brighton, England, 1 Nov. 1980.

21. "Coast to Coast on Homemade Fuel," *Mother Earth News* 178 (Jan. 1982).

22. B. Russell, "Evaluation of a Biomass Gasification Experiment for Vehicle Power," in *Energy from Biomass and Wastes IV*, edited by D. Klass (Institute of Gas Technology, Chicago, Ill., 1980), p. 819.

23. *The International Bio-Energy Directory and Handbook–1984*, edited by Paul F. Bente, Jr. (The Bio-Energy Council, Washington, D. C., 1984).

24. Biomass Energy Institute, Winnipeg, Manitoba, Canada, Publishers of Bio-Joule Newsletter.

25. *Energy from Biomass Conference*, International Congress Center, Berlin, W. Germany, Commission of the European Communities, 20 Sept. 1982).

26. *Producer Gas 1982*, conference in Sri Lanka, 8-12 Nov. 1982, sponsored by Swedish International Dev. Agency, edited by B. Kjellstrom. (Beijer Inst., Stockholm, 1983). Previous publications also available.

27. T. Imai, "Compact Gasifier-Burner Operated by Special Charcoal," in *Proceedings of the 6th Industrial Wood Energy Forum, Forum '82*[15].

28. Andco-Torax, in "Solid Wastes and Residue Conversion by Advanced Thermal Processes," *ACS Symposium Series 76* (ACS, Washington, D. C., 1978). Also see Klass[16], IV, p. 577.

29. T. Masuda and T. F. Fisher, "Purox System Demonstration Test on Simulated Japanese Refuse," in *Thermal Conversion of Solid Biomass Wastes*[12], p. 573.

30. M. D. Groeneveld and W. P. M. Van Swaaij, "The Design of Co-Current, Moving Bed Gasifiers Fueled by Biomass," in *Thermal Conversion of Solid Biomass Wastes*[12], p. 463.

31. The Buck Rogers Co., Olathe, Kans.

32. Environmental Energy Engineering, *Report AH-8-1077-1 to Solar Energy Research Institute*, (1979).

33. T. B. Reed, M. Markson, and M. Graboski, "The SERI High Pressure Oxygen Gasifier," in *Biomass-to-Methanol Specialists Workshop*[10], p. 151, also "The SERI Oxygen Gasifier, Phase II," T. B. Reed, B. Levie, M. Markson, SERI/PR-234-2571, 1985.

34. "International Harvester Builds Methanol Units Around SERI Gasifier," *Biomass Digest* **4** (Nov. 1982).

35. W. P. Walawender. F. M. Chern, and L. T. Fan, "Wood Gasification in a Commercial Downdraft Gasifier," in *Fundamentals of Thermochemical Biomass Conversion: An International Conference*[14].

36. L. K. Mudge, et al., "Steam Gasification of Wood in the Presence of Catalysts," in *Proceedings of the 16th Biomass Thermochemical Conversion Contractor's Meeting*[11], p. 9. Also, L. K. Mudge and D. H. Mitchell, "Catalytic Generation of Methanol Synthesis Gas from Wood," in *Biomass-to-Methanol Specialists Workshop*[10], p. 179.

37. See *Engineering & Economic Characteristics of Commercial Wood Gasifiers in North America*[7], p. 50.

38. J. W. Black, "The Pressurized Fluidized Bed Gasifier," in *The Synthesis of Methanol from Wood*, in *Biomass-to-Methanol Specialists Workshop*[10], p. 231.

39. T. B. Reed and D. Jantzen, in *A Survey of Biomass Gasification, Vol. I-III*[6], p. III-64.

40. W. A. Stevenson, "Large Wood to Methanol Plants," in *Thermal Conversion of Solid Biomass Wastes*[12], p. 175.

41. T. Reed, "Oxygen Production for Biomass Gasification," in *Biomass-to-Methanol Specialists Workshop*[10], p. 79.

42. M. Kagayama, et al., "Gasification of Solid Waste in Dual Fluidized Bed Reactors," in *Thermal Conversion of Solid Biomass Wastes*[12], p. 525.

43. H. F. Feldman, M. A. Paisley, and H. R. Appelbaum, "Gasification of Forest Residues in a High-Throughput Gasifier," in *Proceedings of the 16th Biomass Thermochemical Conversion Contractor's Meeting*[11], p. 46.

44. J. A. Coffman, "Steam Gasification of Biomass," in *Proceedings of the 16th Biomass Thermochemical Conversion Contractor's Meeting*[11], p. 32.

45. The Pyrenco Co., Prosser, Wash.

46. *Engineering & Economic Characteristics of Commercial Wood Gasifiers in North America*[7], p. 72.

47. J. L. Campbell, "Gas Producers," *Automobile Engineer* **32** 156 (April 1942).

48. W. Strauss, *Industrial Gas Cleaning* (Pergamon Press, New York, 1975).

49. S. Calvert, *Wet Scrubber System Study*, EPA R2.72-118A&B (U. S. Environmental Protection Agency, Washington, D. C., 1972). Also available from NTIS, PB 213016, 213017, July 1972).

50. Chapter 6 in *State of the Art for Small (2-50 kW) Gas Producer-Engine Systems*[8].

51. R. Bennett, Chapter 12 in *A Survey of Biomass Gasification, Vol I-III*[6].

52. "Combustion and Flame," Journal of the Combustion Institute.

53. M. Graboski and R. Bain, Chapter 2 in *A Survey of Biomass Gasification, Vol I-III*[6], p. 25.

54. R. Desrosiers, "Thermodynamics of Gas-Char Reactions," Chapter 6 in *A Survey of Biomass Gasification, Vol I-III*[6].

55. M. Graboski and R. Bain, "Properties of Biomass Relevant to Gasification," Chapter 2 in *A Survey of Biomass Gasification, Vol I-III*[6], p.33.

56. M. Graboski, "Reactivity of Char," Chapter 7 in *A Survey of Biomass Gasification, Vol I-III*[6].

57. T. Reed, Chapter 8 in *A Survey of Biomass Gasification, Vol I-III*[6].

58. M. Steinberg and P. Fallon, "Flash Pyrolysis of Biomass with Reactive and Non-Reactive Gases," in *Proceedings of the 16th Biomass Thermochemical Conversion Contractor's Meeting*[11], p. 452.

59. J. Diebold, in *Fundamentals of Thermochemical Biomass Conversion: An International Conference*[14], p. 237.

60. R. Overend. "Wood Gasification—Review of Recent Canadian Experience," National Research Council (Canada) Report NRCC 20094, March 1982.

61. R. Desrosiers, "Fundamental Air-Gasification Engineering Parameters," SERI-/PR-234-1470, (February 1982). Also "Gasification Engineering—Quarterly Report," SERI/PR-622-1348 (December 1981) and SERI/PR-622-1153 (April 1981).

62. D. R. Stull, "JANAF Tables of Thermochemical Data," 2nd ed. NBS 37 (U. S. Chamber of Commerce, 1971).

63. O. Kubaschevsky, E. L. Evans, and C. B. Alcock, *Metallurgical Thermochemistry*, 4th ed. (Pergamon Press, Oxford, 1967).

64. T. Reed, *Free Energy of Formation of Binary Compounds*, (MIT Press, Cambridge, Mass., 1971).

65. V. Yang, in *Biomass-to-Methanol Specialists Workshop*[10], p. 219.

66. F. Williams, "Chemical Kinetics of Pyrolysis" in *Heat Transfer in Fires*, P. L. Blackshear, ed. (J. Wiley, New York, 1974), p. 197.

67. T. Milne, Chapter 5 in *Biomass-to-Methanol Specialists Workshop*[10].

68. M. Antal in *Advances in Solar Energy: An Annual Review of Research and Development* edited by K. W. Böer and J. A. Duffie (American Solar Energy Society, Boulder, Colo.) Vol. 1, pp. 61–112.

69. F. Shafizadeh, in *Fundamentals of Thermochemical Biomass Conversion: An International Conference*[14].

70. E. Soltes, A. T. Wiley, and S.-C K. Lin, in *Biotechnology and Bioengineering Symposium No. 11* (John Wiley, New York, 1981), p. 125.

71. T. Reed, Chapter 8 in *A Survey of Biomass Gasification, Vol I-III*[6].

72. J. Diebold (private communication).

73. See *Proceedings of the Specialists' Workshop on the Fast Pyrolysis of Biomass*[13], pp. 137–347.

74. D. R. Herman, "The Rate of Pyrolysis of Densified Ponderosa Pine," Masters' thesis (T-2525), Colorado School of Mines, Dept. of Petroleum-Refining Engineering, Sept. 1981.

75. D. Q. Tran, "Kinetic Modelling of Pyrolysis and Hydrogasification of Carbonaceous Materials," Ph.D. thesis, University of Wyoming, 1978.

76. F. Shafizadeh, "Pyrolytic Reactions and Products of Biomass," in *Fundamentals of Thermochemical Biomass Conversion: An International Conference*[14].

77. A. G. Beukens and J. G. Schoeters, "Modelling of Biomass Gasification," in *Fundamentals of Thermochemical Biomass Conversion: An International Conference*[14].

78. E. R. Huff, "Effect of Size, Shape, Density, Moisture and Furnace Wall Temperature on Burning Times of Wood Pieces," in *Fundamentals of Thermochemical Biomass Conversion: An International Conference*[14].

79. W. W. Simmons and K. W. Ragland, "Single Particle Combustion Analysis of Wood," in *Fundamentals of Thermochemical Biomass Conversion: An International Conference*[14].

80. J. J. Reuther and G. G. Karsner, "Plane Flame Furnace Combustion Studies of Pulverized Wood," in *Fundamentals of Thermochemical Biomass Conversion: An International Conference*[14].

81. M. J. Antal, "The Effects of Residence Time, Temperature and Pressure on the Steam Gasification of Biomass," in ACS Symposium "Biomass as a Nonfossil Fuel Source," Honolulu Meeting, 1 April 1979.

82. D. H. Bangham and J. G. Bennett, "The Chemistry of Gasification with Reference to Small Producers," *Fuel* **19** 95 (1940).

83. R. Edrich, T. Bradley, and M. Graboski, "The Gasification of Ponderosa Pine Charcoal," in *Fundamentals of Thermochemical Biomass Conversion: An International Conference*[14].

84. M. J. Groeneveld and W. P. M. Van Swaaij, "Gasification of Char Particles with CO_2 and H_2O," *Chem. Eng. Sci.* **35** (1-2) (collected papers of Chemical Reaction Engineering Conference, Nice, Vol. 1, Contrib. pap. 1980), pp. 307–13.

85. T. Reed and M. Markson, "Reaction Velocities in a Downdraft Gasifier," in *Fundamentals of Thermochemical Biomass Conversion: An International Conference*[14].

86. "An Annotated Bibliography of Gasifier Modelling," D. W. Bacon et al., Dept. of Chem. Eng. Rept., Queen's University, Kingston, Ontario K7L 3N6, Jan. 1982.

87. F. G. van den Aarsen, A. A. C. M. Beenackers, and W. P. M. Van Swaaij, "Wood Pyrolysis and CO_2 Char Gasification in a Fluidized Bed," in *Fundamentals of Thermochemical Biomass Conversion: An International Conference*[14].

CHAPTER 4

BIOMASS PYROLYSIS: A REVIEW OF THE LITERATURE

PART 2—LIGNOCELLULOSE PYROLYSIS

Michael Jerry Antal, Jr.

This paper is dedicated to the memory of Fred Shafizadeh, Professor of Chemistry and Director of the Wood Chemistry Laboratory, University of Montana.

4.1 ABSTRACT

A normative review of the literature describing the products, mechanisms, and rates of lignin and whole biomass pyrolysis is presented. The role of a complex sequence of competing solid- and vapor-phase pyrolysis pathways is elucidated.

4.2 INTRODUCTION

In stark contrast with the rest of the United States, the state of Hawaii has maintained its commitment to the development of alternative, renewable sources of energy. Since U. S. national policy now presumes that shortages of energy are unlikely to undermine the growth and prosperity of the industrialized world, why has Hawaii with its limited resources not abandoned its quixotic search for energy independence?

Here in Hawaii memories of World War II remain fresh in the thought of many of its people. Pearl Harbor and Punch Bowl offer constant reminders of the outcome of flawed national policy. Many recall how on 26 July 1941 President Roosevelt announced an embargo on American exports of oil, iron, and rubber to Japan in response to Japanese landings in Saigon and their seizure of Cam Ranh Bay. Great Britain and the Netherlands immediately followed suit. Japan was completely

dependent on imported oil to fuel its economy and its war machine. Japanese response to the embargo was quick and decisive. Less than five months passed before the Japanese attacked and destroyed the U. S. Asiatic fleet in Pearl Harbor. Here in Hawaii these memories provoke unsettling questions concerning the continuing dependence of the industrialized world on imported oil and the foreseeable outcome of any interruptions in its flow.

In Hawaii bagasse is burned on most of the islands to raise steam and generate electric power. Other forms of biomass are available at remote locations around the world, and much interest is presently directed toward the use of these materials to generate electric power. However, as photovoltaic and wind power technologies evolve and become more economical, biomass will be increasingly sought as a source of fluid fuels for vehicular transportation. Because the biomass resource is not large, efficient conversion processes will be requisite. Thermochemical processes, such as those utilized in petroleum refining, are characterized by very high first law (thermal) and second law (availability or exergy[1]) efficiencies, which often exceed 95%. The focus of this review is the underlying science of the thermochemical conversion of biomass to higher value fuels and chemicals.

4.2.1 Scope

Part 1 of this Review in Volume 1 of this series[2] summarized the chemical and physical properties of biomass materials, critiqued existing experimental techniques for studying biomass pyrolysis, and detailed our knowledge of the pyrolysis of carbohydrates. Following a brief summary and update on the pyrolysis of carbohydrates, this Review (Part 2) describes lignin and whole biomass pyrolysis, and offers an original perspective on the generic technologies and economics of biomass conversion.

4.2.2 Goals

The primary goal of this Review continues to be the identification of chemical pathways which lead to the selective formation of a few high value products. Until such pathways are identified, biomass pyrolysis is unlikely to enjoy significant commercial development.

A second major goal of this Review has evolved since Part 1 was written. During the preparation of Part 2, it became increasingly hard to ignore the similarities which appear to exist between the responses of all biomass materials to pyrolysis. Although these similarities remain speculative, a major goal of this work is to suggest those similarities as working hypotheses to guide the design of future experiments. The essence of this second goal is embodied in Fig. 4.1, which was first used by Professor Fred Shafizadeh to describe the pyrolysis of cellulosic carbohydrates. A major conclusion of this Review is that Fig. 4.1 offers useful insight into the pyrolysis of all carbohydrates and lignins. Similar, more detailed, conclusions also follow.

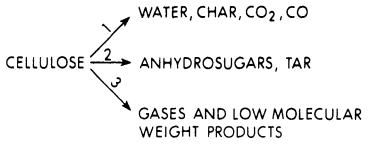


Fig. 4.1: Competing low (1), moderate (2), and high (3) temperature pathways of cellulose pyrolysis.[5]

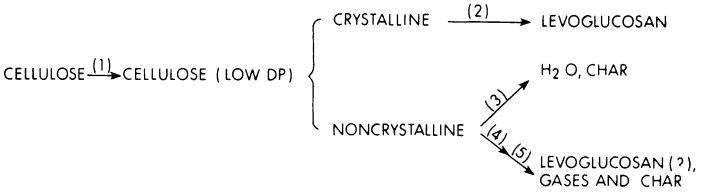


Fig. 4.2: Low temperature pathways of cellulose pyrolysis.[2] (1) Sharp drop in DP to 200 preceding weight loss and occurring at crystalline boundaries, $E \simeq 27$ kcal/gmol, zero order reaction. (2) Levoglucosan peels off crystal face, $E \simeq 60$ kcal/gmol. (3) Fast thermal auto cross linking, char = 1.6 times the initial weight loss, may be concurrent with (1). (4) Rate-limiting initiation step of slower bulk reaction, $E \simeq 30$ kcal/gmol. (5) Propagation step of slower bulk reaction, char = 20% of noncrystalline mass.

4.3 PYROLYSIS OF CARBOHYDRATES

The pyrolysis of both cellulosic and noncellulosic carbohydrates was described in Part 1 of this Review.[2] In order to provide continuity between the two parts, a brief summary of Part 1 has been included here. Other reviews of carbohydrate pyrolysis are also available.[3-20]

4.3.1 Low Temperature Phenomena

The literature contains extensive descriptions of the thermodynamic properties of many carbohydrates,[21-24] including detailed insights into the crystalline structure and other properties of anhydrosugars.[25-29] An understanding of these properties is often prerequisite to the interpretation of complex data derived from thermal analysis experiments.

4.3.1.1 Products. Polymerization phenomena dominate the low temperature chemistry of noncellulosic carbohydrates. They are accompanied by the evolution of H_2O, CO_2, CO, and lesser amounts of furans and low molecular weight carbonyl compounds. Anhydrosugars,[30-38] mono- and di-saccharides,[39-48] oligo- and polysaccharides[11-15,49-58] all polymerize at temperatures below 200°C to form dextrins and related, highly branched, polymers.

Cellulosic materials decompose slowly at temperatures below 250°C to H_2O, CO, CO_2, and a carbonaceous residue.[59-65] At somewhat higher temperatures, a sirup composed primarily of levoglucosan is also obtained following an initial 5%

to 10% weight loss.[66] After a 20% to 30% weight loss, the remaining residue is composed primarily of oligosaccharides.[67]

4.3.1.2 Mechanisms and Kinetics. The complex polymerization chemistry of noncellulosic carbohydrates involves the formation of new ether linkages by transglycosylation reactions[15,20,54] through a carbonium ion intermediate. Acid catalysts are known to favorably influence the polymerization rate and improve the yields of polysaccharides.

Upon heating, cellulosic materials undergo a dramatic drop in degree of polymerization (DP) to about 200 with little weight loss.[66,68-74] As summarized in Fig. 4.2, different pathways play a role in the pyrolysis of the crystalline and amorphous regions of cellulosic materials.[66-68,72,75-80] The preferential formation of levoglucosan from crystalline cellulose, and the competitive degradation mechanism of noncrystalline cellulose emphasizing dehydration reactions, has largely been confirmed by the recent research of Cabradilla and Zeronian.[81] The degree of stretch (orientation) affects the cross linking reactions* which lead to char formation in the amorphous regions of cellulose fibers.[82,83] The mechanism of the cross linking reactions remains an open question, with research supporting both the roles of homolytic[84-89] and heterolytic[90-93] mechanisms. The presence of trace quantities of ash in a cellulose sample can exert a dramatic effect on its weight loss behavior by catalyzing the char forming dehydration reactions.[71] This problem is known to have undermined the research of some early workers in the field.[71]

4.3.2 Moderate Temperature Phenomena

Above 200°C all noncellulosic carbohydrates undergo rapid decomposition, producing a variety of anhydrosugars, light volatile organic compounds, H_2O, CO_2, CO, and char. The crystalline structure of most cellulosic carbohydrates permits them to retain much of their integrity to somewhat higher temperatures and facilitates the selective evolution of anhydrosugars (especially levoglucosan) above 250°C. The effects of fine structure on the pyrolysis of cellulose, evident at low temperatures, are largely masked by the rapid decomposition of the polymer at higher temperatures. However, fine structure will influence the yields of levoglucosan from cellulose throughout this temperature range.[94]

4.3.2.1 Products. At low pressure, anhydrosugars distill when heated to temperatures above 200°C and can be quantitatively recovered.[94-99] However, levoglucosenone is known to polymerize in the liquid phase even at low temperatures,[100] and many anhydrosugars undergo vapor phase dehydration reactions in the presence of acid catalysts.[95,97,101,102] In addition to distillation,[103,104] monosaccharides also polymerize to high molecular weight glucans which undergo subsequent fragmentation reactions, resulting in the formation of a wide variety of furans, ketones, aldehydes, and acids.[105,106] Heyns and his coworkers[105,106] observed the same degradation products from sixteen mono-, di-, oligo-, and polysaccharides. This caused

* In Part 1 of this Review, concern was expressed over the apparent use of a (mis)identity $A^2 - B^2 =^? (A - B)^2$ by Lewin and Basch[68]. Lewin and Basch have independently checked their work and report that their analysis did not involve the use of the (mis)identity given above.

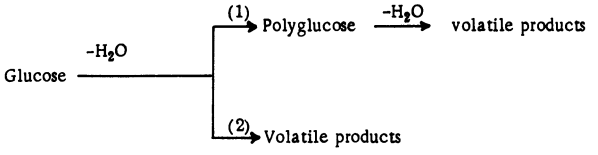


Fig. 4.3: Competitive pathways of glucose pyrolysis.[104]

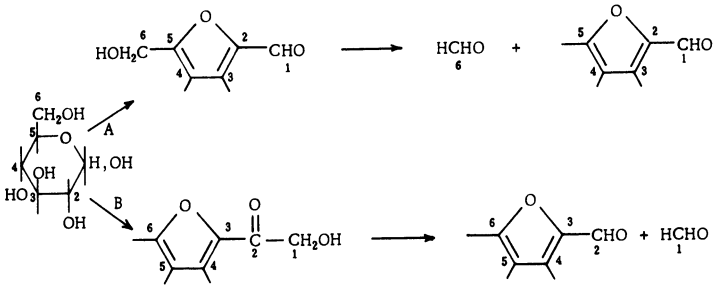


Fig. 4.4: Competitive pathways for furfural formation from glucose.[104]

them to conclude that by dehydration, degradation, and condensation reactions all carbohydrates form similar polymeric intermediates, which then undergo secondary thermal degradation. These findings were in concord with the earlier work of Bryce and Greenwood.[107-109]

Above 250°C transglycosylation reactions effectively compete with the polymerization reactions, and result in the formation of a condensable sirup composed primarily of levoglucosan.[110-117] The crystalline structure and 1 → 4 β-glucosidic linkages of cellulose[106,118,119] favor anhydrosugar formation; however, Furneaux and Shafizadeh[120] were able to obtain a dramatic increase in the yield of 1,6-anhydro-β-D-mannopyranose from ivory nut meal by an acid pretreatment, which reduced the nut meal's ash content from 1.29% to 0.12%. Similar pretreatments,[121,122] as well as reduced pressure[123] and a steam environment,[124-128] improve the yield of levoglucosan from cellulose.

4.3.2.2 Mechanisms and Kinetics. In the absence of catalysts, most anhydrosugars are stable in the vapor-phase at low pressures to temperatures of about 500°C. Figures 4.3 to 4.5 summarize the extensive mechanistic studies of D-glucose thermolysis by Houminer and Patai.[103,104] They found the main source of anhydrosugar formation to be the depolymerization of the glucan (polyglucose) intermediate, which was formed at lower temperatures. Dehydration reactions were shown to play an important role in (1) the formation of glucan, (2) the formation of an enol which rearranges to a carbonyl function, and (3) secondary reactions. Fragmentation reactions were facilitated by the formation of carbonyl and methylene groups, which weaken the carbon skeleton.

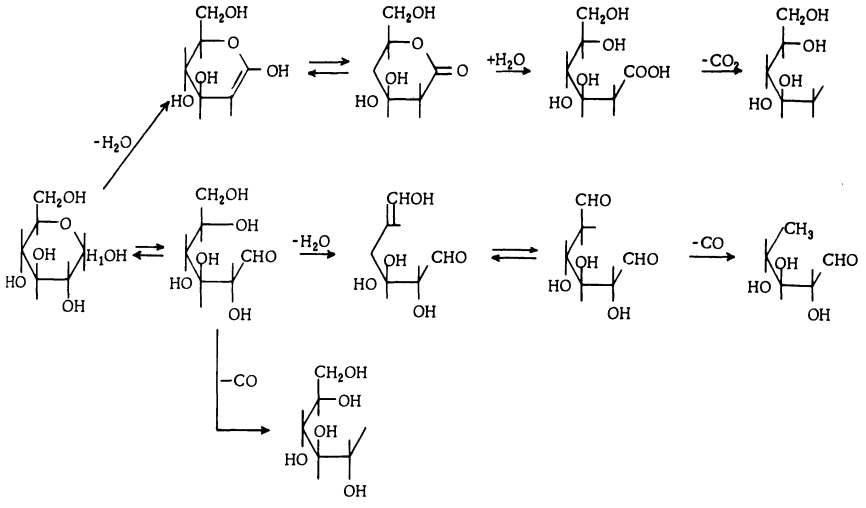


Fig. 4.5: Competitive fragmentation reactions of glucose pyrolysis.[104]

The mechanism of starch decomposition at moderate temperatures is not well understood. Like noncrystalline cellulose, starch undergoes an initial rapid first order reaction followed by a slower first order reaction.[107,119] Above 270°C, Bryce and Greenwood[107] found the molar ratio of CO_2:CO evolution to be independent of temperature with a value between 3:1 and 4:1 for starch, amylose, amylopectin, and cellulose. Similarly, the molar ratio CO:H_2O was observed to lie between 3:1 and 4:1 for the three starches, but reached 5:1 for cellulose. Also, like noncrystalline cellulose, Bryce and Greenwood[107] observed the apparent activation energy for decomposition of the three starches to lie between 29 and 30 kcal/g mol.

Figure 4.6 summarizes the findings of Mok and Antal[129,130] and many previous workers[131-136] concerning the multiple pathways available for cellulose pyrolysis at moderate temperatures. Kinetic data due to Bradbury and Shafizadeh[136] for some of the reactions depicted in Fig. 4.6 are summarized in Fig. 4.7. Other lumped parameter models which do not embody a competitive mechanism have also been proposed.[137-140] Because these models are unable to account for the influence of heating rate on char formation, they must necessarily be specious.

As described later in this review, in order to obtain increased throughputs and improved product selectivity, engineers developing advanced biomass pyrolysis reactors have emphasized high temperatures and heating rates in their designs. Consequently, a knowledge of high temperature reaction rates is critically important to the biomass community. Two approaches have been followed to estimate high temperature reaction rates. The most popular involves an attempt to accurately measure these rates in the temperature regime of interest. The second involves an extrapolation of low temperature data to high temperatures. Differences in these two approaches are indicated in Fig. 4.8, which summarizes results of the kinetic studies of Hajaligol et al.,[141] Antal et al.,[142] and Bradbury et al.[136] At the higher temperatures, the model developed by Bradbury et al.[136] projects reaction rates

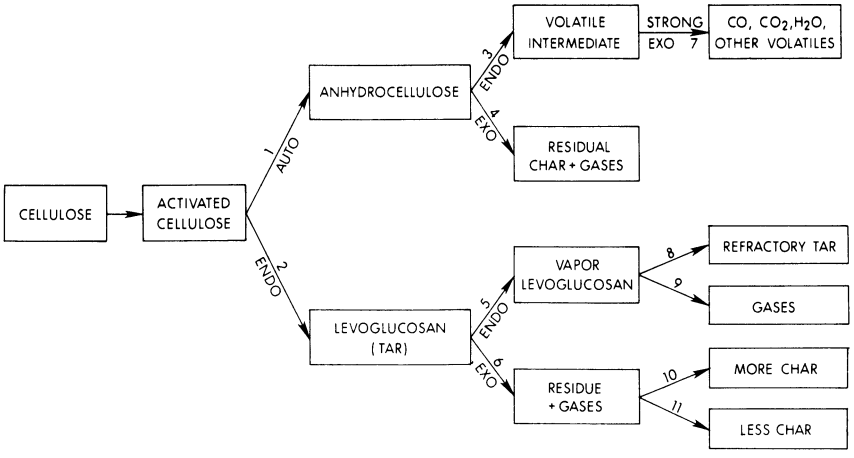


Fig. 4.6: Pathways of cellulose pyrolysis.[130] Pathways 10 and 11 are not well understood. At present all that is known is that the polymer product of Pathway 6 undergoes pyrolysis by at least two competing pathways, one of which produces more solid residue (char) than the second.

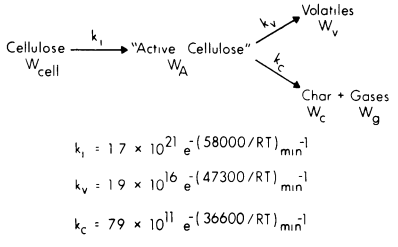


Fig. 4.7: Lumped parameter model for cellulose pyrolysis.[136]

10^5 more rapid than those measured by Hajaligol et al.[141] or projected by Antal et al.[142] How can these disparate results be reconciled?

Research in my laboratory has pointed to the value of Friedman signatures* as a means for gaining insight into the mechanisms and rates of complex solid phase pyrolysis reactions.[143] Figure 4.9 presents the low and high heating rate Friedman signatures of the Avicel microcrystalline cellulose undergoing pyrolysis in flowing N_2, obtained using a standard Dupont TGA system. The U-shaped curve is a signature of the ABCBD mechanism used by Bradbury et al.[136] to model cellulose pyrolysis (see Fig. 4.7). The downward shift of the U at higher heating rates is a signature of the influence of heat transfer on the kinetic measurements.

* As defined in *Thermal Analysis Proceedings of the 7th ICTA*[143], a Friedman signature is a graph of apparent activation energy E versus degree of conversion. The signature is obtained by application of the Friedman mapping function to nonisothermal TG data.

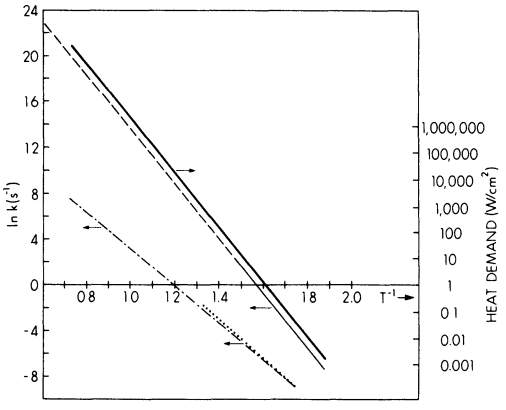


Fig. 4.8: Arrhenius plot of cellulose pyrolysis rate data. Apparent rates $k(s^{-1})$ of weight loss, dashed and dotted line, and dotted line, from Hajaligol[141] and Antal[142] (respectively) vs $10^3 T^{-1}(K^{-1})$. Rate of sirup formation, solid line (measured) and dashed line (extrapolated), from Bradbury et al.[136]. Heat demand, dark solid line, calculated using the relation Heat Demand = $\rho \Delta H k$ where $\rho(g/cm^2)$ is the apparent density of a single sheet of filter paper, ΔH = 335 J/g, and k is the rate of sirup formation from Bradbury et al.[136]

Figure 4.10 presents synthetic Friedman signatures for cellulose exposed to the same set of experimental conditions, using the rate law of Bradbury et al.[136] together with a rate law describing heat transfer within the TGA. Although not absolutely conclusive, the agreement of Figs. 4.9 and 4.10 offers strong support for the rate measurements of Bradbury et al.[136] This agreement also clearly demonstrates the earlier rate measurements of Antal et al.[142] to have been flawed by the influence of poor heat transfer within the TGA.*

In a recent meeting, Lédé[144] speculated that at very high heating rates biomass pyrolysis occurs at about 500°C. Figure 4.8 suggests a rationale for this speculation. Above 550°C the fast, endothermic depolymerization pathway makes extraordinary heat transfer demands on any apparatus attempting to cause pyrolysis at high temperatures.† These heat transfer demands are usually satisfied only if large local temperature gradients exist near the sample. Large local temperature gradients give rise to flawed temperature measurements. Hence the difficulty in studying pyrolysis kinetics at high temperatures. Data on heat demands of cellulose

* Other possible explanations for the differences between my rate data and that of Bradbury et al.[136] include the potential influence of a) the "high" ash content of Whatman No. 1 filter paper, and b) the gaseous environment (flowing N_2 versus vacuum). These influences are now being studied in my laboratory; however, results to date suggest heat transfer limitations to be the primary contributor to the differences noted in Fig. 4.8.

† Nota Bene: Flux densities in excess of 1000 W/cm^2, the equivalent of 10,000 suns, are usually obtained only in high quality solar furnaces. Flux densities of 10000 W/cm^2 can only be obtained using xenon flash tubes or focused laser radiation.

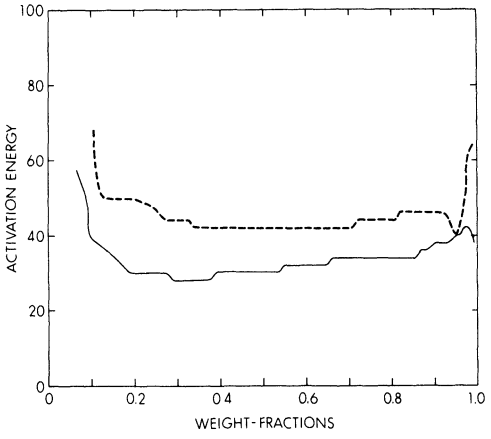


Fig. 4.9: Friedman signatures[143] for Avicel PH102 microcrystalline cellulose pyrolysis in flowing N_2 using a Dupont 951 TGA with heating rates of 2°, 6° and 11°C/min (dashed line), and 20°, 55° and 116°C/min (solid line).

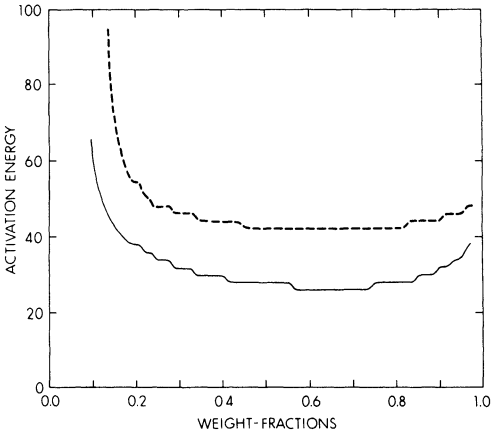


Fig. 4.10: Friedman signatures[143] derived from synthetic weight loss data calculated using the mechanism and kinetics of Bradbury et al.,[186] together with a heat transfer step describing heat transfer in the Dupont 951 TGA. with heating rates of 2°, 6° and 11°C/min (dashed line), and 20°, 55°and 116°C/min (solid line).

pyrolysis given in Fig. 4.8 cause this reviewer to conclude that it is not possible to cause cellulose to undergo pyrolysis at temperatures above about 500°C, in agreement with the speculation of Lédé.

One may still inquire if alternative pathways become active at higher temperatures, which could explain the differences displayed in Fig. 4.8. In response to this question, Broido[145] observes: "remember that you cannot get to a high temperature

regime without going through the low temperature region. *If* the low temperature model is accurate, it *must* apply to the pyrolysis process which goes through the low temperature region on its way to higher temperatures." Since the low temperature rate law of Bradbury et al.[136] projects much more rapid weight loss than the higher temperature data given in Fig. 4.8, the low temperature pathways must take precedence over any new pathways supposed to be responsible for the high temperature weight loss.* These findings cause this reviewer to conclude that accurate estimates of high temperature biomass pyrolysis reaction rates can best be obtained by extrapolating low temperature kinetic data applicable to the reaction pathway(s) of interest at the higher temperatures. Future work should employ high heating rate TGA apparati and sophisticated numerical simulation algorithms to search for new high temperature pyrolysis pathways. These algorithms should explicitly account for the effects of heat and mass transfer limitations on the observed experimental data.

A considerable controversy exists in the literature concerning the mechanism of levoglucosan formation from cellulose (reaction 2 in Fig. 4.6). Homolytic bond cleavage was first proposed by Golova and her wo-workers,[66,146-153] and supported by Madorsky,[62,72] and Arthur and Hinojosa.[84] Heterolytic bond cleavage was favored by Gardiner,[114] Bryne et al.,[115] and Shafizadeh.[121,123,154] Recent studies by Shafizadeh and his colleagues[155] indicate that free radicals play an important role in char formation, which occurs simultaneously with levoglucosan evolution. Hence, Shafizadeh's findings may reconcile the two schools of thought and provide important insight into practical ways for influencing the formation of anhydrosugars from carbohydrates.

One important experimental study whose results seem to contradict the presumption that levoglucosan is the immediate product of cellulose pyrolysis at moderate temperatures was reported by Broido et al.[95] Initial studies[100] confirmed levoglucosenone to be the major component of condensables derived from the vapor phase dehydration of D-glucose, levoglucosan, and cellulose volatiles. However, later work,[95] involving the use of $NaHSO_4$ as a catalyst of either the solid or vapor phase pyrolysis reactions, evidenced significant differences between product distributions derived from cellulose and levoglucosan. These differences seem to imply that a volatile precursor to levoglucosan is formed from cellulose, which further reacts in the vapor phase to form levoglucosan after some delay. Unfortunately, Broido's elegant work has not been pursued by others and no resolution of this question seems to be forthcoming.

A second study whose conclusions differ from the general consensus presented above was described by Ohnishi et al.[156] Their Curie-Point pyrolysis of acid pretreated Avicel cellulose revealed the abundant formation of several furan compounds which were not observed as principal products of levoglucosan pyrolysis. This finding

* One can argue that a perverse, undetected, low activation energy, sequential reaction preceding the levoglucosan forming step could become rate limiting at the higher temperatures. Other arguments based on presumed, arcane reaction pathways also exist. Without further evidence, the most convincing reconciliation of the available rate data is provided by a recognition of the effects of heat transfer on those data.

caused Ohnishi et al.[156] to propose a second major pathway for cellulose pyrolysis involving the formation of 3-deoxyglycosulose as a significant intermediate, which subsequently decomposed to form the furan species observed. However, it can be argued that trace amounts of acid remaining from the acid pretreatment of the Avicel cellulose influenced the results of Ohnishi et al.[156] by catalyzing the degradation of the intermediate levoglucosan to form the observed furan compounds.[157] Studies similar to those of Ohnishi et al.[156] but avoiding the acid pretreatment of the cellulose substrate, could offer an interesting insight into the potential role of 3-deoxyglycosulose as a major intermediate in the direct formation of furan compounds from cellulose.

4.3.3 High Temperature Phenomena

When pyrolysis temperatures exceed 500°C, vapor-phase thermolysis reactions begin to affect observed pyrolysis product distributions. These secondary reactions involve the anhydrosugar and other lower molecular weight species which were evolved as primary products of the solid-phase pyrolysis reactions. In spite of their importance as a source of olefins and other high value products, very few studies of the vapor-phase pyrolysis reactions have been reported in the literature.

The solid-phase pyrolysis of carbohydrates can be effected at temperatures approaching 500°C only under conditions of very rapid heating. Although it is sometimes claimed that pyrolysis occurred at 500°C because the sample was embedded in a furnace, flame, heated screen, or other hot environment at 500°C, the discussion accompanying Fig. 4.8 clearly demonstrates the limitations of heat transfer above about 300°C. Only in carefully engineered situations is it possible to study pyrolysis at temperatures approaching 500°C. As discussed in the following sections, limited experimental data suggests the existence of a new pyrolysis pathway at high temperatures, whose products are (unfortunately) not very interesting.

4.3.3.1 Products. Above 500°C anhydrosugars and related species undergo rapid thermolysis in the vapor phase to form a hydrocarbon rich synthesis gas containing, $CO, CO_2, H_2, H_2O, CH_4, C_2H_4, C_2H_6, C_3H_6$, and heavier hydrocarbons, as well as some condensable, refractory organic compounds.[158-162] Sophisticated techniques[163-170] involving mass spectroscopy (MS) have been used to observe many transient intermediate species at higher temperatures, which offer insight into the pathways of high temperature thermolysis (see following section).

Table 4.1 exhibits the major stable products of the very rapid solid-phase pyrolysis of a variety of biomass materials.[159] Small, finely divided samples of substrates listed in Table 4.1 were exposed to a flux density of 8000 W/cm^2 (visible light) in a vacuum of 10^{-3} torr. Interestingly, solid char was a primary product of several of the substrates, in spite of the very high heating rate.* At somewhat lower heating rates, other workers[171-176] have also observed the formation of levoglucosan and a variety of polar organic compounds from cellulose.

* The formation of char at such high fluxes may reflect the influence of photolytic effects derived from the ultraviolet light present in the Xenon bulb's spectrum.

Table 4.1

Products of the Flash Pyrolysis of Various Biomass Materials
(Mass % Yields of Original Sample)

Sample	CO_2	CO	H_2	C_2H_2	CH_4	C_4H_4	HC	GAS	LIQ	SOL*	TOTAL
Lignin	2.1	31.6	4.0	3.9	0.2	0.2	0.0	42.0	13.2	20.0S	75.2
Lignin	1.1	26.9	3.3	3.9	0.3	0.2	0.0	35.7	18.2	26.1S	80.0
Redwood	4.2	42.3	3.3	5.9	0.6	0.5	0.0	56.8	15.1	6.7SP	78.6
Redwood	2.3	26.2	1.8	4.7	0.7	0.5	0.0	36.2	16.3	31.7SP	78.6
Dextrose	3.6	24.8	1.9	4.1	0.7	0.0	0.0	35.1	54.2	10.1U	99.4
Dextrose	2.5	14.6	0.9	2.2	0.9	0.4	0.0	21.5	62.7	16.5U	100.7
D-cellobiose	4.5	23.9	1.5	3.2	1.0	1.6	0.0	35.7	49.4	8.9U	94.0
D-cellobiose	4.3	22.1	1.5	3.3	1.1	1.6	0.1	34.0	52.0	38.7U	124.7
Kraft paper	2.4	22.8	1.3	1.8	0.7	0.5	0.1	29.6	10.0	43.1PA	82.7
Kraft paper	2.1	14.9	0.7	1.3	0.6	0.5	0.1	20.1	12.3	52.0PA	84.4
Leucaena	3.2	49.3	3.3	5.5	0.7	0.5	0.1	62.6	10.1	8.2P	80.9
Corn Cob	4.0	4.3	0.3	0.7	0.2	0.0	0.0	9.5	11.2	73.4U	94.1
Corn Cob	1.5	4.3	0.4	0.7	0.2	0.3	0.3	7.7	7.6	83.0U	98.3
Calatropis	3.1	20.8	1.6	2.1	0.5	0.1	0.1	28.4	26.0	34.3P	88.8
Newsprint	1.9	20.3	1.1	2.5	0.7	0.5	0.5	27.5	17.1	42.3PA	86.9
Cow Manure	3.0	37.9	2.8	3.0	0.5	0.0	0.0	47.2	8.9	21.4P	77.5

*Note: The form of the solid left in the reactor after the flash varied with the substrate. The letters to the right of the solid mass yield indicate the type of solid found in the reactor cell after the experiment. They are as follows: P–primary char, S–secondary char, A–ash, U–unpyrolyzed substrate.

4.3.3.2 Mechanisms and Kinetics

4.3.3.2.1 Vapor-phase Phenomena. Early studies of Shafizadeh and Lai,[177] involving the vapor-phase pyrolysis of synthetic levoglucosan labeled at the C1, C2, and C6 positions, offered some insight into the pathways responsible for the formation of CO, CO_2, and several minor organic products. In general, several pathways contributed to the evolution of each species. The observed contribution of both the C2 and C6 position to CO_2 formation indicates that extraordinary rearrangements must occur during vapor-phase pyrolysis.

Results obtained in my laboratory[160,161] for the vapor-phase pyrolysis of cellulose derived volatile matter, consisting primarily of levoglucosan and its furanose isomer, evidence temperature-dependent asymptotic yields for all the major gaseous species (except perhaps CO_2). This behavior was explained by a lumped parameter model consisting of the simple competitive mechanism given in Fig. 4.11. This mechanism posits (1) the vapor phase cracking of primary volatiles to permanent gas species with activation energy $E_1 = 49$ kcal/gmol and (2) the gas phase polymerization of pyrolysis fragments to refractory condensable materials (tars and water soluble organics) with $E_2 = 15$ kcal/gmol. The competitive rate law and rate constants summarized in Table 4.2 correlated the experimental data for carbon efficiency given in Antal[161] with a correlation coefficient $R = 0.999$.

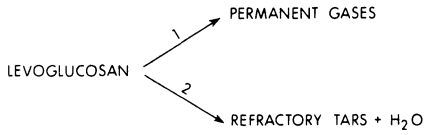


Fig. 4.11: Lumped parameter model of the vapor-phase pyrolysis of levoglucosan.[161]

Table 4.2

Kinetic Parameters for the Gas Phase Pryolysis of Various Biomass Materials

	HW	SW	Cellulose	Lignin
$E_1 - E_2$ (cal/mol)	19700	22900	34300	13000
E_1	23600	24100	48800	NA
E_2	3900	1200	14500	NA
A_1/A_2	4.39×10^4	7.50×10^5	2.07×10^8	1.05×10^2

As discussed in Antal[161], competitive reactions give rise to the following functional dependence of product yields on temperature:

$$\ln\left(\frac{\text{yield of species i}}{\text{yield of species j}}\right) \propto T^{-1} \tag{4.1}$$

Figure 4.12 displays the yield data for various gas species plotted according to Eq. (4.1). In Fig. 4.12, the yield ΔCH_4 of methane (for example) represents the asymptotic (long residence time) formation of methane by the vapor phase pyrolysis of cellulose-derived volatile matter (believed to be primarily levoglucosan). Thus, ΔCH_4 equals observed asymptotic yield of methane exiting the gas-phase reactor minus the yield of methane entering the gas-phase reactor due to the primary pyrolysis reactions. The linearity of the data displayed in Fig. 4.12 provides evidence for the role of competitive reactions in the formation of the various permanent gas species.

The variable $\Delta\eta$, indicated in Fig. 4.12, represents the asymptotic (long residence time) increase in the carbon efficiency due to the vapor phase pyrolysis reactions. Thus. $\Delta\eta$ equals carbon contained in the asymptotic yield of permanent gases exiting the gas-phase reactor divided by the carbon in the cellulose substrate minus carbon contained in the permanent gases entering the gas-phase reactor

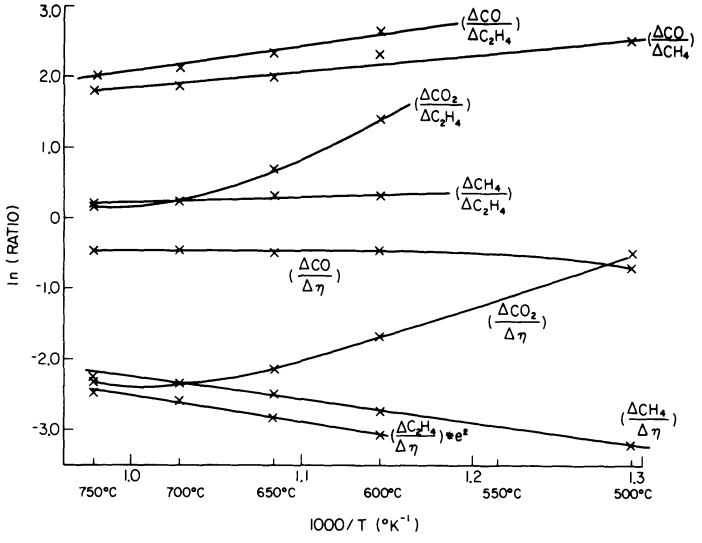


Fig. 4.12: ln of the ratios of various gas-phase yields, and the gas-phase carbon efficiency, η vs T^{-1} for cellulose-derived volatile matter.[161]

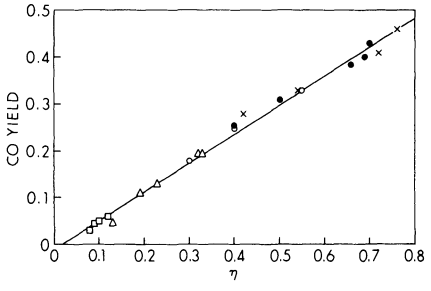


Fig. 4.13: CO yield vs η for the vapor-phase pyrolysis of cellulose derived volatile matter.

divided by the carbon in the cellulose substrate. The value of $\Delta\eta$ can be viewed to be a reaction coordinate for the gas-phase cracking reactions which lead to the formation of permanent gases. The constant value of $\ln(\Delta CO/\Delta\eta)$, evidenced in Fig. 4.12 indicates that as $\Delta\eta$ increases with increasing temperature, ΔCO increases equally rapidly. Apparently, a fraction of the carbon atoms composing the volatile matter are dedicated to CO formation by the gas-phase cracking reaction, irrespective of the prevalent gas-phase temperature. This conclusion is supported by Fig. 4.13, which displays the linear dependence of ΔCO on η for a wide range of vapor-phase temperatures and residence times.

A thoughtful examination of the trends evidenced in Fig. 4.12 suggests that for cellulose-derived volatile matter, the gas-phase cracking reaction involves competition between dehydration reactions, resulting in methane and ethylene formation, and decarboxylation reactions. The role of dehydration reactions may not be imme-

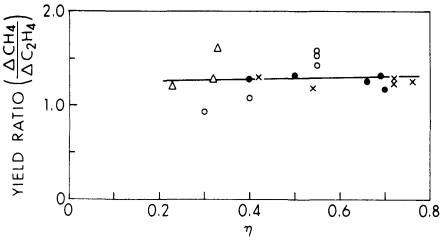


Fig. 4.14: CH_4 to C_2H_4 yield ratio vs η for the vapor-phase pyrolysis of cellulose derived volatile matter.

diately evident from Fig. 4.12, but it can be deduced from the declining yields of CO_2 and the constant relative yield of CO. Recognizing (1) an oxygen balance must be maintained, (2) the presence of oxygen in the permanent gas species declined with increasing temperature, and (3) water was the only major species undetected in the experiments, it seems necessary to conclude that higher, gas-phase temperatures favorably influence the formation of water over CO_2. This deduction is supported by the fact that for vapor-phase temperatures exceeding 700°C the carbon balance was closed by including only char and permanent gases as products; whereas only about 80% of the hydrogen and oxygen could be accounted for (see Fig. 32 in Part 1 and Table I in Antal[161]). Although other explanations are possible, it seems likely that water is a major product of the vapor-phase pyrolysis reactions at temperatures above 700°C. This conclusion is also consistent with the important role played by dehydration reactions in the high temperature vapor-phase decomposition of glycerol to acrolein.[178] Future research should give emphasis to definitive studies of the role of dehydration and decarboxylation reactions in the formation of hydrocarbons from biomass.

No clear mechanistic explanation for these results presently exists; however, the constant value of $\Delta CH_4/\Delta C_2H_4$ evidenced in Fig. 4.14 is in a molar ratio of 2:1 for the vapor-phase pyrolysis reactions. As shown in Fig. 4.12, the increased formation of CH_4 and C_2H_4 at higher temperatures comes at the expense of CO_2 formation, suggesting the existence of a second low temperature pathway involving decarboxylation chemistry. Ongoing research in my laboratory is aimed at interpreting these trends at the molecular level through a systematic study of the vapor-phase thermolysis of model compounds.[162] Results of these studies will be published in the near future.

4.3.3.2.2 Solid-Phase Phenomena. Table 4.3 summarizes the mass ratios of the gaseous products CO, CH_4, and C_2H_4 derived from the high temperature pyrolysis of various biomass materials.[159] As discussed in detail in Part 1 of this Review, the abnormally high values of the CO/CH_4 and CO/C_2H_4 mass ratios obtained by flash heating cannot be explained by some combination of the low and moderate temperature pyrolysis pathways (see Fig. 4.1) discussed earlier. Nor can they be explained by the influence of high temperature, vapor-phase chemistry. The data given in Table 4.3 appears to demand the inclusion of a third, high temperature pyrolysis pathway in any treatment of the global mechanisms of carbohydrate

pyrolysis. As mentioned by some earlier workers,[173,174,179,180] this pathway must involve the catastropic fragmentation of the glycosidic ring into CO and H_2. Presumably, this pathway is Reaction 3 in Fig. 4.1, referred to by Shafizadeh[7] as the fission reaction. With so much energy available for the cleavage of molecular bonds at these high temperatures, it is not surprising to find similar behavior among all the solid phase carbohydrate materials. Unfortunately, CO commands a low value in the marketplace; consequently, the high temperature fragmentation pathway would appear to be of little commercial interest.

4.3.4 Effects of Various Parameters

The primary emphasis of this Review is the effects of temperature on pyrolysis products, mechanisms and kinetics. Other reaction parameters also influence pyrolysis phenomena, including heating rate, pressure, particle size, ambient gas environment, and the presence of ash or additives. This section offers an overview of the effects of these parameters on the pyrolysis of carbohydrates.

The influence of heating rate on pyrolysis is best described by the lumped parameter rate laws presented earlier. In general, condensation reactions are characterized by low apparent activation energies; hence the formation of char and refractory tars is favored by low heating rates. Conversely, high heating rates favor the formation of anhydrosugars and permanent gases.

Increasing pressure also favors char formation and dramatically reduces the yields of permanent gases.[129,130] It is probable that this effect primarily results from a decrease in the rate of mass transfer of sirup precursors from the liquid into the vapor-phase. Associated with an increase in char yield is an increase in reaction exothermicity; hence the solid-phase pyrolysis reactions of most carbohydrates become exothermic at sufficiently high pressures. In addition, some of the vapor-phase reactions also become exothermic at elevated pressures. Conversely, low pressures favor the endothermic sirup forming pathways. These effects are summarized in Fig. 4.6.

Agrawal and McCluskey[181,182] have recently described important studies of the low and moderate pressure pyrolysis of newsprint. In common with the results of some earlier workers[60-62,136,183-185] they observed an initial, very rapid degradation step followed by a slower, first order reaction. Pretreatment using a 1% HCl acid wash increased the sirup yield by 40%, while decreasing the associated gas yield an equivalent amount. The acid wash also increased the extent of the initial, rapid reaction. Their findings that the initial reaction accounted for 68% of the sirups formed and that the sirup precursors formed from acid washed newsprint are more reactive than those from untreated newsprint, have important implications for the design of sirup forming pyrolysis reactors. Measured values of the apparent activation energies associated with the initial rapid (9-25 kcal/gmol) and slow (27-29 kcal/gmol) reactions seem low but may reflect the catalytic influence of ash on the reaction chemistry. The details of the initial rapid reaction, including the influences of pressure, ash and acid washing on sirup formation, merit much more attention by the biomass pyrolysis community.

Table 4.3

A Summary of Reported Carbon Monoxide, Methane, and Ethylene Mass Ratios
The original references indicated in the table can be found in Hopkins et al.[159]

Author	Material	Temperature (°C)	gas (mass %)	ln of CO/CH_4	ln of CO/C_2H_4
	Electric Furnace Heating				
Antal	Cellulose	750	84	1.9	2.2
	Kraft Lignin	750	36	0.6	2.1
	D-Mannose	750	65	1.9	2.4
Hajaligot et al.	Cellulose	750	36	2.7	2.7
	Cellulose	1000	47	2.2	2.3
Mudge et al.	Wood (with K_2CO_3)	650	—	1.5	3.9
	Wood (no catalyst)	650	—	2.2	3.7
Deglise et al.	Dry Beech	700	—	1.8	2.1
	Dry Beech	1000	—	1.9	2.6
Lede et al.	Douglas Pine Sawdust	700	52	1.8	2.1
	Douglas Pine Sawdust	1000	62	1.9	2.7
Diebold & Scahill	Softwood Sawdust		34	2.1	2.3
(cyclone)	Birch Flour		—	2.1	2.2
Halligan et al.	Bovine Waste	728	—	1.4	1.8
(fluid bed)	Bovine Waste	796	—	1.1	1.1
Iatridis & Gavalas	Kraft Lignin	400	—	0.8	4.1
(screen)	Kraft Lignin	650	—	0.6	3.3
	Radiant Heating	Flux (W/cm^2)			
Antal et al.	Cellulose	70	—	2.1	2.3
	Corn Cob	70	—	2.3	2.3
	Hardwood	70	46	1.9	2.4
Hopkins et al.	Cellulose	200	26	2.6	2.2
	Kraft paper	200	55	2.3	2.3
Lede et al.	Douglas Pine Sawdust		60	2.9	3.2
	Douglas Pine Sawdust		—	3.1	3.1
	Douglas Pine Sawdust		75	2.6	3.7
Martin	Cellulose (8 sec)	17	—	3.2	3.2
	Cellulose (4 sec)	46	—	2.6	2.6
	Cellulose (.965 sec)	46	—	4.4	3.8
	Cellulose (.49 sec)	46	—	4.6	4.2
Hopkins et al.	Lignin	8000	42	5.1	5.1
	Lignin	8000	36	4.5	4.9
	Redwood	8000	57	4.3	4.4
	Redwood	8000	36	3.6	4.0
	Dextrose	8000	35	3.6	—
	Dextrose	8000	21	2.8	3.6
	D-cellobiose	8000	36	3.2	2.7
	D-cellobiose	8000	34	3.0	2.6
	Kraft paper	8000	30	3.5	3.8
	Kraft paper	8000	20	3.2	3.6
	Leucaena	8000	63	4.3	4.6
	Corn Cob	8000	10	3.1	—
	Corn Cob	8000	8	3.1	2.7
	Calatropis Procera	8000	28	3.7	4.6
	Newsprint	8000	28	3.4	3.7
	Cow Manure	8000	47	4.3	—

Note: "—" indicates that the gas yield was not listed or if found in the third column, that no C_2H_4 was observed in the pyrolysis products.

The effect of particle size on pyrolysis is not well understood. Because large particles reduce the rate of mass transfer of reactive volatiles (sirup precursors) into the vapor phase, it seems likely that increasing particle size should favor char formation. However, as noted by Martin,[171] volatile products evolved at the center of a hot, pyrolyzing particle must pass through the hot char/ash layer as they exit the particle. Heterogeneous reactions are likely to oxidize the char layer by reducing the oxygen-rich volatile matter, thus gasifying the char.

Of the various gaseous environments used for biomass pyrolysis, steam is thought to be the most practical. The presence of steam has been variously reported to accelerate,[89,186-188] decelerate,[78,84,86,89] and not affect[189] the rate of cellulose pyrolysis. Steam is the preferred environment for the pyrolytic formation of levoglucosan from carbohydrates.[124-128] At low and moderate pressures, other gases (N_2, Ar, H_2, CH_4, etc.) are not believed to exert a major influence on the pyrolysis chemistry.

Even the presence of trace quantities of ash is known to exert an extraordinary influence on the formation of char during the pyrolysis of carbohydrates.[71,120,121] Much of the early Russian work on cellulose pyrolysis was undermined by their use of samples which were not ash free.[190] The removal of ash appears to be a prerequisite for effecting high yields of sirups from carbohydrates. An extensive literature exists on the use of various additives, known as fire retardants, to catalyze the formation of nonflammable char at the expense of flammable volatiles production.[3,8,191] Acid catalysts strongly influence the dehydration reactions and favor the formation of levoglucosenone,[95,97,98,100] furan derivatives,[192] dextrins[15] and char. Acid catalysts also increase the rates of the condensation reactions of levoglucosan,[33,101] and glucose[39] at low temperatures. The research group of Basch and Lewin has also called attention to the swelling effect of acids on cellulose at pyrolysis temperatures, which is superimposed upon other effects.[193-195] Alkaline catalysts enhance the fission and disproportionation reactions[99,106,177] as evidenced in increased yields of glyoxal, acetaldehyde and other low molecular weight carbonyl compounds, and char.[3] Trace amounts of acids[196] and calcium salts, sodium hydrogensulfate, and sulfur dioxide treatment[20] improve the yields of levoglucosan from cellulose pyrolysis. Larger quantities of acid do not serve as catalysts but instead react with the cellulose.[197]

4.3.5 Summary and Critique

The major pathways active during the pyrolysis of carbohydrate materials are depicted in Fig. 4.15. The "reaction tree" portrayed in Fig. 4.15 is somewhat speculative in that not all the branches have been observed to play an active role in the pyrolysis of all carbohydrate materials. Because the relative importance of each pathway is influenced by the pyrolysis temperature, the molecular structure of the individual carbohydrate, the presence of additives, the pressure and chemical composition of the environment, and so forth, it is not surprising that some of the branches would be "dormant" for some materials under certain sets of pyrolysis conditions. The chief strength of the reaction tree is that it offers a unified picture of the very complex phenomena reviewed in this section. One particularly

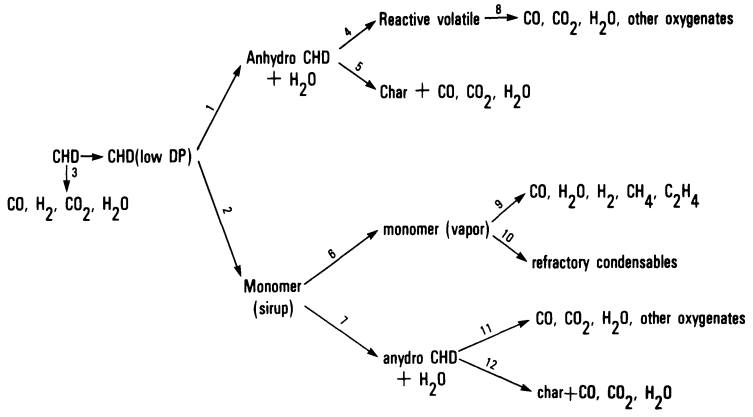


Fig. 4.15: A reaction tree describing the concurrent and consecutive pathways of cellulose pyrolysis.

satisfying aspect of the tree is the similarities which exist between its various families of branches. The following paragraphs call attention to these similarities, while summarizing what is known of the pyrolysis of carbohydrates.

To maintain continuity with the preceding nomenclature of this Review, the low, moderate, and high temperatures pathways are numbered 1, 2, and 3 (respectively) in Fig. 4.15. The un-numbered initiation reaction, which results in a reduction in DP of the polymer, is known to play a significant role in the pyrolysis of crystalline cellulose. The fact that it has not been observed to be active in the pyrolysis of other carbohydrates may be indicative of the paucity of sophisticated mechanistic/kinetic studies of the pyrolysis of these materials.

Pathway 1 in Fig. 4.15 involves intermolecular cross linking reactions: the formation of ether links across polymer chains accompanied by the evolution of water. Pathway 1 is virtually inactive in crystalline cellulose, but Pathway 1 plays a dominant role in product formation during the low temperature pyrolysis of all noncrystalline carbohydrates. Dextrin formation from starch is one example of the commercial utility of Pathway 1. Almost all additives, including acid and alkali catalysts and simple salts, increase the rate of reactions associated with Pathway 1. Subsequent degradation of the anhydrocarbohydrate (dextrin) occurs by the two competing Pathways 4 (followed by 8) and 5. Pathway 4 involves the endothermic evolution of volatile species which subsequently undergo strongly exothermic reactions in the vapor-phase to produce a variety of carbonyl compounds, CO, CO_2, and water. Low pressures and long residence times favor Pathways 4 and 8. Conversely, high pressures favor the exothermic condensation reactions comprising Pathway 5, which ultimately result in the formation of char, H_2O, CO_2 and CO from the anhydrocarbohydrate.

At somewhat higher temperatures transglycosylation reactions (Pathway 2) result in the depolymerization of carbohydrates and the formation of sugar and anhydrosugar monomers. Pathway 2 is the dominant source of products during the

pyrolysis of crystalline cellulose, but plays a less important role with amorphous carbohydrates. Dilute acid washing of the carbohydrate substrate prior to pyrolysis, which removes the catalytic influence of ash on the parasitic reaction Pathway 1, increases the yields of monomeric species. Following their formation, the monomers may either enter the vapor-phase (Pathway 6) or condense to form a dextrin (Pathway 7). Conditions which enhance the ease of mass transfer to the vapor-phase (low pressures and small particle sizes) favor Pathway 6. If the vapor-phase monomers encounter temperatures in excess of 600°C they rapidly crack to form a hydrocarbon rich synthesis gas (Pathway 9); whereas lower temperatures and extended residence times lead to the formation of refractory condensable materials (tars). Mass transfer limiting conditions, or the presence of acid or alkali catalysts, favor Pathway 7. Pathways 1 and 7 are similar in that they lead to the formation of an anhydrocarbohydrate (dextrin), and are favored by low temperatures and acid catalysts. Pathways 11 and 12 are analogous to Pathways 4 and 5 discussed earlier. Alkali catalysts favor Pathway 11, whereas acid catalysts favor Pathway 12. No evidence is available for the analogue (not shown) of Pathway 8; nevertheless the reaction tree leads one to anticipate its existence.

The catastropic fragmentation of glycosidic and other carbohydrate ring structures (Pathway 3) is only observed under conditions which effect very rapid heating (10000°C/s) of the solid-phase carbohydrate. Such extraordinary conditions are usually obtained using focused laser radiation, solar furnaces, arc image furnaces, or Xenon flash tubes. Some researchers have mistakenly attributed the evolution of CO, H_2, and CO_2 during their experiments to Pathway 3, when in fact these gases were formed by some combination of Pathways 11, 9, 8 and 5. Unless very rapid heating of the solid is realized, the role of Pathway 3 can safely be neglected.

The mechanisms of the individual reactions composing the various pathways depicted in Fig. 4.15 are not well understood. Evidence supporting the roles of both homolytic and heterolytic reactions in Pathways 1, 2 and 7 exist in the literature. Little or no information exists concerning the mechanisms of the remaining pathways.

Any kinetic study which hopes to describe the rates of carbohydrate pyrolysis must adequately account for the reaction tree given in Fig. 4.15. Only one kinetic study[136] is available in the literature which incorporates the implications of the reaction tree into its rate law. Future research efforts should emphasize measurements of the rates and heats of reaction associated with the individual branches of the reaction tree.

The reaction tree portrayed in Fig. 4.15 should not be regarded to be a final description of carbohydrate pyrolysis. In addition to the need for a better understanding of the molecular details underlying the indicated pathways, one also anticipates the existence of additional competitive reactions composing Pathways 5, 10 and 12. Competitive reactions probably underlie Pathway 9 and may play a role in Pathway 8. Further insights will be gained by detailed pyrolysis studies of the reactive intermediates indicated in Fig. 4.15, which are ongoing in my laboratory and elsewhere.

What are the implications of Fig. 4.15 for the engineering community concerned with the selective pyrolytic production of high value products from carbohydrates? Only two products can potentially be obtained with high yields: char (via Pathways 5 and 12) and monomer (via Pathway 2). Charcoal is a delightful solid fuel which can be burned directly or gasified as a surrogate, highly reactive coal. Since the pathways which form char (charcoal) are so easily catalyzed, its production from biomass may merit emphasis by those nations which lack adequate reserves of coal or peat. The alternative product, monomer, is not so easily obtained. Parasitic reactions, which are easily catalyzed, reduce the yield of monomer dramatically. Nevertheless, the monomers are intermediate to the formation of high value hydrocarbons (especially alkenes). Reactions of the form[198]

$$(C_6H_{10}O_5)_n \rightarrow n(C_6H_{10}O_5) \rightarrow n(2C_2H_4 + 2CO_2 + H_2O) \tag{4.2}$$

preserve the enthalpy of the carbohydrate substrate in the alkene product. Hence such reactions potentially enjoy a first law efficiency of 100%. The ability to selectively effect the decarboxylation and dehydration of carbohydrate derived monomer as indicated above, could have a major impact on the worldwide hydrocarbon processing industry. The magnitude of this potential impact justifies detailed, long range, fundamental studies of Pathways 2, 6 and 9 (Fig. 4.15) to enhance hydrocarbon (especially alkene) yields. More will be said on this subject later.

4.4 PYROLYSIS OF LIGNIN

Of the major components of biomass, lignin is the least well understood. In part reflecting the extraordinary complexity of the solid phase pyrolysis chemistry, the global mechanisms of lignin pyrolysis are virtually unknown. This lack of understanding is not a result of inadequate experimental work. Indeed, hundreds of papers have appeared on the subject, and several fine reviews are available.[199-206] To date, the most detailed insights into lignin pyrolysis have been obtained from model compound studies, as described in the sections on moderate and high temperature phenomena below.

Lignin is the most refractory of the major components of biomass.* Below 200°C its rate of thermal degradation is very slow. Consequently most studies of lignin pyrolysis have emphasized moderate and high temperature thermolysis. The following section reviews those few studies of low temperature degradation which offer insight into the mechanisms of lignin pyrolysis.

4.4.1 Low Temperature Phenomena

Both IR[210] and NMR[211] studies of the effects of low temperature heat ($\leq$200°C) on lignin have been reported by Hatakeyama and his colleagues. The

* Lignin is often described as the most thermally stable of the major componends of biomass.[207] However, lignin does undergo slight decomposition at very low temperatures,[208] and at least one authority[209] states that during the initial stages of decomposition lignin is less stable than cellulose, but more stable than hemicellulose.

IR studies provided evidence for the breaking of hydrogen bonds between 60° and 80°C and for the influence of aromatic skeletal vibrations on the thermal softening of several isolated lignins between 100° and 180°C.

The NMR studies,[211] and later research[212-215] using DSC, DTA, and TGA, showed the dependence of the glass transition temperatures of a variety of lignins on their molecular weights. The relatively high glass transition temperatures were attributed in part to the role of intermolecular hydrogen bonding due to the *OH* groups present in the lignin.[214]

As noted in the following section, the method of lignin isolation is known to affect its pyrolysis behavior. However, detailed MS studies of the low temperature pyrolysis of oak lignin by Domburgs et al.[216] revealed that the method of lignin isolation only influenced the distribution of products formed in the 100° to 200°C temperature range. Products formed at 300°C were identical for all types of lignin studied.

Using a sophisticated FTIR-EGA technique, Fenner and Lephardt[217,218] have reported the evolution of $H_2O, CO_2, HCHO$ and $HCOOH$ below 200°C from Kraft pine lignin heated at 6°C/min in flowing N_2. Surprisingly, the evolution of formic acid commenced at 80°C, perhaps reflecting sample spoilage or contamination.[218] Fenner and Lephardt attribute the low temperature evolution of these volatiles to single bond fragmentation in the phenyl propane side chains of the lignin polymer.

At somewhat higher temperatures (240° to 380°C) Domburgs et al.[219] have described DTA studies of the thermal decomposition of four phenylpropanoic monomers with different side chain structures. The three model compounds with side chains involving only single bonded carbon atoms underwent cleavage and dehydration; whereas 3-(p-hydroxy-m-methoxyphenyl)-2-propen-1-ol polymerized-polycondensed at 240° to 300°C and the condensation products underwent further degradation at 380°C in a manner reminiscent of noncellulosic carbohydrate pyrolysis. The influence of low temperature condensation reactions is supported by the work of Sandermann and Augustin,[208] who proposed that lignin is transformed into a thermally more stable form by condensation reactions during heating. In his classic review, Wenzl[220] notes that exposed lignin undergoes condensation after activation by low temperature thermal treatment. Unfortunately, little data is available describing the influence of temperature on the solid-phase products formed during the low temperature pyrolysis of lignin. Consequently, much less is known about the condensation reactions than was discussed earlier for the carbohydrates.

Two studies have been reported on the low temperature photodegradation of lignin. Studies of the photochemical cleavage of phenacyl aryl ether linkages led Gierer and Lin[221] to conclude that lignin is stabilized by the etherification of phenolic hydroxyls and the reduction or epoxidation of carbonyl groups. Hon's[222] ESR studies of the effect lignin has on the formation of free radicals in photoirradiated cellulose evidenced a decrease in the ESR spectrum intensity with increasing lignin content in pulp, suggesting that lignin may act as a free radical trap.

4.4.2 Moderate Temperature Phenomena

The variety of gases, condensable materials, and char formed during the moderate temperature pyrolysis of lignin is reminiscent of the behavior of carbohydrates. However, one very important difference exists. No set of conditions has been discovered which leads to the depolymerization of lignin and the formation of a high yield of "monomer." Thus, no analogue for levoglucosan has been reported in the literature. The apparent absence of a pathway leading to monomer formation is a major impediment to the development of pyrolytic processes for lignin utilization.

Much more than cellulose, the products of lignin pyrolysis reflect the source of the lignin. Thus, the products of bamboo, beech and spruce milled wood lignin differ considerably from one another.[223] Moreover, as noted earlier, the method of isolation has a considerable influence on the pyrolytic behavior of lignin. Thus sulfuric acid lignin is more thermostable than several other lignin preparations.[224] Domburgs et al.[224] have shown that DTA curves of softwood and hardwood lignins isolated by the same method were more similar than DTA curves of different lignin preparations from the same wood species. In light of the research of Domburgs et al.[216], differences in the behavior of various lignin preparations presumably reflect products evolved below 300°C. These caveats should be borne in mind throughout the following discussion and point to the depth of complexity underlying pyrolytic studies of lignin.

4.4.2.1 Products. At moderate temperatures lignin undergoes pyrolysis to form a char, an aqueous distillate, a viscous tar, and a mixture of permanent gases.[225] As noted earlier, this wide variety of low value products is reminiscent of the pyrolytic behavior of carbohydrates; however, the chemical composition of the condensable products differs considerably between the two materials. Little emphasis is given in this Review to a detailed statement of the variety of products resulting from lignin thermolysis. This reflects the fact that historically few researchers have carefully characterized both the solid- and vapor-phase conditions prevalent in their studies; consequently compendiums of product distributions offer little insight into the mechanistic details of lignin pyrolysis. The interested reader is referred to Allan and Matilla[202] and Klein and Virk[226] for detailed catalogs of lignin pyrolysis products.

At low heating rates the char yield from lignin exceeds 50% by weight, with elemental formula[227] $(C_{40}H_{38}O_{11})_2$. The research of Elder[228] indicates the char to be insoluble in organic solvents, acids and alkalis, and not susceptible to hydrocracking or hydrotreating, even in the presence of catalysts.

Water, methanol, acetic acid, acetone, and acetaldehyde are major components of the aqueous distillate,[229] which can represent as much as 25% by weight of lignin substrate. The yields of methanol are influenced by the source of lignin—hardwood lignin providing higher yields than softwoods.[202]

Unlike the pyrolytic sirups derived from carbohydrates, the tars derived from lignin contain no single predominant compound. Instead, the tars are composed of a wide variety of compounds usually related to phenol, 2,6-dimethoxyphenol and

guaiacol,[202] all in low concentrations. Tar yields of about 15% to 20% are typically obtained, but values as high as 51% have been reported from the rapid pyrolysis of alkali lignin in vacuum at 600°C.[230]

Gaseous species include $CO_2, CO, CH_4, H_2, C_2H_6, C_4H_8$, and other hydrocarbons.[229] The gas yields are strongly influenced by the gas phase temperature and residence time, as discussed later in this Review.

4.4.2.2 Mechanisms and Kinetics. Domburgs and co-workers[219,224,230-245] have reported extensive studies of the pyrolysis of lignin and various model compounds. These studies have shown that lignins isolated from softwoods are more stable than those from hardwoods, reflecting the higher thermal stability of the more predominant $C_{aryl} - C_{alkyl}$ bonds in the softwood lignins.[224] In addition, the high stability of sulfuric acid lignin was shown to result from intra- and inter-molecular condensation bonds formed during the isolation process. Further evidence for the influence of the method of lignin isolation on its pyrolysis behavior has been presented by Arima.[246]

The likely role of homolytic reactions in lignin pyrolysis was emphasized by Domburgs in his research. ESR studies of the thermal decomposition of alkaline aspen lignin evidenced the influence of both pyrolysis temperature and heating rate on the ESR spectrum. This led Domburgs et al.[232] to conclude that the pyrolysis reactions involve several different free radical species, whose relative abundance depends upon the prevalent thermal conditions.

Because of the complexity of lignin pyrolysis, Domburgs directed much of his work toward the study of model compounds. Other workers have more recently followed his lead. The next section summarizes the results of lignin specific* model compound studies as a prelude to the following section on whole lignin pyrolysis.

4.4.2.2.1 Model Compound Studies. Domburgs' work with model compounds emphasized the techniques and instrumentation of thermal analysis. His DTA studies of the decomposition of dehydrodivanillin[233] (a model of the biphenyl-type structures in lignin) revealed the presence of vanillin as a volatile product following the main endothermic reaction between 300° and 325°C. The high rate constant associated with the reaction, and the vanillin product, suggested the cleavage of the biphenyl bond by a free radical mechanism.[233] Thermal analysis studies of the thermal degradation of dehydrodiisoeugenol at 370° to 400°C at rates of 6° to 12°C/min revealed the formation of phenol, o-cresol, p-cresol, guaiacol, pyrocatechol, eugenol and several other related compounds.[234] Domburgs et al.[234] concluded that the decomposition of dehydrodiisoeugenol commenced with the homolytic formation of free benzoyl and phenoxy radicals, and was attended by the cleavage of C–C bonds. Further studies revealed the homolytic thermal degradation of the model compounds benzyl 4-acetyl-3-methoxyphenyl ether and pinoresinol.[235]

* Due to the chemical similarities of lignin and coal, the literature describing thermolysis of model compounds related to coals is pertinent to lignin pyrolysis. Unfortunately, the coal pyrolysis literature necessarily lies beyond the scope of this Review. The interested reader is referred to Klein and Virk[226] for a partial review of coal related model compound thermolysis studies.

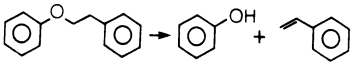
The former began to degrade at 260°C, and evidenced an apparent activation energy of 16 kcal/gmol between 300° and 350°C. The latter was more stable with an activation energy of 35 kcal/gmol between 290° and 350°C. The model compounds ethyldiguaicylmethane and diisoeugenol also underwent homolytic cleavage and gave all of the compounds formed from lignin pyrolysis.[236] The ethyldiguaicylmethane decomposed between 350° and 390°C with an activation energy of 34 kcal/gmol; whereas the diisoeugenol decomposed at 370° to 390°C with an activation energy of 67 kcal/gmol. Related studies by Ceylan and Brandenberg[247] of the thermolysis of guaiacol revealed homolytic bond cleavage satisfying a first order rate law.

According to the research of Domburgs, heterolytic bond cleavage plays a lesser role in lignin thermolysis, often involving formation of epoxide and guaiacyl derivatives from the cleavage of ether aryl-allyl bonds.[237] DTA studies of the two β-guaiacylethers, 1-(3-methoxy-4-hydroxyphenyl)-2-(2-methoxyphenoxy)-1-propanol and 1-(3,4-dimethoxyphenyl)-2-(2-methoxyphenoxy)-1-propanol, both evidenced the heterolytic-hydrolytic breaking off of guaiacol below 200°C.[238] However, a third β -guaiacyl ether, 1-(3-methoxy-4-hydroxyphenyl)-2-(2-methoxyphenoxy)-1-propanol, evidenced an activation energy of 82 kcal/gmol above 270°C, indicating homolytic cleavage of the ether.[238]

Perhaps the most thorough study of model pathways in lignin thermolysis has been recently described by Klein and Virk.[226,248] Their analysis of the molecular structure of lignin (as detailed by Freudenberg[249]) led to the identification of 20 model compounds which mimic many of the reactive functional groups of whole lignin. Phenethyl phenyl ether (PPE), the model compound given greatest emphasis in their studies, is representative of the guaiacyl-glycerol-β -ether linkages which account for almost half of the total number of interunit links in the Freudenberg model.[249] Their experimental work included gas-phase thermolysis studies of both neat PPE and PPE dissolved in tetralin. They spanned a temperature range of 300° to 550°C, concentrations of 0.083 to 1.66 mol/L, and retention times of 1 to 240 min. The two batch reactors used in their studies were fabricated using stainless steel Swagelok fittings and accommodated 0.2 and 0.6 cm^3 of pyrolysis substrate. The reactors were immersed in a constant temperature fluidized sand bath for the duration of an experiment, and quenched in an ice-water bath.

The experimental results led Klein and Virk[248] to conclude that the primary reaction of neat PPE is first order in PPE and results in the formation of one mole each of phenol and styrene, as shown below.

$$A = b + c \tag{4.3}$$



The reaction obeyed an Arrhenius type rate law with parameters $\log A(s^{-1}) = 11.1 \pm 0.9$ and E (kcal/gmol) $= 45.0 \pm 2.7$. The pyrolysis of PPE in tetralin followed the same reaction path with essentially identical kinetics. The styrene product of the primary pathway underwent secondary degradation reactions and formed various hydrocarbons including ethylbenzene and toluene.

The careful elucidation of the primary pyrolysis mechanism of PPE by Klein and Virk is a singular contribution to the field. Using a presumed free radical mechanism which accounted for both the observed reaction products and order, Klein and Virk calculated a rate constant for the overall reaction based on a composite of the elementary rate constants estimated by the methods of thermochemical kinetics. Because the estimated values $\log A = 14 \pm 1$ and $E = 55 \pm 5$ did not agree with the experimentally determined values, Klein and Virk were led to consider both 4- and 6-center concerted mechanisms for PPE pyrolysis. Agreement of estimates of both $\log A$ and E for the 6-center concerted retro-ene mechanism with the experimental values led Klein and Virk to conclude that the concerted retro-ene reaction was the most reasonable mechanism for PPE pyrolysis, with the radical chain mechanism accounting for most of the minor products. They also showed that the product spectra of the lignin related β -ethers studied by Domburgs[243] and Savinykh[250,251] could be understood in terms of the PPE results.

In spite of the extraordinary care evidenced in the work of Klein and Virk, several caveats merit some consideration. Their method of estimating the error bound on $\log A$ and E involved a least squares fit to experimental data on an Arrhenius diagram, with a 95% confidence envelope calculation. A more accurate evaluation of the error bound, accounting for the influence of measurement errors at each step of the experimental procedure as illustrated by Cutler et al.,[252] would probably result in a larger error bound. It is also possible that their results were influenced by heat transfer effects, especially at the higher temperatures, since batch reactors are notorious in this regard. In this reviewer's experience the poor heat transfer characteristics of most TGAs often result in underestimates of E exceeding 10 kcal/gmol. A similar underestimate could bring the experimental values of the Arrhenius parameters of Klein and Virk into agreement with the estimated values for the free radical mechanism. Further experimental work, involving ESR studies such as those of Hon,[222] Domburgs et al.[232] and (more recently) Kim et al.,[253] and the introduction of free radical sensitizers into the reaction media,[254] could provide conclusive evidence for or against the role of homolysis in the thermolysis chemistry of PPE.

In addition to PPE, Klein and Virk have studied the mechanisms and rate laws associated with the thermolysis of 19 other model compounds (see Table 4.4).[226] The details of their results are beyond the scope of this Review; however, some general comments can be made. Of all the model compounds studied, only saligenol evolved water. Dehydration occurred at very low temperatures (175° to 225°C) and was accompanied by the formation of a polymer. The posited intermediate precursor to the polymer, quinonemethide, resembles levoglucosenone (and acrolein), which

quinonemethide

are likely precursors to the refractory tar products of cellulose pyrolysis. This low temperature dehydration reaction must be one element of the low temperature condensation pathway discussed earlier.

Most of the model compounds underwent thermolysis between 200° and 500°C with the formation of CO, H_2, CO_2, CH_4 and lower molecular weight monomers. A few [acetophenone, cinnamaldehyde, cinnamyl alcohol, and orthohydroxydiphenylmethane (OHD)] formed dimers and polymers, apparently due to the presence of a double bond in a product species.

Phenyl ether, biphenyl and biphenol were all stable at temperatures up to about 500°C. Both the phenyl ether and the biphenyl decomposed to form polymers at these high temperatures. Related compounds must play an important role in the high temperature vapor-phase pyrolysis pathways discussed later in this Review.

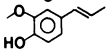
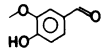
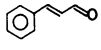
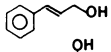
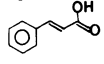
In addition to the foregoing studies, the pyrolysis of various model compounds has also been examined to gain insight into the cleavage of lignin-carbohydrate bonds. As summarized in Part 1 of this Review,[2] Shafizadeh and his coworkers have described extensive studies of the pyrolysis of several phenyl glucosides and related model compounds.[255-258] Subsequent research on the thermolysis of several 3-0-benzyl ethers of D-glucose and 4-0-benzyl ethers of methyl α-D-glucopyranoside has been reported by Kosikova et al.[259-261] The results of these studies will be reviewed in the following section on ligno-cellulose pyrolysis.

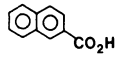
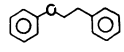
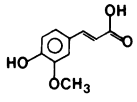
4.4.2.2.2 Whole Lignin Studies. The model compound research of Klein and Virk[226,248,262] was integrated by those authors into a mathematical model for whole lignin pyrolysis based on the product distributions and rate data obtained from their studies of 20 model compounds. The formation of CO, CH_4, H_2O, CH_3OH, and 50 phenolic compounds from a prototypical spruce lignin undergoing pyrolysis at 300°, 400°, 500° and 600°C was followed by numerical integration of the set of rate equations governing pyrolysis of the model compounds. Although the agreement of their simulations with the experimental results of other workers was claimed to be good, in this reviewer's experience the agreement was not good in comparison to simulations based on reasonably sophisticated multi-step lumped parameter models, such as those described by Bradbury et al.,[136] Urban and Antal,[263] and Antal.[161] This comment is not intended as a criticism, but simply indicates the need for a more complete elucidation of the molecular interactions during pyrolysis, especially concerning condensation and high temperature vapor-phase reactions, which remain poorly understood.

Solomon, Coughlin and their colleagues[264,265] have recently described the development of a mathematical model which predicts the molecular weight distribution of the chemical species present in lignin tars. The depolymerization model can represent a polymer composed of seven different and randomly distributed monomers, and accounts for random cleavage of the weak bonds followed by transport of the reactive fragments by diffusion and vaporization. Using a single rate constant representing the cleavage of the α-0-4 bond in lignin (which the authors assert to be the

Table 4.4

Lignin Model Compounds[226]

COMPOUND	STRUCTURE
GUAIACOL	
ANISOLE	
VERATROLE	
ISOEUGENOL	
2,6-DIMETHOXY-PHENOL	
VANILLIN	
BENZALDEHYDE	
ACETOPHENONE	
CINNAMALDEHYDE	
CINNAMYL ALCOHOL	
CINNAMIC ACID	
SALIGENOL	
PHENYL ETHER	
O-HYDROXY-DIPHENYL METHANE	

COMPOUND	STRUCTURE
BIPHENYL	
BIPHENOL	
1-NAPHTHOIC ACID	
2-NAPHTHOIC ACID	
PHENETHYL PHENYL ETHER	
FERULIC ACID	

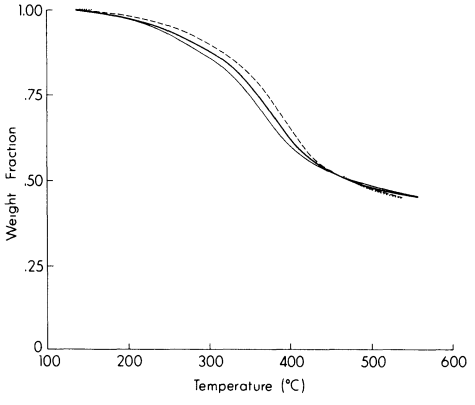


Fig. 4.16: TG Curves of Indulin AT Kraft Lignin pyrolysis at heating rates of 5.4° (solid line), 10.6° (dark solid line), 21.6° (dashed line) and 70°C/min (dotted line).

most labile[264]), the theory predicts molecular weight distributions at low pressures and isothermal conditions which enjoy remarkably good agreement with experiment.

A model which simulates the evolution of gas species by the removal of functional groups, assuming independent parallel reactions in competition with tar formation, has also been described by Avni et al.[265] During thermal decomposition, the rate of formation of various permanent gases is tied to the functional group structure of the lignin substrate by an Arrhenius rate law. The parameters used in the rate law were determined by a comparison of the simulated results with experimental measurements obtained using an FTIR/heated grid pyrolysis apparatus. In general, the agreement of the model with experimental measurements of species yields at a single heating rate of 30°C/min, and *CO* yields at isothermal temperatures of 400°, 500°, 700° and 800°C, was quite good. If the future work of Solomon, Coughlin and their colleagues proves the model's ability to quantitatively predict species yields at a variety of heating rates, their work will be a singular contribution to the field.

The role of a competitive mechanism in the thermolysis of whole lignin, alluded to in the work of Avni et al.,[265] receives further support from a careful study of TGA data. Figures 4.16 through 4.18, which display TG and DTG curves, as well as a Friedman analysis of the foregoing data, offer some insight into the global mechanisms of lignin pyrolysis. Each of these figures present evidence for the role of a competitive mechanisms in the thermal degradation of lignin. This mechanism consists of a low activation energy pathway involving condensation reactions and a high activation energy step producing lignin monomers (see Fig. 4.19).

In Fig. 4.16, the greater weight loss evidenced at higher heating rates, reminiscent of similar phenomena associated with the pyrolysis of carbohydrates, suggests the presence of a competitive mechanism. This trend is substantiated by the results of high temperature pyrolysis experiments, discussed below, and the work of Domburgs.[230] The suggestion of a competitive mechanism is underlined by Fig. 4.17,

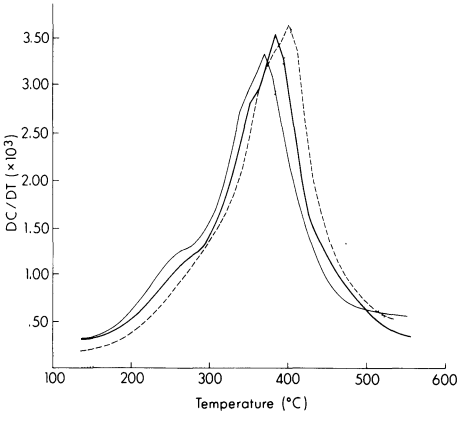


Fig. 4.17: DTG Curves of Indulin AT Kraft Lignin pyrolysis. Rate of conversion DC/DT ($°C^{-1}$) vs Temperature (°C) at heating rates of 5.4° (solid line), 10.6° (dark solid line), 21.6° (dashed line) and 70°C/min (dotted line).

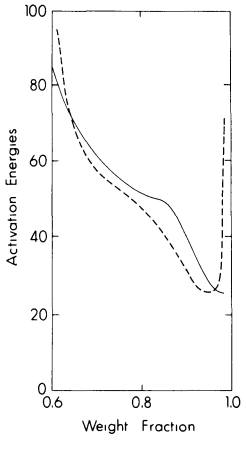


Fig. 4.18: Friedman signatures[143] of Indulin AT Kraft Lignin pyrolysis in flowing N_2 with heating rates of 5.4°, 10.6° and 21.6°C/min (dashed line), and 10.6°, 21.6° and 70°C/min (solid line).

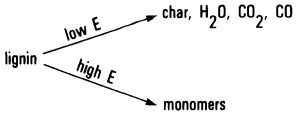


Fig. 4.19: Lumped parameter model for lignin pyrolysis.

which displays a rising DTG peak with increasing heating rate. This behavior is a clear signature of a competitive mechanism.[266] Shoulders on the lower temperature side of the DTG curves suggest additional reactions which cannot be defined without further experimentation. Finally, the rapidly rising apparent activation energy curve given in the Friedman analysis (Fig. 4.18) is yet another clear signature of a competitive mechanism.[143] The U-shaped curve evidenced at the lower heating rates is very similar to that displayed by cellulose, and may indicate the presence of an initiation reaction prior to the competitive reactions.[143] The shift of the signature to higher values of E at higher heating rates is a consequence of the poor heat transfer characteristics and temperature insensitivity of the commerical TGA instrument employed in the experiments coupled with the exothermic lignin pyrolysis reactions.

The foregoing discussion does not rule out the possible need for additional reaction pathways in a lumped parameter model of the solid-phase pyrolysis of lignin. However, evidence presented in this section does indicate the necessity of including two competitive pathways in any posited lumped parameter model of lignin pyrolysis. In addition to these two major pathways, both the DTG and the Friedman curves indicate the presence of other reactions, which are hidden by the two major competitive weight loss steps.

The results of Fenner and Lephardt[218] offer additional support for the role of a relatively simple global mechanism in the solid-phase pyrolysis of lignin. Their measured evolution profiles versus temperature for CH_4, CH_3OH, guaiacol, and a 2-methoxy-4-alkyl-substituted phenol all exhibited similar, simple symmetric peaks at about 400°C. The near coincidence of the evolution profiles of CH_3OH and the 2-methoxy-4-alkyl-substituted phenol suggests their formation by a common mechanism. The findings of Fenner and Lephardt suggest that the evolution profiles of all the major volatile components could be explained by a mechanism with three to five degradation pathways. The more recent work of Avni et al.,[265] which also evidences relatively simple, often coincidental species evolution rate profiles, substantiates the research of Fenner and Lephardt and further encourages work on a lumped parameter model.

Treating lignin pyrolysis as a single reaction, apparent activation energies and other kinetic parameters have been reported by some researchers.[224,239] Others, noting a shift in mechanism, have calculated sets of rate parameters within specified temperature ranges[224,243] where a single pathway was presumed to predominate. As reviewed earlier, kinetic parameters associated with the pyrolysis of model compounds are also available. In general, activation energies reported for whole lignin pyrolysis range from 6 to 55 kcal/gmol; however, most values lie below 30 kcal/gmol. Recognizing the presence of at least two competitive pathways in the mechanism of lignin pyrolysis, attempts to describe its thermal degradation in terms of a single rate law must necessarily be futile.

The results of Iatrides and Gavalas[267] offer an interesting insight into the rates of the two major competing weight loss steps. If the lumped parameter model in Fig. 4.19 is presumed to govern lignin pyrolysis in the temperature range 400° to

650°C, then a graph of ln (tar yield/char yield) vs T^{-1} should result in a straight line whose slope has the value $(E_1 - E_2)/R$, where E_i is the apparent activation energy associated with reaction i of the global mechanism. When plotted in this fashion, the results of Iatrides and Gavalas[267] do fall on a straight line with correlation coefficient R = 0.998 and value $E_1 - E_2 = 7$ kcal/gmol. Unfortunately, Iatrides and Gavalas[267] do not present sufficient data at short residence times to permit the calculation of the individual values of E_1 and E_2. Nevertheless, the ability of the posited competitive mechanism to explain the temperature dependent char and tar yields of Iatrides and Gavalas[267] offers further evidence supporting the role of competitive reactions in lignin pyrolysis.

4.4.3 High Temperature Phenomena

As with the pyrolysis of carbohydrates, two major questions must be addressed concerning the high temperature thermal degradation of lignin:

(1) Are additional solid-phase pyrolysis pathways open at high temperatures which are not observed at lower temperatures?

(2) What role do secondary vapor-phase pyrolysis reactions play in the formation of stable products at high temperatures?

The following sections offer some insight into the answers to these questions.

4.4.3.1 Products. Increasing pyrolysis temperatures decrease the observed yields of char, while enhancing the formation of condensable products and/or gases, depending upon reactor conditions. Furthermore, high temperatures favor the formation of unsaturated hydrocarbons. In 1978 Goheen and Henderson[268] reported the formation of up to 23% by weight C_2H_2 from kraft lignin at 2000° to 2500°C. Using microwave radiation to rapidly heat lignin pellets, Chan and Krieger[269] observed the following yields: 33.5% char, 19.6% volatiles, and 38.3% permanent gases, of which 4.9% was C_2H_2.

Iatridis and Gavalas[267] have described a careful study of the effects of temperature and reaction time on the solid-phase pyrolysis products of kraft lignin using a "captive sample" electrically heated wire mesh reactor. Yields of char, tar, permanent gases, and all major volatile species evidence temperature dependent asymptotic yields (see Figs. 4.20–4.22), as was observed earlier by Antal for the gas-phase pyrolysis of cellulose volatile matter.[161]

A complementary study from this laboratory[161] details the influence of temperature and residence time on the products of the vapor-phase pyrolysis of kraft lignin-derived volatile matter. Again, temperature dependent asymptotic yields were observed (see Fig. 4.23–4.26), with a low reduction in tar yields and an increase in gas yields with increasing temperature.

Table 4.1 details some of the results reported by Hopkins et al.,[159] who used a Xenon flash tube to vaporize kraft lignin and other biomass materials in vacuum. Heating rates were estimated to be in excess of 400,000°C/s, and no primary char was detected in the reactor after a single flash. However, a vapor-phase condensation reaction formed a delicate soot which filled the volume of the reactor.

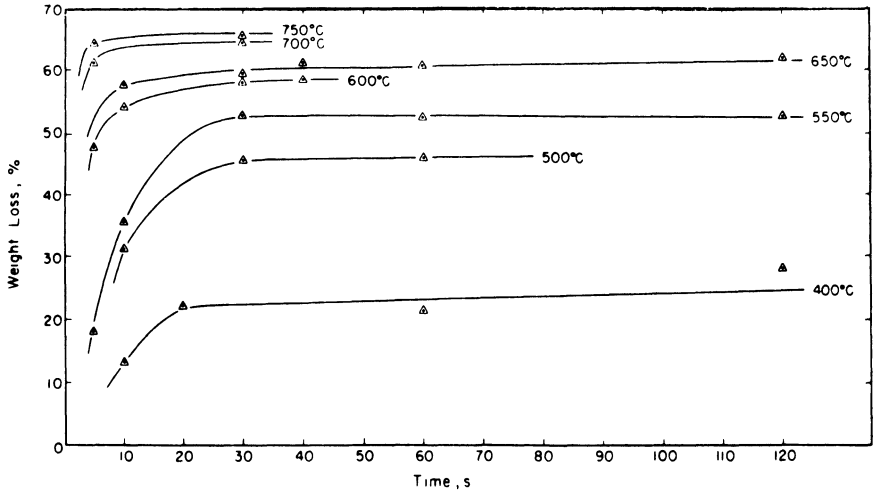


Fig. 4.20: Kraft lignin pyrolysis weight loss vs. pyrolysis time.[267]

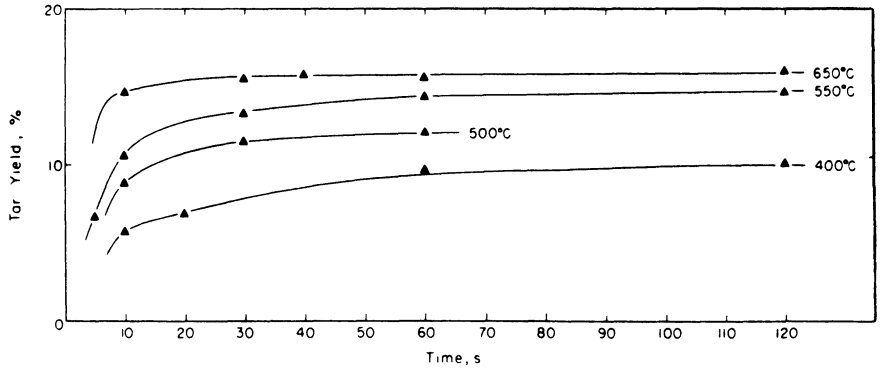


Fig. 4.21: Yield of tar from Kraft lignin vs. pyrolysis time.[267]

4.4.3.2 Mechanisms and Kinetics. As indicated in Table 4.3, the ratios of gaseous product yields observed by Iatrides and Gavalas[267] enjoy good agreement with values obtained by other workers at lower temperatures and heating rates. This good agreement of moderate and high temperature data provides no evidence for the existence of additional high temperature, solid-phase pyrolysis pathways active during the high temperature thermolysis of lignin.

The findings of Hopkins et al.[159] given in Tables 4.1 and 4.3 can be summarized as follows: (1) a high temperature lignin pyrolysis pathway exists which forms no primary char and (2) the high temperature CO/CH_4 and CO/C_2H_4 yield ratios are extraordinarily large. The extraordinary values of the yield ratios probably do not result from the moderate temperature pathways, since the data of Iatridis and Gavalas[267] indicate that increasing temperature above 600°C decreases (instead

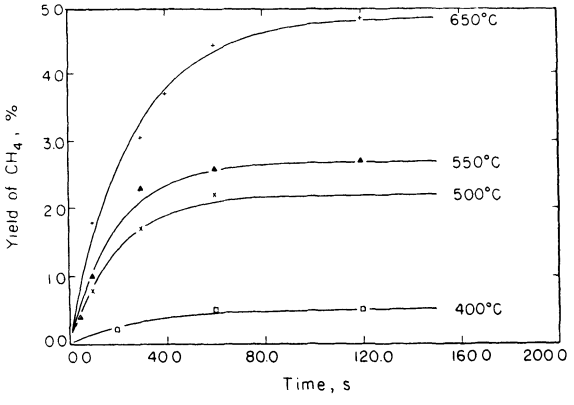


Fig. 4.22: Yield of methane from Kraft lignin vs. pyrolysis time.[267]

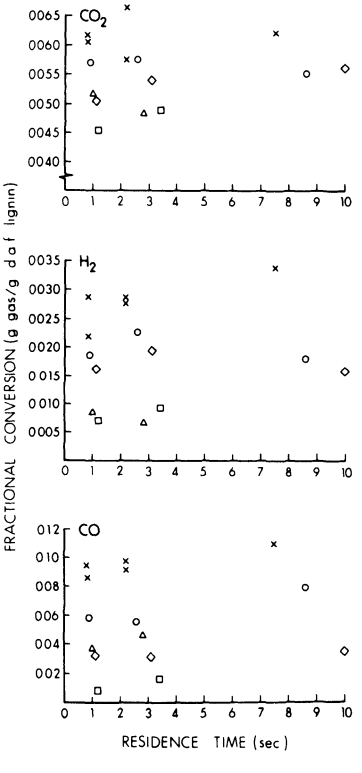


Fig. 4.23: Fractional conversion of Kraft lignin to hydrogen and carbon oxides[161](°C): (◇) 550; (△) 600; (○) 650; (×) 750.

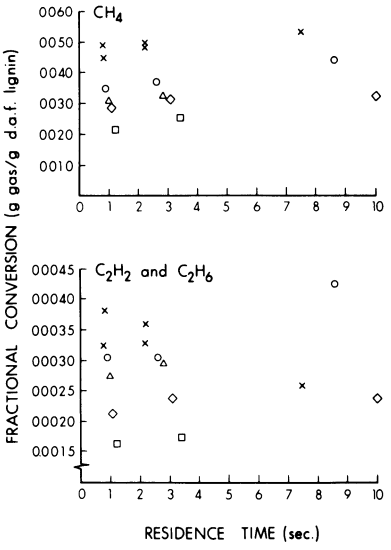


Fig. 4.24: Fractional conversion of Kraft lignin to alkanes and alkynes[161](°C): (□) 500; (◇) 550; (△) 600; (○) 650; (×) 750.[161]

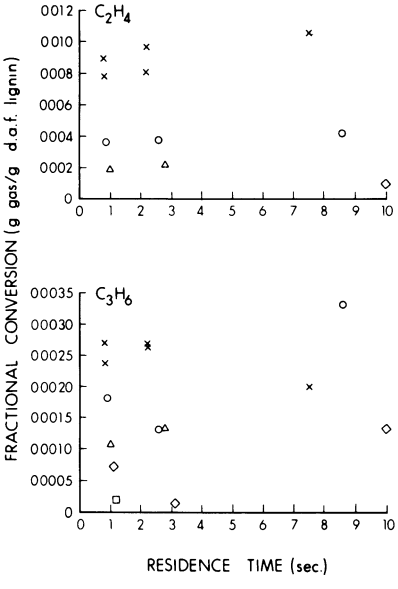


Fig. 4.25: Fractional conversion of Kraft lignin to alkenes[161] (°C): (□) 500; (◇) 550; (△) 600; (○) 650; (×) 750.

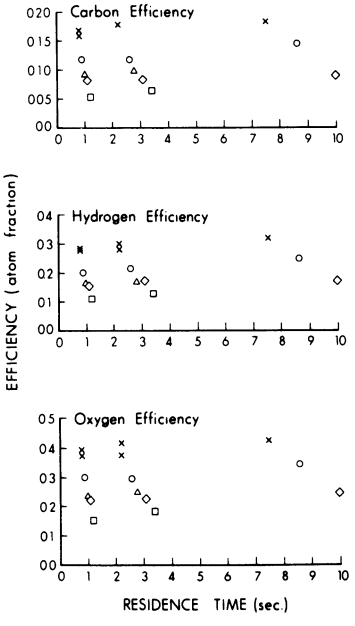


Fig. 4.26: Carbon, hydrogen, and oxygen efficiencies for Kraft lignin[161](°C): (□) 500; (◇) 550; (△) 600; (○) 650; (×) 750.

of increases) the values of the ratios CO/CH_4 and CO/C_2H_4. These observations strongly suggest the existence of a new high temperature pathway for the solid-phase pyrolysis of lignin which forms no char and few hydrocarbons. This reaction must compete with the lower temperature pathways discussed earlier; however, its rate must be very low at moderate temperatures.

The research of my laboratory[161] offers some insight into the global mechanisms of the gas-phase pyrolysis phenomena of lignin-derived volatile matter. The temperature-dependent asymptotic yields of permanent gases and condensables (see Figs. 4.23–4.27) were well explained by a posited competitive, vapor-phase, lumped parameter rate law involving a higher activation energy cracking reaction and a lower activation energy condensation (polymerization) reaction. The former reaction was responsible for the formation of permanent gases; whereas the latter reaction evolved refractory condensables. The activation energies of the two reactions were shown to differ by 13 kcal/gmol (correlation coefficient $R = 0.993$), but insufficient data were available to evaluate the individual values of E_1 and E_2. As discussed earlier, competitive reactions give rise to the following functional dependence of product yields on temperatures:

$$\ln\left(\frac{\text{yield of species i}}{\text{yield of species j}}\right) \propto T^{-1} \tag{4.4}$$

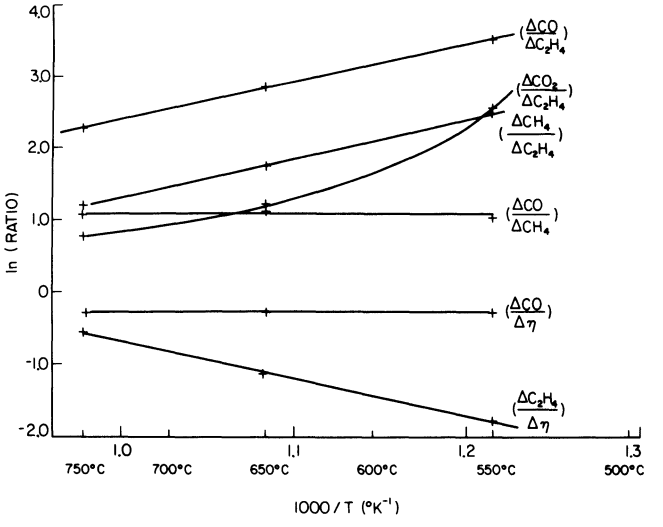


Fig. 4.27: Plots of ln of the ratios of various gas-phase yields and gas-phase carbon efficiency vs. T^{-1} for Kraft lignin.[161]

Figure 4.27 displays yield data for various gas species plotted according to the relationship above. The linearity of the data displayed in Fig. 4.27 offers evidence for the role of competitive reactions in the formation of the major permanent gas species. Thus the details of the vapor-phase cracking reaction appear to include competitive pathways which favor the formation of C_2H_4 at higher temperatures, and CO_2 at lower temperatures.

The constant value of $\ln(\Delta CO/\Delta\eta)$ and $\ln(\Delta CH_4/\Delta\eta)$, evidenced in Fig. 4.27 indicates that as $\Delta\eta$ increased with increasing temperature, the values ΔCO and ΔCH_4 increased equally rapidly. Apparently, a fraction of the carbon atoms composing the lignin-derived volatile matter are dedicated to CO and CH_4 formation by the gas-phase cracking reaction, irrespective of the prevalent gas-phase temperature. This observation parallels a similar phenomenon for cellulose pyrolysis, where a fraction of the carbon atoms composing the reactive primary pyrolysis products (believed to consist primarily of levoglucosan) were dedicated to CO formation by the vapor-phase cracking reaction. These trends are also clearly revealed in Figs. 4.28 and 4.29, which display the linear dependence of ΔCO and ΔCH_4 on η for a wide range of vapor-phase temperatures and residence times.

No clear mechanistic explanation presently exists for the ability of the lumped parameter model to explain the experimental results described here. At the molecular level, certain rate-limiting steps must predominate in such a way as to simplify the apparent behavior of a complex reaction sequence. In view of the results of Klein and Virk, it seems likely that these steps may be related to the pyrolysis of phenyl ether, biphenyl, and biphenol, as discussed earlier in this Review. The unexpected similarities of the vapor-phase pyrolysis of cellulose (see Fig. 4.12) and lignin-derived volatile matter are worthy of note. These similarities may reflect the role of hydroxyl functionalities attached to the more stable rings making up the

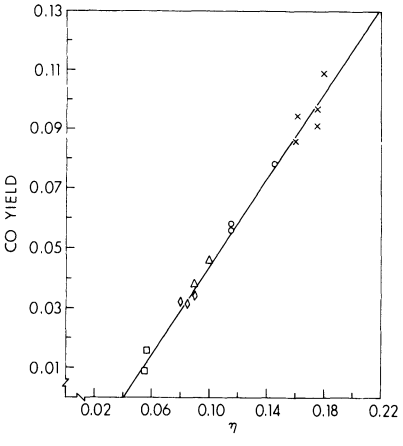


Fig. 4.28: *CO* yield vs. η for the vapor-phase pyrolysis of Kraft lignin derived volatile matter.

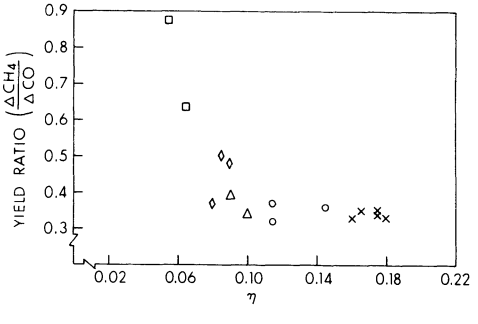


Fig. 4.29: CH_4 to *CO* yield ratio vs. η for the vapor-phase pyrolysis of Kraft lignin derived volatile matter.

primary cellulose and lignin fragments, or they may bespeak the influence of the steam pyrolysis medium used in all the experiments. Further research is needed to explain the fascinating similarities manifest by the two rather different materials.

4.4.4 Effects of Various Parameters

The effects of heating rate, pressure, particle size, additives, and pyrolysis medium on the pyrolysis of lignin are less well established than for the carbohydrates. The following paragraphs review what is known of these effects.

4.4.4.1 Heating Rate. Increasing the heating rate (or pyrolysis temperature) enhances the formation of tar and permanent gases.[230,244] At sufficiently high heating rates, no primary char is formed.[159] The distribution of volatile components also depends heavily upon external pressure, particle size, and the presence of additives.

4.4.4.2 Pressure. The vaccum pyrolysis of lignin enhances the formation of tars and volatiles at the expense of char formation.[230,242] Domburgs et al.[230] obtained a yield of 51% tar, 23% permanent gases and 8% charcoal from the vacuum pyrolysis of alkali aspenwood lignin at 600°C. The 6.8% yield of phenolics was twice the amount obtained at atmospheric pressure. Domburgs et al.[243] also note that vacuum pyrolysis results in the formation of fewer high molecular weight phenolic compounds. Goheen and Henderson[268] report the favorable influence of increased dilution (reduced partial pressure) on the yields of C_2H_2 obtained from the flash pyrolysis of lignin.

4.4.4.3 Particle Size. No studies of the effect of particle size on lignin pyrolysis are known to the reviewer.

4.4.4.4 Additives. As reviewed by Allen and Matilla,[202] acidic or basic salts, or oxides of elemental metals (Ag, Ni, Zn, Pd) exert little influence on the pyrolysis reactions.

4.4.4.5 Pyrolysis Medium. At low temperatures the complex reactions of kraft lignin with fatty acids have been studied by Traitler and Kratzl.[270,271] Much commercial research on the pyrolysis of lignin in superheated steam has been reported in the Russian literature.[244] Domburgs[242] states that normally reactive hydrogen gas does not react with lignin pyrolysis products at atmospheric pressure but simply acts as a diluent. Klein and Virk[226,248] found the pyrolysis medium tetralin to have little influence on the thermolysis of PPE, but at higher pressures Conners et al.[272] used tetralin as a hydrogen donor for their studies of the hydrocracking of kraft lignin and two related model compounds. At extreme pressures (10 MPa) in the presence of active catalysts, lignin is known to react with hydrogen to form a liquid product consisting primarily of phenols.[273] Goheen[273] gives an excellent review of the catalytic hydrogenation of lignin.

4.4.5 Summary and Critique

Figure 4.30 offers a tentative lumped parameter model for lignin pyrolysis which correlates much of the data discussed earlier in this section. As before, the relative importance of a given pathway is strongly influenced by the lignin substrate and the prevalent experimental conditions. One virtue of the "reaction tree" given in Fig. 4.30 is its similarity to the earlier tree (Fig. 4.15) describing cellulose pyrolysis. Since both materials are composed of thermally stable rings interconnected (primarily) by ether linkages with thermally labile hydroxyl functions (and other side chains attached to the rings), some similarities in their pyrolysis behavior are naturally anticipated. Using Fig. 4.30 as a guide, the following paragraphs summarize the findings of literature reviewed in this section.

Low temperature studies of lignin thermolysis have offered insight into the influence of hydrogen bonding on the thermal softening of lignin and on the glass transition temperatures of various lignins. The products of the low temperature pyrolysis Pathway 1 in Fig. 4.30 appear to primarily reflect the method of lignin preparation. At least one model compound study points to the role of dehydration reactions in Pathway 1; however, conclusive evidence is not available. This situation

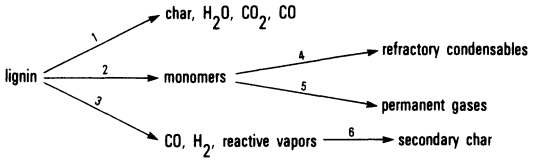


Fig. 4.30: A tentative reaction tree describing the concurrent and consecutive pathways of lignin pyrolysis.

contrasts with the wealth of knowledge available on the role of low temperature condensation reactions during the pyrolysis of carbohydrates. Noting that recent ESR studies[253] have pointed to the role of phenalenyl-like radical polymerization during the pyrolysis of two subbituminous coals at 510°C, more emphasis should be given to modeling the polymerization/condensation phenomena in this temperature range.

At somewhat higher temperatures, the formation of a variety of lignin monomers occurs via Pathway 2. Evidence for the existence of this pathway includes the research of Solomon and his colleagues,[264,265] the findings of Fenner and Lephardt,[217,218] the kinetic interpretation of the data of Iatridis and Gavalas,[267] and the thermal analysis studies of lignin pyrolysis discussed earlier in this section. Unlike the analogous pathway for cellulose pyrolysis, the pathway governing monomer formation from lignin is not well differentiated from the lower temperature condensation/polymerization pathway. Nor is a high yield of a single monomeric species observed from lignin. High heating rates and low pressures favor the formation of tars via Pathway 2. The yield of 51% tar from aspenwood lignin at 600°C under vacuum reported by Domburgs et al.[230] compares favorably with sirup yields obtained from carbohydrates.

At temperatures above 500°C the monomeric species evolved by Pathway 2 begin to undergo degradation by the vapor phase thermolysis reactions composing Pathways 4 and 5 of Fig. 4.30. The lower temperature vapor phase condensation Pathway 4 results in the formation of refractory, condensable materials; whereas the higher temperature vapor phase "cracking" Pathway 5 accounts for the evolution of much of the CO, CH_4, and other gaseous hydrocarbons obtained from the high temperature pyrolysis of lignin. A lumped parameter model incorporating a competitive rate law based on the presumed existence of Pathways 4 and 5 provides an excellent correlation of experimental data on total gas yields over a wide range of temperatures and residence times.[161] Evidence also exists for the competitive role of dehydration reactions (which accompany the formation of C_2H_4) and decarboxylation reactions within Pathway 5. Paralleling the vapor-phase reaction chemistry of carbohydrate derived volatile matter, a fraction of the lignin-derived volatile matter carbon atoms participating in the vapor-phase cracking Pathway 5 are dedicated to CO formation irrespective of temperature. In addition, a fraction is also dedicated to CH_4 formation. This may indicate that CO and CH_4 are coproducts of the same pathway. Adroit studies of carefully selected model compounds could furnish more insight into these phenomena.

At very high heating rates a pathway exists for lignin thermolysis which does not include char formation. However, char formation was observed to occur by the condensation of vapor-phase species, apparently left in a supersaturated state after the rapid heating of the lignin substrate. These two pathways are designated as 3 and 6 in Fig. 4.30. As with the carbohydrates, the high temperature lignin pyrolysis Pathway 3 involves fragmentation reactions and the formation of very high relative yields of *CO*.

The author is aware that the foregoing lumped parameter model is controversial. Sprouse and Schuman[274] recently concluded that multiple parallel reaction models better fit coal devolatilization data than competitive reaction models. Although their conclusions may reflect differences between the pyrolytic behavior of coal and lignin, this reviewer notes that the competitive model used in their studies was a tentative rate law based on an unsophisticated ad hoc analysis of often crude literature data. Sprouse and Schuman might have reached a different conclusion had they used a more sophisticated rate law, developed using methods outlined in this Review.

The similarities of lignin and coal pyrolysis phenomena have caused many researchers to adopt electrically heated screen techniques as a means for studying rapid weight loss and species evolution in a variety of environments. Although great effort has been made to prove that the heated screen does not influence the pyrolysis chemistry, the results are not entirely convincing to this reviewer, especially with regard to the influence of the screen on secondary, vapor-phase reactions. Conclusive evidence could be offered by a careful study of the influence of the screen on the vaporization of various representative model compounds. For example, it is known that levoglucosan (the primary volatile intermediate of cellulose pyrolysis) can be vaporized in vacuum and quantitatively recovered.[104] Studies of the ability of electrically heated screens to vaporize levoglucosan and lignin-related model compounds, without engendering secondary reactions, would offer stronger support for the utility of the experimental technique than has been put forward to date.

This section has emphasized similarities which exist between the pyrolytic behavior of carbohydrates and lignins. While it is important that the analogy not be carried too far, the carbohydrate literature suggests many experiments which could offer further insight into lignin pyrolysis. In addition, the reaction tree given in Fig. 4.30 suggests obvious experimental challenges to further test the existence of the individual pathways, to study the details of the various branches, and to incorporate additional pathways into the tree as required by new experimental data. Well defined kinetic data specific to the individual branches would assist researchers in identifying temperature regimes appropriate for the study of each pathway. More effort should be directed toward the identification of catalysts which enhance the rates of the individual pathways, especially Pathways 2 and 5. Finally, a detailed understanding of the elementary reactions occurring at the molecular level should be developed to explain the reaction tree given in Fig. 4.30. Although the detailed mechanism composed of elementary reactions must be very complex, a few of these reactions must be rate limiting; thereby causing the complex mechanism to manifest itself in the form of the simple reaction tree given in Fig. 4.30.

As was observed with the pyrolysis of carbohydrates, the concurrent reaction pathways of lignin thermolysis are nonspecific. The extraordinary variety of products which result from most lignin pyrolysis experiments are of little more than academic interest. The research community must direct much more effort toward the identification of conditions which result in the selective formation of a few high value products.

4.5 PYROLYSIS OF LIGNOCELLULOSIC MATERIALS

Until the turn of the century, wood was used to meet much of the demand for energy and chemicals in the industrialized world. Consequently, a large body of literature exists on the thermochemical utilization of wood, and several book length reviews are available.[275,276] Until recently the existence of this large body of literature was used by some to argue against the need for basic research on wood pyrolysis. Unfortunately, almost none of this literature reflects the depth of critical thinking and sophisticated experimental techniques which have been employed in the study of cellulose pyrolysis since the early 1960s. The recent growth of interest in the use of biomass as a source of high value chemicals[277-279] seems likely to correct this situation in the next few years, especially if a few successful new industries are established.

The critical question to be addressed in this section is whether the pyrolysis of a lignocellulosic material can be represented as a simple summation of the behavior of its components or if the components interact chemically or physically, thereby causing the lignocellulose to respond in its own unique way to thermal degradation. Unfortunately, only a few studies have been reported in the literature which address this question. The following sections offer a general review of the more recent literature, with special emphasis on those studies which offer insight into the more fundamental chemistry of wood pyrolysis.

4.5.1 Low Temperature Phenomena

Although some work on low temperature thermolysis has been reported in conjunction with higher temperature studies, this reviewer is not aware of any research on the products, mechanisms or rates of the low temperature condensation/polymerization reactions which must occur during wood pyrolysis. These reactions play an important role in char formation, and should be of engineering interest to those countries (Brazil, Finland, the Phillipines) seeking to manufacture charcoal from wood. In this regard, the Russian literature[280] describes the "optimum" industrial conditions for the production of reducing charcoal (used in the manufacture of crystalline silicon) at low heating rates. Research on low temperature cellulose pyrolysis, described in Part 1 of this Review, provides a menu of worthwile experiments which should be included in any investigation of the low temperature pyrolytic behavior of wood.

4.5.2 Moderate Temperature Phenomena

4.5.2.1 Products. In his review of the literature, Goos[276] lists 213 compounds which have been identified as products of wood pyrolysis. Modern instrumentation

could probably lengthen his list by at least an order of magnitude. It seems likely that some of these products do not result from pyrolysis of the isolated wood components, thereby implying chemical interactions which would signal the special identity of wood as feedstock. Unfortunately, no studies of this kind have been reported. The following paragraphs review some of the more recent literature describing the physical and chemical characteristics of the major product classes of lignocellulose pyrolysis.

Studies of the microporosity and yields of lignocellulosic chars have been reported by Blackenhorn et al.[281] and Mackay and Roberts.[282] The latter workers found that both heating rate and substrate composition exert their main influence on the char yield below 500° C. The char microstructure was found to be independent of the source material, moreover the open pore volume was not influenced by the source material or heating rate. More recent studies by Mackay and Roberts[283] have detailed the total mass and carbon yields of char from a wide variety of lignocellulose materials.

Detailed analyses of the chemical composition of tars derived from wood pyrolysis have been presented in the Russian literature,[284] and the use of sediment resin obtained from precipitated tar has been discussed.[285,286] In Canada, Beall and Duncan[287] and Roy et al.,[288] and in Japan Miura et al.[289,290] have described chemical methods for the analysis of tars and pyroligneous liquors.

Wood is regarded to be a promising source of furfural by some workers.[291-293] Karlivans et al.[292] reported that the highest yield (7.6%) of furfural was obtained from the pyrolysis of wood wastes in the presence of 1.8% sulfuric acid at 210° C. However, studies of black cherry heartwood pyrolysis by Barnes et al.[293] evidenced a steady increase in furfural yield from 250° to 500° C.

The formation of permanent gases from lignocellulosic materials is strongly influenced by temperature.[294,295] As discussed later, permanent gas yields increase by a factor of ten or more as the vapor-phase temperature is raised from 500° to 750° C.

Susott, Shafizadeh[296-299] and their colleagues have described the development of novel instrumentation which measures the heat of combustion of volatile matter derived from wood pyrolysis. Their studies of the heat content of pyrolytic vapors offer significant insight into the flammability of natural fuels. For example, Susott et al.[298] found that fresh fir needles release substantial quantities of combustible volatile matter below 300° C, which makes them a significant fire hazard.

4.5.2.2 Mechanisms and Kinetics. The techniques of thermal analysis have been identified as particularly well suited to study the mechanisms of wood pyrolysis,[300,301] and several reviews are available.[302-306] Because the DTG, DTA and DSC data obtained from wood often evidence three peaks, many researchers have been tempted to conclude that the mechanism of wood pyrolysis is simply the superposition of the mechanisms of its three components.[307-312] Other workers were unable to correlate the whole wood behavior with that of its components.[313,314] Many careful studies have made no attempt to obtain a correlation because of the gross

variation in behavior between the "same" wood components isolated by different methods[315] and other complexities.[316]

A recent study by Shafizadeh and McGinnis[317] displays DTA curves for cottonwood and its three isolated components. Because the whole wood curve exhibited peaks which qualitatively corresponded to peaks obtained from its components, Shafizadeh and McGinnis concluded that the thermal behavior of the whole wood reflected the behavior of its components. This conclusion would have been more convincing if the authors had shown the whole wood curve to be a mathematical superposition of the component curves.

A more quantitative approach has been reported by Duvvuri et al.[318] and Mattocks.[319] Duvvuri et al.[318] employed a mass fraction weighted sum of simple Arrhenius rate laws for cellulose and lignin (punky wood) to simulate the weight loss at 10° C/min of an excelsior wood. All experimental data were obtained by TGA, and the Arrhenius parameters were determined by a novel least squares technique. However, the values of all the Arrhenius parameters were observed to be temperature dependent, reflecting both the overly simplified rate law used to interpret the weight loss measurements and the heat transfer limitations of the thermobalance used in their studies. Moderate agreement of the simulation with experimentally determined weight loss measurements was reported. A similar study by Mattocks[319] employed three simple rate laws, representing the thermal degradation of cellulose, hemicellulose, and lignin, to simulate the weight loss of a hardwood and softwood as measured by TGA. Although good agreement was obtained between a simulation and experimental results at a single heating rate, poor agreement was evidenced when the model was used to simulate experimental data obtained at other heating rates.

A generic fault of all models based on a linear superposition of simple (one step) rate laws is their inability to account for the fact that higher heating rates result in the formation of less char from wood. Quite recently Thurner and Mann[320] attempted to overcome this problem through an ambitious use of a kinetic model based on Shafizadeh's proposed mechanism for cellulose pyrolysis (see Fig. 4.1). Between 325° and 385° C the values of the activation energies associated with the primary reactions forming gas, tar and char were reported to be 21.2, 26.9 and 25.4 kcal/gmol (respectively). The gaseous products consisted primarily of O_2, CO, CO_2, C_3 and higher hydrocarbons; whereas the sirup contained seven compounds with levoglucosan accounting for more than half the yield. In spite of the evident sophistication of this approach, a number of questions remain. Free oxygen is not a product of wood pyrolysis. Some explanation of its source and evaluation of its effects seems requisite. More importantly, char cannot be formed without the simultaneous evolution of tar and/or gas. Thus the rate equations governing gas and/or tar evolution must necessarily be coupled with the rate equation for char formation. Future work should incorporate this coupling into the rate equations. Finally, it seems likely that three sets of rate equations, similar to those employed by Thurner and Mann,[320] but accounting for the different responses of cellulose, hemicellulose, and lignin to pyrolysis, will ultimately be required to properly simulate the pyrolytic behavior of wood. This last observation concurs with the work of Roberts,[321] who pointed to the role of competitive reactions in the thermal degradation of the cellulose and

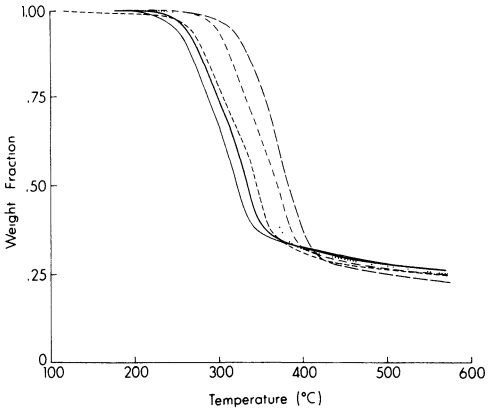


Fig. 4.31: TG Curves of ground corncob pyrolysis at heating rates of 2.1° (solid line), 5.1° (dark solid line), 10.5° (dashed line), 20.0° (dotted line), 60.2° (dashed and dotted line), and 153.0 °C/min (long dashed line).

lignin components of a whole wood. Roberts used this observation to explain the discrepancies in the values for $\Delta H_{\text{pyrolysis}}$ reported by many workers.

Further evidence substantiating the role of competitive reactions in the pyrolysis of lignocellulosic materials is given in Figs. 4.31–4.33, which display TG, DTG, and Friedman signatures[143] for a ground corncob material. Evidence for the role of several competitive reactions includes the decline in char yields with increasing heating rate (Fig. 4.31), the increasing DTG peak heights with increasing heating rate (Fig. 4.32), and the sinusoidal behavior of the Friedman signatures (Fig. 4.33). It seems likely that these curves could be simulated by a combination of rate laws describing three competitive mechanisms, but a considerable effort would be required to identify all the constants associated with the various rate equations.

Hornof et al.[322] have described a very interesting study of the effect of lignin content on the thermal degradation of wood pulp which hints of the complexities of wood pyrolysis. Their results show that the apparent activation energy associated with pulp pyrolysis is strongly affected by the presence of lignin. An increase in the lignin content of pulp from 3.7% to 23% was observed to dramatically reduce the apparent activation energy from 67.0 kcal/gmol to 41.4 kcal/gmol. It seems unlikely that this behavior could be explained using a linear superposition of cellulose and lignin weight loss models. Since phenols are known to be active free radical traps, and recalling the role of free radicals in cellulose pyrolysis, the complex nonlinear effects of lignin content on the pyrolytic behavior of lignocellulosic materials are not surprising.

Model compound studies offer further insight into the mechanisms of lignocellulose pyrolysis at the molecular level. The extensive research of Shafizadeh and his colleagues[323–329] on the thermolysis of various phenyl glucosides and related model compounds was described in Part 1 of this Review.[2] Among their results was

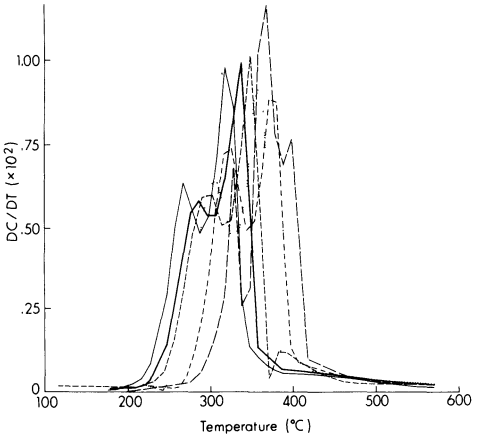


Fig. 4.32: DTG Curves of ground corncob pyrolysis. Rate of conversion DC/DT ($°C^{-1}$) vs. temperature (°C) at heating rates of 2.1° (solid line), 5.1° (dark solid line), 10.5° (dashed line), 20.0° (dotted line), 60.2° (dashed and dotted line), and 153.0°C/min (long dashed line).

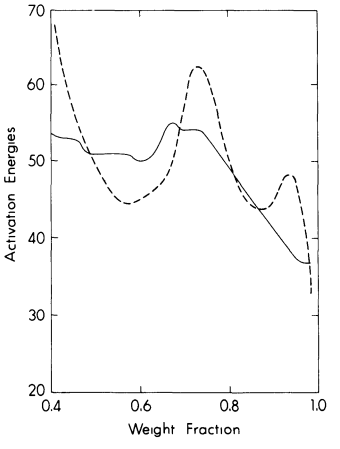


Fig. 4.33: Friedman Signatures of ground corncob pyrolysis with heating rates of 2.1°, 5.1°, and 10.5°C/min (dashed line) and 10.5°, 20.0°, and 60.2°C/min (solid line).

the finding that substituents on the aglycone moiety affected the cleavage of the glycosidic group, thereby influencing the reaction products. This finding would lead one to anticipate interactive effects between the components of whole lignocellulose materials. More recently Kosikova and his coworkers[259-261] have confirmed and extended the earlier work of Shafizadeh. Kosikova's studies of 3-0-benzyl ethers of D glucose[260] showed that *MeO* substituents on the aromatic moiety decreased the rate of pyrolysis and formed active sites for condensation reactions. Similar studies

of 4-0-benzyl ethers of methyl-α-D-glucopyranoside[261] revealed a remarkable decrease in the thermal stability of the aryl ethers when OH groups were found in the para position. The results of the model compound studies led Shafizadeh and his colleagues[323-329] to conclude that transglycosylation proceeds by the heterolytic cleavage of glycosidic bonds. By way of contrast, the recent work of Domburgs et al.[330] involving ESR, GC and thermal analysis studies of lignocellulose pyrolysis emphasizes the formation of free radicals during the rapid formation of levoglucosan.

A number of studies have been reported on the pyrolysis of large wood cylinders.[331-333] Typically these studies have been more concerned with the effects of heat and mass transfer on pyrolysis phenomena than the details of the pyrolysis chemistry. Pyrolysis product compositions were not reported, and only simple rate laws with low activation energies (~20 kcal/gmol) were employed to interpret the results. Mathematical models presented in these studies should be of interest to researchers studying the production of charcoal from wood.

Two interesting thermogravimetric studies of the pyrolysis of wood bark have been described in the recent literature.[334,335] Fairbridge et al.[334] concluded that the Prout Tompkins equation, used previously to describe inorganic decompositions of the type solid → solid + gas, best represented their weight loss data. They also noted that their data fit a reaction rate compensation curve, evidencing a linear dependence of $\ln A$ on E. Similarly, a dependence of E on weight loss was observed by Tran and Rai.[335] Unfortunately, neither study attempts to account for the influence of heating rate on char formation. As with other studies,[336-338] the role of competitive reactions may explain the observed compensation effect and offer a better insight into the mechanisms of bark pyrolysis. Gas yields from the pyrolysis of poplar bark have been described by Kumar and Mann.[339,340] Their kinetic analysis ignores the effects of secondary reactions at temperatures exceeding 500° C; consequently the reported rate constants are of limited value.

Exploratory studies of the pyrolysis of bagasse[341,342] and rice hulls[343] have been described in the very recent literature. The low tar yields obtained from bagasse[341] may offer evidence for the influence of mineral matter on the pyrolysis mechanism.

4.5.3 High Temperature Phenomena

The high temperature (> 500° C) chemistry of wood pyrolysis involves both primary solid phase and secondary gas phase pyrolysis phenomena. In addition, heterogeneous char gasification reactions become increasingly important at temperatures above 700° C. The following sections focus attention on reports in the literature describing carefully chosen experiments which attempt to isolate the solid phase, gas phase, and char gasification phenomena.

4.5.3.1 Products. Products of the high temperature pyrolysis of biomass typically include CO, CO_2, H_2, H_2O, a variety of light alkanes, alkenes and alkynes, some condensable refractory tars and light water-soluble organic compounds, and char. Studies of the influence of temperature on product yields have been reported by Goheen and Henderson,[268] Mattocks,[319] Diebold and Smith,[344] Diebold and Sc-

ahill,[345,346] Lédé et al.[347,348] Bohn and Benham,[349,350] Milne and Soltys,[351] Deglise et al.,[352] Caubet et al.,[353] and Scott and Piskorz.[354] Typically, higher temperatures reduce the formation of char and tar and favor the formation of olefins and acetylene.

4.5.3.2 Mechanisms and Kinetics. Very little information is available on the high temperature mechanisms of the solid-phase pyrolysis of lignocellulosic materials. The results of Hopkins et al.[159] indicate extraordinary yields of CO relative to CH_4 and C_2H_4 from redwood, Leucaena, and corncob materials (see Table 4.3). This behavior mirrors the high temperature pyrolysis of cellulose, hemicellulose, and lignin discussed earlier. The only plausible explanation for these extraordinary yields is the existence of a high temperature pathway involving the catastropic fragmentation of the complex lignocellulosic polymer into the elementary functional groups which compose it.

The high temperature gas phase pyrolysis of hardwood- and softwood-derived volatile matter has been extensively studied by Mattocks.[319] Patterning the behavior of cellulose and lignin, a pair of competing reactions (see Fig. 4.17) appear to describe the thermal degradation of both the softwood- and hardwood-derived volatile matter into either permanent gases or refractory condensable materials. Values of the rate constants associated with the cracking reaction 1 and polymerization reaction 2 are given in Table 4.2 and compared with the similar values for cellulose and lignin-derived volatile matter. The values for the two woods represent an average of high values for cellulose and low values for lignin, with the influence of the lignin-derived volatile matter being out of proportion with its actual presence in the gas phase. This may provide further evidence for the role of lignin-derived volatile matter as a free radical trap in the pyrolysis process.

Figures 4.34 and 4.35 indicate the influence of temperature on the vapor-phase carbon efficiency $\Delta\eta$ and the relative values of $\Delta CO, \Delta CH_4$ and ΔC_2H_4 from hardwood- and softwood-derived volatile matter.[295,319] As with cellulose, increasing temperature favors the vapor-phase formation of C_2H_4 over CO and CH_4 for both woods. However, the temperature dependence of the ratio of ΔCO to ΔCH_4 more closely resembles lignin than cellulose. As with both cellulose- and lignin-derived volatile matter, a fraction of the carbon atoms of both woods appears to be dedicated to CO formation.

The similarity in the gas-phase behavior of both softwood- and hardwood-derived volatile matter to that of cellulose and lignin might lead one to speculate that the gas-phase phenomena simply pattern the phenomena of the cellulose, hemicellulose and lignin constituents. The falsity of this speculation was demonstrated by Mattocks,[295,319] who showed that the linear superposition of product yields derived from two (cellulose and kraft lignin) and three (cellulose, D-Mannose and kraft lignin) lignocellulose components did not represent the product yields obtained from the hardwood and softwood. Moreover, an attempt to describe the wood yields using a linear programming optimization model which varied the relative contributions of the three components to the overall gas yields led Mattocks to conclude that the wood yields were best simulated by the cellulose yields alone. This result appears to conclusively indicate that wood-derived volatile matter does not behave as the

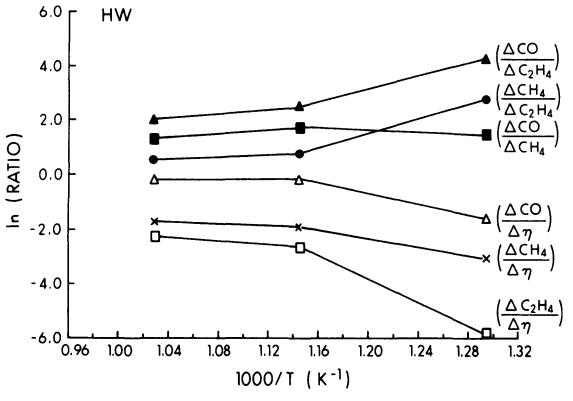


Fig. 4.34: Ln of the ratio of various gas yields and $\Delta\eta$ vs. T^{-1} for hardwood (HW) pyrolysis.[319]

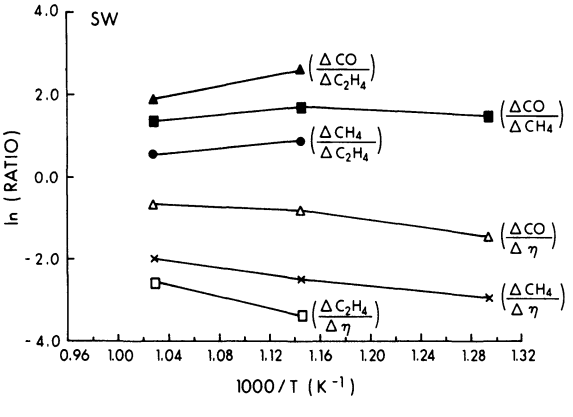


Fig. 4.35: Ln of the ratio of various gas yields and $\Delta\eta$ vs. T^{-1} for softwood (SW) pyrolysis.[319]

simple linear superposition of volatile matter derived from its components. To test this conclusion, future work should emphasize native lignins which may behave differently than the highly altered kraft lignin studied by Mattocks.[319]*

A discussion of the gasification phenomena of lignocellulosic char is beyond the scope of this Review. The interested reader is referred to Reed,[355] Rensfelt et al.,[356] and Richard et al.[357] for a description of recent research on the gasification properties of various biomass-derived chars.

* Note however that Antal[161] provides evidence suggesting only minor differences in the response of native and highly altered lignins to high temperature vapor-phase pyrolysis conditions.

4.5.4 Effects of Various Parameters

4.5.4.1 Heating Rate. Reflecting the behavior of each of its components, higher heating rates favor the formation of volatiles from lignocellulosic materials.[319] Sufficiently rapid heating can effect the complete vaporization of wood.[345] Scott and Piskorz[358] have demonstrated tar yields as high as 65% from hybrid aspen-poplar sawdust rapidly heated ($\sim 10^\circ$ C/s) in a fluidized bed.

4.5.4.2 Pressure. The research of Mok and Antal[129,130] discussed earlier has revealed the favorable influence of increased pressure on char formation. Enhanced char formation is accompanied by enhanced exothermicity of the solid-phase pyrolysis reactions. This behavior resembles the response of the individual biomass components to increasing pressure.

Complementary studies by Roy and Chornet[359,360] have demonstrated that vacuum pyrolysis increases the yields of tars from many lignocellulosic materials. Tar yields in excess of 50% by weight were obtained from poplar wood, with a char yield of 15% when the wood was heated in vacuum to 950° C.

4.5.4.3 Particle Size. The response of lignocellulosic materials to increasing particle size is not well understood. One interesting aspect of increasing particle size is its influence on the apparent heat of reaction of wood pyrolysis ΔH_{pyr}. The lore of the field recognizes both positive and negative values of ΔH_{pyr}. Unpublished research from this laboratory has shown that for Red Oak wood an increase in particle diameter from <1 mm to ~5 mm decreased the composite value of ΔH_{pyr} from +60 to +38 cal/g when heated at 2° C/min, and from +53 to +28 cal/g when heated at 10° C/min. Apparently, negative values of ΔH_{pyr} reflect the influence of secondary reactions prevalent in large particles at low heating rates.

Krieger and her colleagues[361-363] have reported an ambitious experimental and theoretical study of the effects of particle size on pryolysis product distributions. They note the complex interactive influence of increasing size on internal heating rates, pressure, absolute temperatures, and secondary reactions. Their finding that an increase in particle size reduces the primary char yield is in concord with the earlier results of Dubova et al.[244]

4.5.4.4 Gaseous Environment. Steinburg and Fallon[364,365] in the U.S.A. and Basu and Stangeby[366,367] in Canada have reported studies of the flash hydropyrolysis of biomass. According to Steinberg and Fallon,[364,365] benzene and C_2H_6 were major products at 800° C, 3.4MPa (hydrogen) and <4 sec residence time; whereas CH_4 and H_2O were the primary products at 900° to 1000° C, 3.4MPa (hydrogen) and <1 sec residence time. Interesting studies of the flash pyrolysis of biomass materials in methane, which evidence high yields of ethylene, have recently been reported by Steinberg and his coworkers.[368]

4.5.4.5 Mineral Matter and Additives. The influence of ash content on the pyrolytic yield of 1,6-anhydro-β-D-mannopyranose from ivory nut meal was discussed earlier. An extensive study of the influence of mineral matter on the pyrolysis of various plant materials has been reported by Philpot.[369,370] His results

indicate that silica has little influence on pyrolysis[369] and that only some of the mineral elements present in plant ash affect the pyrolysis reactions.[369] In general, higher char yields were obtained from plant materials with higher ash contents, but the influence of ash content on char yield was less strong above 5% to 7% ash. The greatest change in the pyrolytic properties occurred as the ash content increased from 0.01% to 4%.[369]

Deglise et al.[371,372] have reported the favorable influence of moisture content on the yields of gases obtained from the high temperature pyrolysis of wood. They discount the role of the gasification reaction $C + H_2O \rightarrow CO + H_2$ in explaining their results. This reviewer believes that their results provide further evidence for the influence of moisture on the heterolytic pathways of biomass pyrolysis.

Research groups in the U.S.A., Canada and Europe have recently evidenced an intense interest in the catalytic gasification of biomass. Extensive research at the Pacific Northwest Laboratory[373,374] has led to the identification of a "long"-lived Ni alloy catalyst for the generation of methanol synthesis gas from the steam gasification of wood. In Canada Fong and Ross[375-377] have described the catalytic effects of potassium carbonate, iron (III) oxide and zinc (II) chromite on solid-phase pyrolysis rates using TGA, and also the $[(CO_2 + H_2)/CO]$ molar ratio of product gases from cellulose gasification in helium and helium/water vapor mixtures at total pressures up to 2500 kPa. Deglise and his colleagues[378] have recently reported an interesting study of the effects of a wide variety of impregnated catalysts on product yields during the flash pyrolysis of Douglas fir wood. The catalyst K_2CO_3 increased the mass fraction yield of gases from *moist* Douglas fir from 0.771 to 0.849 at 900° C. Similarly, using $Ni(CH_3COO)_2$ the gas yield from the moist wood increased to 0.989! The significant influence of moisture on the pyrolysis product distribution, first noted by Deglise and his coworkers some years ago, is not easily understood. A mechanistic explanation of these observations seems likely to provide important insights into biomass pyrolysis phenomena.

4.5.5 Summary and Critique

Only a limited number of studies have been directed at gaining insight into the interactions of the cellulose, hemicellulose, and lignin components of wood during pyrolysis. Evidence presented in this Review supports the hypothesis that these components do interact; hence the pyrolytic behavior of wood is not simply a linear superposition of its components. Unfortunately, these interactions are very complex and manifest themselves in a diverse product slate which challenges the analytic capabilities of our best laboratories. Future research efforts in this area should attempt to quantitatively identify these interactions, seeking mechanistic explanation of the phenomena at the molecular level.

The overall pattern of three major competing pathways, noted to play a role in the pyrolysis of cellulose, hemicellulose, and lignin materials, is preserved in the pyrolysis of composite lignocellulose materials. At low temperatures, condensation reactions give rise to the formation of charcoal. At higher temperatures, transglycosylation and similar reactions result in the formation of sirups, tars and some gases. Above 500° C vapor-phase cracking reactions convert the reactive vapors evolved by

the solid-phase pyrolysis reactions into a hydrocarbon rich synthesis gas, while competing vapor-phase condensation reactions form refractory condensable materials. At very high heating rates, lignocellulose polymers undergo catastropic solid-phase fragmentation reactions, resulting in high yields of *CO*. Unfortunately, complex reactions between the major components of lignocellulose complicate the picture. Consequently, lumped parameter models based on a few competing/consecutive reactions do not enjoy the same level of success in describing the temperature/time history of product yields from lignocellulose as they did in simulating the product yields from its isolated components. Some combination of lumped parameter models will no doubt be developed as an academic exercise, but the insight such models will offer into ways of reducing complexity and increasing product selectivity seems likely to be limited.

This reviewer was among the first to identify alkali metal carbonates (especially K_2CO_3) as active catalysts for biomass gasification.[379,380] Nevertheless, he is skeptical concerning the desirability and utility of catalysis in the context of biomass conversion. Unlike coal, biomass undergoes rapid pyrolysis at relatively low temperatures and produces little char. Reactors characterized by high rates of heat transfer produce negligible char. Consequently, catalytic gasification of biomass chars is not of great importance. Instead of char gasification, one may seek a catalytic means to gain enhanced product selectivity. If the desired product is charcoal or a high value chemical such as levoglucosan, catalysis may be justifiable. But if the desired product is a synthesis gas, the catalytic reactor will have to compete with simpler oxygen-blown downdraft gasifiers and coal gasifiers. The life and recoverability of the catalyst will have critical bearing on the outcome of this competition. Since biomass gasifiers must be small (usually <1000 tons/day), economies of scale will favor coal gasifiers. Moreover, the small scale of biomass gasifiers does not appear to be well suited to the use of complex catalytic reactors. These arguments do not augur well for catalytic biomass gasifiers. In any case, as emphasized by this reviewer four years ago,[381] those seeking to study catalysis of biomass pyrolysis and gasification reactions would be well advised to become thoroughly familiar with the vast literature on fire retardants. An excellent review of fire retardants has been given by Lyons.[191]

As with its components, the key to wood utilization lies in product selectivity. But unlike its components, the literature contains no reports of the pyrolysis of whole wood leading to high yields of a few *valuable* products. High yields of charcoal are easily obtained, but charcoal is an unlikely substitute for fluid fuels. High yields of synthesis gas are also possible, but the economic benefits are unclear. Processes which attempt to utilize the components separately, either following chemical fractionation of the lignocellulose or by selective chemical attack, appear to be more promising. In this regard, Shafizadeh and Stevenson[382] recently described a process for the saccharification of wood involving prehydrolysis of the hemicelluloses, followed by pyrolysis of the residual lignocellulose to produce a levoglucosan-rich sirup. Hydrolysis of the sirup resulted in a 59% recovery of the hexoses. Similar approaches have seen commercial development in the Soviet Union.[383] While it is too early to foresee which thermochemical processes will enjoy success in the

marketplace, it seems likely to this reviewer that the more sophisticated technologies will not utilize whole lignocellulose materials in the conversion step.

4.6 COMMERCIAL DEVELOPMENT

* Processes for the thermal conversion of coal, peat and oil shale to more useful fuel forms have advanced markedly since their inception a century (or more) ago. Consequently, it is of interest to consider their progress from one generic technology to another, and to anticipate the implications of that progress for the evolution of biomass conversion processes. One goal of this section is to review from a generic standpoint the orderly technological evolution of processes for the thermal conversion of coal, peat and oil shale to more useful fuel forms, and to examine the implications of that evolution for biomass conversion processes. A second goal is to consider the generic economics of thermal conversion processes. Since capital costs often dominate the economics of synthetic fuel production, a "figure of merit" is used to make a reasonably meaningful comparison of the capital costs of various conversion processes. This figure of merit suggests ways to improve the economics of future fuel conversion processes.

4.6.1 Generic Technologies

Numerous reviews exist which catalog the various technologies for conversion of biomass, coal, oil shale, and peat into fluid fuels. The review by Reed[384] in this volume gives an excellent summary of the state of the art of biomass conversion processes. The *McGraw-Hill Encyclopedia of Energy* offers an overview of fossil fuel conversion systems. This section presumes familiarity with the field at the level of these two reviews. Its goal is not to provide a catalog description of commercialization activities; rather it offers the reader a special insight into the generic aspects of reactor development. These aspects underlie future progress in the field.

Processes for thermal conversion rely on temperature dependent chemical steps evoked by heat transfer to accomplish the desired transformation of biomass to more useful fuel forms. Consequently, thermal conversion reactors are best classified by the manner in which they accomplish heat transfer. Other classifications have been suggested;[385,386] however, this approach offers the greatest insight into the generic aspects of the conversion process.

The classification scheme given in Table 4.5 and used throughout the remainder of this review was originally proposed five years ago,[389] and is a modification of a scheme used by Guthrie[387] to classify oil shale retorts. It is equally useful for classifying coal, peat and biomass conversion processes. Table 4.6 presents a classification of some of the better known coal gasification processes[388] based on the scheme given in Table 4.5, and Table 4.7 gives the analogous classification for oil shale retorts.[388]

* This section is an updated version of material originally presented by Antal.[389]

Table 4.5

Classes of thermal conversion reactions[389]

Class I:	Indirect heating by radiation and conduction through the reactor walls.
IA	Long solids resident times (minutes or hours), low solids throughput as in a batch still.
IB	Short solids residence times (seconds or less), high solids throughput as in an entrained flow, transport reactor.
Class II:	Direct heating by circulation of hot gases resulting from the combustion of a portion of the feedstock or reaction products.
IIA	Combustion internal to the reactor.
IIB	Combustion external to the reactor.
Class III:	Heat transfer due to recycled solids, where the heat is derived from.
IIIA	Sensible heat carried by the hot recycled solids.
IIIB	Exothermic heat of reaction of the recycled solids within the primary gasification reactor.

The evolution of conversion technologies through the various classifications is of some interest. Coal gasification processes progressed from batch stills and large retorts employing large railroad cars moving slowly through heated tunnels (Class IA), to air- and oxygen-blown gasifiers (Class IIA) using traveling packed bed and (later) fluidized bed reactors, to fluidized bed hydrogasification processes (Class IIB), to reactors which utilize the exchange of either hot or reactive solids to achieve heat transfer (Class III). Finally, "third generation" coal gasifiers presently being developed rely on rapid heat transfer by radiation and convection in an externally heated tubular transport reactor (Class IB) to achieve high product yields and economy of operation.[386,388]

Similarly, oil shale retorts evolved from batch stills and vertical (Pumpherson) retorts (Class IA), to simple stoves and more sophisticated vertical air-blown downdraft retorts which rely on hot flue gases from an external combuster to retort the shale (Class IIB). The most advanced retorts use hot solid heat carriers (ceramic balls or spent shale) to heat and retort the shale feed (Class III).[386,390] Because of the low volatile matter content of oil shale, it is unlikely that Class IB reactors will be employed to retort shale on a commercial scale.

Those readers knowledgeable in reactor design will recognize the important role heat transfer plays in the classification scheme. Heat transfer is slowest and

Table 4.6

Classification of various coal gasifiers[389]

Class IA:	Coke Ovens
	Fischer Assay Retort
Class IB:	Cities Service Research Reactor
Class IIA:	Lurgi Gasifier
	Bi-Gas Reactor
	Hygas Reactor
	Synthane Reactor
	U-Gas Reactor
Class IIB:	GOED Reactor
	Hydrane Reactor
Class IIIA:	Westinghouse Double Fluidized Bed System
	TOSCOAL Reactor
	Garrett Coal Pyrolysis Reactor
Class IIIB:	CO_2 Acceptor Process

least efficient in Class IA, somewhat better in Classes IIA and IIB, relatively rapid in Class III and very rapid in Class IB. Improved heat transfer leads to larger throughputs which have a critical bearing on process economics (as will be discussed in the following section). Since biomass conversion is often limited by the rate of heat transfer and not chemical kinetics, it is only natural to expect heat transfer to play an important role in the design of biomass pyrolysis reactors.

Table 4.8 presents a classification of some of the better known biomass conversion processes. As can be seen from Table 4.8, a "full slate" has already developed. Following the discussion of generic economics given in the next seciton, examples of each generic technology will be discussed.

4.6.2 Generic Economics

The detailed economic analysis of fuel conversion processes is beyond the scope of this Review. Instead, this section describes a methodology for gaining insight into the economics of fuel conversion based on an approach analogous to the use of nondimensional numbers in engineering design. This approach not only allows one to compare technologies which differ in scale and application, it also offers insight into how the economics of a process can be dramatically improved. The format

Table 4.7

Classification of various coal retorts[389]

Class IA:	Pumpherson Vertical Retort Fischer Assay Retort
Class IB:	None
Class IIA:	Bureau of Mines Gas Combustion Retort Union "A" Retort Paraho Retort
Class IIB:	Petrosix Retort Union "B" Retort Union SGR Retort
Class IIIA:	Tosco II Retort Lurgi–Ruhrgas Retort Soviet UTT–500 Retort
Class IIIB:	None

prescribed by this methodology is suggested for future, in-depth economic studies of biomass conversion processes.

Capital costs often dominate the economics of synthetic fuel production. One convenient measure of the capital cost of a process is its total capital investment ($) divided by its daily fuel production rate (in thermal watts or equivalent barrels of oil per day).* Table 4.9 lists the approximate $ per W_{th} of several representative fossil fuel conversion technologies. The reader should note that SNG from coal and related technologies are over 100 times more capital intensive than oil from the Persian Gulf. This gross, unprecedented increase of the capital cost of fuel production largely explains the current energy crisis. If a 20% annual rate of return on capital investment is required, the capital cost component of SNG from coal is $16 per barrel (= 0.20 ÷ 365 x $30,000) or almost $3 per MBtu ($3 per GJ).

The $ per kW_{th} figure of merit suggests two ways of technologically lowering the capital cost component of fuel conversion processes. The first is to simply reduce

* Equivalent barrels of oil per day are calculated using the rough equality 1 bbl/d = 6 MBtu/d. This does not account for the quality of the fuel. For example, the production of 6 MBtu/d of gasoline is not economically equivalent to the production of 6 MBtu/d of low Btu gas since gasoline is much more valuable than gas.

Table 4.8

Classification of selected biomass conversion process[389]

Class IA:	Wood Distillation Ovens
	Garrett Multiple Hearth Reactor*
	Albany, Oregon Liquefaction Process
Class IB:	China Lake Transport Reactor
	SERI Ablative Fast Pyrolysis Reactor
	University of Hawaii Radiant Flash Pyrolysis Research Reactor
	University of Washington Microwave Research Reactor
Class IIA:	SERI Downdraft Oxygen Blown Gasifier
	Union Carbide Purox Process
	Monsanto Landgard System
	Andco–Torrax Gasifier
	Tech–Air Gasifier
	PNL Gassifier
	University of California (Davis) Gasifier
	ERCO Fluidized Bed
	Texas Tech Fluidized Bed
Class IIB:	Syngas Reactor
Class IIIA:	Bailie (University of West Virginia) Twin Fluidized Bed Reactor
	Battelle Twin Fluidized Bed Reactor
	Pyrox Reactor
Class IIIB:	None

*Multiple hearth furnaces rely on a number of mechanisms for heat transfer; consequently some engineers might reason that they should be classified as either IIA or IIB.

the $ value of the capital associated with a unit of production, usually by taking advantage of inherent economies of scale. The second is to increase the reactor throughput per unit of capital investment, usually through improved reactor design.

Economies of scale cannot be relied upon to greatly reduce capital costs for biomass conversion processes since it is difficult to accumulate large quantities of biomass at a single location. Increasing the scale of a conversion facility from 100 to 1000 tons per day* increases capital costs by a factor of five (using a 0.7 scaling law),

* Conversion facilities smaller than 100 tons per day or larger than 1000 tons per day are of limited interest.

Table 4.9

Some capital requirements for fuel production[389]

Source	Approximate Capital Cost (1978 dollars)	
Persian Gulf	\$4 per kW_{th}	(\$ 300 per bbl per day)
North Sea	55	(4 000)
Alaskan North Slope	55	(4 000)
Imported LNG	140	(10 000)
Oil from Shale by Retort	>270*	(>20 000*)
SNG from Coal	>410*	(>30 000*)
Oil from Alberta Tar Sands	>410*	(>30 000*)

*Note that capital costs are directly related to the degree of commercialization for these developing technologies, i.e., oil from shale by retort has a lower projected cost than SNG from coal and is also further from commercialization.

resulting in a net decrease in capital costs measured in \$ per kW_{th} of a factor of two. This factor is small compared with variations in throughput among the various reactor classes listed in Table 4.10, as will be seen shortly. A better approach to reducing the \$ value of the capital investment is the projected development of small scale standardized reactors which could be mass manufactured on an assembly line. As discussed in the following section, research activities at SERI in conjunction with International Harvester were aimed at this goal.

Since the pyrolysis/gasification reactions of biomass occur so rapidly, a conversion reactor's throughput is determined primarily by its inherent heat transfer characteristics. As noted in the preceding section, heat transfer rates increase when progressing from Class IA (worst heat transfer), to Classes II and III, and finally to Class IB (best heat transfer). More generally, entrained flow transport reactors and fluidized bed reactors are noted for rapid heat transfer. Table 4.10 lists the throughputs of various biomass conversion reactors. The reader will note that throughput rates can be increased by a factor of about 100 through an intelligent choice of reactor design. This large increase in throughput could manifest itself in a reduction in capital costs (\$ per kW_{th}) by as much as 100,* as contrasted with a factor 2 due to

* This value assumes that the costs in \$ per m^2 of the reactors listed in Table 4.10 are equal. Of course, this is not true. However, it seems likely that large variations in throughput

Table 4.10

Throughputs of various biomass conversion systems[389]

Reactor	Class	Throughput* Mg/M²-day	ton/ft²-day
Garret Multiple Hearth[394]	IA or II	5.	0.5
China Lake[395]	IB	450.0	43.0
Purox[391]	IIA	26.0	2.5
Tech-Air[391]	IIA	42.0	4.0
ERCO[396]	IIA	66.0	6.3
U.C. (Davis)[397]	IIA	10.0	1.0
Bailie[398]	IIIA	?**	
Occidental Flash Pyrolysis[391]	IIIA	101.0	9.7

* Usually reported in terms of *dry organic* solids mass · unit cross sectional area^{-1} · time^{-1}, however, not all references were clear on this. A better measure would employ the total surface area of the reactor; rather than its cross sectional area.

** Probably similar to ERCO.

economies of scale. Clearly, improved reactor design merits heavier emphasis than large scale reactor development.

Projected capital costs ($ per kW_{th}) of various biomass conversion processes are listed in Table 4.11. These projections have been made by independent, respected engineering firms which have no vested interest in marketing the technology in question. Several studies of this nature were supported by U.S.E.P.A.,[391,392,393] and have been quite helpful in obtaining a less biased perspective of biomass conversion economics.

4.6.3 State of the Art Reactors

Various state of the art reactors are under development which hold promise for the improved utilization of biomass fuels. This section reviews recent activity in this field with emphasis on those processes which take advantage of the unique thermal properties of biomass fuels.

($Mg \cdot m^{-2} \cdot d^{-1}$) offset small increases in $ per m^2 to yield more improved capital costs ($ per kW_{th}) than can be realized by economies of scale.

Table 4.11

Throughputs of various biomass conversion systems[389]

Reactor	Product	Capital ($/kW)	Cost ($/bbl/d)
Albany, Oregon	liquefaction oil	?*	
	gasoline	1360**	99 000**
Purox[391]	medium Btu gas	490	36 000
Tech-Air[391]	pyrolysis oil and char	55	4 000
ERCO[396]	low Btu gas	190	14 000
Bailie[398]	intermediate Btu gas	180	13 000
Occidental Flash Pyrolysis[391]	pyrolysis oil	510	37 000

* Liquefaction oil price in excess of $30 per bbl indicates extraordinary capital requirements.

**Due to the high value of the product, the apparently high capital cost of this system appears to be justified.

4.6.3.1 Class I Reactors. Since 1973 research activities by the Wright Malta Corporation have emphasized the development of a large augur kiln designed to gasify biomass in a pressurized steam environment.[399] Relying on the exothermicity of the pressurized steam pyrolysis reactions, as well as systematic recuperation of all "waste" heat, biomass to methane conversion efficiencies of 70% are projected. The process also takes advantage of the extraordinary reactivity of biomass chars[400] to secure virtually complete pyrolytic gasification of the biomass feedstock at relatively low temperatures. Unlike the Class IB reactors discussed in the following paragraphs, the Wright Malta Process seeks to achieve chemical equilibrium within the reactor while optimizing the first law efficiency of the conversion process.

Diebold[401] and his colleagues at SERI were the first to identify the extraordinary potential of conductive heat transfer to completely vaporize biomass materials. Diebold's display[402] of an electrically heated red hot wire rapidly slicing through wood seems likely to be remembered as the classic demonstration of the unique thermal properties of biomass fuels. Pursuing the phenomena evidenced in this experiment, Diebold and his colleagues at SERI are developing a cyclonic flow reactor which utilizes conductive heat transfer between the reactor walls and coarse, sawdust size particles of biomass to effect ablative flash pyrolysis of the particle and vaporize it. Pyrolysis vapors are entrained and diluted by steam and carried through a gas-phase pyrolysis reactor where the vapors are cracked to an olefin-rich

synthesis gas. The development of a similar reactor system has also been described by Lédé, Villermaux, and their colleagues in France.[403]

Paralleling the work of Diebold, research in the author's laboratory has focused on the use of radiative heat transfer to rapidly vaporize biomass materials. Early numerical simulations established the ablative nature of the radiant flash pyrolysis process.[404,405] More recent experimental results[406-409] and detailed numerical simulations[410,411] demonstrated the ability of radiative heat transfer to flash heat the opaque solid biomass particles in a relatively cold gaseous environment. By decoupling the solid-phase from the gas-phase temperatures in this way, high yields of pyrolytic sirups have been obtained.[407-409,412] The spouted bed reactor used in this work can accommodate a wide variety of particle sizes and requires less steam than a comparable fluidized bed reactor. These features indicate its promise for further development and scaleup.

The flash pyrolysis reactors being developed at SERI and the University of Hawaii differ from the Wright Malta design by their emphasis on the use of rapid heat transfer and short residence times to effect chemical disequilibrium within the reactor. Because the pyrolysis reactions are essentially authothermic, both reactors have the potential of achieving a high first law efficiency.

4.6.3.2 Class II Reactors. The development of a pressurized, downdraft, oxygen blown gasifier has been ongoing at SERI under the leadership of Reed[413] and his colleagues since 1978. (See Reed's article in this volume.) The downdraft design carries the pyrolytic vapors evolved in the upper part of the reactor through a hot charcoal layer in the bottom of the reactor, thereby ensuring the absence of tars from the reactor effluent. The high temperatures and pressures present effect chemical equilibrium in the reactor, and transform the biomass into a synthesis gas which is well-suited for the production of methanol or methane. As mentioned earlier, Reed[413] anticipates that this reactor design will be mass produced on an assembly line, leading to a dramatic reduction in the capital and operating costs of the reactor.

The Swedish biomass gasification effort is among the most agressive and technologically sophisticated programs in the world. They seek to develop a pressured (1.0 to 3.0 MPa), oxygen-blown moderate temperature (700° to 900° C) fluid bed gasifier.[414] Products of the gasifier pass through a high temperature metallic dust filter before entering a secondary, oxygen-blown, catalytic reformer which converts tar and methane to synthesis gas. Methanol is then produced using conventional techniques.

In light of the fact that automobiles can be inexpensively* converted to run on either methane or conventional gasoline with the "flick of a switch," it is somewhat of a mystery to this reviewer why alcohol fuels continue to receive exclusive attention as a transportation fuel. Convincing arguments asserting methane to be the ideal transportation fuel have been made by knowledgeable scientists.[415] In Europe, several hundred thousand vehicles are powered by methane. Since the synthesis gas

* Conversion kits are available for about $1500

used to produce methanol can also be used to synthesize methane, some comparison of the market potential of these two products would seem to be a desirable prerequisite to further commercial developments.

The Enerco reactor[416] has been developed using private capital; consequently, less information is available on it than some of the other systems. The reactor concept involves recirculating a fraction of the evolved pyrolytic gases through a heat exchanger where the gas temperature is raised above 600°C. These hot gases are recycled through the pyrolysis reactor where their sensible heat is used to pyrolyze the biomass feedstock. The heat demand for the pyrolysis reactions is minimized by using large chips of biomass in the reactor which evidence an exotherm while undergoing pyrolysis. The heat exchanger is fired by burning a fraction of the combustible gas effluent of the reactor. Products of the reactor include combustible char, tar and a medium Btu gas.[416]

4.6.3.3 Class III Reactors. Both Bailie[417] and Feldman[418] have been active in the development of twin fluidized bed reactors. One fluidized bed combusts biomass derived pyrolytic char while hot sand is withdrawn from the combustor and circulated to the second fluidized bed. In this second reactor the biomass feedstock is pyrolyzed in a fluidizing steam or hot recycle gas environment. Char produced in the pyrolysis reactor is carried to the combustor where it is burned in air. The advantage of this reactor concept is the production of a high quality medium Btu gas with no oxygen requirement. The commercial Pyrox system in Japan uses this reactor concept to convert municipal solid wastes into a valuable gaseous fuel.

A novel Class III reactor concept being developed at the University of Western Ontario[419] uses finely divided solid heat carriers to rapidly heat biomass particles in a "vortical" reactor. Vapors evolved in the vortical reactor pass through an entrained flow reactor prior to rapid quenching of the product gases in another vortical reactor which uses cold, finely divided solids to effect rapid heat transfer and quench. This concept could be particularly attractive if inexpensive heterogeneous catalysts were developed to direct reaction pathways toward more desirable products.

4.6.3.4 Summary and Critique. A consideration of basic modes of heat transfer leads to the development of a classification system for thermochemical conversion processes. In terms of the generic technologies which comprise the various classes, innovative reactor concepts are being developed which represent each technology. This apparent breadth of research and technology development augurs well for the future of biomass energy utilization.

A consideration of the generic economics of fuel conversion processes points to the desirability of high reactor throughputs and low capital costs. Because biomass materials are extraordinarily reactive, high reactor throughputs are primarily a question of good engineering design. Many of the reactors presently being developed enjoy high throughputs. Unfortunately, economies of scale cannot be used to significantly reduce the capital costs of biomass conversion processes. Reduction in the capital cost seems more likely to result from the mass manufacture of standardized reactors, which could occur in the near future if the demand for such reactors by the developing world continues to grow.

Four years ago this reviewer was among the most outspoken critics[389] of the U.S.A.'s biomass thermochemical conversion program. Evidence presented in this section indicates that a profound change has occurred in that program during the past four years. Now technologies are being developed which hold much promise because they take advantage of the peculiar properties of biomass fuels. Although it is too soon to foretell which of these technologies will replace the oil refineries of today, there is now good reason to hope that a replacement is indeed being developed.

4.7 CONCLUSIONS

Evidence presented in this Review supports the hypothesis that the same general set of reaction pathways governs the pyrolysis of all carbohydrates. The relative role of each pathway in the formation of product species is influenced by a variety of experimental parameters, including the heating rate and final temperature, the pressure and composition of the surrounding environment, the presence of ash or additives, and the detailed molecular structure of the carbohydrate. Evidence also supports the hypothesis that a similar general set of reaction pathways governs the pyrolysis of all lignins and lignocellulose materials. However, several experimental studies clearly show that the response of lignocellulose materials to pyrolysis is not the simple composition of responses of the individual components.

Reaction trees (Figs. 4.15 and 4.30), proposed to summarize the dominant pyrolysis pathways, suggest definitive tests of the hypotheses discussed above. Lumped parameter kinetic models based on the reaction trees can be developed to describe product evolution as a function of reactor conditions. The ability of these models to interpolate and predict species formation will offer insight into their validity and suggest potential modifications to improve their utility. The existence of lumped parameter models which accurately describe pyrolysis phenomena also suggests a search for those rate-limiting steps at the molecular level which cause the complex phenomena of pyrolysis to be manifest in a relatively simple manner.

The existence of a reaction tree can offer considerable guidance to the chemical engineer seeking to design reactors which promote the selective formation of desired products. The reaction trees hypothesized here indicate that a combination of pathways is available which can lead to the selective formation of charcoal, a hydrocarbon-rich synthesis gas, or monomeric carbohydrate sirups and lignin tars. As described in this Review, the design and operation of several advanced pyrolytic reactors implicitly recognize the influence of reactor conditions on the rates of various pathways composing the reaction tree and their contribution to the reactor's product slate. Unfortunately, neither the charcoal, synthesis gas, nor the sirups and tars produced by these reactors command a high value in the marketplace. Consequently, commercial development of pyrolytic processes has reached an apparent impasse. What research activities seem likely to circumvent this predicament and offer new reactors with more attractive product slates?

In the author's laboratory a major research effort has been initiated to study the hydration/dehydration chemistry of carbohydrate and lignin monomers (and related model compounds) in supercritical water. The reason for this interest lies

in the potential importance of the generic reaction stoichiometry

$$\text{monomer} \rightarrow a CO_2 + b H_2O + \text{hydrocarbons} \tag{4.5}$$

For example, using levoglucosan as a representative cellulose monomer we have

$$C_6H_{10}O_5 \rightarrow 2CO_2 + H_2O + 2C_2H_4 \tag{4.6}$$

Similar stoichiometries can easily be written for other anhydrosugars and lignin-derived monomers. In addition to the high value products generated by this broad class of vapor-phase dehydration/decarboxylation reactions, such reactors preserve the heat of combustion of the reactants in the hydrocarbon product. This reflects the fact that these reactions are autoergic ($\Delta H = 0$). Hence processes based on dehydration/decarboxylation chemistry can enjoy thermal (first law) efficiencies approaching 100%.

Unfortunately, the highest yields of ethylene obtained to date are only about 25% of the theoretical yield indicated above.[198] Although the dehydration reactions occur rapidly in the vapor phase,[178] CO_2 evolution is much less than CO formation under all experimental conditions reported in the literature. This reflects the fact that, following dehydration and rearrangement, carbonyl functions are created (as in levoglucosenone), not carboxyl functions. In fact, chemists who have performed detailed mechanistic studies of vapor-phase model compound thermolysis[162] are hard pressed to explain by conventional free radical chemistry the "high" yields of CO_2 experimentally observed! The key step in improving the yields of olefins and other hydrocarbons from biomass is thus seen to depend upon molecular rearrangements which result in the formation of carboxyl functions, which can be cleaved from the hydrocarbon skeleton at higher temperatures. How can this feat of reaction engineering be accomplished?

One answer is given in Fig. 4.5, where Houminer and Patai[104] observed that the hydration of a carbonyl function can result in the formation of a carboxylic acid while opening the ring. Thus dehydration/hydration chemistry may underlie both the formation of olefins and other hydrocarbons and the molecular rearrangements required to assemble carboxyl functions. Supercritical water is a convenient medium to effect high temperature, fluid (vapor) phase, dehydration/hydration/decarboxylation reactions involving biomass monomers. A study of these reactions and the influence of acid and alkai catalysts on their kinetics, is now a major research thrust of the author's laboratory.

As discussed in Part 1,[2] one goal of this Review has been to assemble in one place the chemical literature concerning biomass pyrolysis in order to nurture new and creative thinking by the engineering community concerned with high value fuel and chemicals production. The preceding paragraphs suggest how the author benefited from this exercise. It is hoped that others will be stimulated by the wealth of knowledge summarized in this Review, and that new research directions based on innovative chemistry will be identified and vigorously pursued during the next decade.

4.8 ACKNOWLEDGMENTS

The preparation of this Review was supported by the National Science Foundation under Grant No. CPE-8304381, the Hawaii Natural Energy Institute, and the Coral Industries Endowment. The author thanks O. Zaborsky (NSF), J. Shupe (HNEI), and D. Chalmers (Coral Industries) for their interest in this work. Much of this Review was written while the author was a Visiting Scholar in Applied Mathematics at Harvard University. The author thanks Professor D.G.M. Anderson with Harvard for his hospitality and patience during the summer months. The author is also indebted to A. Basch, A. Broido, M. Lewin, T. Reed, M. Seibert, F. Shafizadeh, F. Williams and the various reviewers for their critical, constructive review of this work. The timely preparation of the manuscript by Mrs. Connie Kawamoto was appreciated.

4.9 NOTATION

A	Pre-exponential constant (apparent frequency factor).
CrI	Crystallinity index.
DP	Degree of polymerization.
DSC	Differential scanning calorimetry.
DTA	Differential thermal analysis.
DTG	Derivative thermogravimetry.
E	Apparent activation energy (kcal/gmol).
EGA	Evolved gas analysis.
FTIR	Fourier transform infrared (spectrometer).
k	Rate constant: k = A exp (-E/RT).
t	Time (sec).
T	Temperature.
TG	Thermogravimetry.
TGA	Thermogravimetric analysis.
$w(t)$	Time dependent sample weight (g).
w_i	Initial sample weight (g).
w_f	Final sample weight (g).
α	Degree of conversion: $\alpha = (w_i - w(t))(w_i - w_f)$.
β	Heating rate (°C/min).
Δ	Increase: ΔCO is increase in CO mass yield due to gas phase pyrolysis.

REFERENCES

1. M. J. Moran, *Availability Analysis: A Guide to Efficient Energy Uses* (Prentice-Hall, Englewood Cliffs, 1982).
2. M. J. Antal, Jr., "Biomass pyrolysis: A review of the literature Part I–Carbohydrate pyrolysis" in *Advances in Solar Energy* edited by K. W. Böer and J. A. Duffie (American Solar Energy Society, Boulder, Colo.) Vol. 1, pp. 61–112 (1982).
3. F. Shafizadeh, *Adv. in Carbohydrate Chem.* **23**, 419 (1968).
4. G. David and M. MacKay, Forestry Branch Dept. Pub. 1201, ODC 813.4 (1967).
5. J. R. Welker, *J. Fire and Flammability* **1**, 12 (1970).
6. F. Shafizadeh, in *Appl. Polymer Symp. No. 28* (J. Wiley and Sons, New York, 1975) p. 153.
7. F. Shafizadeh, in *Proceedings of the Specialists Workshop on Fast Pyrolysis of Biomass*, SERI/CP-622-1096 (Solar Energy Research Institute, Golden, Colo., 1980).
8. P. M. Molton and T. F. Demmitt, BNWL-2297 UC (Battelle Pacific Northwest Lab., Richland, Wash., 1977).
9. T. Milne, in *Biomass Gasification: Principles and Technology* edited by T. Reed (Noyes Data Corp., Park Ridge, N. J., 1981).
10. F. J. Kilzer, in *Cellulose and Cellulose Derivatives*, Vol. 5, 2nd Ed. (Wiley-Interscience, New York, 1971) pp. 1015–1046.
11. O. Wolf, *Staerke* **2**, 273 (1950).
12. J. A. Radley, *Starch and Its Derivatives*, Vol. 2, 3rd Ed. (Chapman Hall Ltd., London, 1953) p. 107.
13. C. C. Gaper and D. M. Rathman, in *Industrial Gums* edited by R. L. Whistler (Academic Press, Inc., New York, 1959) p. 699.
14. G. V. Caesar, in *Chemistry and Industry of Starch*, 2nd Ed., edited by R. W. Kerr (Academic Press, Inc., New York, 1950) p. 345.
15. C. T. Greenwood, *Adv. Carbohyd. Chem.* **22**, 483 (1967).
16. O. Theander, in *Carbohydrate Sweeteners in Foods and Nutrition* edited by P. Koivistoinen and C. Hyvonen (Academic Press, Inc., London, 1980) p. 185.
17. S. Peat et al., *J. Chem. Soc.*, 586 (1958).
18. M. Cerny and J. Stanek, *Fortschr. Chem. Forsch.* **14**, 526–555 (1970).
19. R. J. Dimler, *Adv. Carbohy. Chem.* **7**, 37 (1952).
20. M. Cerny and J. Stanek, Jr., *Adv. Carbohyd Chem.* **34**, 23 (1977).
21. F. Shafizadeh et al., *Carbohyd. Res.* **13**, 184 (1970).
22. F. Shafizadeh et al., *Carbohyd. Res.* **15**, 165 (1970).

23. F. Shafizadeh, R. A. Susott, and C. R. McIntyre, *Carbohyd. Res.* **41**, 351 (1975).

24. F. Shafizadeh and R. A. Susott, *J. Org. Chem.* **38**, 3710 (1973).

25. Y. J. Park, H. S. Kim, and G. A. Jeffrey, *Acta Cryst.* **B27**, 220 (1971).

26. W. Sandermann and H. Augustin, *Holz Roh Werkst.* **22**, 377–386 (1964); *Chem. Abstr.* **66**, 11994 (1967).

27. M. S. Bains, *Carbohydr. Res.* **34**, 169–173 (1979).

28. *International Critical Tables*, Vol. 5 (McGraw-Hill, New York, 1929) p. 166.

29. D. R. Stull, E. F. Westrum, and G. C. Sinke, *The Chemical Thermodynamics of Organic Compounds* (Wiley, New York, 1969).

30. A. Pictet, *Helv. Chim. Acta* **1**, 276 (1918).

31. H. Pringsheim and K. Schmalz, *Ber.* **55B**, 3001 (1922).

32. J. C. Irvine and J. W. H. Oldham, *J. Chem. Soc.* **127**, 2903 (1925).

33. J. daS Carvalho, W. Prins, and C. Schuerch, *J. Am. Chem. Soc.* **81**, 4054 (1959).

34. M. L. Wolfrom, A. Thompson, and R. B. Ward, *J. Am. Chem. Soc.* **81**, 4623 (1959).

35. M. L. Wolfrom et al., *J. Org. Chem.* **26**, 4617 (1961).

36. H. Abe and W. Prins, *Makromol. Chem.* **42**, 216 (1961).

37. A. Bhattecharya and C. Schuerch, *J. Org. Chem.* **26**, 3101 (1961).

38. F. Shafizadeh and Y. Z. Lai, *Carbohyd. Res.* **40**, 263 (1975).

39. P. T. Mora and J. W. Wood, *J. Am. Chem. Soc.* **80**, 685 (1958).

40. H. W. Durand, M. F. Dull, and R. S. Tipson, *J. Am. Chem. Soc.* **80**, 3691 (1958).

41. P. T. Mora et al., *J. Am. Chem. Soc.* **80**, 693 (1958).

42. H. Sugisawa and H. Edo, *Chem. Ind.* (London, 1964) p. 892.

43. G. G. S. Dutton and A. M. Unrau, *Can. J. Chem.* **40**, 1196 (1962).

44. G. G. S. Dutton and A. M. Unrau, *Can. J. Chem.* **41**, 2439 (1963).

45. G. G. S. Dutton and A. M. Unrau, *Can. J. Chem.* **42**, 2048 (1964).

46. H. Sugisawa, *J. Food Sci.* **31**, 381 (1966).

47. K. Kato and H. Komorita, *Agr. Biol. Chem.* **32**, 715 (Tokyo, 1968).

48. G. N. Richards and F. Shafizadeh, *Aust. J. Chem.* **31**, 1825 (1978).

49. J. R. Katz., *Rec. Trav. Chim.* **53**, 554 (1934).

50. J. R. Katz and A. Weidinger, *Z. Physik, Chem.* **A148**, 100 (1939).

51. B. Brimhall, *Ind. Eng. Chem.* **36**, 72 (1944).

52. R. W. Kerr and F. C. Cleveland, *Staerke* **5**, 261 (1953).

53. A. Thompson and M. L. Wolfrom, *J. Am. Chem. Soc.* **80**, 6618 (1958).

54. M. L. Wolfrom, A. Thompson, and R. B. Ward, *Ind. Eng. Chem.* **53**, 217 (1961).

55. G. V. Caesar, N. S. Gruenhut, and M. L. Cushing, *J. Am. Chem. Soc.* **69**, 617 (1947).

56. H. Rüggeberg, *Staerke* **4**, 78 (1952).

57. J. D. Geerdes, B. A. Lewis, and F. Smith, *J. Am. Chem. Soc.* **79**, 4209 (1957).

58. G. M. Christensen and F. Smith, *J. Am. Chem. Soc.* **79**, 4492 (1957).

59. F. Shafizadeh and A. G. W. Bradbury, *J. Appl. Polym. Sci.* **23**, 1431 (1979).

60. S. L. Madorsky, V. E. Hart, and S. Straus, *J. Res. Nat. Bur. Stand.* **56**, 343 (1956).

61. S. L. Madorsky, *J. Res. Nat. Bur. Stand.* **60**, 343 (1958).

62. S. L. Madorsky, *Thermal Degradation of Organic Polymers* (Interscience Publishers, J. Wiley and Sons, Inc., New York, 1964).

63. M. Lewin and A. Basch, In "Fire retardancy," in *The Encyclopedia of Polymer Science and Technology*, Suppl. Vol. 2 (J. Wiley and Sons, Inc., New York, 1977) p. 340.

64. W. D. Major, *Tappi* **41**, 530 (1958).

65. S. T. Kosiewicz, *Thermochimica Acta* **40**, 322 (1980).

66. O. P. Golova and R. G. Krylova, *Dokl. Akad. Nauk, SSSR* **116**, 419 (1957).

67. S. Patai and Y. Halpern, *Israel J. Chem.* **8**, 655 (1970).

68. T. V. Gatovskaya et al., *Zh. Fiz. Khim.* **33**, 1418 (1959); *Chem. Abs.* **54**, 8212 (1960).

69. A. Paucault and G. Sauret, *Compt. Rend. Acad. Sci.* **246**, 608 (1958).

70. D. P. C. Fung, *Tappi* **52**, 319 (1969).

71. A. Broido and F. J. Kilzer, *Fire Res. Abstr. Rev.* **5**, 157 (1963).

72. Y. Halpern and S. Patai, *Israel J. of Chem.* **7**, 673 (1969).

73. A. Broido et al., *J. Appl. Polym. Sci.* **17**, 3627 (1973).

74. A. Broido and H. Yow, *J. Appl. Polym. Sci.* **21**, 1677 (1977).

75. F. J. Kilzer and A. Broido, *Pyrodynamics* **2**, 151 (1965).

76. M. Weinstein and A. Broido, *Combus. Sci. & Tech.* **1**, 287–292 (1970).

77. A. Broido and M. Weinstein, *Combus. Sci. & Tech.* **1**, 279 (1970).

78. F. Shimazu and C. Sterling, *J. Food Sci.* **31**, 548 (1966).

79. A. Basch and M. Lewin, *J. Polym. Sci.* **11**, 2071 (1973).

80. K. Kato and H. Komorita, *Agr. Biol. Chem.* **32**, 21 (1968).

81. K. E. Cabradilla and S. H. Zeronian, in *Thermal Uses and Properties of Carbohydrates and Lignins* edited by F. Shafizadeh, K. V. Sarkanen, and D. A. Tillman (Academic Press, New York, 1976).

82. A. Basch and M. Lewin, *J. Polym. Sci.* **11**, 3071 (1973).

83. A. Basch and M. Lewin, *J. Polym. Sci.* **12**, 2053 (1974).

84. J. C. Arthur, Jr. and O. Hinojosa, *Textile Res. J.* **36**, 385 (1966).

85. J. C. Arthur, Jr., T. Mares, and O. Hinojosa, *Textile Res. J.* **36**, 630 (1966).

86. P. J. Baugh, O. Hinojosa, and J. C. Arthur, Jr., *J. Appl. Polym. Sci.* **11**, 1139 (1967).

87. T. Maresad and J. C. Arthur, Jr., *J. Polym. Sci.* **7**, 419 (1969).

88. N. S. Hon, *J. Polym. Sci., Polym. Chem. Ed.* **13**, 1347 (1975).

89. N. S. Hon, *J. Polym. Sci., Polym. Chem. Ed.* **13**, 955 (1975).

90. W. A. Reeves, *J. Text. Inst.* **53**, 22 (1962).

91. H. Mehta, *Text. Res. J.* **31**, 387 (1969).

92. R. S. Parikh, *Text. Res. J.* **37**, 538 (1967).

93. R. H. Barker and S. L. Vail, *Text. Res. J.* **37**, 1077 (1967).

94. A. Basch, private communication.

95. A. Broido, M. Evett, and C. C. Hodges, *Carbohyd. Res.* **44**, 267 (1975).

96. G. R. Bedford and D. Gardiner, *Chem. Commun.* **13**, 287 (1965).

97. F. Shafizadeh et al., *Carbohyd. Res.* **61**, 519 (1978).

98. F. Shafizadeh and P. P. S. Chin, *Carbohyd. Res.* **46**, 149 (1976).

99. F. Shafizadeh et al., *Carbohyd. Res.* **64**, 433 (1978).

100. Y. Halpern, R. Riffen, and A. Broido, *J. Org. Chem.* **38**, 204 (1973).

101. F. Shafizadeh, C. W. Philpot, and N. Ostojie, *Carbohyd. Res.* **16**, 279 (1971).

102. F. Shafizadeh, R. H. Furneaux, and T. T. Stevenson, *Carbohyd. Res.* **71**, 169 (1979).

103. Y. Houminer and S. Patai, *Tetrahedron Letters* **14**, 1297 (1967).

104. Y. Houminer and S. Patai, *Israel J. Chem.* **7**, 513 (1969).

105. K. Heyns, R. Stute, and H. Paulsen, *Carbohyd. Res.* **2**, 132 (1966).

106. K. Heyns and M. Klier, *Carbohyd. Res.* **6**, 436 (1968).

107. D. J. Bryce and C. T. Greenwood, *Appl. Polym. Symp.* **2**, 149 (1966).

108. D. J. Bryce and C. T. Greenwood, *Staerke* **15**, 285 (1963).

109. D. J. Bryce and C. T. Greenwood, *Staerke* **15**, 359 (1963).

110. A. Pictet and J. Sarasin, *Compt. Rend.* **166**, 38; *Helv. Chim. Act.* **1**, 87 (1918).

111. S. Glassner and A. R. Pierce, III, *Anal. Chem.* **37**, 525 (1965).

112. S. B. Martin and R. W. Ramstad, *Anal. Chem.* **33**, 982 (1961).

113. C. T. Greenwood, J. H. Knox, and E. Milne, *Chem. Ind.* **46**, 1878 (London, 1961).

114. D. Gardiner, *J. Chem. Soc.* **(C)**, 1473 (1966).

115. G. A. Byrne, D. Gardiner, and F. H. Holmes, *J. Appl. Chem.* **16**, 81 (1966).

116. A. E. Lipska and F. A. Wodley, *J. Appl. Polym. Sci.* **13**, 851 (1969).

117. F. A. Wodley, *J. Appl. Polym. Sci.* **15**, 835 (1971).

118. F. Shafizadeh, Y. Z. Lai, and R. A. Susott, *Carbohyd. Res.* **25**, 387 (1972).

119. I. A. Puddington, *Can. J. Res.* **B26**, 415 (1948).

120. R. H. Furneaux and F. Shafizadeh, *Carbohyd. Res.* **74**, 354 (1979).

121. F. Shafizadeh et al., *J. Appl. Polym. Sci.* **23**, 3525 (1979).

122. F. Shafizadeh and T. T. Stevenson, *J. Appl. Polym. Sci.*, in press.

123. F. Shafizadeh, T. Cochran, and Y. Sakai, *AIChE Symposium Series 184* **75**, 24 (1979).

124. C. M. Lakshmanan and H. E. Hoelscher, *Ind. Eng. Chem. Prod. Res. Dev.* **9**, 57 (1970).

125. C. M. Lakshmanan and H. E. Hoelscher, *Staerke* **22**, 261 (1970).

126. L. J. Carlson, U. S. Pat. 3,235,541, *Chem. Abstr.* **64**, 16,122 (1966).

127. A. K. Esterer, U. S. Pat. 3,298,928, *Chem. Abstr.* **66**, 56,886 (1967).

128. A. K. Esterer, U. S. Pat. 3,309,356, *Chem. Abstr.* **66**, 106,072 (1967).

129. W. Mok and M. J. Antal, Jr., *Thermochim. Acta.* **68**, 155 (1983).

130. W. Mok and M. J. Antal, Jr., *Thermochim. Acta.* **68**, 165 (1983).

131. D. F. Arseneau, *Can. J. Chem.* **49**, 632 (1971).

132. D. F. Arseneau, in *Thermal Analysis Proceedings of the Third JCTA* edited by Hans G. Wiedemann (Birkhauer, Basel, 1971).

133. A. Broido and M. Weinstein, in *Thermal Analysis Proceedings of the Third ICTA* edited by Hans G. Wiedemann (Birkhauer, Basel, 1971) p. 285.

134. A. Broido and M. A. Nelson, *Comb. Flame* **24**, 263 (1975).

135. A. Broido, in *Thermal Uses and Properties of Carbohydrates and Lignins* edited by F. Shafizadeh, K. V. Sarkanen, and D. A. Tillman (Academic Press, New York, 1976) p. 19.

136. A. G. W. Bradbury, Y. Sakai, and F. Shafizadeh, *J. Appl. Polym. Sci.* **23**, 3271 (1979).

137. P. K. Chatterjee and C. M. Conrad, *Text. Res. J.* **36**, 487 (1966).

138. C. Dollimore and B. Holt, *J. Polym Sci.* **A11**, 1703 (1973).

139. C. Fairbridge, R. A. Ross, and S. P. Sood, *J. Appl. Polym. Sci.* **22**, 497 (1978).

140. C. Kala, I. S. Gur, and H. L. Bhatnagar, *Indian J. Chem.* **19A**, 641 (1980).

141. M. R. Hajaligol et al., *Ind. Eng. Chem. Process Des. Dev.* **21**, 457 (1982).

142. M. J. Antal, Jr., H. L. Friedman, and F. E. Rogers, *Comb. Sci. and Tech.* **21**, 141 (1980).

143. M. J. Antal, Jr., in *Thermal Analysis Proceedings of the 7th ICTA* (Kingston, Canada, 1982).

144. J. Lédé, comments during presentation, American Chemical Society Annual Meeting, Washington, D. C., Sept. 1983.

145. A. Broido, private communication.

146. O. P. Golova, R. G. Krylova, and I. I. Nikolaeva, *Vysokomol. Soedin* **1**, 1235 (1959).

147. A. M. Pakhomov, O. P. Golova, and I. I. Nikolaeva, *Izv. Akad. Nauk SSSR, Otdel. Khim. Nauk. 521, Chem. Abs.* **52**, 3811 (1958).

148. O. P. Golova et al., *Dokl. Akad. Nauk SSSR* **115**, 1122; *Chem. Abs.* **52**, 4165 (1958).

149. A. N. Kislitsyn et al., *Zh Prikl. Khim.* **44**, 2587 (Leningrad, 1971).

150. A. N. Kislitsyn and S. M. Rodionova, *Tr. Tsentr. Nauchn.-Issled. Inst. Leokhim. Prom.* **20** (1969).

151. A. M. Pakhomov, *Izv Akad. Nauk SSSR Otdel. Khim. Nauk. (1957) 1497, Chem. Abs.* **52**, 5811 (1958).

152. O. P. Golova, A. M. Pakhomov, and E. A. Andrievskaya, *Izv. Akad. Nauk. SSSR, Otdel. Khim. Nauk. 1499 (1957), Chem. Abs.* **52**, 5811 (1958).

153. O. P. Golova, *Russian Chemical Reviews* **44**, 687 (1975).

154. F. Shafizadeh and Y. L. Fu, *Carbohydrate Res.* **29**, 113 (1973).

155. F.Shafizadeh, A. G. W. Bradbury, and W. F. DeGroot, "Formation and reactivity of cellulosic chars," **29**, 2163 (1984).

156. A. Ohnishi, K. Kato, and E. Takagi, *Polym. J.* **7**, 431 (1975).

157. E. F. L. J. Anet, *Adv. Carbohyd. Chem.* **19**, 181 (1964).

158. J. Lédé et al., *Revue Phys. Appl.* **15**, 545 (1980).

159. M. Hopkins, M. J. Antal, Jr., and J. Kay, *J. Applied Polym. Sci.*, **29**, 2163 (1984).

160. M. J. Antal, Jr., in *Biomass as a Nonfossil Fuel Source* edited by D. E. Klass (American Chemical Society, Washington, D. C., 1981).

161. M. J. Antal, Jr., *Ind. Eng. Chem. Prod. Res. Dev.* **22**, 366 (1983).

162. A. Cutler, Ph.D. thesis, Princeton University, 1984.

163. T. A. Milne and M. N. Soltys, *Fundamental Pyrolysis Studies. Annual Report for Fiscal Year 1981*, SERI/PR-234-1454 (Solar Energy Research Institute, Golden, Colo., 1981).

164. T. A. Milne and M. N. Soltys, *J. Anal. Appl. Pyrol.* **5**, 93 (1983).

165. T. A. Milne and M. N. Soltys, *J. Anal. Appl. Pyrol.* **5**, 111 (1983).

166. H. R. Schulten and W. Gortz, *Anal. Chem.* **50**, 428 (1978).

167. H. R. Schulten, U. Bahr, and W. Gortz, *J. Anal. Appl. Pyrolysis* **3**, 137 (1981).

168. H. R. Schulten and W. Görtz, *Anal. Chem.* **50**, 428 (1978).

169. H. R. Schulten, U. Bahr, and W. Görtz, *J. Anal. Appl. Pyrol.* **3**, 137 (1981).

170. H. R. Schulten, U. Bahr, and W. Görtz, *J. Anal. Appl. Pyrol.* **3**, 229 (1981).

171. S. B. Martin, presented at the Tenth Symposium (International) on Combustion, *The Comb. Inst.*, 877 (1965).

172. J. B. Berkowitz-Mattuck and T. Noguchi, *J. Appl. Polym. Sci.* **7**, 709 (1963).

173. K. A. Lincoln *Pyrodynamics* **2**, 133 (1965).

174. K. A. Lincoln, in *Proceedings of the Specialists' Workshop on Fast Pyrolysis of Biomass*, SERI/CP-622-1096 (Solar Energy Research Institute, Golden, Colo., 1980).

175. P. C. Lewellen, W. A. Peters, and J. B. Howard, presented at the Sixteenth International Symposium on Combustion, MIT (1976).

176. M. R. Hajaligol et al., *Ind. Eng. Chem. Process Des. Dev.* **21**, 457 (1982).

177. F. Shafizadeh and Y. Z. Lai, *J. Org. Chem.* **37**, 278 (1972).

178. Y. S. Stein, M. J. Antal, and M. Jones, *J. Anal. Appl. Pyrol.* **4**, 283 (1983).

179. R. R. Baker, *J. Thermal Analysis* **8**, 163 (1975).

180. I. L. Eventova et al., *Khim. Volokra* **4**, 29 (1974).

181. R. K. Agrawal and R. J. McCluskey, *J. Appl. Polym. Sci.* **27**, 367 (1983).

182. R. K. Agrawal and R. J. McCluskey, Division of Fuel Chemistry preprints, Vol. 28, p. 5 (American Chemical Society, Washington, D. C., 1983).

183. P. K. Chatterjee, *J. Appl. Polym. Sci.* **12**, 1859 (1968).

184. A. E. Lipska and W. J. Parker, *J. Appl. Polym. Sci.* **13**, 851 (1966).

185. W. K. Tang and W. K. Neill, *J. Polym. Sci.* **C:6**, 65 (1964).

186. J. G. Wiegerink, *J. Res. Natl. Bur. Stand.* **25**, 435 (1940).

187. R. C. Waller, K. C. Bass, and W. E. Roseveare, . *Ind. Eng. Chem.* **40**, 138 (1948).

188. A. J. Stamm, *Ind. Eng. Chem.* **48**, 413 (1956).

189. M. J. Antal, Jr., H. L. Friedman, and F. E. Rogers, *Comb. Sci. and Tech.* **21**, 141 (1980).

190. O. P. Golova and R. G. Krylova, *Dokl. Akad. Nauk SSR* **135**, 1391 (1960).

191. J. W. Lyons, *The Chemistry and Uses of Fire Retardants*, (Wiley-Interscience, New York, 1970).

192. F. H. Newth, *Advances in Carbohydrate Chemistry* **6**, 83 (1951).

193. A. Basch et al., *Cell. Chem. Tech.* **5**, 353 (1973).

194. A. Basch, B. Hirschman, and M. Lewin, *Cell. Chem. Tech.* **7**, 255 (1973).

195. B. Schaffer, A. Basch, and M. Lewin, *Proc. 6th Intl. Wool Text. Conf.* **5**, 267 (1980).

196. A. Basch and M. Lewin, *Text. Res. J.* **43**, 693 (1973).

197. A. Basch and M. Lewin, *Text. Res. J.* **45**, 246 (1975).

198. M. J. Antal, Jr., in *Proceedings, Bio-Energy '80* (BioEnergy Council, Washington, D. C., 1980).

199. D. W. Goheen, D. W. Glennie, private communication.

200. D. W. Goheen, in *Lignins* edited by K. V. Sarkanen and C. H. Ludwig (Wiley-Interscience, New York, 1971).

201. D. W. Goheen, in *Organic Chemicals from Biomass* edited by I. S. Goldstein (CRC Press, 1981).

202. G. G. Allan and T. Matilla, in *Lignins* edited by K. V. Sarkanen and C. H. Ludwig (Wiley-Interscience, New York, 1971).

203. E. Adler, *Wood Sci. Technol.* **2**, 169 (1977).

204. S. Drew, AIChE Symposium Series *181* **79**, 21 (1978).

205. T. Nguyen, E. Zavarin, and E. M. Barral, III, *J. Macromol. Sci. Rev. Macromol. Chem.* **C20**, 1 (1981).

206. R. W. Coughlin, E. Avni, and D. W. Sundstrom, *Conversion of Lignin to Useful Products*, in press.

207. W. Sandermann and H. Augustin, *Holz. Roh-Werkstoff* **21**, 305 (1963).

208. W. Sandermann and H. Augustin, *Holz. Roh-Werkstoff* **22**, 377 (1964).

209. M. V. Ramiah and D. A. I. Goring, *Cellul. Chem. Technol.* **1**, 277 (1967).

210. H. Hatakeyama et al., *Tappi* **52**, 1724 (1969).

211. H. Hatakeyama and J. Nakano, *Tappi* **53**, 472 (1970).

212. H. Hatakeyama, K. Kubota, and J. Nakano, *Cellul. Chem. Technol.* **6**, 521 (1972).

213. H. Hatakeyama et al., *Mokuzai Gakkaishi.* **21**, 618 (1975).

214. T. Hatakeyama, K. Nakamura, and H. Hatakeyama, *Polymer* **19**, 593 (1978).

215. S. Hirose, T. Hatakeyama, and H. Hatakeyama, *Cellul. Chem. Technol.* **12**, 713 (1978).

216. G. Domburgs et al., *Khim. Drev.* **10**, 97 (1971).

217. J. O. Lephardt and R. A. Fenner, . *Polym. Prepr.* **22**, 282 (1981).

218. R. A. Fenner and J. D. Lephardt, *J. Agric. Food Chem.* **29**, 846 (1981).

219. G. Domburgs, V. N. Sergeeva, and G. Zeibe, *Khim. Drev.* **7**, 59 (1971).

220. H. F. J. Wenzl, *The Chemical Technology of Wood* (Academic Press, New York, 1970).

221. J. Gierer and S. Y. Lin, *Sv. Papperstidn.* **75**, 239 (1972).

222. N. S. Hon, *J. Polym. Sci., Polym. Chem. Ed.* **13**, 2641 (1975).

223. F. Martin, C. Saiz-Jimenez, and F. J. Gonzalez-Vila, *Holzforschung* **33**, 210 (1979).

224. G. Domburgs, V. N. Sergeeva, and A. Kalnins, *Therm. Anal. Proc. Int. Conf., 3rd 1971* **3**, 327 (1972).

225. T. L. Fletcher and E. E. Harris, *Tappi* **35**, 536 (1952).

226. M. T. Klein and P. S. Virk, Report, MIT-EL-81-005 (1981).

227. A. Gillet and J. Urlings, *Chimie & Industrie.* **67**, 909 (1952).

228. T. J. J. Elder, Ph.D. thesis, Texas A & M University (1979).

229. R. J. Dimler, *Adv. Carbohyd. Chem.* **7**, 37 (1952).

230. G. Domburgs, I. Kirsbaums, and V. N. Sergeeva, *Khim. Drev.* **7**, 51 (1971).

231. G. Domburgs et al., *Khim. Drev.* **10**, 97 (1971).

232. G. Domburgs, V. N. Sergeeva, and Z. Trops, *Khim. Drev.* **7**, 55 (1971).

233. G. Domburgs, V. N. Sergeeva, and G. Zeibe, *Khim. Drev.* **9**, 119 (1971).

234. G. Domburgs et al., *Khim. Drev.* **14**, 109 (1973).

235. G. Domburgs, G. A. Rossinskaya, and G. Dobele, *Khim. Drev.* **4**, 87 (1975).

236. G. Domburgs, G. A. Rossinskaya, and G. Dobele, *Khim. Drev.* **5**, 103 (1975).

237. G. Domburgs et al., in *Khim. Ispol'z. Lignina* edited by V. N. Sergeeva (Rigga., USSR, 1974).

238. G. A. Rossinskaya and G. Domburgs, *Khim. Drev.* **4**, 95 (1975).

239. G. Domburgs, V. N. Sergeeva, and A. N. Popov, *Khim. Drev.* **6**, 133 (1970).

240. G. Domburgs, *Latv. PSR Zinat. Akad. Vestis, Kim. Serv.* **6**, 734 (1974).

241. G. Domburgs et al., in *Khim. Ispol'z. Lignina* edited by V. N. Sergeeva (Rigga., USSR, 1974).

242. G. Domburgs, I. Kirsbaums, and V. N. Sergeeva, *Khim. Drev.* **7**, 43 (1971).

243. G. Domburgs et al., *Khim. Drev.* **1**, 51 (1974).

244. L. K. Dubova et al., *Khim. Drev.* **2**, 83 (1981).

245. G. Domburgs et al., *Khim. Drev.* **15**, 101 (1974).

246. T. Arima, *Mokuzai Gakkaishi* **19**, 443 (1973).

247. R. Ceylan and J. B. Bredenberg, *Fuel* **61**, 377 (1982).

248. M. T. Klein and P. S. Virk, *Ind. Eng. Chem. Fundam.* **22**, 35 (1983).

249. K. J. Freudenberg and A. C. Neish, *Constitution and Biosynthesis of Lignin* (Springer, New York, 1968).

250. V. I. Savinykh et al., *Khim. Drev.* **5**, 100 (1975).

251. V. I. Savinykh et al., *Khim. Drev.* **3**, 91 (1976).

252. A. H. Cutler, M. J. Antal, Jr., and M. Jones, submitted to *Ind. Eng. Chem. Fundam.*

253. S. S. Kim, M. L. Jarand, and K. Durai-Swamy, *Fuel* **61(2)**, 1124 (1982).

254. Y. S. Stein, M. J. Antal, Jr., and M. Jones, *J. Anal. and Appl. Pyrol.* **4**, 283 (1983).

255. F. Shafizadeh, R. A. Susott, and G. D. McGinnis, *Carbohyd. Res.* **22**, 63 (1972).

256. F. Shafizadeh, Y. Z. Lai, and R. A. Susott, . *Carbohyd. Res.* **25**, 387 (1972).

257. F. Shafizadeh, M. H. Meshreki, and R. A. Susott, *J. Org. Chem.* **38**, 1190 (1973).

258. Y. Z. Lai and F. Shafizadeh, *Carbohyd. Res.* **38**, 177 (1974).

259. B. Kosikova, D. Joniak, and M. Kosk, *Cellul. Chem. Technol.* **9**, 61 (1975).

260. B. Kosikova et al., *Cellul. Chem. Technol.* **12**, 657 (1978).

261. B. Kosikova et al., *Cellul. Chem. Technol.* **12**, 665 (1978).

262. M. T. Klein and P. S. Virk, Preprints of Div. Fuel Chem. Am. Chem. Soc., Vol. 26, p. 77 (1981).

263. D. L. Urban and M. J. Antal, Jr., *Fuel* **61**, 799 (1982).

264. H. H. King et al., Preprints of Div. Fuel Chem. Am. Chem. Soc., Vol. 28, p. 319 (1983).

265. E. Avni et al., Preprints of Div. Fuel Chem. Am. Chem. Soc., Vol. 28, p. 307 (1983).

266. J. H. Flynn, *Thermochim. Acta.* **37**, 225 (1980).

267. B. Iatridis and G. R. Gavalas, *Ind. Eng. Chem. Prod. Res. Dev.* **18**, 127 (1979).

268. D. W. Goheen and J. T. Henderson, *Cellulose Chem. Technol.* **12**, 363 (1978).

269. R. W. Chan and B. B. Krieger, *J. Appl. Polym. Sci.* **26**, 1533 (1981).

270. H. Traitler and K. Kratzl, *JAOCS*, 153 (1980).

271. H. Traitler, E. Lorbeer, and K. Kratzl, *JAOCS*, 335 (1980).

272. W. J. Connors et al., *Holzforschung.* **34**, 29 (1980).

273. D. W. Goheen, *Advan. Chem. Ser.* **59**, 205 (1966).

274. K. M. Sprouse and M. D. Schuman, *Combust. Flame* **43**, 265 (1981).

275. L. F. Hawley, *Wood Distillation* (Chemical Catalog Co., Inc., New York, 1923).

276. A. W. Goos, *Wood Chemistry* **2**, ACS Nomograph Series No. 97, 2nd ed. (1952).

277. E. S. Lipinsky, *Science* **212**, 465 (1981).

278. Y. Fahmy, *Cellul. Chem. Technol.* **16**, 347 (1982).

279. T. K. Ng et al., *Science* **219**, 733 (1983).

280. G. I. Sorokina and V. S. Petrov, *Izv. Vyssh. Uchebn. Zaved., Lesn. Zh.* **6**, 90 (1981).

281. P. R. Blackenhorn et al., *Wood Science* **2**, 90 (1981).

282. D. Mackay and P. V. Roberts, *Ext. Abstr. Program— Bienn. Conf. Carbon* **15**, 218 (1981).

283. D. M. Mackay and P. V. Roberts, *Carbon* **20**, 87 (1982).

284. S. M. Krutov and V. E. Kovalev, *Gidroliz. Lesokhim. Prom.* **21**, 22 (1968).

285. M. Dalbina et al., *Khim. Drev.* **2**, 78 (1980).

286. M. Dalbina et al., *Khim. Drev.* **5**, 99 (1980).

287. K. Beall and D. Duncan, in *Proc.—Bioenergy R & D Semin. 2nd*, 147 (1980).

288. C. Roy et al., in *Proc.—Bioenergy R & D Semin. 3rd*, 191 (1981).

289. M. Miura et al., *Mokuzai Gakkaishi* **28**, 649 (1982).

290. M. Miura, H. Nishizaki, and H. Kaga, *Mokuzai Gakkaishi* **28**, 727 (1982).

291. M. Kosik and V. Reiser, *Holztechnolgie* **14**, 179 (1973).

292. V. Karlivans et al., *Gidroliz. Lesokhim. Prom-st.* **3**, 6 (1981).

293. D. P. Barnes, P. R. Blankenhorn, and W. K. Murphey, *Wood Science* **12**, 122 (1979).

294. T. W. Mattocks, MSE thesis, Princeton University, 1981.

295. M. J. Antal, Jr., in *Fundamentals of Thermochemical Biomass Conversion* edited by R. P. Overend, T. A. Milne and L. K. Mudge (Elsevier, London, 1985).

296. R. A. Susott, W. F. DeGroot, and F. Shafizadeh, *Journal of Fire and Flammability* **6**, 311 (1975).

297. F. Shafizadeh, P. P. S. Chin, and W. F. DeGroot, *Forest Science* **23**, 81 (1977).

298. R. A. Susott, F. Shafizadeh, and T. W. Aanerud, *Journal of Fire and Flammability* **10**, 94 (1979).

299. R. A. Susott, *Forest Science* **28**, 404 (1982).

300. F. Beall, Ph.D. thesis, Syracuse University, 1968.

301. A. Broido and M. Weinstein, in *Thermal Analysis Proceedings 3rd ICTA*, edited by Hans G. Wiedemann (Birkhauer, Basel, 1971) p. 285.

302. H. W. Eickner, *Forest Prod. J.* **12**, 194 (1962).

303. F. C. Beall and H. W. Eickner, *Thermal Degradation of Wood Components. A Review of the Literature* (U. S. D. A. Forest Serv. Res. Pap. FPL130, 1970).

304. A. F. Roberts, *Flame* **14**, 261 (1970).

305. J. A. Havens, Ph.D. thesis, University of Oklahoma, 1969.

306. J. A. Havens, J. R. Welker, and C. M. Sliepcevich, *J. Fire and Flammability* **2**, 321 (1971).

307. K. Akita, *Rep. Fire Res. Inst.* **9**, 10 Japan (1956).

308. A. J. Stamm, *Ind. Eng. Chem.* **48**, 413 (1956).

309. D. F. Arseneau, *Can. J. Chem.* **39**, 1915 (1961).

310. D. Radislav and R. Frantisek, *Holz. Roh-Werkstoff* **20**, 473 (1962).

311. W. K. Tang, U. S. Dept. Agriculture Forest Serv. Res. Paper, FPL 71 (1967).

312. F. Shafizadeh and W. F. DeGroot, in *Fuels and Energy from Renewable Resources* edited by F. Shafizadeh, K. Sarkanen, and D. Tillman (Academic Press, New York, 1977).

313. V. N. Sergeeva and A. Valvads, *Latvijas PSR Zinatnu Akad. Vestis* **86**, 102 (1954).

314. R. Keylwerth and N. Christoph, *Materialpruf* **2**, 281 (1960).

315. F. C. Beall, *Wood Science and Technology* **5**, 159 (1971).

316. A. F. Roberts, in *13th International Symp. of Combustion* (The Combustion Institute, Pittsburgh, Pa., 1971) p. 893.

317. F. Shafizadeh and G. D. McGinnis, *Carbohyd. Res.* **16**, 273 (1971).

318. M. S. Duvvuri et al., *J. Fire and Flammability* **6**, 468 (1975).

319. T. W. Mattocks, MSE thesis, Princeton University, 1981.

320. F. Thurner, U. Mann and S. R. Beck, *Report*, DOE/ET/20041-T2 (1980).

321. A. F. Roberts, *Combustion and Flame* **17**, 79 (1971).

322. V. Hornof et al., *Thermochimica Acta.* **19**, 63 (1977).

323. F. Shafizadeh et al., *J. Org. Chem.* **36**, 2813 (1971).

324. F. Shafizadeh, G. D. McGinnis and C. W. Philpot, *Carbohyd. Res.* **25**, 23 (1972).

325. F. Shafizadeh, R. A. Susott, and G. D. McGinnis, *Carbohyd. Res.* **22**, 63 (1972).

326. F. Shafizadeh, Y. Z. Lai, and R. A. Susott, *Carbohyd. Res.* **25**, 387 (1972).

327. F. Shafizadeh, M. H. Meshreki, and R. A. Susott, *J. Org. Chem.* **38**, 1190 (1973).

328. F. Shafizadeh and Y. Z. Lai, *Carbohyd. Res.* **31**, 57 (1973).

329. Y. Z. Lai and F. Shafizadeh, *Carbohyd. Res.* **38**, 177 (1974).

330. G. Domburgs, T. N. Skripchenoko, and I. Z. Kirshbaum, in *Therm. Anal., Tezisy Dokl., Soveshch. 7th* edited by V. Karlivan and P. Zinatne (Riga, USSR, 1979).

331. A. F. Roberts and G. Clough, in *9th Symposium on Combustion* (Academic Press, New York, 1963).

332. A. M. Kanury and P. L. Blackshear, Jr., *Pyrodynamics* **4**, 285 (1966).

333. J. A. Havens et al., *Combustion Science and Technology* **5**, 91 (1972).

334. C. Fairbridge, R. A. Ross, and P. Spooner, *Wood Sci. Technol.* **9**, 257 (1975).

335. D. Q. Tran and C. Rai, *Fuel* **57**, 293 (1978).

336. C. Fairbridge, R. A. Ross, and S. P. Sood, *J. Appl. Polym. Sci.* **22**, 497 (1978).

337. R. D. Cardwell and P. Luner, *Tappi* **61**, 81 (1978).

338. E. Chornet and C. Roy, *Thermochimica Acta.* **35**, 389 (1980).

339. A. Kumar and R. S. Mann, in *Proc.—Bio-energy R & D Semin. 3rd*, 231 (1981).

340. A. Kumar and R. S. Mann, *J. Anal. Appl. Pyrolysis* **4**, 219 (1982).

341. F. Mobarak, Y. Fahmy, and W. Schweers, *Wood Sci. Technology* **16**, 59 (1982).

342. J. N. Foussard, B. Talayrach, and J. Besombes Vailhe, *Afinidad* **36**, 335 (1979).

343. M. A. Hamad, *J. Chem. Tech. Biotechnol.* **31**, 624 (1981).

344. J. P. Diebold and G. D. Smith, *Noncatalytic Conversion of Biomass to Gasoline*, 79-Sol-29 (American Society of Mechanical Engineers, 1978).

345. J. P. Diebold and J. Scahill, in SERI/PR-234-1456 (1982).

346. J. P. Diebold and J. Scahill, in *14th Biomass Thermochemical Conversion Contractors' Review Meeting*, SERI/TP-234-1654 (1982).

347. J. Lédé et al., *Rev. Phys. Appl.* **15**, 545 (1980).

348. J. Lédé and J. Villermaux, *Recherche* **13**, 786 (1982).

349. M. Bohn and C. Benham, in *Proc. Specialists Workshop on Fast Pyrolysis of Biomass*, SERI/CP-622-1096 (Solar Energy Research Institute, Golden, Colo., 1980).

350. M. Bohn and C. Benham, SERI/TP-252-1398 (Solar Energy Research Institute, Golden, Colo., 1982).

351. T. A. Milne and M. N. Soltys, Quarterly Report, SERI/PR-234-1923 (Solar Energy Research Institute, Golden, Colo., 1982).

352. X. Deglise et al., *Comm. Eur. Communities* in 1st Energy Biomass Conf. EUR:7091:548.

353. S. Caubet et al., *Solar Energy* **29**, 565 (1982).

354. D. S. Scott and J. Piskorz, *Canadian Journal of Chem. Eng.* **60**, 666 (1982).

355. T. Reed, *A Survey of Biomass Gasification* (Noyes Data Corp., Park Ridge, N. J., 1981).

356. E. Rensfelt et al., in *Energy from Biomass and Wastes* edited by D. E. Klass (Washington, D. C., 1978).

357. J. R. Richard, M. Cathonnet, and J. P. Rouan, *Gasification of Charcoal Influence of Water Vapor*, to appear.

358. D. S. Scott and J. Piskorz, *Canadian Journal of Chem. Eng.* **60**, 666 (1982).

359. C. Roy and E. Chornet, to appear.

360. C. Roy et al., to appear.

361. W. C. Chan, M. Kelbon, and B. B. Krieger, in *Proceedings, Fundamentals of Thermochemical Biomass Conversion: An International Conference* (Estes Park, Colo., 1982).

362. R. Chan et al., presented at Western States Combustion Inst. Meeting (1981).

363. R. Chan and B. B. Krieger, in *Chemical Reaction Engineering* (American Chemical Society, Washington, D. C., 1982).

364. M. Steinberg and P. Fallon, in *Proceedings of the 12th Biomass Thermochemical Conversion Contractors' Meeting* (Washington, D. C., 1981).

365. M. Steinberg and P. Fallon, in *Proceedings of the 13th Biomass Thermochemical Conversion Contractors' Meeting*, PNL-SA-10093 (Arlington, Va., 1981).

366. P. C. Stangeby and S. N. Basu, in *Proc.—Bio-energy R & D Seminar*, 2nd, 239 (1980).

367. S. N. Basu and P. C. Stangeby, in *Proc.—Bio-energy R & D Seminar*, 3rd, 257 (1981).

368. M. Steinberg, P. Fallon, and M. S. Sundaram, in *Proceedings, Fundamentals of Thermochemical Biomass Conversion: An International Conference*, Estes Park, Colo (1982).

369. C. W. Philpot, *Forest Sci.* **16**, 461 (1970).

370. C. W. Philpot, USDA Forest Service, Research Note INT-139 (1971).

371. X. Deglise et al., in *Energy from Biomass, 1st E. C. Conference* edited by W. Palz, P. Chartier, and D. O. Hall (Applied Science Publishers, London, 1980).

372. X. Deglise et al., *Rev. Gen. Therm. Fr.* **19**, 871 (1980).

373. L. K. Mudge, L. J. Sealock, Jr., and S. L. Weber, *J. Anal. Appl. Pyrolysis* **1**, 165 (1979).

374. D. H. Mitchell, L. K. Mudge, and E. G. Baker, in *Proceedings of the 13th Biomass Thermoconversion Contractors' Meeting*, PNL-SA-10093 (Arlington, Va., 1981).

375. P. Fong and R. A. Ross, *Wood Sci.* **13**, 87 (1980).

376. P. Fong and R. A. Ross, *Applied Energy* **9**, 211 (1981).

377. R. A. Ross and P. Fong *Ind. Eng. Chem. Prod. Res. Dev.* **20**, 197 (1981).

378. A. Rolin et al., *J. Anal. Appl. Pyrolysis* **5**, 151 (1983).

379. M. J. Antal, Jr., *Method and Apparatus for Producing Synthetic Fuels from Solid Wastes*, U. S. Patent 3,993,458.

380. R. C. Feber and M. J. Antal, Jr., Environmental Protection Technology Series, EPA-600/2-77-147 (1977).

381. M. J. Antal, Jr., *Thermochemical Conversion of Biomass: The Scientific Aspects*, NTIS DE 081904166 (PB-81-134793, 1980).

382. F. Shafizadeh and T. T. Stevenson, *J. Appl. Polym. Sci.* **27**, 4577 (1982).

383. O. P. Golova, *Russian Chemical Reviews* **44**, 687 (1975).

384. T. B. Reed, "Principles and technology of biomass gasification," *Advances in Solar energy* edited by K. W. Böer and J. A. Duffie (Plenum Publishing Corp., New York, 1985).

385. J. L. Jones, in *Solid Wastes and Residues* edited by J. L. Jones and S. B. Radding (American Chemical Society, Washington, D. C., 1978).

386. R. F. Probstein and R. E. Hicks, *Synthetic Fuels* (McGraw-Hill, New York, 1982).

387. B. Guthrie, *Colorado School Mines Quart.* **3**, 7 (1964).

388. G. L. Baughman, *Synthetic Fuels Handbook*, 2nd ed. (Cameron Engineers, Inc., Denver, Colo., 1978).

389. M. J. Antal, Jr., in *Energy from Biological Processes. Appendices* **3**, NTIS/DE 081904166 (PB-81-134793, 1980).

390. Office of Technology Assessment, "An assessment of oil shale technologies," 052-003-00759-2 (1980).

391. E. M. Wilson et al., *Engineering and Economic Analysis of Waste to Energy Systems*, NTIS/EPA-600/7-78-086 (1978).

392. T. E. Ctvitnicek, R. J. McCormick, and A. Wojtowicz, *Technical, Environmental and Economic Feasibility of Multiwaste Gasification* (U. S. E. P. A, Monsanto Research Corp., Dayton, 1978).

393. Dow Chemical, U. S. A. *Technical, Econonmic and Environmental Feasibility of China Lake Pyrolysis System* (U. S. E. P. A, Freeport, 1978).

394. R. D. Mikesell, D. E. Garrett, and D. C. Hoag, in *Solid Wastes and Residues* edited by J. L. Jones and S. B. Radding (American Chemical Society, Washington, D. C., 1978).

395. J. Diebold and G. Smith, *Conversion of Trash to Gasoline*, NWC Tech. Pub. 6022 (1978).

396. E. Epstein, H. Kosstrin, and J. Alpert, in *IGT's Symposium on Energy from Biomass and Wastes* (IGT, Washington, D. C., 1978).

397. R. O. Williams et al., in *Solid Wastes and Residues* edited by J. L Jones and S. B. Radding (American Chemical Society, Washington, D. C., 1978).

398. R. Bailie and C. A. Richmond, in *Solid Wastes and Residues* edited by J. L. Jones and S. B. Radding, (American Chemical Society, Washington, D. C., 1978).

399. J. A. Coffman, in *Proceedings of the 14th Biomass Thermochemical Conversion Contractors' Meeting*, CONF-820685/PNL-SA-10646 (Arlington, Va., 1982).

400. *Biomass Gasification: Principles and Technology* edited by T. B. Reed, Energy Technology Review No. 67 (Noyes Data Corp., Park Ridge, N. J., 1981).

401. J. Diebold and J. Scahill, *Ablative Pyrolysis of Biomass in the Entrained Flow Cyclonic Reactor at SERI*, SERI/PR-234-1883 (Solar Energy Research Institute, Golden, Colo., 1983).

402. J. Diebold, in *Proceedings of the Specialists' Workshop on Fast Pyrolysis of Biomass*, SERI/CP-622-1096 (Solar Energy Research Institute, Golden, Colo., 1980).

403. J. Lédé et al., in *Proceedings of the Specialists' Workshop on Fast Pyrolysis of Biomass*, SERI/CP-622-1096 (Solar Energy Research Institute, Golden, Colo., 1980).

404. V. S. Kothari, MSE thesis, Princeton University, 1980.

405. V. S. Kothari and M. J. Antal, Jr., Division of Fuel Chemistry preprints, Vol. 28, p. 5 (American Chemical Society, Washington, D. C., 1983).

406. M. J. Antal, Jr. et al., *Solar Energy* **30**, 299 (1983).

407. M. W. Hopkins, MSE thesis, Princeton University, 1982.

408. M. W. Hopkins, C. I. DeJenga, and M. J. Antal, Jr., *Progress in Solar Energy*. In Press.

409. M. W. Hopkins, C. I. DeJenga, and M. J. Antal, Jr., *Solar Energy* **32**, 547 (1984).

410. L. Hoffman, MSE thesis, Princeton University, 1981.

411. L. Hofmann and M. J. Antal, Jr., *Solar Energy*. In press.

412. C. K. DeJenga, M. J. Antal, Jr., and M. Jones, *J. Applied Polym. Sci.* **27**, 4313 (1982).

413. T. B. Reed, in *Proceedings of the 13th Biomass Thermochemical Conversion Contractors' Meeting*, CONF 8110115/PNL-SA-10093 (Arlington, Va., 1981).

414. L. Waldheim and E. Rensfelt, in *Proceedings of the 11th Biomass Thermochemical Conversation Contractors' Meeting* (Richland, Wa., 1980).

415. P. McGeer and E. Durbin, *Methane Fuel for the Future* (Plenum Press, New York, 1982).

416. Myles Thompson, private communication.

417. R. C. Bailie, in *Proceedings of the 14th Biomass Thermochemical Conversion Contractors' Meeting*, CONF-820685/PNL-SA-10646 (Arlington, Va., 1982).

418. H. F. Feldman, M. A. Paisley, and H. R. Appelbaum, in *Proceedings of the 14th Biomass Thermochemical Conversion Contractors' Meeting*, CONF-820685/PNL-SA-10646 (Arlington, Va., 1982).

419. L. K. Mok et al., Forest Products Research Society Annual Forum, Washington, D. C., 1982.

CHAPTER 5

THERMAL COMFORT AND PASSIVE DESIGN

S.V. Szokolay

Is it pleasant to have it cool in summer and warm in winter? Now, in houses with a southern aspect the sun's rays penetrate into the porticoes in winter, but in summer the path of the sun is right over our heads and above the roof, so that there is shade. If then this is the best arrangement, we should build the south side loftier to get the winter sun and the north side lower, to keep out the cold winds. In short: the house in which the dweller can find a pleasant retreat at all seasons and store his belongings safely, is presumably at once the pleasantest and the most beautiful.

(Socrates, as quoted by Xenophon, in Memorabilia, III, VIII, 9.)

5.1 HISTORICAL NOTES

Already Socrates and later Vitruvius had some thoughts on the climatic suitability of buildings, but these had never seriously influenced architects and builders. Up to the Industrial Revolution, thermal comfort was largely an academic question and, as such, rarely raised. When it was cold, a fire was lit to ameliorate the conditions. The potential of the available controls was the limiting factor, rather than the risk of overheating. When it was hot, the use of hand-operated fans was about the only possibility for relief. Only in exceptional situations was the heat storage capacity of caves used for cooling, or—in some other cultures—man-made tunnels and wind towers used for a similar purpose. In none of these cases was it necessary to specify the desirable environmental conditions.

Nevertheless, environmental temperatures were already considered in the late eighteenth century. Heberden[1] recognized that factors other than air temperature have an influence on thermal sensation. The first serious study was carried out only at the beginning of this century in England by Haldane.[2]

The development of highly effective heating systems and the appearance of the new air-conditioning equipment made it quite possible to overheat in winter or overcool in summer. It became necessary to define what conditions the designer should aim for. Systematic empirical work began in the early 1920s by Houghten and Yagloglou[3] and was continued by Yagloglou and Drinker.[4] In England, Vernon[5] approached the problem from the direction of occupational hygiene and industrial health. Vernon and Warner[6] as well as Bedford[7] carried out empirical work amongst factory workers. Analytical work began in the U.S. in the mid-1930s and Winslow, Herrington, and Gagge[8] made a significant contribution.

During and after the war, research activity increased to such an extent that the lines of investigation became quite divergent within the framework not only of engineering, but also within many established disciplines from physiology to geography and climatology. It was left to an architect, Victor Olgyay,[9] to bring together the scattered findings and interpret these for practicing users.

Until quite recently we had two extreme working methods:

(a) architectural design normally considered thermal factors in qualitative terms only;

(b) engineering design of thermal control equipment relied on "design temperatures" given in statutory regulations, standards, codes of practice or engineering handbooks (e.g., *IHVE Guide*,[10] *ASHRAE Handbook of Fundamentals*,[11] *Deutsches Institut für Normung* 4701/1959, *Carrier Design Manual*, SI metric ed., 1974[12]).

This rather simplistic approach was justified on the grounds that

(a) the thermal control performance of the building is never accurately guaranteed, the design can only aim at making the indoor conditions "as good as practicable";

(b) plant design is aimed at establishing the maximum required capacity, it can then be regulated down, or cater for the worst conditions so that an overload capacity is normally provided.

Current thinking suggests that the thermal performance of the building can be predicted with reasonable accuracy and the building itself can control the indoor conditions (by passive means) to an extent far greater than previously believed and that any active control systems should work in sympathy with the building, aimed at energy economy. The task of calculation is not only system sizing, but also the prediction of annual energy usage.

The U.S. Office of Emergency Preparedness[13] coined the terms "leak plugging" and "belt tightening," as potential methods of energy conservation. Whilst the former would mean measures such as improved thermal insulation, better system efficiency and reduction of wastage, the latter may mean the "reduction of environmental standards." The Emergency Building Temperature Restrictions[14] did in fact reduce thermostat settings to a maximum of 65°F (18.3°C) for winter and to not less than 78°F (25.5°C) for summer. The ensuing debate renewed the interest in thermal comfort.

5.2 PHYSIOLOGICAL BASIS

The deep tissue temperature of the human body is remarkably constant at about 98.5°F (37°C), while the skin temperature can vary between 88°F (31°C) and 93°F (34°C) under comfort conditions. Variations can occur in time but also at any given time between various parts of the body, depending on the degree of cover provided by clothing and on the local blood circulation rate. Living tissue constantly produces heat. This metabolic heat production can be of two kinds:

(a) basal metabolism, the heat production of biological processes which are continuous and unconscious;

(b) muscular metabolism, the heat production of muscles whilst carrying out work. This is consciously controllable (except in shivering). Thermal equilibrium exists when the heat produced is fully dissipated to the environment. The heat produced in deep body tissues in continuously transported to the skin surface (and lungs) from where it is emitted to the environment by convection, radiation, and evaporative heat transfer (possibly also by conduction, if in contact with cooler solid bodies). Table 5.9 shows average metabolic rates at various activity levels.

If, due to an increase in metabolic heat production or because of a change in environmental conditions, the heat produced is not fully dissipated, the equilibrium is disturbed, and some heat goes into storage in the body and elevates its temperature. The maximum temperature increase the body would normally tolerate is 2.5°F (1.4°C). In order to restore the balance, the vasoregulatory mechanism is activated. Blood vessels near the surface are dilated (vasodilation). Thus, the heat transport to the surface is increased, the skin temperature is elevated, and both radiant and convective losses are increased.[15]

When vasodilation is insufficient to restore or maintain a thermal equilibrium, the sweat glands are activated, thus, the evaporative cooling mechanism is brought into action. Sweat can be produced for short periods at a rate of up to approximately 4 L per hour but the mechanism is fatigable.[16] The sustainable rate is about 1 L per hour. The evaporation of moisture is an endothermic process, absorbing some 1030 Btu heat from the environment for every pound evaporated (2400 kJ/kg).

Short-term adjustment is achieved in 20 to 30 minutes, but there are also long-term, endocrine adjustments extending beyond six months which constitute the acclimatization process. Both the vasomotor and the evaporative regulation mechanisms are subject to acclimatization. In hot climates, the volume of blood circulating can be increased by up to 20% to maintain a constant vasodilation. The sweat secretion rate also increases over a period of several weeks. It is believed that the forward section of the hypothalamus regulates these changes through a complex neuroendocrine process.[17]

When all the above control mechanisms are unable to restore thermal equilibrium, hyperthermia (inevitable body heating) occurs. When the deep body temperature rises to about 104°F (40°C), heat stroke may develop. This is a circulatory failure (venous return to the heart is reduced) leading to fainting. Early symptoms

are: fatigue, headache, dizziness when standing, loss of appetite, nausea, vomiting, shortness of breath, flushing of face and neck, pulse rate can reach 150/min, glazed eyes, and mental disturbances including apathy, poor judgment, irritability. Heat cramps or painful muscle spasms can occur much earlier, largely due to loss of salt through sustained sweating. At heat stroke the body temperature rapidly climbs to above 106°F (41°C), sweating stops, coma sets in, and death is imminent. Even if a person is saved at this point, the brain may have suffered irreparable damage. At about 108°F (42°C) death normally occurs.[17-19]

Under cold conditions, when the heat dissipation rate exceeds the heat production rate, the first physiological response is vasoconstriction. Blood vessels near the body surface contract, heat transport from the deep tisses is reduced, the skin temperature is lowered, thus diminishing both convective and radiant heat dissipation. Deep body temperature remains constant at about 98.5°F (37°C) as the insulating value of the tissues is increased. Body extremities, such as fingers, toes, and ears may be starved of blood and may reach a temperature below 68°F (20°C), and in severe exposure may even freeze before deep body temperature would be affected.[18]

Goose pimples (atavistic remains of hair-erection) may be associated with vasoconstriction. With an adequate fur this would increase the insulation of this skin cover. When vasoconstriction cannot sufficiently reduce heat losses, a metabolic regulatory mechanism is activated. Shivering can produce a tenfold increase in metabolic heat production.[15] In acclimatization to cold conditions the thyroid gland is activated and the hormones produced increase the overall rate of metabolism. Full adjustment is achieved in about 21 days.[17] When the physiological control mechanisms fail to restore thermal equilibrium, hypothermia (inevitable body cooling) occurs. The deep body temperature is lowered to below 95°F (35°C). Death occurs between 77°F (25°C) and 86°F (30°C), (except under medically controlled conditions). Even if hypothermia is not reached, continued exposure to cold conditions requiring full operation of vasomotor and metabolic controls can cause mental disturbances (inadequate blood supply to the brain). Willpower is softened and conscious control gives way to hallucinations, drowsiness, and stupor.

5.3 SUMMARY

The limits of existence can be defined in terms of deep body temperature as 95°F and 104°F (35°C and 40°C), the normal being about 98.5°F (37°C). The skin temperature should always be less than these values and the temperature of the environment should be slightly less than the skin temperature in order to allow adequate, but not excessive, heat dissipation. Table 5.1 gives a summary of critical temperatures. The range of environmental conditions, which will allow such adequate, but not excessive, heat dissipation, will allow a "sense of physical well-being," will be judged as comfortable and is referred to as the "comfort zone."

Table 5.1

Critical body temperature (a crude guide)

Skin Temperature	Deep Body Temperature	Physiological Zone
	108°F (42°C)	death
pain: 113°F(45°C)		
	104°F (40°C)	hyperthermia, evaporative regulation, vasodilation
88°–93°F(31–34°C)	98.5°F (37°C)	comfort
		vasoconstriction, metabolic regulation
	95°F (35°C)	hypothermia
pain: 50°C (10°C)		
	77°F (25°C)	death

5.4 EMPIRICAL STUDIES

The question of the definition of the comfort zone in terms of environmental parameters can be approached from two directions, using two different methods:

(1) empirical methods, using social survey techniques, such as comfort votes, under defined environmental conditions and

(2) analytical methods, tracing the heat flow paths from metabolic heat production to the environment and considering resistances to this flow.

The classic example of empirical studies is the work of Houghten and Yagloglou.[20] This led to the recognition of three environmental parameters which contribute to thermal sensation:

(1) air temperature or dry bulb temperature (DBT);

(2) humidity, measured by the wet bulb temperature (WBT), but expressible as relative humidity (RH), or percentage saturation, in %, or as absolute humidity (AH), or moisture content grams of moisture per kg of dry air, or humidity ratio (kg/kg, therefore nondimensional), or as vapor pressure (VP), in any pressure unit. The relationship of these quantities is normally shown on the psychrometric chart;

(3) air movement, measured by its velocity in m/s.

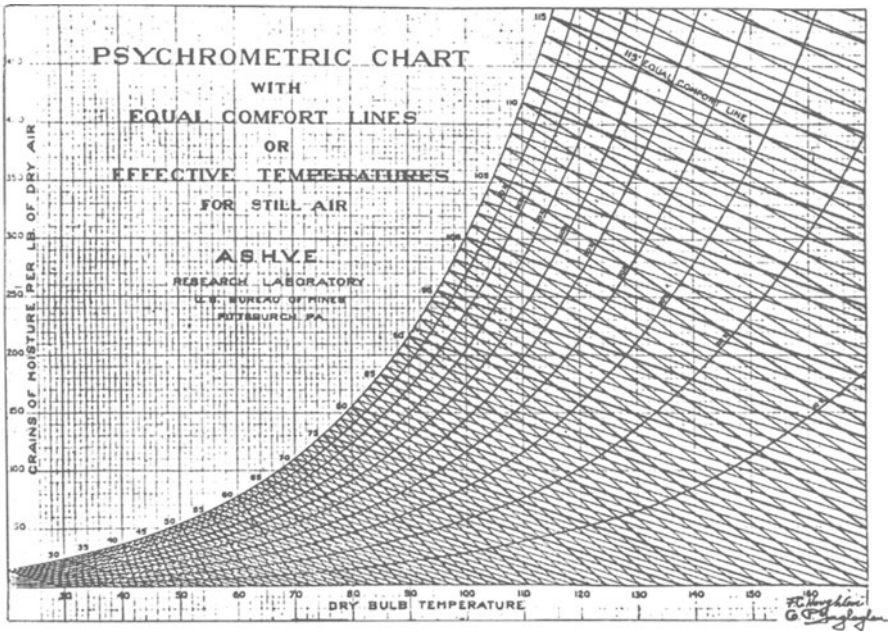


Fig. 5.1: Yagloglou's original representation of the "effective temperature."

In their experimental work they controlled the air velocity at zero examining first the relationship of air temperature and humidity. They constructed a set of "equal comfort lines" superimposed on the psychrometric chart, designated as "effective temperature" lines (ET), as shown in Fig. 5.1. ET has been defined as the temperature of a still, saturated atmosphere, which would, in the absence of radiation, produce the same effect as the atmosphere in question. The same authors went on to determine the comfort zone in terms of this effective temperature. The results were reported by Yagloglou[21] and Yagloglou and Drinker.[4] Other work in the meantime had shown that the relationship of DBT and RH is not quite linear and Bedford[22] published a nomogram for a revised definition of ET (Fig. 5.2). He took into account the fourth environmental parameter:

(4) radiation, measured by using a globe thermometer, and substituting these readings for DBT. The scale thus produced became known as the "corrected effective" temperature (CET).

The same nomogram will give CET values if globe thermometer readings (GT) are used in lieu of DBT. Clothing has been found to have a great influence on people's thermal preferences and on the relative influence of environmental parameters, particularly of wind. The nomogram shown in Fig. 5.2 is valid for people wearing a normal business suit with cotton underwear. This has been taken as a unit of clothing: 1 clo, which corresponds to an average U-value of 1.14 Btu/h

$ft^2°F$ (6.45 in W/m^2K). Bedford[20] also gave an ET nomogram for people stripped to the waist (Fig. 5.3). Yaglou* also revised his original index.[23]

As the four environmental parameters contribute to a singular thermal sensation. there were many attempts to produce a single-figure thermal comfort index expressing the joint effect of all or several of the four parameters. The ET and CET mentioned above are already such indices, but some others, at least the more generally known ones, warrant a brief examination.

The concept of equivalent temperature (EqT) was introduced by Dufton[24,25] and its use is described by Bedford.[26] Its conceptual definition is: the temperature of a thermally uniform enclosure, with still air, in which a sizable black body at 75°F (24°C) would lose heat at the same rate as that observed. Mathematically, the definition is (quoted in original units)

$$EqT = 0.522t_a + 0.478t_w - 0.01474v^{0.5}(100 - t_a) \tag{5.1}$$

where t_a = air temperature (°F), t_w = mean radiant temperature, and v = air velocity (ft/min).

$$EqT = 0.522t_a + 0.478t_g + v^{0.5}(0.0808t_g - 0.0661t_a - 1.474) \tag{5.2}$$

where t_g = globe temperature.

It does not take into account the humidity and thus is unsuitable for temperatures above approximately 75°F (24°C) since at such levels evaporative dissipation of heat becomes important.

The "equivalent warmth" (EqW) scale was devised by Bedford.[7] He used some 2,000 factory workers engaged in light work under varying indoor conditions as his sample. Air temperatures, humidity, and mean radiant temperature (MRT) were measured and recorded, as well as the subjective responses of workers. Surface temperatures of skin and clothing were also measured. Statistical analysis produced the EqW scale. It is now thought to be reliable within the comfort zone and up to 95°F (35°C) with low RH, or up to 86°F (30°C) with high RH. It underestimates the cooling effect of air movement in the case of high humidities.

The "operative temperature" (OT) index was produced by Winslow, Herrington, and Gagge[8] as a result of work very similar to Bedford's. It is defined by the expression

$$OT = \frac{k_r T_w + k_c T_a}{k_r + k_c} \tag{5.3}$$

where k_r and k_c are radiation and convection constants and T_w and T_a are wall surface and air temperatures.

It integrates air temperature and radiation effects, ignoring humidity and air movement. The study was carried out in a cool climate, where humidity was small

* In the intervening period, Yagloglou seems to have shortened his name to Yaglou.

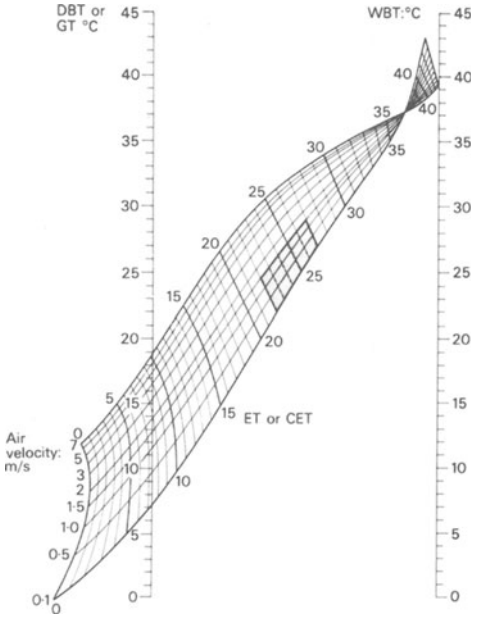


Fig. 5.2: ET nomogram for people wearing 1 clo.

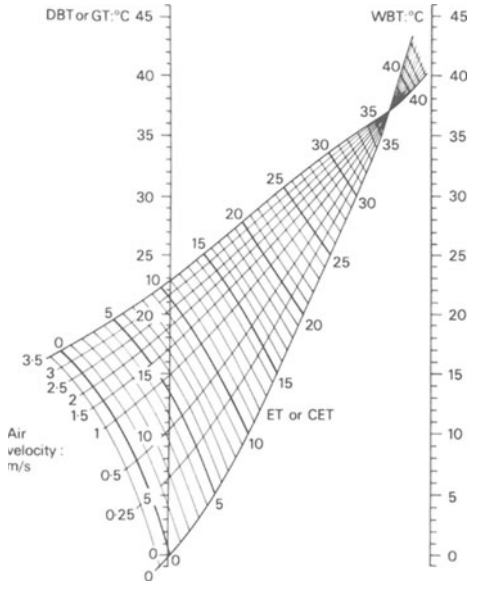


Fig. 5.3: ET nomogram for people stripped to waist.

and indoor air movement negligible. The index is thought to be unsuitable above 80°F (27°C). It is, however, significant in that it is a first attempt to create a rational index, considering at least some of the heat exchange processes between the body and its environment.

The "resultant temperature" (RT) index was developed by Missénard in France.[27] It is a slight improvement on the ET scale. The nomogram defining it is almost identical to the ET nomogram. It is thought to be reliable for moderate climates but not for tropical conditions as it underestimates the cooling effect of air movements at temperatures above 95°F (35°C) and humidities over 80%.[28]

Webb[29,30] produced the "equatorial comfort index" (ECI), working in Singapore and Malaysia. Subjective responses of fully acclimatized subjects were recorded, together with measurements of air temperature, humidity, and air movement. The relationships found were organized to produce a formula and expressed in a graph similar to the ET nomogram.

All seven indices (ET, CET, EqT, OT, RT, and ECI) were developed empirically without attempting to relate the observed effects to physiological causes. They all suffer from the fact that the experimental work was carried out in one particular location, thus they are applicable only under similar cicumstances. The ET scale was based on experiments in test rooms and instant reactions. The EqW scale was developed from tests under normal working conditions. The RT scale was based on comfort votes in test rooms, after 0.5 hour of adjustment. Consequently, any comparison or generalization is difficult.

5.5 ANALYTICAL WORK

All analytical work is based on some variant of the following expression of the body's thermal equilibrium:

$$M \pm Cd \pm Cv \pm R - E = 0 \tag{5.4}$$

where M is metabolic heat production, Cd is conduction heat gain or loss, Cv is convection heat gain or loss, R is radiation heat gain or loss, and E is evaporation heat loss.

Thermal equilibrium exists when the sum total of heat inputs and outputs is zero. For a given metabolic rate and environmental conditions, the Cd, Cv, and R terms can be modified partly by a change in clothing, partly by vasomotor adjustments. When such adjustments fail to produce sufficient heat dissipation, the low level of E (due to imperceptible perspiration) can be increased quite substantialy by sweating, thus a zero sum can still be maintained. Thermal equilibrium, however, does not necessarily mean thermal comfort. Most authors agree that the onset of sweating indicates the beginning of thermal stress. A number of indices have been developed to measure the magnitude of such stress.

The "thermal strain index" (TSI) was developed by Lee[31] partly on the basis of observation, partly by analyzing the heat transfer mechanisms. He plotted a

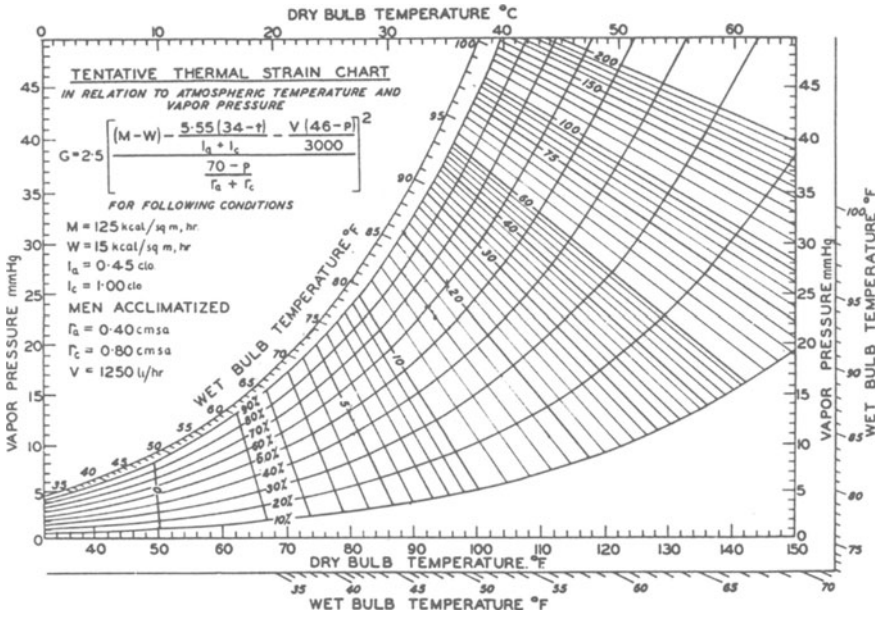


Fig. 5.4: Lee's equal strain lines on the psychrometric chart.

set of "equal strain lines" on the psychrometric chart (Fig. 5.4). At high levels of strain these are almost parallel with the WBT lines, whilst at low levels they are vertical, i.e. coincide with th DBT lines. He also developed an equation to calculate this index from temperature, humidity, air velocity, metabolic rate, and clothing data. McArdle and his collaborators[16] and later Smith[32] reported on work carried out initialy for the British naval authorities. The intention was to determine the physical stress in objective terms on the basis of sweat rate, pulse, and internal body temperature. The "predicted 4-hour sweat rate" (P4SR) scale was developed on the basis of many different combinations of air temperature, humidity, air movement, mean radiant temperature, metabolic rate, and amount of clothing worn, producing the same sweat rate, thus—presumably—the same degree of physiological stress. It seems to be the most reliable index for high temperature conditions but not suitable for temperatures below 82°F (28°C). The cooling effect of air movement at high humidities is underestimated.[33]

The "heat strain index" (HSI) was devised by Belding and Hatch.[34] It is defined as the ratio

$$HSI = \frac{\text{evaporative cooling required}}{\text{max. evaporative cooling possible}} \tag{5.5}$$

This can be expressed in terms of metabolic rates, air and wall temperatures, air movement, and vapor pressure. The theoretical formulation was corrected by experimental findings where the metabolic rate of subjects doing various kinds of

work was measured as well as the environmental parameters. The scale is thought to be reliable for still air between 81°F and 95°F (27°C and 35°C), 30-80% RH, and for higher temperatures with low humidities, but it exaggerates the effect of air movement at low humidities and the effect of high humidities at medium to high temperatures. It is not suitable for the comfort zone or for conditions below the comfort zone.

Lee and Henschel,[35] working for the U.S Department of Health, have further developed the work of Belding and Hatch and produced the "relative strain index" (RSI), defined (quoted in original units) as

$$RSI = \frac{M(I_{cw} + I_a) + 5155(t_a - 35) + RI_a}{7.5(44 - P_a)} \tag{5.6}$$

where M is metabolic rate (kcal/m^{2}h), t_a is air temperature (°C), R is mean radiant energy incidence (kcal/m^{2}h), I_a is insulation of air (clo units), I_{cw} is insulation of wet clothing (clo units), and P_a is vapor pressure of air (mmHg).

In the 75°F to 81°F (24°C to 27°C) range, the RSI values closely agree with effective temperatures (ET). At higher temperatures in the high humidity range, the ET scale underestimates the stress compared with RSI. Values of RSI can be interpreted as follows—below 0.1: all comfortable; at 0.2: 85% comfortable, 15% too warm; at 0.25: 50% comfortable, 50% too warm; at 0.3: none comfortable, 50% too hot; at 0.4: 75% show distress, some failure; at 0.5: all too hot, all show signs of distress; above 0.5: too hot to endure.

Givoni, working during the same period, published his new "index of thermal stress" (ITS) in 1963[36] (later included in his book.[28]) This is the calculated cooling rate produced by sweating which would maintain thermal balance under the given conditions. The calculation is based on a refined biophysical model of the man-environment thermal system. It takes into account all objective (environmental) and subjective factors. The general expression is (quoted in original non-SI metric units)

$$ITS - [(M - W) \pm C \pm R](1/f) \tag{5.7}$$

where ITS = required sweat rate in equivalent kcal/h; M = metabolic rate, (kcal/h); W = part of the above transferred to mechanical work (kcal/h); C = convective heat exchange (kcal/h); R = radiant exchange (kcal/h); and f = cooling efficiency of sweating (dimensionless).

Each term is evaluated by complex anaytical formulas.

Whilst there has been a proliferation of comfort indices, there was also a large amount of work on thermal comfort which did not lead to any single figure indices. Only a few examples will be mentioned.

Drysdale conducted a series of studies between 1948 and 1953 measuring the skin impedance to indicate the onset of sweating.[37] As a result he suggested the "discomfort limits" shown in Table 5.2.

Table 5.2

Environmental conditions at onset of discomfort

Humidity	Moderate (vp.13 mmHg)			High (vp.22 mmHg)		
Air Movement (ft/min)	10	75	200	10	75	200
DBT (°F)	88	89	95	86	88	94

Morse and Kowalczewski[38] produced a thermal comfort equation to calculate the required air temperature by tracing the heat flow processes involved (quoted in original units)

$$T_a = T_s - \frac{M - E_o}{A}\left(\frac{1}{h_r + h_c} + r\,I\right) \tag{5.8}$$

where T_a is temperature preferred for comfort (°F); T_s is skin temperature, 91.5°F; M is metabolic heat production (Btu/h), a function of activity; E_o is insensible perspiration heat loss, 75 Btu/h (including respiration); A is body surface area, taken as 18 ft^2; h_r is radiation coefficient, taken as 1 Btu/h ft^2°F; h_c is convection coefficient, 0.24 $\sqrt{v}$ Btu/h ft^2°F; I is insulation index of clothing (clo units); and r is 0.317, a factor to convert clo units into Btu/h ft^2°F.

Their results were shown in a series of thermal comfort charts, an example of which is shown in Fig. 5.5. Humphreys[39] obtained very similar results (Fig. 5.6). Kowalczewski[40] extended the use of the P4SR index into a PYSR (predicted yearly sweat rate) index for the purposes of characterizing various climates. Givoni and Goldman[41] developed a biophysical model which predicts dynamic variations of deep body temperature. Fanger[42] developed a "comfort equation" which summarizes all heat exchange processes between body and environment. His approach is similar to that of Morse and Kowalczewski or of Givoni, but it appears to be more detailed and complicated. It is a good basis for further research.

The ASHRAE thermal comfort laboratory was transferred from Cleveland to Kansas State University (Manhattan) in 1963. Extensive empirical testing[43] was paralleled by analytical work which produced a new effective temperature scale.[44] A two-node model of body temperature regulation has been developed in which skin and core temperatures are the controllers and which is based on a series of heat balance equations.

First the "operative temperature" scale was revised to include the effect of humidity and named "humid operative temperature" (T_{oh}), then a "new effective temperature" scale was constructed, based on plots of constant skin wettedness. Values of points on this scale are set by the intersection of the "equal wettedness" lines with the 50% RH line (rather than the saturation line in the old scale).

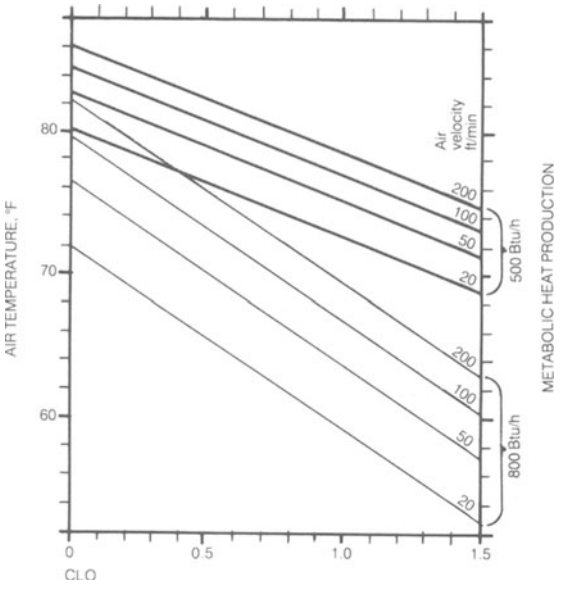


Fig. 5.5: Kowalozewski's thermal comfort chart.

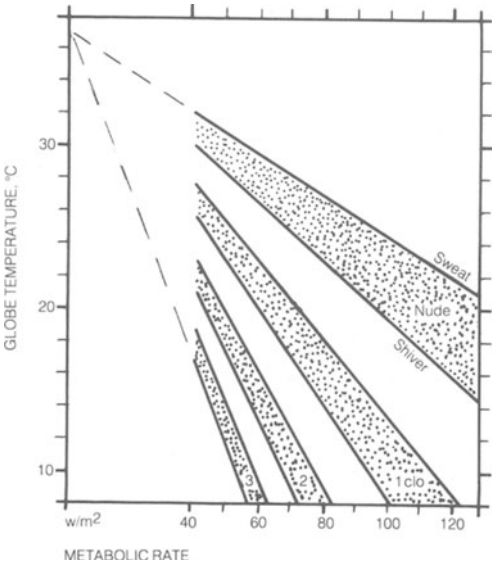


Fig. 5.6: Humphrey's thermal comfort chart.

The state of the art is best summarized in the very thorough work of McIntyre.[45] He gives the mathematical expressions relating to the body's heat dissipation processes (which can be related to the equation given at the beginning of this section) in the following form:

Radiation

$$R = \varepsilon \cdot 0.72 \cdot f_{cl} \cdot h_r \left[T_{cl} - T_r\right] \quad \text{in W/m}^2$$

where:

$\varepsilon = 0.95$ (emissivity)

$f_{cl} = 1 + 0.15(\text{clo})$ (clothing area factor)

$h_r = 4.6\left[1 + 0.01T_r\right]$ (radiative transfer coefficient) in $\text{W/m}^2\text{K}$

T_{cl} = surface temperature of clothed body in °C

T_r = mean radiant temperature (MRT) in °C

Convection

$$Cv = h_c\left[T_{cl} - T_a\right] \quad \text{in W/m}^2$$

where:

$h_c = 8.3\sqrt{v}$ (if $v > 0.2$ m/s) in $\text{W/m}^2\text{K}$

$h_c = 4.0$ (if $v < 0.2$ m/s)

(convective transfer coefficient)

T_a = air temperature (DBT) in °C

Evaporation

regulatory:

$$E = w \cdot h_e\left[p_{sk} - p_a\right] \quad \text{in W/m}^2$$

where:

w = skin wettedness

$h_e = 1.65h_c$ (evaporative transfer coefficient) in $\text{W/m}^2\text{mb}$

p_a = saturated water vapour pressure at T_a temperature in mb

p_{sk} = saturated water vapour pressure at T_{skin}

skin diffusion:

$$E_s = 4 + 0.12\left[p_{sk} - p_a\right] \quad \text{in W/m}^2$$

respiration, latent:

$$E_r = 0.0017M\left[59 - p_a\right] \quad \text{in W/m}^2$$

where:

M = metabolic free heat production in W/m^2

respiration, sensible heat:

$$C_r = 0.0014 \cdot M \cdot \left[34 - T_a\right] \quad \text{in W/m}^2$$

Through clothing

conduction:

$$K = R + Cv = \frac{\left[T_{skin} - T_{cl}\right]}{0.155\ (\text{clo})}$$

evaporation:

$$E_c + f_{pcl} \cdot E \quad \text{in W/m}^2$$

where:

$$f_{pcl} = \frac{1}{1 + 0.143\ h_c\ (\text{clo})} \quad \text{(clothing permeation factor)}$$

Sophisticated computer models of human response have been constructed on the basis of such algorithms (e.g., NBS PMV, the National Bureau of Standards Predicted Mean Vote thermal comfort analysis program, reported by Kusuda et al.)[46] Such models are now almost universally accepted and current research work is directed at dynamic factors in thermal response under transient conditions[47] or at the refinement of details of such models. Such work, in several directions, now exists well beyond the immediate interests of practicing architects and designers.

5.6 THE BIOCLIMATIC CHART

Olgyay first published his chart in 1953 in a paper entitled "Bioclimatic approach to architecture"[48] and it was included in his famous book *Design with Climate*. Although Olgyay based his chart on the best available published scientific information, some scientsts would consider it as intelligent guesswork which has never been fully validated. However, it seems to work and both his approach and his chart had a seminal influence.

The chart takes into account all four atmospheric factors of thermal sensation without attempting to construct a single-figure index. Such indices conceal the contribution of individual factors. For example, a statement such as "it is 25°C ET," could mean any of the following (and many other) combinations:

DBT 32°C,with	WBT 28°C	and v = 7 m/s
32°C	19°C	0.1 m/s
30°C	27°C	3.0 m/s
25°C	25°C	0.1 m/s

In architecture and building design, the various factors may be controlled by different means, therefore we want to recognize each individually.

Olgyay first defines a "comfort zone" in terms of dry bulb temperataure (vertical axis) and relative humidity (horizontal axis). Humidity between 30% and 65% is preferred, but under some temperature conditions it is acceptable down to 18% and up to 77%. The temperature limits (the parallel part of the comfort zone) are given for United States moderate climates, around 40° latitude, for persons wearing 1 clo clothing, as 70°F (21°C) and 82°F (28°C) for summer, 68.5°F (20°C) and 76°F (24.5°C) for winter. For lower latitudes, the bottom limit of the zone may be elevated by 0.75°F (0.42°C) for every 5° latitude change. The top would be similarly elevated, but not beyond 85°F (29.5°C). He then examines the cooling effect of air movements and by a family of curves above the comfort zone, (labeled in ft/min), shows the limits to which the comfort zone would be moved up if wind of that velocity were present.

Radiation would have the inverse effect: it would make lower temperatures acceptable. Referring to Yaglou, he suggests that every degree F drop in temperature would be counteracterd by 13 Btu/h (by which he probably means Btu/h ft^2) solar radiation (= 74 W/m^2 for every °C). A series of curves labeled in Btu/h shows the limits to which the comfort zone would be lowered if that amount of radiation were present. For indoor conditions he suggests that every degree drop in DBT (air temperature) can be counteracted by a 0.8 degree increase in MRT (mean radiant

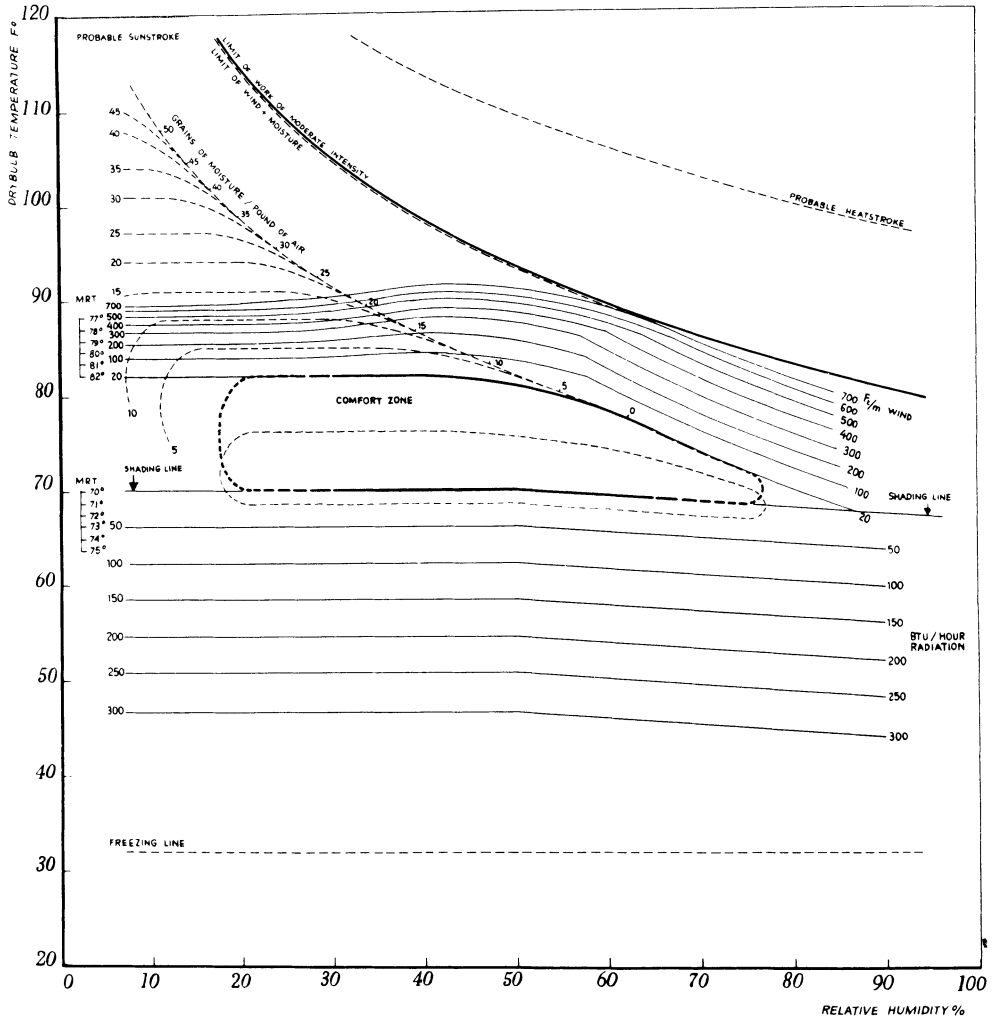


Fig. 5.7: Olgyay's bioclimatic chart.

temperature). The practical limit seems to be about 5°F (3°C) difference between DBT and MRT.

Figure 5.7 shows the bioclimatic chart and Figure 5.8 gives its "architectural" version. The chart is a useful tool for an initial climate analysis. The yearly climatic data (temperature and humidity) can be plotted on the chart. This can be done at several levels of accuracy: one dot for each hour of the year; one dot for each hour for an average day of each month (possibly connected by a line, giving a figure-eight form loop for each month); or two points for each month (the mean minimum temperature with early morning mean humidity and the mean maximum temperature with the afternoon humidity, then the two points connected by a line).

The position of the area thus defined in relation to the comfort zone would indicate the nature of the climatic problem and suggest some remedial measure.

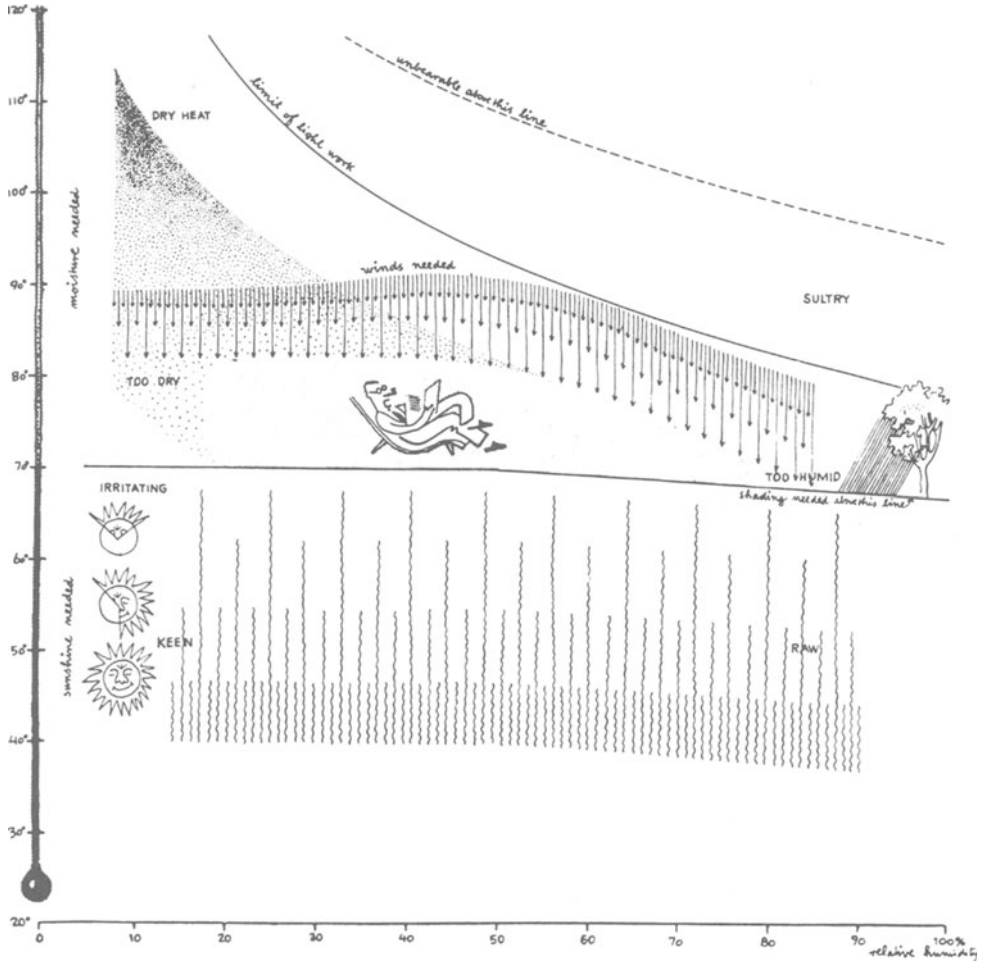


Fig. 5.8: Schematic explanation of the bioclimatic chart.

5.7 VARIABILITY OF COMFORT

Fanger's comfort equation attempts to predict a single neutrality temperature, valid for all climates, regardless of acclimatization, dependent on clothing and activity level only. This is contrary to accumulated knowledge as well as to the most recent research findings. Perhaps the most interesting development in this subject over the last few years is the work of Humphreys[49] and Auliciems[50]. Humphreys carried out a detailed examination of over 60 thermal comfort studies previously reported from various parts of the world, where the mean neutrality temperature varied from 62°F to 88°F (17°C to 31°C). He then collected climatic data for all the locations for the period of the original studies and examined the correlation of neutralities with these outdoor temperatures. He found that for "free-running"

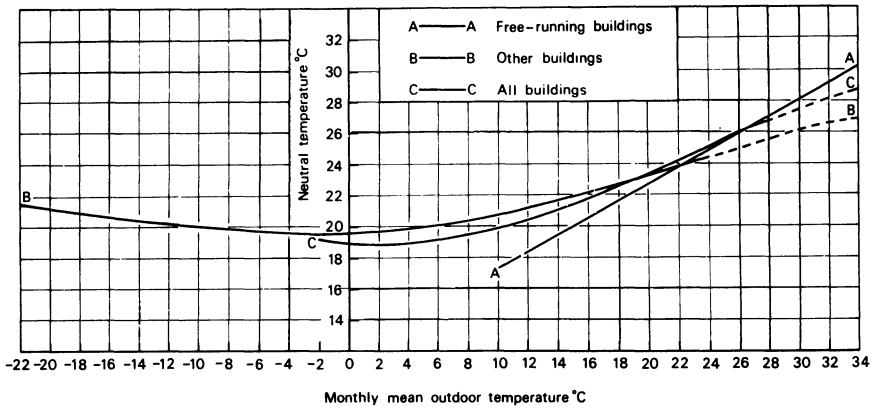


Fig. 5.9: Humphreys' correlation of neutrality and outdoor temperature.

(non-heated, non-airconditioned) buildings the neutrality temperature is

$$T_n = 11.9 + 0.534\overline{T}_o \tag{5.9}$$

where $\overline{T}_o$ is the mean outdoor temperature for the month(s) considered. The correlation coefficient was found to be 0.97 and the residual standard deviation about the line defined by this regression function is less than 1°C. All data points were within the 95% confidence limits.

For air-conditioned or heated buildings the regression function was found to be

$$T_n = 23.9 + 0.295(\overline{T}_o - 22)\exp - \left(\frac{\overline{T}_o - 22}{24\sqrt{2}}\right)^2 \tag{5.10}$$

The correlation coefficient is 0.72 and the residual standard deviation is 1.5°C (see Fig. 5.9).

Combining free-running and conditioned buildings the relationship was found to be

$$T_n = 24.2 + 0.43(\overline{T}_o - 22)\exp - \left(\frac{\overline{T}_o - 22}{20\sqrt{2}}\right)^2 \tag{5.11}$$

with a correlation coefficient of 0.88.

Auliciems, examining Humphrey's results, found that the strange curve at the cold end of the above graph is due to the inclusion of some data from the north of the USSR, which showed an increase in preferred temperatures in the colder climates. As this was probably due to the unusual asymmetrical scaling method employed by Goromosov, this data was excluded. At the same time Auliciems made use of more recent data from Australia. The final data base included 52 studies and some 250,000 individual votes. By an extensive statistical analysis, Auliciems found that for all buildings the regression of neutrality on mean outdoor temperature gives

$$T_n = 17.6 + 0.31\overline{T}_o(°C)$$

or

$$T_n = 63.7 + 0.31(\overline{T}_o - 32)(°\mathrm{F})$$

and that the range of acceptable conditions is from $T_n - 2$ to $T_n + 2$ (°C) or ±3.6°F. The correlation coefficient is 0.88 so there is no loss of apparent accuracy compared with Humphreys's results (for all buildings).

It is suggested that this simple formula would adequately predict thermal neutrality* and consequently the 4°C wide comfort zone. This would be valid not only for different locations (using annual mean temperature), but also for different seasons of the year in the same location, if for each month's comfort limits the mean temperature of the preceding month were used.

Some authors suggest that variations in temperature are quite acceptable. Humphreys and Nicol[39] for example gave a summary of their findings in a graph (Fig. 5.10), which shows the percentage of time when conditions would be judged as comfortable and as a function of the variability of temperature (expressed as monthly standard deviation of globe temperature, in degree C). The heavy line shows the mean and the shaded area would include two-thirds of the population. It shows that even with the optimum temperature the average person would be dissatisfied some 3 to 4% of the time and that a standard deviation of 1°C would make practically no difference.

Sprague and McNall[51] concerned themselves with acceptable limits of short-term variability of dry bulb temperature, such as that caused by thermostat cycling. They found that a peak-to-peak variation (ΔT) is acceptable up to the limit of

$$\Delta T = \sqrt{15 \cdot P}\ (°\mathrm{F}) \tag{5.12}$$

where P is the cycle period in hours. In metric units this would be

$$\Delta T = \sqrt{4.63 \cdot P}\ (°\mathrm{K}) \tag{5.13}$$

Later Kusuda et al.[46] reviewed several studies showing substantial energy savings if a swing of internal temperature is permitted, Shavit[52] found that a swing between 70° and 76°F (21.1° and 24.4°C) would give a 37% energy saving in Pittsburgh and in San Diego some 61% in cooling and 93% in heating. McNall et al.[53] examined four locations and concluded that allowing a swing of 8 K (14.4°F) up to 75% of the building energy use can be saved. Berglund and Gonzales[54] suggested that a drift of 0.9°F/h (0.5°C/h) is almost imperceptible up to a limit of 3.6°F (2°C) deviation from the optimum.

Other authors think that short-term variations in temperature are not only acceptable, but also desirable. Givoni[28] suggests that "slight fluctuations in the indoor conditions, such as temperature and air velocity, are beneficial, as they prevent a monotonous feeling and have an invigorating effect." Loudon and Petherbridge[55] argued that maintaining a constant indoor temperature in air-conditioned buildings

* With probable pragmatic limits of 18.3°C and 29.5°C (65°F and 85°F).

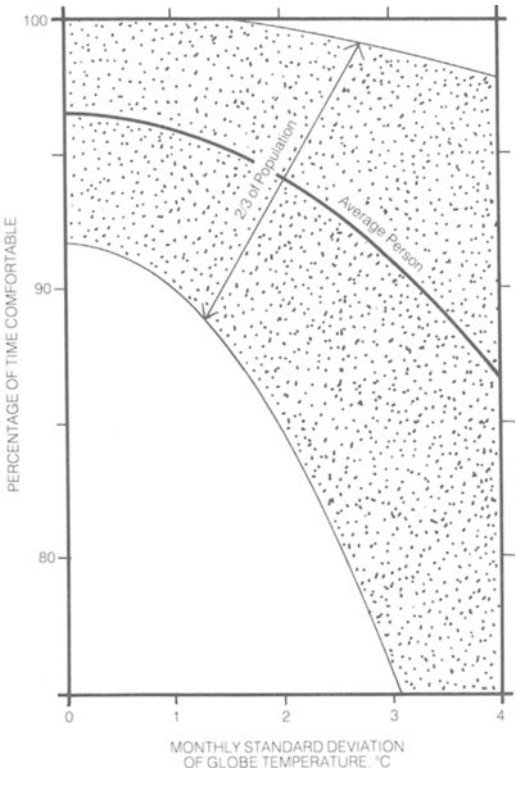


Fig. 5.10: Expectation of comfort as a function of variability (after Humphreys).

is not only uneconomical and wasteful in energy but also undesirable. They suggested that, if the mean temperature for the working day is between 68° to 75.2°F (20° to 24°C) then a swing of 7.2°F (4°C), that is ±3.6°F (±2°C) during the working day is desirable.

This opinion is supported by physiological evidence. It is well documented that humans show a distinct pattern of "circadian rhythm." This became the subject of intense discussion in relation to jet lag (the disturbance of this rhythm) but our interest here is the body temperature variation over the daily cycle, which is in the order of 1.8°F (1°C). The minimum usually occurs in the early hours of the morning (4 to 5 am) and the maximum in the afternoon, around 4 to 6 pm. This seems to be caused by changes in the metabolic rate. A logical consequence is the preference for a slightly cooler environment, when the metabolic rate is higher.[45]

5.8 PSYCHOLOGICAL EXTENSIONS

It was over a hundred years ago that Weber (1795-1878) and Fechner (1801-1887) laid the foundations of the new science of psychophysics, the study of rela-

Table 5.3

Thermal comfort scales

	ASHRAE	Bedford
+3	hot	much too warm
+2	warm	too warm
+1	slightly warm	comfortably warm
0	neutral	comfortable
−1	slightly cool	comfortably cool
−2	cool	too cool
−3	cold	much too cool

tionships between physical and corresponding psychological events. Fechner's law

$$R = k \log S \tag{5.14}$$

relates a stimulus to a response, where both S and R are measureable on some scale.

The best-known scales for the measurement of thermal sensation (response) are those of ASHRAE and Bedford, given in Table 5.3.

The semantics imply that the ASHRAE scale measures a thermal sensation, but the Bedford scale brings in the broader concept of comfort. However, even the use of the ASHRAE scale requires a judgment. As Stevens[56] refers to it: we measure a "judgment continuum" against a "stimulus continuum," and only assume that the judgment corresponds to a sensation. His "power law" is put in the form

$$J = k.S^n \tag{5.15}$$

where J is judgment; S is stimulus; k is a constant; and n is an exponent, suggested as $n = 1$ for cold; $n = 1.6$ for warmth; and $n = 0.6$ for loudness, etc.

In a later version, $(S - S_o)$ is substituted for S, where S_o is the stimulus at threshold level, or, in case of thermal stimuli, the neutrality temperature. He also suggests that there is no simple way of knowing whether the judgment expressed is analogous with the sensation experienced.

Stevens and Hall[57] showed in relation to light and sound perception that both n and k depend on the duration of exposure and that the senses integrate the effect of stimulation over time. Marks and Stevens[58] reported on temporal summation of thermal stimuli: 3 and 6 second exposures are well described by the power law, but shorter durations are influenced by the delaying action of skin resistance. If the change in skin temperature (ΔT) is taken as the stimulus, over short durations

only partial summation occurs and over longer durations there appears to be no summation. At low levels of ΔT, a "reverse summation" occurs.

Mark and Stevens[59] also considered spatial summation. Earlier work showed that the same thermal sensation can be produced (above a threshold) by a larger stimulus acting on a smaller skin area, as by a lesser stimulus over a greater area. However, the degree of spatial summation decreases with the increase of warmth and disappears at the threshold of pain. Reaction time also becomes shorter with increased area of increased heat input. For very short exposure times the threshold of pain is higher.

The authors in both papers suggest two hypotheses:

(1) two sets of thermal receptors operate, one at low intensity, which has good summation properties, the other at higher intensities, displaying little or no summation;

(2) an adaptation mechanism may be at work, (a lateral neural inhibition) which increases with the intensity, thus reduces spatial summation.

This ongoing work may be of interest from the physiological point of view, but has little immediate relevance to architects and designers.

There is no doubt that the basic mechanism of thermal sensation is physiological, but it is overlaid by discriminatory, affective, and cognitive layers. Cabanac[60] suggested the term "alliesthesia" to describe the variability of sensation caused by the same stimulus, depending on internal signals of the subject. He suggested the following mechanism of thermal perception: external stimulus → skin receptors → neural signal → sensation (in consciousness).

The affective aspect of this sensation (pleasure or displeasure) is not entirely stimulus-bound; it depends on internal signals. To a body in hyperthermia, a cold stimulus (even below the normal pain threshold) will cause pleasure. In hypothermia a hot stimulus, which would otherwise be unpleasant or even painful, would cause pleasure. He concluded that what is "useful" will be judged as "pleasant."

The picture is further complicated by the fact that in real-life situations, man is not a passive experimental subject; he or she will react to the thermal stimulus by behavioral or even technological adjustments. The simple vote cast on a seven-point scale is "verbal behavior," involving all the above levels.

Such a reasoning led Auliciems[61] to put forward the psychophysiological model of thermal perception shown in Fig. 5.11. He also suggests that the term "thermal comfort" has become hackneyed to such an extent that it has lost its precise meaning. It is today no more than a general functional term, covering (hiding?) a whole range of parameters and processes. The term "neutrality" should be used at the "sensation" level. What the ASHRAE scale measures is "satisfaction," which is at the "affective level" of integration, being influenced by past experience and thermal expectation and involving cognitive processes as well as hitherto inadequately explained psychological factors. This is why the color of surroundings or noise,[62] personality,[63] and physical arrangements of rooms[64] may have an influence of thermal perception. "Thermopreferendum" is the term that

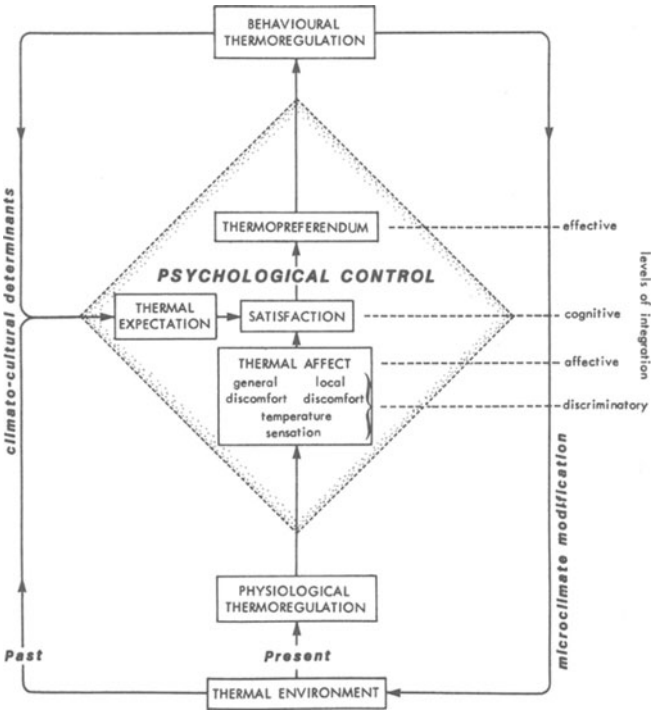


Fig. 5.11: Psychophysiological model of thermoregulation proposed by Auliciems.

should be employed for the still higher "effective" level of integration (what people want) which should be the designer's target.

It is questionable, however, whether the thermal response (be it thermal comfort or thermopreferendum) can be considered in isolation from the total physical and social environment, from satisfaction in its totality. Roethlisberger and Dickson in their "Management and the Worker" (referred to by Sommer[65]) give an account of what became known as the "Hawthorne effect."

The work first carried out by the Western Electric Company (but since then repeated and confirmed many times by independent work) showed that employee reaction to changes in environmental conditions depends on the prevailing labor relations. If trust and understanding exist, the reaction will be positive to a wide variety of changes. In a hostile situation, the employee will be suspicious of even the best-intentioned changes and react negatively.

Sommer concluded that there is no simple relationship between single environmental elements and human behavior. The latter is very complex, the environmental effects would be filtered and mediated by individual attitudes, which in turn may be determined by group processes. So, the thermopreferendum is dependent not only on physical environmental stimuli, but also on social and psychological factors.

Table 5.4

Recommended design temperature °F (°C)

	MoHLG		DIN. '59		IHVE '59		IHVE '70		ASHRAE '74	
	°F	°C	°F	°C	°F	°C	°F	°C	°F	°C
Living-dining	64.4	(18)	68	(20)	65	(18.3)	69.8	(21)	75–80	(23.8–26.6)
Bedroom	no heat'g		68	(20)	50	(10)	64.4	(18)	75	(23.8)
Kitchen	55.4	(13)	68	(20)	60	(15.5)	60.8	(16)	65–70	(18.3–21.1)
Bathroom	no heat'g		71.6	(22)	60	(15.5)	71.6	(22)	80	(26.6)
Communication spaces	55.4	(13)	59	(15)	60	(15.5)	60.8	(16)		—

McHLG: UK Ministry of Housing and Local Government, circular 1/68.

IHVE: Institution of Heating and Ventilating Engineers (UK), Guide, 1959 and 1970.

DIN: Deutsches Institut für Normung, standard 4701:1959.

American Society of Heating, Refrigeration and Air-conditioning Engineers Guide and Data Book, Applications volume, 1974.

5.9 CONSEQUENCES

The last two sections have some important consequences from the point of view of passive solar building design, because they indicate the emergence of an entirely new attitude toward thermal comfort and the setting of indoor design temperatures.

In the 1950s and 1960s it was almost general practice in the U.S.A. to keep buildings heated in winter and cooled in summer to a steady and unchanging temperature of 75°F (23.8°C) or even in many cases going overboard and cooling to 72°F (22.2°C) in summer. It is salutary to compare indoor temperature (winter) recommendations, e.g., for residential buildings by various bodies and their changes over the years.

Whilst in the United Kingdom the trend was upwards, they never caught up with the United States where the ASHRAE general recommendation in 1974 was still 75°F (23.8°C) as the optimum design temperature for both winter and summer. However, the 1973 energy crisis initiated a change of attitude and led up to the 1979 Building Temperature Restrictions (BTR). These stipulate an upper limit of 65°F (19.3°C) for winter and a lower limit of 78°F (25.5°C) for summer.

The motivation for these was obviously energy conservation, and quite justifiably so. Many studies show the potential for energy saving by such measures. Woodhead and Scanes[66], for example, examined the effect of house construction and thermostat setting on annual air conditioning load in a hot climate location in Western Australia (Port Headland). Their results are summarized in Table 5.5.

Table 5.5

Annual air conditioning load of an average house in 10^6Btu (MWh), as a function of thermostat setting

T_i =	70°F	(21.1°C)	74°F	(23.3°C)	78°F	(25.6°C)
Timber frame, a/c sheeting	162	(47.5)	115	(33.6)	75	(21.9)
Brick veneer	147	(43.6)	105	(30.8)	67	(19.7)
Cavity brick	140	(41.1)	97	(28.3)	62	(18.1)

It is apparent that thermostat setting has a far greater influence than change of construction. A change from 21.1°C to 25.6°C would reduce energy use to less than half. (Similar studies were mentioned above.) The validity of BTR from the point of view of energy conservation was not really doubted by anyone, but many considered it as a reduction of standards, as "belt tightening," as a sacrifice. This is questionable.

The problem can best be approached by examining a few sample locations in the U.S.A. In Table 5.6 we list the summer (July, column 1) and winter (January, column 5) mean temperatures and find the corresponding "neutrality" temperatures, as defined by Auliciems, quoted above, in columns 2 and 6—reminding ourselves that the comfort range will be ±3.6°F (±2°C) about these values. In the last two columns, the ASHRAE'74 recommendation and the BTR limits are shown.

From the comparison, the following points emerge:

(1) For summer, the 75°F (23.8°C) would be below the neutrality temperatures in all cases, and in Las Vegas it would be below the lower comfort limit. In all cases the result would be overcooling.

(2) The 78°F (25.5°C) BTR limit would still be below the neutrality temperature for Atlanta and Las Vegas and slightly above it for the other two locations, but in all cases well below the upper limit of comfort.

(3) For winter, the 75°F (23.8°C) is above the neutrali ty temperatures in all cases and would result in a heating requirement even in Miami. For the upper three locations it would be well above the upper comfort limits.

(4) The 65°F (18.3°C) BTR limit would be more than adequate for New York and slightly less than the neutrality but well within the comfort range for Atlanta and Las Vegas. For Miami, it would result in no heating requirement, possibly producing discomfort in the evening hours in very lightweight, inadequate buildings.

The generalized conclusion can then be drawn, that the building temperature restrictions do not imply a sacrifice or a lowering of standards; in most cases they

Table 5.6

Comparison of neutralities and recommendations

		Summer: July				Winter: January			
				ASHRAE				ASHRAE	
		T_o (1)	T_n (2)	'74 (3)	BTR (4)	T_o (5)	T_n (6)	'74 (7)	BTR (8)
In °F	Miami. Fla.	72.5	76.3			67.1	74.7		
	Atlanta, Ga.	78.8	78.3			43.7	67.3		
				75	78			75	65
	Las Vegas, Nev.	72.5	76.3			67.1	74.7		
	New York, N.Y.	75.9	77.4			32.0	63.7		
In °C	Miami, Fla.	22.5	24.6			19.5	23.7		
	Atlanta, Ga.	26.0	25.7			6.5	19.6		
				23.8	25.5			23.8	18.3
	Las Vegas, Nev.	29.5	26.7			7.0	19.8		
	New York, N.Y.	24.4	25.2			0	17.6		

would produce a more comfortable indoor environment than the previous standards. The other obvious conclusion is that the use of one set of recommended temperatures for all climatic regions of the United States is an oversimplification and will not only lead to waste of energy but will not produce comfort either.

5.10 BEHAVIORAL AND CLOTHING DIFFERENCES

The reasons for the variation of thermal preferences are probably twofold:

(1) behavioral and clothing differences,

(2) acclimatization and habit.

The first of these is no longer disputed. The ASHRAE comfort standard 55-66 gave the comfort zone as bounded by the 20% and 60% relative humidity lines and by the 73.2°F and 7.5°F (22.9°C and 25.3°C) dry bulb temperature lines for persons wearing 0.8-1 clo clothing. (For explanation of clo units, see Tables 5.7 and 5.8.)

This has been revised and published as 55-74. The comfort envelope is now defined for persons wearing 0.6 to 0.8 clo clothing as between dew-point temperature lines (35°F and 62°F (1.7°C and 16.7°C) corresponding to 5 mmHg and 14 mmHg vapor pressure. The temperature range at the lower humidity level was 72.6° to

79.7°F (22.6° to 26.5°C) and at the upper level 71.5° to 77.6°F (21.9° to 25.3°C). At the 50% RH line the limits were 72° and 78°F (22.2° and 25.6°C). The mean of these two values, 75°F (23.9°C), was the design temperature mentioned earlier. Both are shown on the psychrometric chart, Fig. 5.12.

However, the clothing level of people varies with the season. Gonzalez and Nishi[68] found that the lower comfort limit of 72°F (22.2°C) set by 55-74 is valid for people wearing 0.85 clo, at an activity level of 1.1 met, resulting in 5% of the population being dissatisfied. Adding a wool pullover, thus increasing clothing to 1.15 clo, 80% of the population would find 68°F (20°C) acceptable.

In summer, Gagge et al.[69] observed the following clothing levels in New York:

0.2 clo: 5% of males and 43% of females

0.4 clo: 50% of males and 50% of females,

and that people dressed for the season, rather than for the expected indoor temperature. They suggested that 80% of the population would be satisfied:

in winter, with 68° to 70°F (20° to 21°C), wearing 1 to 1.2 clo,

in summer, with 78° to 80°F (25.5° to 26.7°C), wearing 0.3 clo or less.

In the latest revision of the ASHRAE comfort standard published last year, the role of acclimatization and habit is still not recognized, but the standard is based on the recognition that people wear different clothes in summer and winter. The vapor pressure (or dew-point temperature) limits are the same as before, but for summer the temperature limits are slightly increased and for winter they are substantially reduced. The basis of this is an assumption of 0.5 clo for summer and 0.9 clo for winter. Along the 50% RH line the limits are:

summer: 73° to 79°F (22.8° to 26.1°C)

winter: 68° to 75°F (20° to 23.9°C)

The new revision permits a variation of 2°F (1.1°C) from peak-to-peak. If the variation exceeds this, its rate must be limited to 4°F/h (2.2°C/h), and the comfort limits should not be exceeded by more than 1°F (0.6°C).

Adjustments are provided for several variables: *Air movements.* The standard is valid for a maximum air velocity of 30 fpm (0.15 m/s) in winter and 50 fpm (0.25 m/s) in summer. For each fpm (0.005 m/s) increase above 30 fpm (0.15 m/s) the comfort zone is to be raised by 1°F (0.6°C). *Activity.* The standard is valid for sedentary work. For every "met" increase in metabolic rate (up to a maximum of 3 met) the temperature limits should be decreased by 4.5°F (2.5°C), but not below 59°F (15°C). (For a list of metabolic rates, see Table 5.9.) *Clothing.* Every change of 0.1 clo can be compensated for by a change of 1°F (0.6°C) in temperature in the opposite direction.

Arens et al.[70] proposed a revised form of the bioclimatic chart which embodies the latest research results discussed above (see Fig. 5.13). It is based on 1.3 met activity rate and 0.8 clo clothing and follows the same structure as Olgyay's. The main differences are:

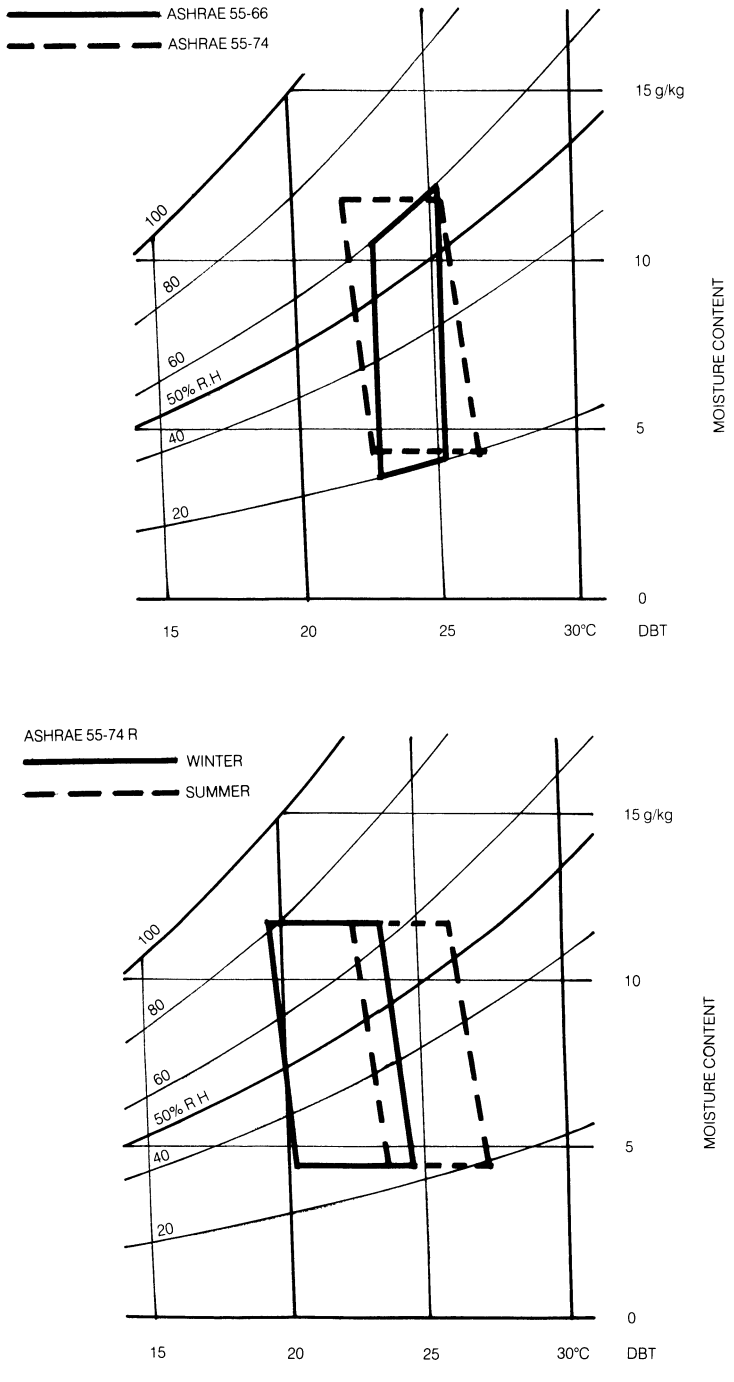


Fig. 5.12: Definitions of the "comfort envelope" by ASHRAE.

Table 5.7

Insulating value of clothing

		U-value		Resistance	
	(clo)	W/m^2K	Btu/h $ft^2°F$	m^2K/W	h $ft^2°F$/Btu
Shorts only	0.1	64.50	11.40	0.016	0.09
Shorts + short-sleeve shirt	0.3	21.50	3.80	0.05	0.09
Slacks + short-sleeve shirt	0.5	12.90	2.28	0.08	0.26
Light business suit + vest	0.7	9.21	1.63	0.11	0.61
Normal business suit + vest	1.0	6.45	1.13	0.16	0.88
Heavy business suit + vest with long underwear	1.5	4.30	0.76	0.23	1.32
Winter clothes + overcoat	2.0	3.23	0.57	0.31	1.75
Heaviest artic clothing	4.5	1.43	0.25	0.70	3.95

(1) vapor pressure (or absolute humidity) defines the side limits, rather than relative humidity;

(2) the radiation lines are far more widely spaced than in Olgyay's chart, apparently attributing a lesser effect to radiation;

(3) the wind lines are completely different.

The first of these is based on general consensus. The previous theory was that the "humidity deficit," i.e., (100%-relative humidity), determines the evaporation potential. It is now accepted that the vapor pressure at the surface of wetted skin being almost constant, the vapor pressure of the atmosphere alone will determine the differential, hence the evaporation potential.

Regarding the second point: the difference is apparent only and is due to a different labeling of the MRT lines. Olgyay gives the actual MRT scale (see Fig. 5.7) e.g., if the DBT was only 64°F and the lower limit of the comfort zone could drop to that level if the MRT was 75°F (i.e., a drop of 6°F in DBT is compensated by an increase of 5°F in MRT). The scale of the new chart is labeled in MRT-Ta, that is, the difference between mean radiant and air temperature. For example, in Fig. 5.13 if the air temperature (DBT) was 18.6°C (at 40% RH), and this is 2.4°C below the comfort boundary, we read from the chart that the MRT-Ta value should be 4.8°C. Therefore the MRT should be 18.6 + 4.8 = 23.4°C to make the condition comfortable.

The authors give a detailed discussion of the equivalence between solar radiation (strongly monodirectional), the "effective radiant field" (ERF) and MRT. The chart is based on the equivalence of 1:1 between MRT and DBT, which is not unrea-

Table 5.8

Insulating value of clothing elements

Shoes:	sandals	0.02 clo
	normal, closed	0.04
	boots	0.08
Underwear:	long, upper	0.10
	long, lower	0.10

Men		**Women**	
Singlets	0.06 clo	Bra + panties	0.05 clo
T-shirts	0.09	Half slip	0.13
Underpants	0.05	Full slip	0.19
Shirt, light, short sleeve	0.14	Blouse, light, short sleeve	0.18
long sleeve	0.22	long sleeve	0.22
heavy, short sleeve	0.25	heavy, short sleeve	0.26
long sleeve	0.29	long sleeve	0.29
+5% for tie or turtle-neck		Dress,* light, short sleeve	0.20
		long sleeve	0.22
Vest, light	0.15	heavy, short sleeve	0.63
heavy	0.29	long sleeve	0.70
Trousers, light	0.26	Skirt,* light	0.10
heavy	0.32	heavy	0.22
Jacket, light	0.22	*+5% if below, -5% if above knee	
heavy	0.49		
Sweater, light, short sleeve	0.18	Sweater, light, short sleeve	0.15
long sleeve	0.20	long sleeve	0.17
heavy, short sleeve	0.33	heavy, short sleeve	0.33
long sleeve	0.37	long sleeve	0.37
Socks, short	0.04	Jacket, light	0.17
heavy	0.10	heavy	0.37
		Stockings or pantyhose	0.01

Total clo = 0.77 Σclo + 0.05 for women and 0.727 Σclo + 0.113 for men.
(After L.G. Berglund: *Habitat International* 5(3/4) : 528.)[67]

Table 5.9

Metabolic rates

Activity	met	Unit Body Surface W/m²	Unit Body Surface Btu/h ft²	For Average Man W	For Average Man Btu/h
Sleeping	0.7	40.7	12.9	70	245
Lying in bed	0.8	46.5	14.7	80	280
Seated at rest	1.0	58.2	18.4	100	350
Sedentary work, standing	1.2	69.8	22.1	120	420
Very light work (e.g., light industry, shopping)	1.6	93.1	29.4	160	560
Medium light work (e.g., house work, machine tools)	2.0	116.4	36.8	200	700
Steady medium work (e.g., garage work, social dancing)	3.0	174.6	55.2	300	1050
Heavy work (e.g., planning by hand, furnace work)	4-6	233-350	74-110	400-600	1400-2450
Vigorous sport, very heavy work (e.g., squash)	5-7	291-407	92-129	500-700	1750-2450

sonable. This is the basis of the U.K. "dry resultant temperature" and differs only slightly from Olgyay's assumption of 0.8:1 or Fanger's[42] finding of 0.48:0.52, which is 0.92:1. The "environmental temperature" index used by the *IHVE Guide*[10] is based on a ratio of 0.5:1, which, in my view, is still valid in overheated situations for people wearing little clothing (shorts, short-sleeved shirts). For winter conditions, a ratio of 1:1 is generally accepted. There is, however, a disagreement regarding the maximum acceptable difference between MRT and DBT. Some authors suggest a limit of 9°F (5°C); others go as far as 36°F (20°C). This is clearly an area which urgently requires further research.

The third difference, the wind lines, is the most controversial. There is a reasonable agreement regarding the cooling effect of air movement (as a function of velocity) of medium-to-high humidities. In the new chart the wind lines keep rising at the low humidity end, whereas in Olgyay's original they drop from about 50% RH down. The effective temperature nomograms show an inversion of wind effect at about 37°C ET, above which the presence of wind actually worsens the situation.

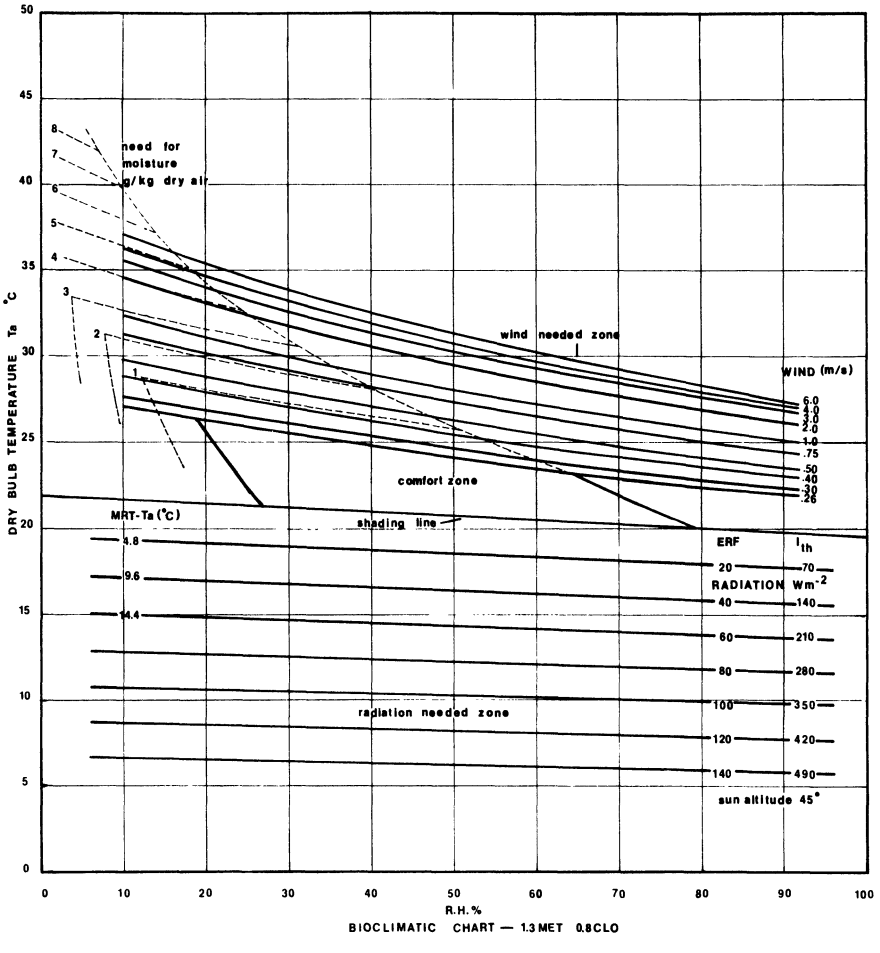


Fig. 5.13: Revised bioclimatic chart, proposed by Arens et al.

The fact is that as soon as the air temperature exceeds skin temperature, air movement will increase the convective heat input. The evaporative cooling still continues, but there will be a crossover point where the convective gain exceeds the evaporative loss, resulting in a net gain. Further research is needed to determine where this crossover occurs. Until then, I suggest that the wind lines of the original Olgyay chart give a more plausible representation of the real effect than the revised chart.

5.11 ACCLIMATIZATION AND HABIT

The second reason we suggest for the variation of thermal preferences is acclimatization and habit. It is interesting that chapter 7 of the *ASHRAE Handbook of Fundamentals*[11] clearly shows the mechanisms of acclimatization (and what they call "accustomization"), but it does not transpire into the recommendations for design temperatures. Givoni[28] gives a summary of acclimatization as it affects

metabolic responses. At rest, comfortable conditions produce the lowest metabolic rate. If an unacclimatized person is exposed to a cold environment, his or her metabolic rate may increase by a factor of 2 or 3. For an acclimatized person this increase is much less, only 30 to 40%. In unacclimatized persons the metabolic rate will also increase by 25 to 30% during exposure to hot conditions. Acclimatization greatly reduces this response.

The role of these mechanisms is empirically well documented. An interesting paper was presented by Bánhidi[71] at the ICBEM in Portugal: in his laboratory he tested the thermal performance of over 20 people, divided into two groups, students in their early twenties and pensioners. To his surprise and contrary to all expectations, he found that students prefer higher temperatures than the pensioners, the mean neutralities being:

	Stated Preference		Measured Preference	
Pensioners	68.9°F	(20.5°C)	73°F	(22.8°C)
Students	73°F	(22.8°C)	75.9°F	(24.4°C)

The explanation was found: students lived in colleges, having a continuously operated district heating system, whilst the pensioners lived on their own and could only afford to heat one room, even that only intermittently. (The findings of Fox et al.[72] were similar.) This is a clearly demonstrated case of the influence of habit on thermal preferences. This throws some doubt on the validity of tests, such as those reported by Gonzalez and Nishi,[68] where the previous conditioning of the experimental subjects is not known.

Much has been written about the problems of the "white man's" adaptation to tropical climates. Lee[73] gave a good summary of many relevant opinions. Few of these had any factual base and when authors such as Markham[74] refer to the "degeneration of energetic whites" in tropical climates, it appears to be pure speculation. (Climate may have an influence, but social and psychological factors are more probable reasons if any such "degeneration" occurs.)[75]

All other factors being satisfactory, the acclimatization process can be completed in four to five months. This was well illustrated by a study of Damm and Fenelon.[76] They carried out a series of physical and psychological tests on a group of European immigrants in Australia, repeated three times:

(1) at the reception center (Bonegilla, Victoria, lat. 37°S) shortly after arrival from a European winter, in May, mean range of temperatures 43-60°F (6-15.5°C);

(2) at Townsville (Northern Queensland, lat. 18°S) shortly after arrival there, for work on sugarcane plantations, in June, mean range of temperature 58.5°–77.9°F (14.7°–25.5°C);

(3) at the same location, after four months of acclimatization, in October, mean range of temperatures 68.9°–84.7°F (20.5°–29.3°C) with very high humidities.

At the second series there was a reduction in physical performance compared to the first and there were complaints of fatigue. At the third series both the physical

and the psychological test results were identical with the first. The acclimization was complete.

5.12 PASSIVE HEATING SYSTEMS AND COMFORT

If we combine any passive heating system with an auxiliary heater controlled by a thermostat set at, say, 75°F (23.8°C), but otherwise switched on all the time, it will be very difficult to achieve a solar fraction greater than 0.4 (and probably less). It is an inherent characteristic of all passive systems that they reduce the diurnal swing of temperatures indoors, compared with that outdoors, but that swing will not disappear. If every time the indoor temperature drops below the set level, the heater starts, the energy consumption will increase and the heat losses from the building will be greater as the temperature differential is greater.

This was very clearly demonstrated by two almost identical passive houses in Hobart (Tasmania, Australia) where one family left the thermostat-controlled auxiliary heater switched on all the time, whilst the other switched it off at night; the energy use of the latter was less than half of that of the former. In the latter case the indoor temperature showed a swing of 7.2°F (4°C, between 21° and 17°C) whilst the outdoor variation was 21.6°F (12°C, between 15° and 3°C).

There is a well-proven and documented general case for allowing the indoor temperature variations for reasons of energy economy, but what would be the effect of these on thermal comfort? In the above example, the neutrality temperature would be about 68.7°F (20.4°C) so the comfort range would be 65.1°–72.3°F (18.4°–22.4°C). The indoor temperature dropped slightly below the lower limit, but:

(a) the living area is not in use at night;

(b) for the bedroom most sources recommend lower temperatures;

(c) the air temperature is low, but the mean radiant temperature is higher. The wall temperature does not drop below 71.6°F (22°C) and the mean radiant temperature remains above 65.3°F (18.5°C).

This last point has a great significance for both direct-gain and mass-wall systems. In buildings incorporating such systems at night, when the building is cooling down, the floor or wall surfaces would be substantially warmer than the room air. This is advantageous for two reasons:

(1) the heat loss is approximately proportionate to the air temperature difference;

(2) the thermal sensation depends, to a large extent, on the mean radiant temperature.

There is a reasonable consensus regarding the degree of this dependence. In the previous section, five sources were quoted giving a relationship between MRT and DBT, showing what increase would be required in MRT to compensate for one degree drop inDBT, or vice versa.

I suggest that the 0.5 deg. MRT = 1 deg. DBT relationship given by the "environmental temperature" would still be applicable in summer, or in warm

climates for people wearing shorts, short-sleeved shirts, or short dresses. But in cold climates, for people wearing 0.8-1 clo, the relationship given by the "dry resultant temperature" (i.e., 1 deg. MRT = 1 deg. DBT) is more applicable. This is the relationship implied by the revised bioclimatic chart.

In both direct-gain and mass-wall systems, the surfaces we rely on may be 10° to 15°F (6° to 8°C) warmer than air temperature. If cold surfaces (e.g., windows) which would reduce MRT are eliminated (e.g., by insulating shutters or curtains), it is reasonable to assume that the spherically averaged surface temperature (i.e., weighted by the solid angle subtended by each surface) which is the MRT, would be halfway between the air temperature and the solar surface temperature (i.e., inside of the mass wall, or the direct-gain floor or wall surface), thus 5° to 7.5°F (3° to 4°C) above air temperature. Which means that we may allow the air temperature to drop 5°to 7.5°F (3° to 4°C) below the lower comfort limit and still maintain comfortable conditions.

5.13 SUMMARY

There are a number of methods available to predict the daily temperature profile in a passive solar building. These are then usually compared with the comfort level or the "design indoor temperature," to examine whether the system is satisfactory or not. The basic recommendation of this chapter is that the design indoor temperature should be considered as a variable, as a function of the monthly mean temperature at the location:

$$T_n = 63.7 + 0.31(\overline{T}_o - 32)\ (°F) \text{ but } 65 < T_n < 85°F$$

$$T_n = 17.6 + 0.31\ \overline{T}_o\ (°C) \text{ but } 18.3 < T_n < 29.5°C$$

and that a variation of ±3.6°F (±2°C) should be taken as acceptable. This should be valid for fully acclimatized people at sedentary work (1.2 met), wearing their normal choice of clothes.

The following interrelationships exist between various thermal factors, which can be used as a basis for compensatory measures:

	Variable	Air Temp.	Note
Air movement	each fpm above 30 fpm	+1.0°F	
	(each 0.005 m/s above 0.15 m/s)	+0.6°C	
Activity	each met increase	−4.5°F	not below 59°F
	(up to 3 met)	−2.5°C	(15°C)
Clothing	each 0.1 clo added	−1.0°F	
		−0.6°C	
Radiation	each +1°F(°C) in MRT	−1.0°F	max. difference 9.0°F
		−1.0°C	(5°C)

The psychological considerations discussed above mean that, in practice, allowances must be made for the attitude of the user. First, we must distinguish be-

tween owner-occupied (or single-family occupancy) buildings and multi-occupancy buildings (or buildings where heating is controlled by others than the users). Second, further subdivisions may take into account the probable attitude of the users.

(1) Owner- or single-family occupancy

(a) The owner-builder-designer will have the greatest tolerance and will adapt his or her behavior where the environment cannot be fully controlled.

(b) The occupant of a house who is involved with the system understands it, and controls it with sympathy, will likely be satisfied even if the objectively defined comfort conditions are not always achieved.

(c) More difficult is to satisfy the commercial house buyer who is used to having an almost instant response at the flick of a switch, who may expect the passive system to provide the same flexibility. In this case, an active component or a conventional auxiliary heater may have to be included, not only to cope with inclement weather, but also with the variable wishes of the user.

(2) Multi-occupancy buildings

(a) Where the landlord provides heating, he may have legal obligations to provide a specified performance. Notwithstanding such specified levels, satisfaction may depend on the tenant/ landlord relationship and, given the unpredictability of humans, it may be safer to allow for some degree of individual control.

(b) In work places, the "Hawthorn effect" may come into full play: the satisfaction of the occupants will depend on employer-employee relationships as much as on actual thermal performance of the building and the system.

5.14 REFERENCES

1. W. Heberden. "An account of the heat of July 1825, together with some remarks on sensible cold," *Phil. Trans.* **2**, 69 (1826).
2. J. S. Haldane, "Influence of high temperatures," *J. of Hygiene Camb.* **5**, 494 (1905).
3. F. C. Houghten and C. P. Yagloglou, "Determination of comfort zone," *Trans. ASHVE.* **29**, 361 (1923).
4. C. P. Yagloglou and P. Drinker, "The summer comfort zone: climate and clothing," *Trans. ASHVE.* **35**, 269 (1929).
5. H. M. Vernon, *An Investigation of Factors... in... Industrial Accidents*, Memorandum 21, Health and Munition Workers Com.Cd.9046. HMSO, 1918.
6. H. M. Vernon and C. G. Warner, "The influence of the humidity of the air on capacity of work at high temperatures," *J. Hyg. Camb.* **32**, 431, (1932).

7. T. Bedford, *Warmth Factor in Comfort at Work*, Med. Res. Council, Industr. Health Res. B'd Report 76 HMSO, (1936).

8. C. E. A. Winslow et al. "Physiological reactions to environmental temperature," *Am. J. Physiology* **120** (1) (1937).

9. V. Olgyay, *Design with Climate*. (Princeton Uni. Press., 1963).

10. *IHVE Guide*, books A, B, C, Inst. of Heat'g. Vent'g. Engrs. (London, 1970).

11. *ASHRAE Handbook of Fundamentals*, Am. Soc. Heat'g. Refrig. and Air cond. Engrs. (New York, 1972).

12. Carrier System Design Manual, Australian (SI metric ed.) 1974.

13. Office of Emergency Preparedness, *The Potential for Energy Conservation*, (Washington, D. C., 1972).

14. Dept. of Energy, Emergency Building Temperature Restrictions, Federal Register AA. No. 107 (June 1, 1979).

15. D. L. Ingram and L. E. Mount, *Man and Animals in Hot Environments*, (Springer Verlag, 1975).

16. B. McArdle et al. *Prediction of the Physiological Effect of Warm and Hot Environments*, Med. Res. Council, RNP 47/391, (London, 1947).

17. W. V. Macfarlane, "Thermal comfort zones," *Arch.Sc.Rev.* **1**, 1–14 (1958).

18. D. H. K. Lee, *Physiological Objectives in Hot Weather Housing*, US Govm't Housing and Home Finance Agency, 1953.

19. R. K. Macpherson, "The assessment of the thermal environment—a review," *Brit. J. of Industr. Medicine* **19** (3), 151 (1962).

20. F. C. Houghten and C. P. Yagloglou, "Determining equal comfort lines," *J. ASHVE.* **29**, 165 (1923).

21. C. P. Yagloglou, "The comfort zone for man," *J. of Indust. Hygiene* **9**, 251 (1927).

22. T. Bedford, *Environmental Warmth and Its Measurement*, Med. Res. Council, War Memorandum 17. HMSO, (1940).

23. C. P. Yaglou, "Indices of comfort," *Physiology of Heat Regulation*, edited by L. H. Newburg, (Saunders, Philadelphia, 1949).

24. A. F. Dufton, *Equivalent Temperature and Its Measurement*, B. R. Techn. Paper 13. HMSO, (1932).

25. A. F. Dufton, "The use of Kata thermometers for the measurement of equivalent temperature," *J.Hygiene Camb.* **33**, 349 (1933).

26. T. Bedford, "Equivalent temperature," *Heat'g. Pip. Air Cond.* **23**, 87 (1951).

27. A. Missénard, "On thermally equivalent environment," *IHVE. J.* **27**, 231 (1959).

28. G. Givoni, *Man, Climate and Architecture* (Elsevier, 1969).

29. C. G. Webb, "An analysis of some observations of thermal comfort in an equatorial climate," *Brit. J. of Industr. Medicine* **16** (3), 297 (1959).

30. C. G. Webb, "Thermal discomfort in an equatorial climate," *IHVE J.* **27**, 297 (1960).

31. D. H. K. Lee, "Proprioclimates of man and domestic animals," *Climatology: Reviews of Research*, UNESCO Conf., (Paris, 1956).

32. F. E. Smith, *Indices of Heat Stress*, Med. Res. Council Report **29** (London, 1955).

33. B. Givoni,*The Nature and Application of Thermal Indices*, Bulletin No. 73–74 (Israel Inst. of Techn., 1962).

34. H. S. Belding and T. F. Hatch. "Index for evaluating heat stress in terms of resulting physiological strains," *Am. J. Heat Pip. Air Cond.* **27** (8), 129 (1955).

35. D. H. K. Lee and A. Henschel, "Evaluation of thermal environment in shelters," *Trans. US Dept. of Health, Educ., and Welfare*, 1963.

36. B. Givoni, *Estimation of the Effect of Climate on Man: Developing a New Thermal Index* (Technion,Haifa, 1963).

37. J. W. Drysdale, *Physiological Study No. 3*, C'wealth Exp. Bldg. Stn. Techn. Study 35, Sydney, (1951).

38. R. N. Morse and J. J. Kowalczewski, "A rational basis for human thermal comfort," *ASHRAE J.* **9** (9), 72 (1967).

39. M. A. Humphreys, "A simple theoretical derivation of thermal comfort conditions," *J. IHVE.* **38**, 95 (1971).

40. J. J. Kowalczewski, "A method of evaluating and comparing human thermal stresses in different climates," Mech. & Chem. Eng. Trans. of Inst. Engrs. Aus. MC4 (1), 55 (1968).

41. B. Givoni and R. F. Goldman, "Predicting effects of heat acclimatisation on rectal temperature and heart rate responses as functions of activity, environment, and clothing conditions," *J. Appl. Physiol.* **35** (6), 875 (1973).

42. P. O. Fanger, *Thermal Comfort* (Danish Technical Press, 1970).

43. F. H. Rohles and R. G. Nevins, "The nature of thermal comfort for sedentary man," *ASHRAE Trans.* **77** (1), 239–246 (1971).

44. A. P. Gagge, J. A. J. Stolwijk, and Y. Nishi, "An effective temperature scale based on a simple model of human physiological regulatory response," *ASHRAE Trans.* **77** (1), 247–262 (1971).

45. D. A. McIntyre, *Indoor Climate*, (Applied Science Publishers, 1980).

46. T. Kusuda, J. W. Bean, and P. E. McNall, "Potential energy savings using comfort index controls for building heating and cooling systems," in *Indoor Climate*, Proc. First International Indoor Climate Symposium, edited by Fanger and Valbjorn, Danish Bldg. Res. Inst., 1979.

47. J. A. J. Stolwijk, "Physiological responses and thermal comfort in changing environmental temperature and humidity," in *Indoor Climate*, Proc. First International Indoor Climate Symposium, edited by Fanger and Valbjorn, Danish Bldg. Res. Inst., 1979.

48. V. Olgyay, "Bioclimatic approach to architecture," *BRAB Conf. Report No. 5*, (National Research Council. Washington, D. C., 1953), pp. 13–23.

49. M. A. Humphreys, "Outdoor temperatures and comfort indoors," *Bld. Res. & Practice* **6** (2), 92–105 (1978).

50. A. Auliciems, "Global differences in indoor thermal requirements," *ANZAAS Conf.* (Brisbane, May 1981).

51. C. H. Sprague and P. E. McNall, "The effects of fluctuating temperature and relative humidity on the thermal sensation (thermal comfort) of sedentary subjects," *ASHRAE Trans.* **76** (1), 146–156 (1970).

52. G. Shavit, "Energy conservation and fan systems: floating space temperature," *ASHRAE J.* (Oct. 1977).

53. P. E. McNall. E. T. Pierce, and J. P. Barnett, "Control strategies for energy conservation in buildings," J. P. Pierce Foundation seminar, (Jan. 1978).

54. L. G. Berglund and R. R. Gonzalez, "Application of acceptable temperature drifts to built environment as a mode of energy conservation." *ASHRAE Trans.* **84** (1), (1978).

55. A. G. Loudon and P. Petherbridge, *Possible Economies in Air-conditioning*, BRA (U. K.), Current Paper 48/68 (1968).

56. S. S. Stevens, "To honour Fechner and repeal his law." *Science* **133**, 80–86 (1961).

57. J. C. Stevens and J. W. Hall, "Brightness and loudness as functions of stimulus duration," *Perception and Psychophysics*, **1**, 319–327 (1966).

58. L. E. Marks and J. C. Stevens, "Temporal summation related to the nature of proximal stimulus for the warmth sense." *Perception and Psychophysics* **14** (3), 570–576 (1973).

59. L. E. Marks and J. C. Stevens, "Spatial summation of warmth: influence of duration and configuration of the stimulus," *Am. J. Psychology* **86**, 251–267 (1973).

60. M. Cabanac, "Physiological role of pleasure," *Science* **173**, 1103–1107 (1971).

61. A. Auliciems, "Towards a psycho-physiological model of thermal perception," *Int. J. Biometer.* **25** (ii), 14pp. (1981).

62. P. O. Fanger, N. D. Breum, and E. Jerking, "Can colours and noise influence man's thermal comfort?" *Ergonomics* **20**, 11–18 (1977).

63. A. Auliciems and J. Parlow, "Thermal comfort and personality," *The Bldg. Serv. Engr.* **43**, 94–97 (1975).

64. F. H. Rohles and W. V. Wells, "The role of environmental antecendents on subsequent thermal comfort," *ASHRAE Trans.* **83**, (ii), 21–29 (1977).

65. R. Sommer, *Personal Space, the Behavioral Basis of Design*, (Prentice-Hall, 1969).

66. W. D. Woodhead and P. S. Scanes, *Living in Remote Communities in Tropical Australia*, Pt. 3, "Airconditioning of dwellings in a tropical climate," CSIRO Div. Bldg. Res. report TB-27/3 (1972).

67. L. C. Berglund, "Revised standards on thermal conditions for human occupancy," *Habitat International* **5** (3/4), 525–532 (1980).

68. R. R. Gonzalez and Y. Nishi, "Effect of cool environments on local thermal sensation, discomfort and clothing selection," *ASHRAE Trans.* **82** (1), 76–84 (1976).

69. A. P. Gagge, T. Nishi, and R. G. Nevins, "The role of clothing in meeting F. E. A. energy conservation guidelines," *ASHRAE Trans.* **82** (2), 234–240 (1976).

70. E. Arens et al. "A new bioclimatic chart for environmental design," *Intern. Conf. Bldg. Energy Management (ICBEM) Portugal.* (1980).

71. L. Bànhidi. "Thermal comfort investigations," *Intern. Conf. on Bldg. Energy Management (ICBEM) Portugal.* (1980).

72. R. H. Fox et al.. "Body temperature in the elderly: a national study of physiological, social and environmental conditions," *Brit. Med. J.* **1**, 200–206 (1973).

73. D. H. K. Lee, *Human Climatology and Tropical Settlement*, John Thomson lecture, Uni. of Queensland, 1947.

74. F. H. Markham, *Climate and the Energy of Nations*, Oxford University Press (London 1947).

75. F. I. Sargent, "Tropical neurasthenia: giant or windmill?" *Environmental Physiology and Psychology in Arid Conditions*, Proc. Lucknow Symp. Arid Zone Res. 24, UNESCO (Paris, 1963) 280–313.

76. J. T. Damm and B. Fenelon, "Psychological adjustment to life in the tropics," Proc. *Man and Animals in the Tropics*, Symp. Aus. Acad. of Sc. at University of Queensland, (1956).

CHAPTER 6

EARTH CONTACT BUILDINGS: APPLICATIONS, THERMAL ANALYSIS AND ENERGY BENEFITS

John C. Carmody
Dr. George D. Meixel
Kenneth B. Labs
Lester S. Shen

6.1 ABSTRACT

Earth contact buildings, also referred to as earth sheltered or underground structures, have been identified as one means of passive cooling that can contribute to reducing total energy requirements as well as peak load demands. Unlike most conservation and passive heating or cooling strategies which are singular in their purpose and effect, the use of earth contact represents a wide range of energy-related characteristics, as well as a number of diverse advantages not directly related to energy use. Earth contact is part of the thermal envelope in virtually all buildings, and presents opportunities for energy conservation in residential-scale and larger structures, in complexes of several structures, and in more innovative coupling of buildings to the ground.

To provide an overview of these various facets of earth contact design and research, this chapter is divided into three parts. The first provides general background material on earth contact structures and their application. The second section summarizes key issues and pertinent research related to thermal analysis of earth contact buildings. The final section includes preliminary results of energy performance analysis for components of small earth contact structures.

6.2 EARTH CONTACT STRUCTURES AND THEIR APPLICABILITY

The background material presented in this section establishes the general relevance and wide range of applications of earth contact structures. With respect to energy conservation, understanding and optimizing earth contact systems can contribute to meeting goals for energy efficiency in the following ways:

- All buildings can be made more efficient since the ground is part of the thermal environment to some degree.
- Specific earth contact strategies such as berming, covering, and complete underground placement can be used on a variety of building types.
- Relatively unexplored strategies for enhanced exploitation of earth contact potential hold great promise for improved energy efficiency in underground and above-grade buildings.
- Improved patterns of land use can be created with earth-integrated structures resulting in a more energy-efficient and environmentally sound approach to building and development.

In order to clarify the broader context in which earth contact buildings and the related thermal analysis research should be considered, these issues are discussed in greater detail in this section. First, the diverse advantages of earth contact structures are examined. These benefits, combined with energy conservation potential, form the retionale for earth-integrated buildings. This is followed by general discussions of the potential energy-related benefits and limitations of earth contact structures. Third, potential applications of earth contact systems are discussed.

6.3 THE GENERAL ADVANTAGES OF EARTH CONTACT STRUCTURES

In addition to energy conservation, there are a number of advantages of earth contact structures over conventional above-grade structures that form the rationale for deciding to place a building partially or completely below grade. In most cases, the decision to place a building underground is based on a number of interrelated benefits. These benefits, listed below, make increases in the use of earth contact structures likely, regardless of their energy performance compared to other conservation strategies. It is important to note that while some of these advantages—land-use efficiencies and environmental benefits—do not directly conserve heating or cooling energy in a building, they can result in significant energy savings in the construction and operation of systems at the community scale. While many of the advantages discussed here generally can be attributed to earth contact buildings, some benefits are more applicable to housing and other smaller near-surface structures, whereas others pertain more to larger underground buildings and community systems. The text in this section is adapted from *Earth Sheltered Residential Design Manual* and *Underground Building Design*[1,2]

6.3.1 Visual Impact/Aesthetics

One of the great attractions of the concept of earth sheltering and, for some people, the major reason for designing earth sheltered buildings, is the aesthetic potential of earth shelter design. Building a structure partially or completely below grade with earth-covered walls and roof offers opportunities for integrating a building with the surrounding landscape to a degree that is simply not possible

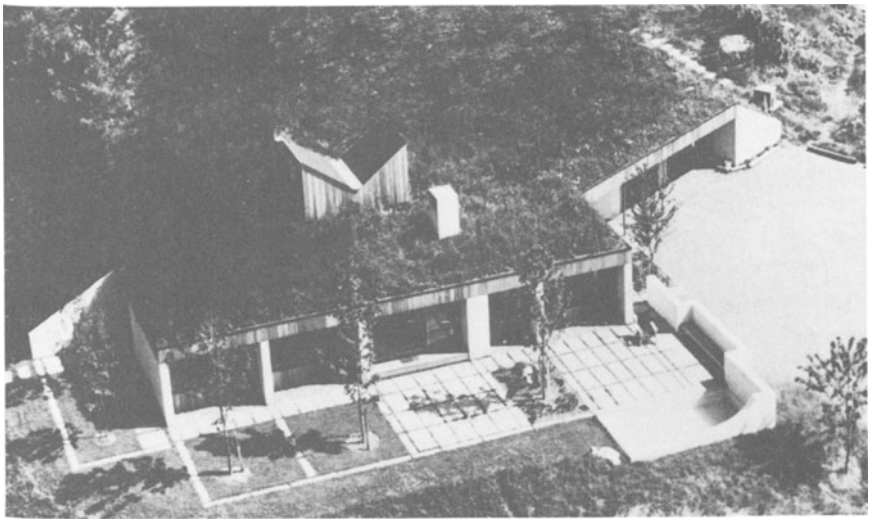


Fig. 6.1: Winston House—Lyme, New Hampshire. Architect: Don Metz.

Fig. 6.2: Pusey Library—Harvard University—Cambridge, Massachusetts. Architect: Hugh Stubbins and Associates.

with conventional buildings. The house shown in Fig. 6.1 is a good example of the unique aesthetic possibilities with earth sheltering.

In a similar way, an underground building is often an appropriate solution for an area with a special historical character where an above grade contemporary building would be disruptive. Examples include libraries, student unions, and bookstores at universities such as Cornell, Yale, Harvard, Michigan, and Minnesota (Fig. 6.2).

Industrial and storage buildings, utilities, transit systems and parking garages are often placed underground because they are quite massive in appearance and are often considered unattractive. There are great benefits for the overall surface environment when certain functions are made less visible.

6.3.2 Preservation of Surface Open Space

Closely related to the advantage of low visibility for underground buildings is the preservation of open space on the surface. By placing a building below grade and allowing the roof to remain as a park or plaza area, no open space is lost. This advantage is particularly important for nonresidential buildings in very built-up areas such as urban commercial centers or college campuses where the open space is a limited but necessary relief in the dense environment.

6.3.3 Land-Use Benefits

A number of potential land-use benefits are associated with earth sheltered buildings. With regard to land use, perhaps the most important characteristic of earth sheltered construction is its ideal suitability to sloping sites. Some sites that might be considered unbuildable using conventional construction may be well suited for earth sheltering. Generally, the steeper the slope, the more densely it can be developed with earth sheltered construction, whereas the inverse is usually true with conventional construction. It is possible to build earth sheltered housing on slopes up to 50%, provided that the soils are appropriate.[3]

Earth contact buildings can also contribute to more efficient land use with the ability to exceed normal setbacks and lot coverage requirements. Usually, underground portions of buildings exceeding these requirements do not violate the intentions of the zoning ordinances and, assuming there are no other circumstances preventing the acquisition of a zoning variance, a partially or completely underground building can be built on a smaller parcel of land than an above grade building can.

Another land-use benefit of earth sheltered construction is the potential development and enhancement of urban sites normally considered undesirable—i.e., sites lacking vegetation or located adjacent to airports, freeways, railways, or industry. Earth sheltered structures can be used on such sites because they can be designed to reduce outside noises and views, as well as enhance and restore vegetation to barren sites. This use of undesirable sites in central locations can result in more compact development with inherent efficiencies in utility and transportation systems.

6.3.4 Environmental Benefits

Many owners and designers of earth sheltered buildings are attracted to the concept as a means of preserving the environment. Techniques for accomplishing this goal range from recessing some of the walls into a hillside, so that the volume of the building has less visual impact, to covering the entire roof with plant material growing out of the earth, thereby consuming carbon dioxide and generating oxygen,

Fig. 6.3: Seward Town Houses—Minneapolis, Minnesota. Architect: Mike Dunn, Close Associates.

absorbing rainwater (rather than draining it off), and providing a habitat for plants and animals.

6.3.5 Noise and Vibration

Another benefit of earth sheltered buildings is the damping effect of the weight of the structure and the earth on noise and vibration. Protection from a specific noise source, such as a busy freeway, can be achieved by placing earth berms adjacent to the freeway and facing the windows and doors of buildings away from it. Protection from outside noises is most useful in dense urban settings, particularly on sites adjacent to freeways, railways, airports, or other localized noise sources. An example of an actual project using earth for noise protection on a site adjacent to a busy freeway is a twelve-unit town house project in Minneapolis, Minnesota (Fig. 6.3).

The isolation from noise and vibration can also work in the opposite fashion for enclosing functions that produce noise. The outside environment would benefit from the isolation of noisy manufacturing facilities or transit systems.

6.3.6 Maintenance

One often-cited advantage of earth sheltered houses is reduced maintenance in comparison to conventional wood-frame houses. This assumption is based on the durability of concrete structures and the protection of much of the exterior surface by the earth mass. Materials buried beneath the earth are not exposed to cycles of freezing and thawing, the intense heat and ultraviolet radiation of the sun, or other severe effects of the local climate.

6.3.7 Fire Protection

Most underground structures are concrete surrounded by soil. These inherently fireproof materials provide a great degree of fire protection and prevent the spread of any fires to or from other buildings. For this reason, underground buildings are likely to sustain less fire damage, a fact reflected in lower insurance rates in some cases. This makes subsurface space suitable for storage of valuable or irreplaceable materials and records. The isolation from the surface also makes underground space appropriate for the storage of volatile or flammable fuels and other liquids.

6.3.8 Protection from Earthquakes

The heavier and stronger structural systems of earth sheltered buildings can also offer protection from the damaging forces of earthquakes. Heavily reinforced walls placed in the earth are generally less likely to be destroyed by these violent lateral forces than are surface structures, which are shaken by their foundation. Although some cracking may occur, the walls and roof can usually be protected from failure by proper engineering.

6.3.9 Suitability for Civil Defense

Just as it provides protection from earthquake forces, the heavier, stronger structure of an earth sheltered building can also protect it from damage from nearby explosions. Moreover, in the case of a nuclear blast, earth berms can offer greater protection from radiation and the lowered levels of infiltration in an earth sheltered house could reduce the effects of fallout. Of course, because exposed walls and windows in an earth sheltered house behave no differently from those in a conventional structure, the house must be properly designed in order to provide effective protection. Larger nonresidential underground structures offer even greater opportunities to provide protection.

6.3.10 Storm and Tornado Protection

Structures that are substantially surrounded by and covered with earth have the advantage over conventional structures of protection from a variety of natural disasters. In areas of the United States subject to frequent high winds, hail storms, and tornadoes, earth sheltered housing is gaining popularity largely because of the protection that is afforded. An atrium-type design, as shown in Fig. 6.4, with all windows facing a courtyard, is likely to be an effective design for protection from a variety of natural disasters.

6.3.11 Security

Because earth sheltered structures often have fewer windows, doors, and walls exposed to the outside, as well as more limited, controlled points of access, they are frequently perceived as being more secure than a conventional above grade building. In addition to being generally desirable to reduce break-ins in all buildings, increased security is a particularly appealing feature for the storage of important records or critical materials such as emergency food and fuel supplies.

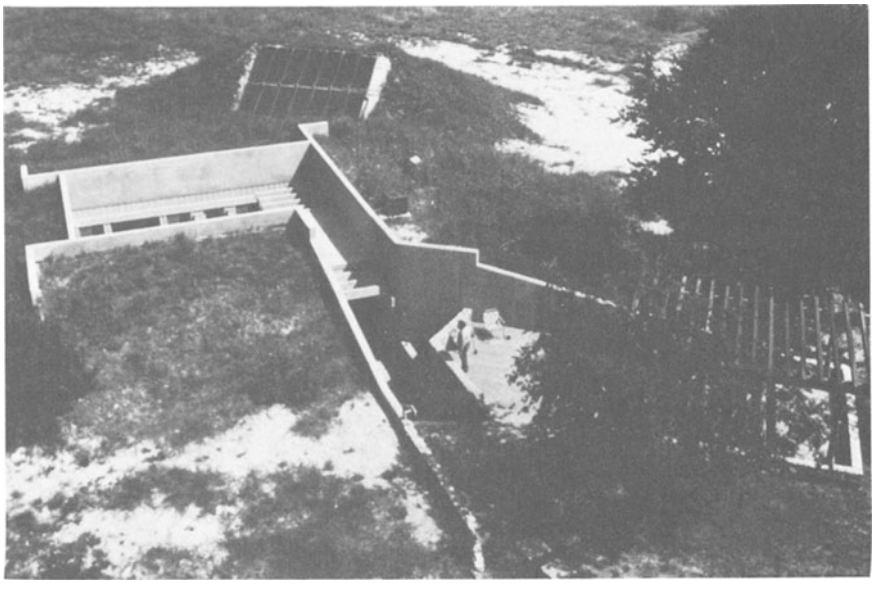


Fig. 6.4: Bordie Residence—Austin, Texas. Architect: Coffee and Crier.

6.3.12 Life Cycle Costs

It has been demonstrated, in a general sense, that an underground building can have a lower life cycle cost than a comparable above grade building, even if its initial construction costs are somewhat higher.[1,4] This is primarily due to the reduced energy and maintenance costs and, in some cases, reduced insurance costs as well. In addition, it can be argued that an underground structure will have a longer useful lifetime than some types of above grade construction.

6.4 POTENTIAL BENEFITS RELATED TO ENERGY CONSERVATION

Structures that are in contact with the earth have unique characteristics that result in a variety of potential benefits with respect to energy conservation for heating and cooling. The word "potential" must be emphasized for two reasons: some of the benefits are not yet easily quantified and the degree to which energy is conserved can vary greatly, depending on the individual design and the climate. Most of the benefits listed below relate to the direct contact of the structure with the earth. Generally, the greater the percentage of surface area in contact with the earth and the deeper the structure penetrates the earth, the more the structure will benefit from these effects in terms of energy conservation. The text in this section is adapted from *Earth Sheltered Residential Design Manual.*[1]

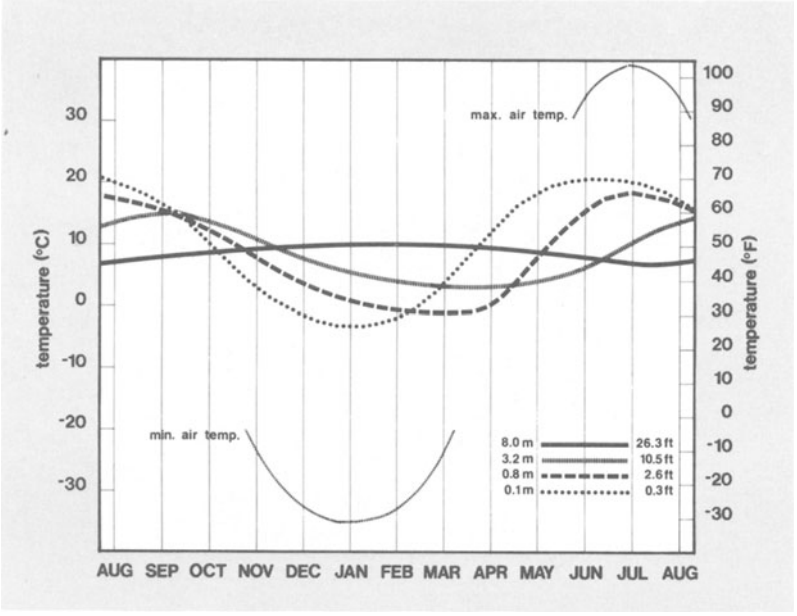


Fig. 6.5: Annual temperature fluctuations—Minneapolis, Minnesota.

6.4.1 Infiltration

By placing earth against walls and on top of the roof, infiltration through those surfaces is completely eliminated, resulting in both heating and cooling load reduction. In addition, if the exposed surfaces of the structure are located away from the prevailing winter winds, the berms can divert wind over and around the structure to reduce infiltration through these areas.

6.4.2 Heat Loss

It is generally assumed that heat loss resulting from conduction will be minimized through surfaces that are in contact with the earth at a reasonable depth below grade. This assumption is based on the fact that temperatures are more moderate below grade than the extremes of above grade air temperatures and that below grade temperatures reach reasonable stability at depths greater than 10 feet (see Fig. 6.5 for temperatures for the Minneapolis area[5]). Ground temperatures will generally approximate the annual average air temperature in warm climates and be a few degrees warmer than the annual average temperature in areas with significant winter snow cover. Urban areas also will raise local ground temperatures.

Although the R-value per unit thickness of earth is not great compared to the R-value of typical insulation materials, the large mass of earth results in a significant reduction and delay in heat exchange with the surface, which in turn results in a net benefit during extremely cold periods.

6.4.3 Cooling

For most climates in the United States, ground temperatures at depths from two to ten feet are lower than the temperatures required for comfort in summer. Thus, when air temperatures within the space of the building exceed the ground temperatures, heat is transferred from the space to the surrounding ground. This phenomenon occurs both by conduction through the envelope and radiation from warmer objects in the space to cooler wall and floor surfaces. Insulation, which separates the interior space from the earth, reduces this cooling effect. Cooling potential is very dependent on the ground temperatures surrounding a building which are affected not only by the climate but by other factors such as ground cover and soil moisture content.

6.4.4 Heat Gain

Earth-covered roofs and earth-bermed walls reduce radiant heat gain from the sun. The massive earth has the capacity to absorb a considerable amount of radiation before it reaches the envelope. In climates with cool summer nights, the earth may release the heat to the cooler air by convection and to the night sky by radiation. A very important component in reducing heat gain by radiation is the use of plant materials. In the process of evapotranspiration, plants can effectively cancel out most, if not all, of the incoming radiation from the sun. This process requires a sufficient level of moisture in the ground to enable the plants to grow.

6.4.5 Daily Temperature Fluctuations

The large mass of the earth surrounding the structure delays and dampens the effect of rapid fluctuations in temperature and typical day-night temperature swings. In addition, because many earth sheltered structures are constructed primarily of concrete and are insulated on the outside, the resultant large mass inside the insulated envelope further stabilizes inside temperature fluctuations. The most significant effect of this phenomenon is that peak loaks are reduced, thus resulting in smaller requirements for heating and cooling equipment. Total energy consumption may also be reduced to some degree, depending on the exact climate characteristics and on the specific design and operation of the house. A final effect associated with reduced temperature fluctuations is that temperatures may remain relatively stable for many days with no heat input. This phenomenon can be a great benefit during a power outage, especially for a building situated in a remote location.

6.4.6 Seasonal Temperature Lag in Ground

As indicated in the chart of ground temperatures (Fig. 6.5), at greater depths the temperature fluctuations on a seasonal basis are not only dampened but are also delayed. This phenomenon presents the potential for saving heating energy because the warmer ground temperatures from summer are carried into the fall and winter, thus reducing heating needs during the early months of the heating season. Conversely, the cooler ground temperatures of the winter, which are carried into the spring and summer, reduce cooling loads.

6.5 POTENTIAL LIMITATIONS RELATED TO ENERGY CONSERVATION

In addition to energy-related benefits, earth sheltered buildings have potential limitations. Although the energy-related limitations of earth sheltered buildings represent a possible reduction in the overall benefits of these structures, they need not negate the benefits entirely. First, neither benefits nor limitations are quantified here, and both can exist to varying degrees. Second, the effect of the potential limitations depends on specific climate conditions as well as on the individual design of the structure. After the potential limitations are identified, it is desirable to design the structure to minimize the effects of the drawbacks while maintaining the positive benefits.[1]

6.5.1 Structural and Economic Limitations

The relatively great weight of earth on flat roofs, combined with the lateral earth pressures on buried walls (these pressures increase with depth), require heavier and more expensive structures than is typical for most housing. This limitation has given rise to two construction methods to avoid the problem. The first is the use of earth-bermed walls and a conventional roof, which is less expensive than a more heavily structured earth-covered roof. The other approach uses structures such as shells or domes, which can support the heavier loads much more efficiently, resulting in lower costs and larger amounts of earth placed over the structure. An example of a house using a metal vaulted structure is shown in Fig. 6.6. In spite of the potential cost advantages of these alternatives, most earth sheltered houses built thus far have flat, earth-covered roofs.

6.5.2 Requirements for Openings

A windowless chamber buried in the earth is unsuitable for certain functions such as housing. The various requirements for access and egress, window openings, and for other exposed portions of the building envelope diminish the area of the envelope in contact with the ground as well as the depth of the structure. For these smaller residential buildings, access must occur from grade; and windows, courtyards, and other openings are only partially set into the earth. Numerous openings on various sides of a structure break the continuity of the earth mass surrounding the house, thereby diminishing the energy-conserving benefits related to structure-earth contact.

6.5.3 Slow Response

Although the mass of the concrete structure within the envelope and the surrounding earth mass are assets in many ways, they result in a structure that cannot respond rapidly to changing conditions. This means that some energy-conserving strategies, such as night thermostat setback, may not work effectively due to an unsatisfactory time lag.

Fig. 6.6: Clark-Nelson House—River Falls, Wisconsin. Architect: Michael McGuire.

6.5.4 Ground Temperatures

Ground temperatures are related to annual climate conditions. During both heating and cooling seasons, a lack of useful ground temperatures can diminish the benefits of a structure's contact with the earth. The parts of the country requiring the most heat have the lowest ground temperatures and the regions with the greatest cooling requirements have the highest ground temperatures. Nevertheless, both heating and cooling season benefits are evident in a variety of climates. Various techniques for ground temperature manipulation could improve the benefits significantly.

6.5.5 Drawbacks of Seasonal Time Lag in Temperatures

Although ground temperatures are often more favorable than outside temperatures throughout the year, they can also be less favorable at certain points in the seasonal cycle. For example, in a cold climate, the ground at a depth of ten feet reaches its lowest temperature in spring. Although this time lag is beneficial in the summer, it is detrimental in the spring because the air temperature is warming faster than the ground, resulting in a higher heating requirement for the building. In a warm climate, benefits from cooler air temperatures in autumn are greater than those from the ground, which has been warmed to its peak by the end of summer and early fall. Two mitigating circumstances may diminish the negative effects of these phenomena. First, because the ground around the structure is not at the same temperature as natural undisturbed ground, these effects may not be

as important in a more constant and usually warmer zone of ground temperatures. Second, strategies can be employed to counteract the potential negative effects of the seasonal time lag; for example, solar gain in the spring or ventilation in the autumn may offset the limitations of the time lag while maintaining its benefits.

6.5.6 Heating/Cooling Compromises

Although a below grade ground environment offers potential benefits in both the heating and cooling season, maximizing these benefits requires insulation in the heating season, but direct earth contact with no insulation in the cooling season. In some cases, the necessary compromise prevents optimizing for either condition alone.

6.5.7 Condensation

Because the surrounding earth is almost always cooler than the indoor air temperature, condensation on interior surfaces may occur, especially in summer. For condensation to occur, wall temperatures must be below the dew point temperature. In such cases, it is likely that dehumidification would be desirable to provide comfort. This potential problem is mitigated by the fact that as the earth and walls adjacent to the building warm up, the temperature difference between indoor air and walls is not very great. Problems with condensation are, of course, dependent on the local climate as well as on the manner in and degree to which outside air is introduced into the building.

6.5.8 Evapotranspiration

Evapotranspiration, the process by which plant materials dissipate incoming solar radiation so that it never warms the ground or the structure, depends on the availability of moisture. In hot, dry climates, the full benefit of this effect from an earth-covered roof may be difficult or costly to realize because water may have to be supplied to the plants by irrigation.

6.5.9 Indoor Air Quality

In smaller, residential, earth contact structures, the extremely well-sealed exterior envelope can result in an air exchange rate that is less than desirable for health and comfort. In addition to odors, there is concern over exposure to a wide range of detrimental pollutants, including radon. This problem can be solved, however, without sacrificing the energy saving benefit of low infiltration by the use of efficient heat exchanging ventilation equipment.

6.6 APPLICATION OF EARTH CONTACT SYSTEMS

Nearly all buildings are in contact with the earth to some extent. For many buildings, the ground is a major component of the thermal environment. For example, in a conventional house with a basement, 40% to 45% of the exterior surface is in contact with the earth. A similar percentage can be found in many commercial,

industrial, or warehouse structures—particularly large, single-story, slab-on-grade buildings. If the interaction with the ground environment is not understood in these buildings, opportunities for energy-efficiency can be missed and significant quantities of energy wasted. For housing in colder climates, inadequately insulated foundation walls and basements can account for as much as one-third of the total heat loss. Some of the research findings in the last section of this chapter pertain to conventional basements, slabs, and foundations.

6.6.1 Residential Structures

One of the most common applications of earth contact systems and the focus of most heat transfer research is in relatively small residential structures. Estimates of the number of earth sheltered housing structures in the United States in 1982 range from 4,000 to 8,000. While examples exist in almost all regions of the country, the most concentrated activity is in the upper midwest and central states; considerable activity is also found in the northeast and southwest. These one- and two-story structures are set into hillsides, placed completely below grade around courtyards, or placed partially below grade with earth bermed around the walls. In some cases, the roof is covered with one to three feet of earth; in other cases, it is not. These structures are distinguished from conventional houses with basements by the fact that the earth is in contact with 50% of the exterior envelope or more, and that the building is intentionally integrated with the earth for one or more reasons.

The advantage of placing earth in contact with these structures is that the heating and cooling loads are usually dominated by transmission, infiltration, and radiative losses and gains through the exterior envelope. The energy conserving characteristics of earth contact structures are well suited to this type of structure. On the other hand, the drawback of these small structures is that they are near the surface and require a certain amount of exposure for access, view, and light. These characteristics can prevent maximizing some of the energy-efficient opportunities below grade; however, this is seldom a real drawback if properly designed. Most of the preliminary research findings in the last section of this chapter pertain mainly to this category of structure.

6.6.2 Nonresidential Structures

Nonresidential buildings represent a wide variety of heating and cooling loads based on variations in occupancy, internal loads, and requirements for windows and access. Virtually all types of nonresidential buildings have been built partially or completely underground in the United States. For many functions, locations 30 to 100 feet beneath the surface have been used. This results in the potential for taking advantage of totally constant temperature and humidity with virtually no contact with the surface climate. Although such examples may seem unusual, hundreds of acres of offices, warehouses, and the largest cold storage area in the country are located over 100 feet beneath the surface of Kansas City, Missouri; a similar extensive development has been considered in Minneapolis. This type of space is not limited to these two locations, and in certain circumstances such development is less expensive than that above grade, with greatly reduced operating costs.

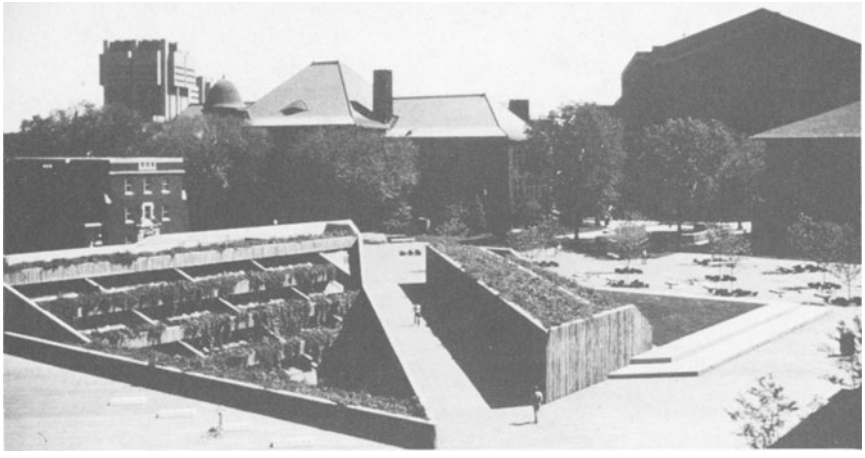


Fig. 6.7: Williamson Hall—University of Minnesota—Minneapolis, Minnesota. Architect: BRW Architects, Inc.

Until recently, there has been little emphasis on building underground to conserve energy. Underground buildings have been constructed for many years for a wide variety of other planning, environmental, and economic reasons. In some cases, unforeseen reduction in heating and cooling energy, as well as cost savings in construction, have resulted. For example:

- A comparison of the above and below grade warehouses and refrigerated storage in Kansas City indicates mechanical equipment installation costs for below grade are one-fourth of those for above grade structures. Moreover, operating costs for the below grade structures are one-tenth of those of the above grade.[6]
- Two elementary schools in Reston, Virginia, were constructed underground resulting in operational costs that are 50% to 75% less than comparable above grade schools. They were built for exactly the same construction cost as above grade schools.[7]
- An analysis of below grade office and warehouse buildings was recently completed for the U.S. Navy by Setter, Leach, and Lindstrom of Minneapolis. The study concluded that earth berms around buildings in many climates not only saved energy but resulted in construction cost savings due to the reduction in mechanical equipment, reduced depth of frost footings, and reduction of exterior finishing materials.[8]
- Actual energy use for Williamson Hall, a 95% underground office and bookstore building at the University of Minnesota, was 47,000 Btu per square foot per year in 1981. This performance is considered good by most standards for buildings of this type in this climate and is notable in that it is approximately 30% better than the estimated performance (see Fig. 6.7).[2,9]

6.6.3 Clusters of Buildings Employing Earth Contact

The energy benefits of earth contact structures are often examined in the context of a single building placed into the natural ground environment. The ground environment is likely to be considerably different with several earth contact structures in relatively close proximity. The buildings tend to confine the earth mass between them, permitting it to act more effectively as a storage medium, and generally would result in raised ground temperatures over a period of time. While this may have negative implications for earth contact cooling, transmission losses during the heating season could be reduced even further than with a single earth contact structure.

6.6.4 Improved Exploitation of Earth Contact Potential

The ground not only represents a moderate climate environment into which a building can be placed, but also an enormous thermal reservoir available for seasonal or long-term energy strategies. With the exception of a few seasonal storage research schemes, such as ACES (Annual Cycle Energy System) or aquifer storage, the emphasis on thermal mass storage research has been on its effect on diurnal cycles or periods of up to a few days. With its moderate temperatures and high thermal mass (by virtue of quantity) the ground represents both a source of low-grade heat energy which can be heat pumped at high coefficients of performance to levels suitable for heating as well as a direct source of cooling.

Examples of effective uses of the ground temperature resource include groundwater heat pumps, aquifer storage, and earth tubes. More innovative examples include a closed-cycle, ground-coupled heat pump system developed in Sweden which uses a shallow, buried, inexpensive plastic pipe grid which circulates glycol and water. During the heating season, the heat pump extracts heat from the ground that has been stored from spring and summer. Heating efficiency can be enhanced by using solar collectors to circulate heat into the ground. The low temperatures enable very efficient solar collection and the large thermal mass allows the acceptance of solar heat whenever it is available.

In addition to improved coupling of the building to the ground, energy savings with passive cooling in earth contact structures can potentially be enhanced by various techniques of ground temperature modification. These include reduction in deep ground temperatures, various surface treatments and insulation, manipulation of soil thermal properties with moisture, and greater displacement of the seasonal time lag.[10,11] Clarifying and developing such opportunities is only beginning. However, these techniques may represent an additional avenue to increasing the benefits and applicability of earth contact structures.

6.7 THERMAL ANALYSIS OF EARTH CONTACT BUILDINGS

The primary goals of earth contact thermal analysis are to understand and predict the heat transfer at earth contact surfaces so that its impact on building energy performance can be assessed. Given these capabilities, the design of the earth contact portion of the building configuration and the associated insulation

placement could be optimized from estimates of seasonal energy requirements and the heating, ventilation, and air-conditioning system could be properly sized. Also, the determination of the temperatures of the earth contact components is necessary for the evaluation of (1) human comfort conditions inside the building, (2) the potential for moisture condensation, and (3) the potential for freeze/thaw or frost heave structural damage.

These goals are being pursued through theoretical studies often aided by computer-based numerical techniques, well-controlled laboratory experiments, full-scale field tests, and monitoring of occupied buildings. Theoretical and experimental analyses work together to increase understanding of the operative physical processes, the basis for accurate detailed predictions as well as for appropriate engineering approximations.

Factors that can strongly influence the magnitude of the earth contact energy exchange from buildings and should therefore be considered in the thermal analysis include:

(1) Climate: air temperature, solar, and longwave radiation, rainfall, wind speed, snow cover.

(2) Site: type of surface cover (grass, gravel, pavement), shading, drainage, groundwater level, neighboring buildings, orientation, soil thermal properties.

(3) Building:

  (a) Configuration, such as amount of building in contact with the earth and plan layout.

  (b) Materials, such as concrete block, poured concrete, all-weather, wood-frame construction, and so forth.

  (c) Insulation placement, such as placement along roof, wall, and floor, inside insulation versus outside insulation, possible extension of insulation into soil mass.

  (d) Operating conditions, such as interior air temperature, humidity, direct solar gain.

Because many of these factors vary both spatially and temporally, accurate predictions of earth contact heat transfer are a formidable task. Subsequent discussion in this section is intended to provide a brief perspective on each of the major effects which should be considered in the analysis.

(1) Climate

For above grade portions of buildings, maintenance of internal thermal comfort is related to heat transfer from current or recent climatically induced temperature differences, wind speed, relative humidity, and solar and longwave radiation. For the below grade portion of a building, the thermal mass of the earth averages the above grade climate changes over time, as outlined above and discussed by several authors.[11-13] While analysis of an earth covered roof may be carried out on an hourly basis with only its recent history

affecting the current heat transfer, the thermal environment near a basement wall may not change significantly for several days and accurate calculations may require examination of climatic conditions for an entire year. Because the building will gain or lose energy to the earth, its presence alters the temperatures of the ground.

Although a reasonable estimate for the mean annual ground temperature, T_m, is the mean annual air temperature TA (e.g., Columbus, Ohio: $T_m = 53.0°$ F, $TA = 52.5°$ F), this is not sufficiently accurate for northern climates when there is snow to reduce winter heat loss from the earth (Minneapolis, Minnesota: $T_m = 48°$ F, $TA = 43°$ F) or for southern climates where there is more incident solar radiation to increase summer heat transfer to the earth (Tucson, Arizona: $T_m = 75°$ F, $TA = 67°$ F). In addition, climatic variations in seasonal rainfall may alter the ground thermal properties.

(2) Site

Different surface treatments at the ground-atmosphere interface may generate significantly different earth temperatures. Experiments at the National Bureau of Standards for five different surface covers show that a grass cover results in cooler summertime earth temperatures than under a bare earth surface, while a black asphalt surface results in markedly higher earth temperatures.[14] Measurements of the magnitude of heat flux at different points on the earth-contact perimeter of an underground building wall at the University of Minnesota correlate well with changes in the surface cover.[15]

The type of ground cover alters the energy balance at the surface of the ground through reflection and absorption of incoming radiation as well as by its influence on energy exchange through conduction, convection, and evaporation. Variations in the slope of a site and shading of the ground surface by trees, hills, or other structures may also alter the effective incident radiation. For example, soil temperature mesurements near the north side of the house (shaded by the house itself) are several degrees cooler than at a comparable location on the south side.[16]

Site drainage and the local water table may affect the soil moisture content which, in turn, has a strong effect on soil thermal conductivity.[17,18] The thermal conductivity of the soil is directly related to the earth contact heat transfer from adjacent surfaces.[15]

(3) Building

Because the earth contact portion of a building may exchange energy with the surrounding earth, thereby modifying the underground thermal environment, the energy performance of the earth contact components of a building may be coupled to each other in more ways than the above grade portions. For example, insulating the upper portion of an earth contact wall may increase the heat loss from the lower wall because less heat is transferred through the upper wall to the soil. Consequently, the surrounding earth temperatures are lower and the temperature gradient across the uninsulated lower wall is

greater than it was without insulation on the upper wall, thus increasing the heat loss.

In addition to the strong radiative coupling between interior room surfaces that exists for all buildings, there may be significant conduction vertically through concrete block or poured concrete walls, induced by the vertical temperature differences maintained by the surrounding soil.[19-21] With hollow core concrete block, there is also the potential for increased vertical energy transport due to convection in the enclosed air space within the wall.[19,22]

Because an earth contact floor is implicitly associated with an earth contact wall, the thermal performance of the floor may be enhanced by that association even without considering direct thermal interaction. As the amount of wall earth contact is increased by submerging the wall below grade, the floor is simultaneously moved deeper. The new thermal environment of the floor, which is further beneath the surface, may be cooler during periods when summer heat loss is helpful or less cold during winter months.

6.8 EXAMPLES OF CURRENT EARTH CONTACT ANALYSIS

Heat losses from underground spaces have traditionally been estimated by manual methods. All those in use prior to 1982 were devised to determine design winter heat losses from basements and are neither very accurate nor capable of estimating total annual and transient heat exchanges with the ground. This status of the manual methods is largely due to the lack of interest in occupying underground spaces in the past and the fact that house basements, as such, represent a small fraction of the whole house heating load.

An assortment of sophisticated mainframe computer programs has been developed to meet the need for more accurate and detailed estimates of energy exchanges with the ground. The majority of these use either implicit or explicit finite difference, finite element or boundary element techniques (see Table 6.1). The cost of computation and the past lack of availability and lack of interfacing of these with other building load analysis programs have limited them almost exclusively to research use. Recent developments in devising an earth coupling subroutine to be incorporated into the DOE-2 energy analysis computer program have removed the latter limitation, but the cost of running the analyses and the small number of users of mainframe programs still, as a practical matter, limit their applications to buildings with large design budgets, and to parametric studies for research and the development of design guidelines.

Although past manual methods were very crude, these are being improved through use of detailed simulation studies. Inasmuch as the thermal properties of the soil vary throughout the year and are difficult to estimate even as average values, very sophisticated analysis methods may not be warranted for many applications. An emerging generation of manual methods which correlates energy exchanges with variations in ground temperature throughout the year promises to provide reasonably good accuracy and ease of use. The problem of estimating heat losses to

Table 6.1

Examples of mainframe computer programs for analyzing heat transfer from underground structures

Explicit Finite Difference	
Kusuda and Achenbach[23]	McBride[26]
GROCS[24]	Speltz[27]
MITAS[25]	HEATING V[44]
Implicit Finite Difference	
Shipp[28]	Underground Space Center[30]
Szydlowski[29]	HEATING V[44]
Finite Element	
ADINAT[31]	Wang[19]
TEMPFEM[32]	HEAT[33]
2-D Response Factor	
Ceylan and Myers[34]	Underground Space Center[30]

the ground are revealed in the manual calculation methods that have been developed over the years. A review of these follows.

6.8.1 Manual Methods

The potential sinks for heat loss from an underground structure consist of the ground surface at daily average temperature, T_{gs}, and the deep ground, assumed to have a constant temperature, T_m, equal to the mean annual ground temperature. The deep ground temperature, T_m, is indicated by the temperature of water drawn from nonthermal wells. The ground surface temperature, T_{gs}, varies periodically over the course of the year; it is loosely approximated by the long term daily average air temperature. Throughout temperate regions, the ground surface is the dominant winter heat sink for losses from basements and other underground spaces built near the surface, so that losses to the deep ground can be ignored. Calculation of winter design heat losses can, therefore, be treated as a simple, two-dimensional steady-state problem. During the overheated season, the deep ground is the only available sink; interactions with the surface still must be taken into account, however, since it is a source of gain that acts in opposition to losses to the deep ground. The

Table 6.2

Thermal energy sinks assumed in manual calculation methods

Method	Energy Sink (designated temperature)		
	Deep Ground (T_m)	Ground Surface (T_{gs})	Adjacent Ground ($\overline{T}_{a,b}$)
A. Estimation of Winter Design Heat Loss Only			
1. "Old ASHRAE"	SS	na	na
2. Elliot and Baker	SS	SS	na
3. Boileau and Latta	na	SS	na
4. Wang	na	SS	na
B. Estimation of Heat Loss Throughout the Year			
5. Mitalas	SS	VR	na
6. *F* Factor	na	na	VR

Notes: SS: sink temperature is assumed constant.

VR: sink temperature is variable.

na: method does not consider heat loss to this type of sink.

combined influence of the two sinks may be represented by a fictional third sink; this can conveniently be taken as the block of soil horizontally adjacent to the wall, having an undistrubed (field) temperature, $\overline{T}_{a,b}$, averaged over its height from depth a to b. All methods for calculating heat losses from underground spaces can be classified according to the sink or combination of sinks assumed in the procedure (see Table 6.2). These are briefly reviewed in the following discussion.

6.8.2 "Old ASHRAE" Method

The "old ASHRAE" method here refers to the procedure described in the ASHRAE *Handbook of Fundamentals* prior to the 1977 volume. It computes the heat loss from basement walls and floors by means of empirically derived conductances (U-values) and the difference in temperature between the indoors and the deep ground:

$$q = U \cdot (T_i - T_m)$$

where $U = 0.2$ Btu/(ft^2·h·°F) for walls and 0.1 Btu/(ft^2·h·°F) for floors; q = average rate for heat loss [Btu/(ft^2·h)]; T_i = indoor air temperature (°F); and T_m = deep ground temperature (°F).

The conductances were inferred and averaged from observations made by Houghton, Taimuty, Gutberlet, and Brown in an earth covered test room near Pittsburgh in the early 1940s.[35] The method was devised for determining winter

design heat losses from uninsulated walls. It is now obsolete and lacks even the sophistication of the observations on which it was based. That experimental work remains an excellent record of monitored annual conditions and is one of the few studied cases of condensation occurrence in an earth covered building.

6.8.3 Method of Elliot and Baker

In 1960, Elliot and Baker published a method for computing winter design basement heat losses based on two-dimensional studies using an electrical analog technique.[36] Surface and deep ground sinks were modeled for a six-foot-deep basement. The steady-state heat loss was expressed as a function of the temperature difference between the indoors and the ground surface. The thermal effect of the ground was represented by the thermal conductivity of the soil and the two-dimensional geometry of the system as a shape factor, S:

$$q = S \cdot k_s \cdot (T_i - T_{gs})$$

where q = lineal heat loss [Btu/(ft·h)]; S = 1.5 for an uninsulated concrete wall (dimensionless) and 1.25 for an uninsulated concrete block wall (dimensionless); and k_s = soil thermal conductivity [Btu/(ft·h·° F)].

In this analysis, the deep ground temperature is assumed to be constant enough throughout the northern regions in which the calculation would be made so that losses to deep ground are embodied in the shape factor, S. The method offers a significant improvement over the "old ASHRAE" method by recognizing the role of both surface and deep ground thermal sinks and the important of soil thermal conductivity. The electrical analog technique could have been applied to summer conditions, but this was not of interest to the researchers at that time.

6.8.4 Method of Boileau and Latta

Boileau and Latta published their method for estimating winter design heat losses in 1968.[37] It appeared in the ASHRAE *Handbook of Fundamentals* nine years later (1977), and is probably the procedure most familiar to users today. Boileau and Latta asserted that two-dimensional heat conduction theory could be applied to an assumed winter steady-state condition of heat loss to the surface (disregarding losses to the deep ground). In their words:

> In the simple condition of steady heat flow between two parallel sources at different temperatures (as for a wall above grade), the paths of heat flow are parallel lines at right angles to the surfaces. When the two surfaces are not parallel but at an angle to one another and separated by a homogeneous material, the paths of heat flow are still parallel, but circular, with the center of the circles at the intersection of the two surfaces. From the symmetry of the situation it may be seen that the isotherms will be radial lines and that for equal increments in temperature they will be spaced an equal angular distance apart. (See Fig. 6.8a.)

By this model, heat losses from the wall at any depth z can be expressed,

$$q_z = \frac{(T_i - T_{gs})}{\frac{r_s \pi z}{2} + R_w} \tag{6.1}$$

where q_z = heat loss rate at depth z [Btu/(ft^2· h)]; z = depth (ft); r_s = thermal resistivity of soil [(ft·h·° F)/Btu]; and R_w = thermal resistance of wall [(ft^2·h·° F)/Btu].

The term $\pi z/2$ represents the length of the path that heat leaving the wall at depth z takes to the surface (i.e., the circumference of one-quarter circle of radius z). It may be thought of as an effective resistance, ER_z, of the soil at depth z. The effective resistance $ER_{a,b}$ of the entire pie-shaped wedge of soil adjacent to the wall through which heat passes to the surface can be expressed by integrating ER_z with respect to depth z between the limits of the top of the wall a and its base b:[38]

$$ER_{a,b} = \int_a^b ER_z dz = \int_a^b \frac{r_s \pi}{2} z dz = \frac{r_s \pi (b^2 - a^2)}{4} \tag{6.2}$$

This can be expressed as an average resistance $\overline{ER}_{a,b}$ for the height of the wall by dividing by the height $(b - a)$,

$$\overline{ER}_{a,b} = \frac{r_s \pi (b^2 - a^2)}{4(b - a)} \tag{6.3}$$

In the special case of a basement, $a = 0$, so $\overline{ER}_{0,b} = r_s \pi b/4$.

The lineal heat loss for the profile (a, b) from this simplified version of Latta and Boileau's method can be written [38]

$$q_{a,b} = \frac{(T_i - T_s)(b - a)}{\frac{r_s \pi (b - a)}{4} + R_w} \tag{6.4}$$

The reasonableness of the method for estimating design winter heat loss has been indicated by subsequent computer simulations[20,39] and experimental studies which show that nearly contstant heat loss rates prevail throughout the winter in northern regions.[40] It can be used to tailor insulation configurations for the winter condition, but it cannot be used to compute heat losses during other seasons or to optimize insulation placements over the annual cycle.

6.8.5 Method of Wang

Wang[19] has suggested that the heat loss from insulated and partially insulated walls under steady-state conditions tends to travel in straight lines toward the surface instead of taking the circular route described by Boileau and Latta (see Fig. 6.8b). This is explained by the fact that the wall insulation reduces the warming

of the soil and that the horizontal field of temperature disturbance is shortened (or, alternatively, that the natural winter upward flow of heat exerts greater influence on the building soil system than in the uninsulated case, in which it is comparatively unimportant). This postulation implies that the method of Boileau and Latta increasingly underestimates the heat transfer as the wall resistance increases. Otherwise, Wang assumes the same conditions as Boileau and Latta and the calculation procedure is identical in concept. Although Wang's method is described in ASHRAE *1981 Handbook of Fundamentals*, no means is given for determining the lengths of the linear paths. It is, therefore, an incomplete method, having the nature of a commentary on that of Boileau and Latta. It can be noted that as the wall resistance, R_w, increases, the effective resistance of the soil becomes less and less important, as do errors in its estimation.

6.8.6 Method of Mitalas

Mitalas[41] has developed a two sink method for predicting transient heat losses throughout the year by relating these to the simultaneous temperatures of the deep ground and the ground surface. His calculation is made for each of five segments of the basement wall-floor section, and has the general form (see Fig. 6.8c)

$$q_n = S_n(T_i - T_m) + V_n(\sigma_n)T_{gs} \tag{6.5}$$

where q_n = average heat loss rate through segment n (W/m^2); S_n = steady-state conductance between the interior and the deep ground through segment n [$W/(m^2 \cdot °C)$]; V_n = periodic conductance between the interior and the ground surface through segment n [$W/(m^2 \cdot °C)$]; and σ_n = amplitude attenuation factor for segment n.

The conductances S_n and V_n are described by Mitalas as shape factors, although he does not use the term in the same way as Elliot and Baker. The value of T_{gs} is assumed to vary sinusoidally throughout the year according to the idealized relation

$$T_{gs} = T_m - A_s \cos \frac{360(t - t_o - L_n)}{365} \tag{6.6}$$

where A_s = annual ground surface temperature amplitude (°F); 360/365 = conversion for days into degrees; t = time of year (days); t_o = phase constant (days); and L_n = lag time for segment n (days).

Mitalas includes the term L_n in the argument of the cosine to account for the lag in the heat flux wave at the basement section segment n behind the phase of the surface temperature wave.

Values for the factors S_n, V_n, σ_n, and L_n and for the corner loss allowance have been determined for two different sets of soil conditions and are presented in tabular form for hand calculation.[41] He also discusses the results of comparisons with monitored house basements in Canada and the limitations of any effort to predict heat losses to soils of varying moisture content and unknown groundwater conditions.

6.8.7 *F* Factor Method

The problem of computing heat transfer between the wall and two different sinks can be simplified by using as the driving temperature the undisturbed temperature of the soil profile horizontally adjacent to the wall. This can be expressed as the integrated average temperature of all depths between the top of the wall at depth a and the base of the wall at depth b. The average profile temperature, $\overline{T}_{a,b}$, varies sinusoidally over the course of the year. It can be computed from the expression

$$\overline{T}_{a,b} = T_m + \frac{A_s}{(b-a)\delta\sqrt{2}}\,(e^{-z\delta})\cos\frac{360(t - t_o - zL - 45.6)}{365}\Bigg|_a^b \tag{6.7}$$

where a = upper limit of soil profile (ft); b = lower limit of soil profile (ft); and z = the depth below grade (ft).[13]

The logarithmic decrement, δ, represents the quantity

$$\delta = \left(\frac{\pi}{365\alpha}\right)^{\frac{1}{2}} \tag{6.8}$$

and the lag time, L, is given by the expression

$$L = \frac{1}{2}\left(\frac{365}{\pi\ \alpha}\right)^{\frac{1}{2}} \tag{6.9}$$

where α is the thermal diffusivity of soil (ft^2/day).

The rate of heat loss to the soil is always less than predicted by simply taking the difference between the indoor and undisturbed ground temperatures, however, since the soil stores heat loss from the building. A correction factor F is introduced to account for the reduction in temperature difference caused by heating of the soil. F factor calculations take the form for the lineal heat loss rate (see Fig. 6.8d),

$$q_{a,b} = F(T_i - \overline{T}_{a,b})\,\frac{b-a}{R_w} \tag{6.10}$$

where $q_{a,b}$ = lineal heat loss rate of wall between depths a and b [Btu/(ft·h)]; a = upper limit of soil profile (ft); b = lower limit of soil profit (ft); F = a correlation factor (dimensionless); and R_w = thermal resistance of wall section [(ft^2·h·°F)/Btu].

A theoretical framework based on the solution for a buried heat exchanger was described by Akridge, but portions of it remain unsolved.[42] Akridge's work reveals that the factor F is closely related to the thermal resistance of the wall and the thermal conductivity of the soil and, to a lesser degree, to the thermal diffusivity of the soil. This relationship has been confirmed by Poulos,[43] who used a series of parametric computer simulations to arrive at a table of F factors (Table 6.3) using

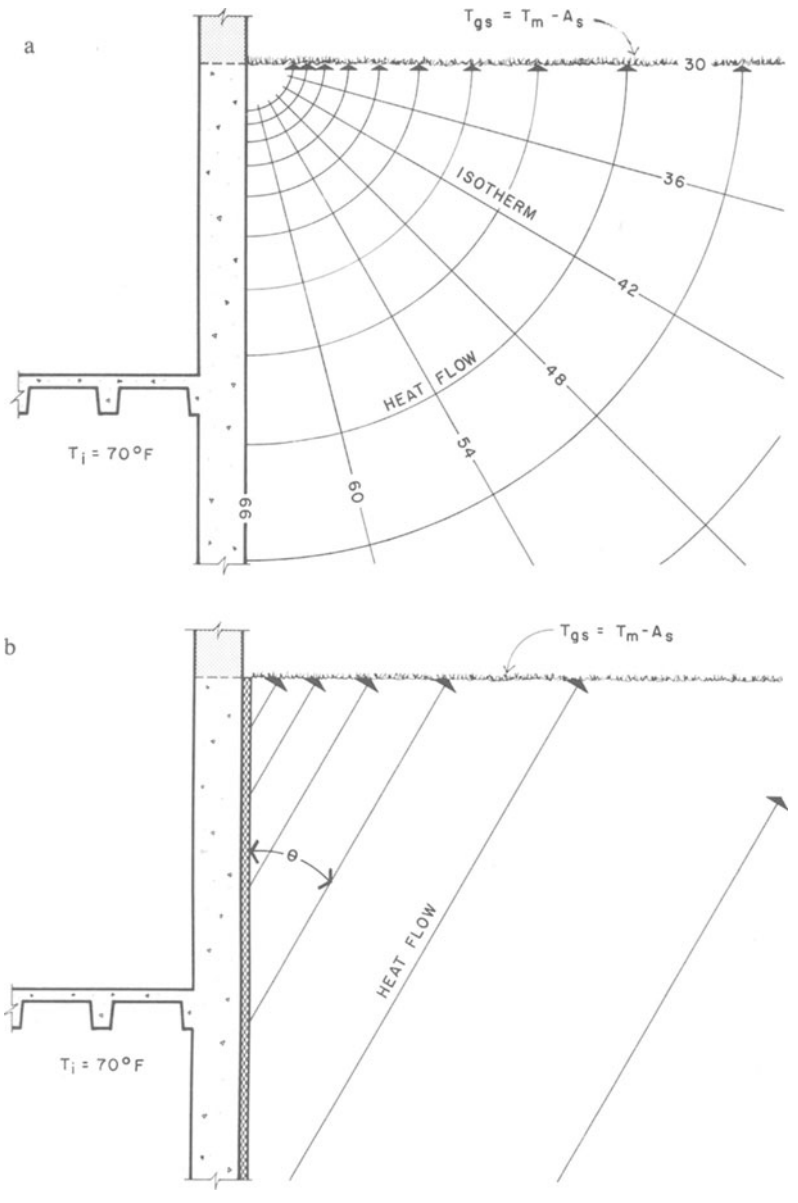


Fig. 6.8: a) The method of Boileau and Latta assumes a steady-state condition in which heat travels along circular paths from the wall to the surface. The heat flow path length increases with depth. It is assumed that no heat travels vertically in the wall. **b)** Wang supposes that heat flows in parallel straight lines (instead of arcs) from insulated walls to the surface in the steady-state condition. The path length depends on the angle θ it makes with the wall. According to Wang's reasoning, angle θ decreases with increasing wall R value, but no means of relating the two has been offered.

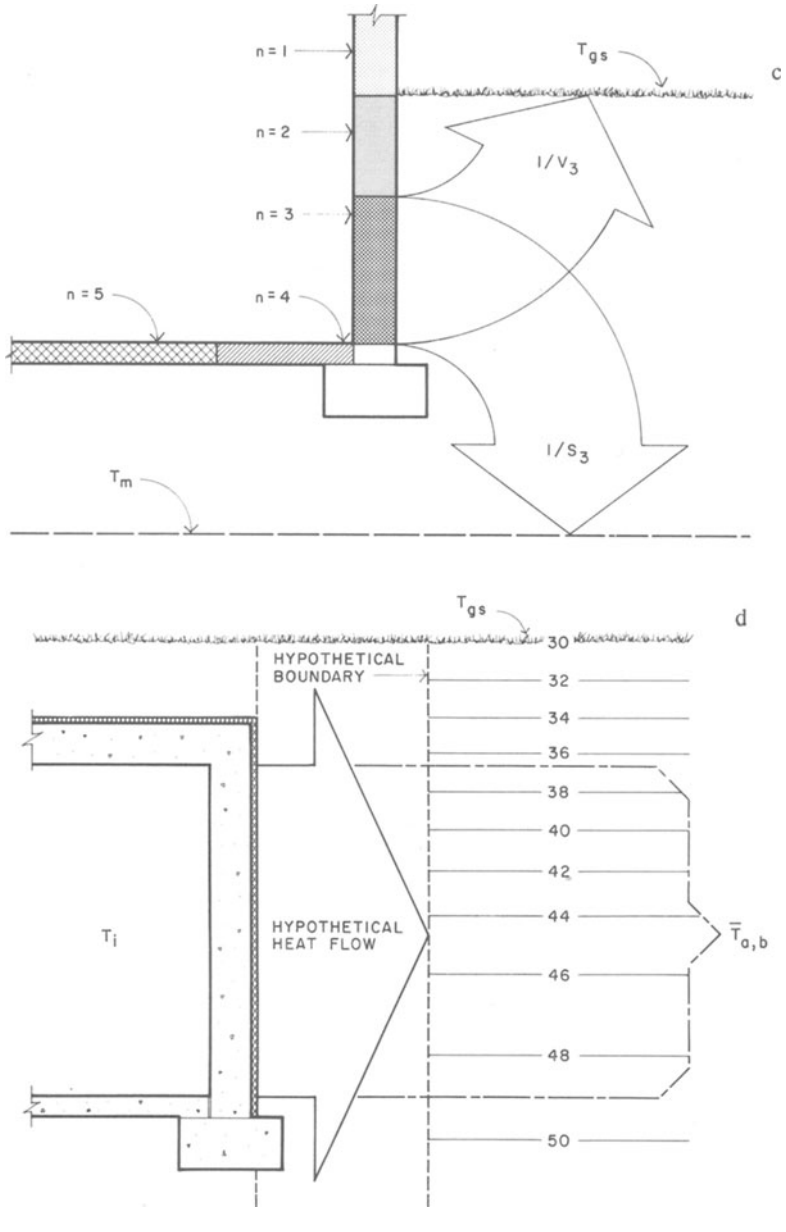


Fig. 6.8: c) Mitalas's method computes transient heat loss in two components, one a steady loss to the deep ground, and the second as a function of sinusoidally varying surface temperature. Each of five basement section segments n is assigned effective resistances $1/S_n$ and $1/V_n$ related to soil thermal conductivity and R value of segment n. **d)** The F Factor method relates transient heat loss to sinusoidally varying soil profile average temperatures. The problem can then be treated as a one-dimensional loss through a fictional thickness of soil to the undisturbed temperature of the soil profile corresponding in depth range to that of the wall.

Table 6.3

Factors Obtained by Poulos for a Subgrade Wall Between 2 and 10 Feet in Depth

R_w [(ft²*h*°F)/Btu]	Thermal Conductivity of Soil [(Btu/(ft*h*°F)]					
	0.25	0.50	0.75	1.00	1.25	1.50
2†	.11	.21	.28	.34	.39	.44
7	.31	.50	.60	.67	.72	.77
12	.43	.61	.70	.78	.83	.86
17	.52	.69	.77	.83	.87	.90
22	.59	.75	.82	.87	.90	.93

†Represents uninsulated concrete wall.

Source: J. F. J. Poulos, "Thermal Performance of Underground Structures: The Development of the Decremented Average Ground Temperature Method for Estimating the Thermal Performance of Underground Walls," M. Architecture thesis, Georgia Institute of Technology, 1982.

the defining relationship

$$F = \frac{\text{actual (simulated or observed) heat loss } q_{a,b}}{\dfrac{(T_i - \overline{T}_{a,b})(b-a)}{R_w}} \tag{6.11}$$

The limitations and full potential of the F factor method require further study. Some work is underway to obtain factors for horizontal and other insulation placements; this should extend its usefulness for estimating losses from buildings configured to maximize summer cooling. The F factor method promises to be a very accessible method for estimating transient and total annual energy exchanges with the ground; it and the method of Mitalas are the only manual methods capable of estimating heat losses during the cooling season.

6.9 INVESTIGATIONS USING COMPUTER TECHNIQUES

Because earth contact heat transfer from buildings is characterized by time-dependent, multi-dimensional heat flow through materials with widely varying thermal properties (e.g., concrete, insulation, and soil), accurate prediction of the heat exchange through floor slabs or underground walls generally requires the use of a computer. Kusuda,[45] for example, obtains solutions for the energy loss from a heated floor slab using numerical methods to evaluate Green's Functions, following the analytical method of Lachenbruch.[46] This method models the three-dimensional relationship between the slab and the earth beneath as well as that between the slab

and neighboring land and buildings. Monthly average earth temperatures at given levels beneath the floor slab are generated and then used in a simplified procedure for determining temperatures below the slab. Application to response-factor determination of floor slab heat loss in comprehensive energy analysis programs such as NBSLD, BLAST, and DOE-2 is discussed.

Two-dimensional, finite difference techniques have been used by several investigators to study the relative performance of different wall insulation placements and earth contact configurations throughout the year.[20,21,39,47] These methods discretize the spatial domain so that appropriate thermal properties can be assigned to, for example, the location of insulation along the inside or outside of the wall, or to insulation extending out into the soil mass. Among the results of these studies are indications of the strong potential for vertical heat flow in a concrete wall to short-circuit some insulation placements.

Research programs for monitoring the ground temperature field surrounding earth contact buildings and measuring the heat transfer to these buildings are vital to the development of the physical understanding of the energy transfer occurring in these structures as well as for validating computer models to be used in preparing design information. Experimental data useful in verifying computer simulations are described by Bligh et al.,[47] Bligh and Knoth,[48] Knoth,[49] McBride et al.,[50] and Szydlowski.[51]

Integration of the earth contact heat transfer calculation into a comprehensive analysis of the energy performance for a building requies careful modeling of the other components of the energy balance including, for example, internal energy gains due to people, lights, and equipment. One of the first computer programs devised to determine internal conditions associated with underground structures was developed by Kusuda and Achenbach for analyzing thermal conditions inside fallout shelters.[23] More recently, Speltz's model[27] couples a two-dimensional finite difference scheme to the NBSLD energy analysis computer program for computation of the HVAC energy requirements and internal building conditions. An elegant method for computing two-dimensional response factors for earth contact surfaces developed by Ceylan and Myers[34] has also been introduced into the DOE-2.1a computer program.[52] This capability for performing comprehensive energy analysis and peak-load determinations for earth contact buildings with realistic internal conditions (i.e., internal gains, ventilation, solar gain, thermostat setback) is making possible the systematic assessments of the energy benefits of earth contact construction.

6.10 EXAMPLE OF A DETAILED COMPUTER ANALYSIS

As an example of the state of the art in earth contact heat transfer computer modeling, some of the work that has been carried out as part of the development of a computer program for the detailed analysis of the earth contact portion of buildings is discussed in this section.[53] The purpose of this computer program is to provide the analytical capability to predict the earth contact heat transfer associated with fundamental building configurations for varying degrees of earth contact, soil thermal properties, insulation placements, and climates. Output from the simulations pro-

vides earth temperature profiles to aid in understanding the underground thermal response and heat flux profiles at the earth contact surfaces to define the component performance. Heat flux values are integrated to determine monthly, seasonal, and annual energy performance. Examples of the energy transfer predictions associated with different earth contact configurations, insulation placement, and climates are found later in this chapter.

The transient, finite difference computer program solves the two-dimensional diffusion equation

$$\rho C \frac{\delta T}{\delta t} = \frac{\delta}{\delta y}\left(k \frac{\delta T}{\delta t}\right) + \frac{\delta}{\delta z}\left(k \frac{\delta T}{\delta z}\right) \tag{6.12}$$

for spatial domains like the one shown for the earth covered building in Fig. 6.9a. In the preceding equation, ρ is the density, C is the specific heat, T is the temperature, t is time, k is the thermal conductivity, and y and z are the Cartesian coordinates. The fully implicit finite difference method used to solve this conduction heat flow problem is clearly presented in the first four chapters of *Numerical Heat Transfer and Fluid Flow* by Suhas V. Patankar.[54] The method includes discretizing the spatial domain into regions that facilitate the numerical analysis (example in "Thermal Characteristics of a Large Underground Building," Part II[15]). For the studies described in this chapter, values for k, ρ, and C corresponding to the concrete, insulation, and soil indicated in Fig. 6.9a are used in the discretization equation for the appropriate region. These values, although varying greatly from region to region, are assumed to be constant in time.

Heat conduction around the earth covered building is approximated by the assumption of a two-dimensional configuration. This assumption is most valid midway between the ends of a building (assuming a rectangular floor plan). Note that only half of the building cross section need be analyzed; symmetry of the below grade configuration is assumed. Energy exchange at the inside surface of the building is modeled by the common approximation that the heat flux, q_s, leaving the surface can be related linearly to the temperature difference between the room, T_{Bldg}, and the wall, floor, or ceiling surface temperature, T_s, by

$$q_s = (h_{c,i} + h_r)(T_s - T_{Bldg}) \tag{6.13}$$

In Eq. (6.2), h_r is the linearized surface radiation coefficient and $h_{c,i}$ is the surface convection coefficient which varies for the different room surfaces, i.

At the deep ground boundary, the temperature is assumed constant. At the interface between the ground surface and the air, there are many components which contribute to the energy balance. In addition to the sensible heat exchange due to convection that is driven by the temperature difference between the air and the ground surface, there is shortwave solar and infrared radiation heat transfer as well as the possibility of latent heat transfer through evapotranspiration from surface plantings. Figure 6.9b is a hypothetical schematic of an earth covered roof indicating the heat flow for a sunny summer morning when the roof is cooling the building. The major components of the heat balance at the inside and outside surfaces are shown with arrows indicating the assumed direction of heat flow. At the upper

surface, where the outside air and the ground cover meet, the incoming energy must be balanced by the energy leaving. Net incident radiation, ROS, combined with the heat conducted up from the inside of the roof mass to the ground surface, GOS, must equal the sensible energy convected into the outside air, HOS, plus the latent energy lost through evapotranspiration, LE, i.e.,

$$ROS + GOS = HOS + LE \tag{6.14}$$

Note that the net incident radiation, ROS, is the algebraic sum of the incident shortwave solar radiation, the outgoing shortwave radiation reflected from the ground surface, the incident infrared radiation, the outgoing infrared radiation reflected at the ground surface, and the infrared radiation emitted from the ground surface.

Each of these components can be significant in the energy balance at the ground surface. Their relationships to one another are described in *Physical Climatology* by William D. Sellers.[55] In determining quantitative values for the components, global solar radiation data is available on SOLMET weather tapes. Statistical methods developed by Clark and Allen[56] can be used to estimate the infrared heat exchange between the surface and the sky. Evapotranspiration is discussed by Kreith and Sellers.[57] Jensen[58] provides many references and some useful data.

Establishment of the validity of this earth contact computer model in real world conditions is demonstrated by comparing computer predictions of the one-dimensional temperature distribution beneath a grass covered field with experimental data. Figure 6.10 compares experimental data taken in Minnesota (symbols) with computer predictions (lines) for the last day of several months. The agreement is very good.

Verification of predictions from finite difference computer models with experimental results for full scale buildings has been presented by Shipp[15,28] and Szydlowski,[29] and their respective work is continuing. Comprehensive measurements on the earth contact heat transfer from an occupied earth sheltered residence have recently been completed by a research team at M.I.T.[47-49] Comparisons of this data with predictions of the computer model discussed in this section are in progress.

6.11 FUTURE RESEARCH

6.11.1 Comprehensive, Integrated Energy Analysis

Because of the significant potential for energy savings through appropriate utilization of earth contact building elements, accurate predictions of the energy performance of specific building configurations are necessary to guide building design and operation. The thermal performance of the earth contact elements of a building can currently be predicted with reasonable confidence and some initial parametric studies for different configurations and insulation placements have been completed (see below). However, most studies to date have, from necessity, separated the analysis of the performance of the earth contact components from that of the rest of the building. Simplified internal building operating conditions have been assumed,

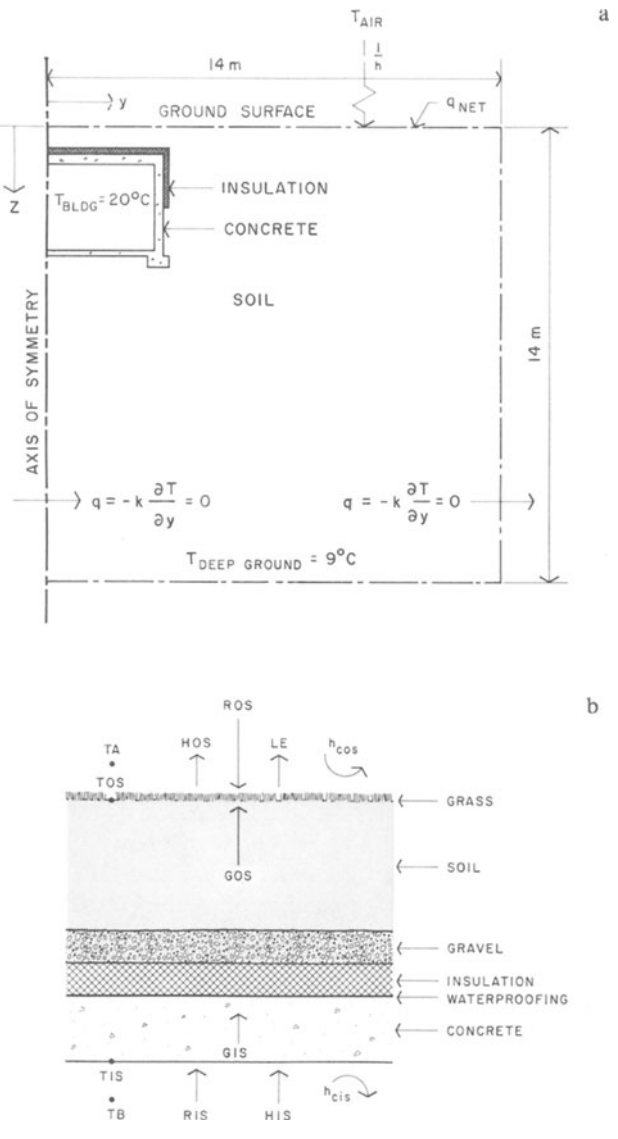


Fig. 6.9: a) The spatial extent of the two-dimensional cartesian calculation domain indicating boundary conditions. **b)** Earth-covered roof cross section defining the major components of heat flow, temperatures, and surface coefficients.

e.g., constant room temperature, no internal load from people or direct solar gain, no thermostat setback.

Critically needed parametric studies would be based on comprehensive, integrated analyses that (1) realistically couple the earth contact facets of a building

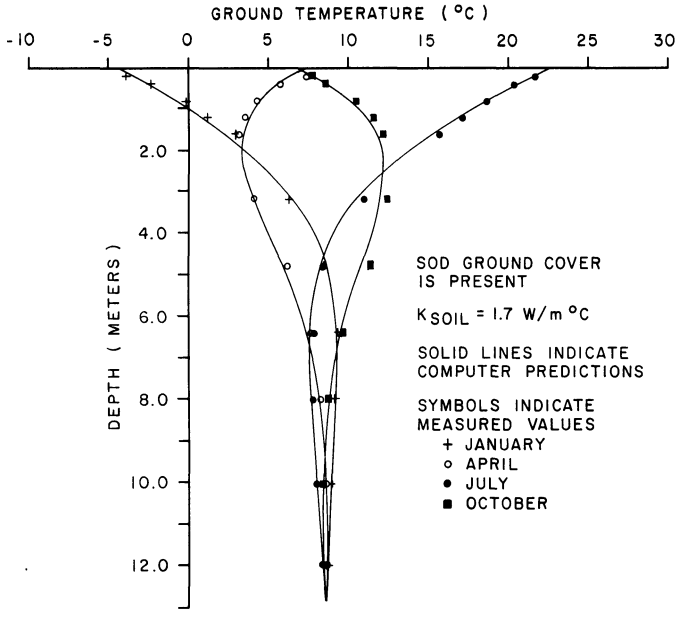


Fig. 6.10: A comparison of numerically predicted soil temperatures with measured values for Minneapolis, Minn.

to the outer building elements including above grade roofs and walls, (2) model the effect of changes in the internal loads due to people, lights, equipment, and direct solar gain on earth contact heat transfer, and (3) account for the interaction between various passive design strategies such as direct gain solar, natural lighting, ventilation, and earth contact cooling as they are coupled to each other and to an appropriate mechanical space conditioning system. The design information generated from these studies must be clearly organized and then communicated to architects, engineers, builders, and the general public.

Prerequisites for conducting comprehensive, accurate predictions of the thermal performance and energy savings associated with earth contact structures are delineated in the following discussions of the needed increases in the sophistication of the analytical models and the needed expansion of the physical understanding of earth contact heat transfer through experimentation.

6.11.2 Analytical Models

The research efforts directed at the understanding and prediction of earth contact heat transfer from buildings have generated analytical models that are helping to provide preliminary insight into the thermal response of these structures. To make more useful predictions of the energy performance of common earth contact configurations (e.g., basements and floor slabs at grade level) and to generate design information for more advanced second generation earth contact systems, the following important details of this modeling need attention:

(a) three-dimensional heat flow from the floor, corners, the building as a whole, and earth tube systems;

(b) radiative and convective coupling of earth contact surfaces to each other and to the room air;

(c) calculation methods for including the earth contact building elements in an hourly energy balance for each zone so that the coupling to the thermal mass of these elements is correctly estimated as well as the coupling designated in (b) above;

(d) accurate, manageable models of the energy balance at the ground-atmosphere interface.

Computational costs or computer storage limitations or both have generally limited the analysis of earth contact heat transfer to two spatial dimensions. Although the conditions at the axis of symmetry of a subgrade wall can be reasonably approximated by numerical methods employing a two-dimensional Cartesian coordinate system, accurate prediction of the temporal and spatial variations in the heat flow through the floor slab, near wall-to-wall corners, or to the surrounding earth from the building as a whole during the summer cooling months, requires appropriate consideration of variations in all three spatial dimensions.

Because earth contact surfaces may be strongly coupled to the thermal mass of the earth, they have the potential to be at significantly different temperatures from other nonearth contact surfaces of a room. For example, during a summer day the ceiling may be considerably warmer than the floor slab directly below. Heat will flow from the ceiling to the floor, as well as to other surfaces by radiation and to the room air, predominantly by convection. Since the radiative exchange may be much larger than the convective transfer, it is obvious that these heat transfer processes must be properly modeled to generate reliable simulations.

Although a few computer models have been developed that attempt to accurately calculate the energy performance of building zones with earth contact[23,27,30], these programs have not been widely exercised and should probably be regarded as somewhat crude initial efforts. Accurate, efficient interfacing of methods for predicting the slowly varying multidimensional heat flow in the soil mass, with the possible rapid changes in heat flow that occur near the earth contact surface, requires careful analysis to ensure the implementation of a realistic energy balance for building zones. The concept of a two-dimensional response factor, as introduced by Ceylan and Myers,[34] deserves considerable attention because once the response factors are determined, the energy analysis can proceed very efficiently. Calculation of three-dimensional response factors should also be feasible.

As outlined in the preceding section, prediction of the energy balance at the ground-atmosphere interface is a considerable challenge because there is sensible heat exchange through radiation, convection and conduction, as well as the possibility of latent energy transfer through evapotranspiration. This situation is further complicated by the likelihood of considerable shading of the ground surface on one side of a building while the other side may be receiving an increased amount due to reflection from the side of the building and radiative exchange from that surface.

6.11.3 Measurements of Earth Contact Heat Transfer

Although there have been several significant research efforts directed at the experimental definition of the heat transfer from earth contact buildings, many aspects of this work require further attention. Because soil thermal conductivity is so strongly influenced by moisture content, factors such as site drainage, soil moisture retention characteristics, water table movement and coupled heat and moisture content (which affect the moisture distribution around buildings), and associated heat transfer must be more clearly understood.

On the interior side of an earth contact surface, heat is radiated to other room surfaces (and people) and radiated and convected into the room air. Earth contact surfaces may be at considerably different temperatures from the other room surfaces (e.g., cooler during the summer), and, consequently, an accurate model of both the radiative and convective heat transfer process at the surfaces is essential for understanding and then predicting the energy balance for an earth contact zone. The current uncertainty in the magnitude of the surface convection heat transfer coefficients must be reduced.

In addition to providing fundamental physical understanding of the operative heat transfer processes occurring with earth contact buildings, these experimental studies will provide the data needed to further verify the computer simulations of thermal performance.

6.11.4 Earth Contact Configurations

As increased understanding of the thermal response of earth contact components is developed, there is the opportunity to explore the potential applications of earth contact systems as outlined above. Sophisticated computer models will aid the architect and engineer in the evolution of more energy efficient second generation earth contact designs. The design information generated from such an experimentally validated simulation presented below is part of an initial effort to facilitate this process, but this is only a beginning.

6.12 ENERGY PERFORMANCE ANALYSIS FOR COMPONENTS OF SMALL EARTH CONTACT STRUCTURES

The purpose of this section is to present preliminary results of energy performance analysis for small earth contact structures and to discuss pertinent design guidelines. This can be a difficult task to achieve since a number of analysis tools exist—each with assets and liabilities—and there is a vast array of variables which can influence energy performance calculations. In order to clearly understand the impact of placing walls, floors, and roofs in contact with the earth, a validated two-dimensional finite difference computer program (discussed in the previous section) has been developed which can predict energy transfer through these components under a variety of conditions. By limiting the analysis to energy transfer through the envelope, the program is simple enough to be used for a wide range of parametric studies, but is also accurate enough to reflect the impact of subtle changes in ground conditions, structural components, and insulation placement. Using this

tool to examine the energy transfer through the wall, floor, and roof components, it is possible to compare between conventional and earth contact structures as well as variations in earth contact designs.

The finite difference computer program used in the parametric studies that follow is designed specifically to examine energy transfer through surfaces in contact with the earth. It is not an appropriate tool for predicting energy transfer through above-grade components. In order to provide a comparison between earth contact and above-grade components, another calculation method has been applied to above-grade walls. The method selected was developed by the National Association of Home Builders Research Foundation.[59]

A comprehensive analysis of energy performance in earth contact structures would have to include the total heating and cooling load for the building along with the interaction of the earth contact components with a variety of other passive and conservation strategies. Some of these broader considerations are part of the more complete energy analysis and design guidelines already reported,[60] while others require further research to develop proper analysis tools. In this discussion, however, the focus is on representative parametric studies which indicate some of the key effects of earth contact design. This is followed by a brief discussion of the limitations of this analysis and preliminary conclusions which will lead to further refinement of design guidelines.

The research findings in this section have been adapted from the "Design Guidelines for Earth Contact Structures," which reports on research sponsored by the U.S. Department of Energy.[60]

6.13 GENERAL DESCRIPTION OF PARAMETRIC STUDIES

In order to provide a broad enough selection of conditions to reflect the variations in earth contact energy performance, energy transfer information is presented for six basic design variations in three diverse climates. The three representative climates are Tucson, Arizona, where providing cooling is the major concern although there is also some heating required (Fig. 6.11); Columbus, Ohio, where there are moderate needs for both heating and cooling energy (Fig. 6.13); and Minneapolis, Minnesota, where providing heat is the major concern although there is a small amount of cooling required (Fig. 6.14).

The six design variations in each climate include two above grade and four earth contact examples. In all three climates, typical and well-insulated above grade structures are the first two cases followed by a structure with fully bermed (basement) walls. The last three cases all include earth covered walls and roofs but are not identical in the three climates in order to demonstrate various alternatives appropriate to each region.

The energy transfer has been calculated for the wall and floor components of each case. Wall energy is expressed for a one-foot-wide, eight-foot-high section of wall; floor energy is expressed for a one-foot-wide, ten-foot-long section of floor. In all cases, the interior temperature is assumed to be constant at 68°F in the winter months and permitted to rise to 78°F in the summer months.

The manner in which this energy transfer is combined and presented in Figs. 6.11, 6.13, and 6.14 is a reflection of two important concepts in analyzing earth contact data. The first is that earth contact systems have important effects in both the heating and the cooling season, thus, total energy performance comparisons should reflect these combined effects. For each of the three climates, energy transfer is shown for the heating season and cooling seasons separately as well as combined. Energy transfer is indicated as a positive number when energy is required from another source (heat loss in winter or heat gain in summer). Energy transfer is indicated as a negative number when energy is being provided through the earth contact surface (heat gain in winter or heat loss in summer). The numbers in the combined heating and cooling season charts are arrived at by adding the energy required and/or provided from the separate heating and cooling season charts. It is important to note that since heating and cooling energy are provided in different manners, often using different fuels, these energy transfer figures do not necessarily translate directly into fuel cost differences.

The second important concept reflected in the presentation of the data is that energy transfer should not only be examined through individual components (i.e., a comparison of various walls) but also as the aggregate of interrelated components—in this case, walls and floors together. Thus, for each case in each climate, energy transfer is shown as a grouping of three bars—one reflecting wall energy transfer, one reflecting floor energy transfer, and one that is the algebraic sum of these two components. For some common sizes and shapes of smaller structures (a 30-foot by 60-foot basement for example), this correlation of an eight-foot-high section of wall to a ten-foot section of floor is generally accurate. For both smaller and larger structures or unusual geometries, adjustments would be required to reflect the correct proportion of wall to floor. It is important to note that the energy transfer predicted by this two-dimensional model does not reflect the three-dimensional effects that occur in actual building components. Nevertheless, looking at these components together rather than separately is necessary and will be discussed further below.

6.14 RESULTS AND HIGHLIGHTS OF PARAMETRIC STUDIES

6.14.1 Tucson, Arizona

For Tucson, Arizona, six cases are illustrated and their wall and floor energy transfer is compared in Fig. 6.11. In this climate, the heating season is considered to be seven months (October through April) and the cooling season five months (May through September).

6.14.1.1 Above Grade Cases. The base case (AG-1) is an above grade wall with R-10 insulation value and a slab-on-grade floor. The surrounding ground surface is bare earth. As expected, the above grade wall represents a heat loss in the winter and heat gain in the summer, resulting in an energy requirement throughout the year. The slab-on-grade floor in contact with the relatively warm earth actually represents a heat gain in the winter that can offset losses elsewhere. During the cooling season, the energy transfer through the floor is negligible—ground temperatures are not low

enough to provide cooling. Thus, the combined annual effects of the wall and floor can be summarized as an energy requirement through the walls in both summer and winter that is only partially offset by some heat provided through the floor in winter.

The second above grade case (AG-2) represents a change in only one variable—the wall insulation is increased to R-20. This reduces wall energy transfer in winter to a point where heating energy gained through the floor offsets the loss through the wall. In summer the heat gain through the wall is still substantial resulting in a net energy requirement for the year.

6.14.1.2 Fully Bermed Case. In the third case (FB-4), the wall and floor are placed into the earth to a depth of seven feet which is typical for a fully bermed structure with a conventional roof or a basement wall. Insulation at a value of R-20 is placed outside the concrete wall so that the earth contact is the main variable being compared between cases AG-2 and FB-4. The wall actually gains a slight amount of heat in the winter, but also gains a slight amount in the summer, resulting in a negligible total energy requirement. There is improvement in winter heat loss compared to the above grade cases and summer heat gain is reduced as well. The floor also represents a notable improvement since it is deeper than in the slab-on-grade cases. Both a greater amount of heat through the floor in winter and some summer cooling are provided. The total annual effect is one of almost eliminating energy transfer through the walls and providing both winter heating and summer cooling through the seven-foot-deep floor.

6.14.1.3 Earth Covered Case. Case ES-2 represents a building with two feet of earth on the roof resulting in a floor level that is eleven feet below grade. Since the previous cases in this climate and other cases not shown here indicate there is potential for both heating and cooling season benefits with earth contact, the insulation in case ES-2 is configured to maximize earth contact while minimizing direct conduction paths to the surface. Thus, there is no insulation placed directly against the walls; instead, the roof insulation (R-30) is extended horizontally eight feet beyond the exterior walls.

The result of this configuration is that the floor does not perform much differently than in the previous case (FB-4) even though it is four feet deeper. There is still heat provided in summer and cooling provided in winter. The major change is that the wall is now performing similar to the floor—actually providing some heat from the earth in the winter and some cooling in the summer. In case ES-2, the net annual effect for the wall and roof components combined is that over 60,000 Btu/lineal foot is provided compared to anywhere from 50,000 to 90,000 Btu/lineal foot required for the two above grade cases (AG-1 and AG-2).

6.14.1.4 Ground Surface Modifications. While the difference in total energy transfer between the previous four cases is impressive, it is important to note that due to the relatively high ground temperatures in Tucson, there is a substantial earth contact benefit in providing winter heat with only a moderate amount of cooling available. However, this situation can be influenced somewhat by changing the ground surface treatment. The bare earth in the first four cases permits the

intense solar radiation in this climate to be readily transferred to the ground, raising its temperature. By placing a different material on the surface—a light-colored gravel or grass, for example—the ground temperatures can be lowered to a level approximating the mean annual temperature. This type of surface treatment is modeled in the last two cases for Tucson.

The only difference between cases ES-2 and ES-3 is that the ground surface is assumed to be gravel or grass in case ES-3 rather than bare earth. The result is that in the winter heating season only a moderate amount of heating energy is provided through the walls and floor in case ES-3, but a more substantial amount of cooling is available—about twice as much as case ES-2. The net energy provided for the heating and cooling seasons combined is slightly less in case ES-3 than ES-2, however, most of it is provided for cooling rather than heating. Decisions on optimizing for heating or cooling will depend on other loads, other opportunities for passive heating and cooling, and the relative cost of auxiliary heating and cooling.

6.14.1.5 Two Story Case. The final case for Tucson is a two story, earth covered building (ES-7). Two feet of earth covers the roof and the lowest floor level is twenty feet below the surface. The upper nine feet of the wall is covered on the outside with insulation at a value of R-20 and the remainder of the wall is insulated in full contact with the earth. The ground surface in case ES-7 is the same as the previous case—covered with gravel or grass—resulting in a lower ground temperature in order to maximize cooling effects.

Since a one-foot section of wall and floor in this two story example would actually enclose twice as much space as the previous one story cases, all energy transfer figures for case ES-7 are divided by two to represent a fair comparison with the other cases. The average effect of the wall is similar to case ES-3: a small amount of heat is provided in winter and a larger amount of cooling in summer. Undoubtedly, the uninsulated lower half of the wall is responsible for most of this energy transfer. The deeper floor in this two story configuration provides a greater amount of both winter heat and summer cooling per unit area than the floor in the one story structure (case ES-3). This is not reflected in Fig. 6.11, however, since the totals are divided in half. In effect, the greater cooling benefits of placing the floor deeper into the earth with a two-story design is offset by the fact that only half of the total floor area is exposed to the ground. Thus, the compact geometry of a two story structure does not appear to be an optimal means for providing earth contact benefits in this climate since maximizing exterior surface area in contact with the earth rather than minimizing it is most desirable.

6.14.1.6 Monthly Distribution. The wall and floor energy transfer in Fig. 6.11 represents totals for heating season months and cooling season months with no reflection of how the energy required or provided is distributed on a monthly basis. In order to clarify these patterns, the monthly energy transfer is shown in Fig. 6.12 for two cases in Tucson (AG-1 and ES-3). The upper chart in Fig. 6.12 is the energy transfer through the walls only and the lower chart is for the wall and floor components combined.

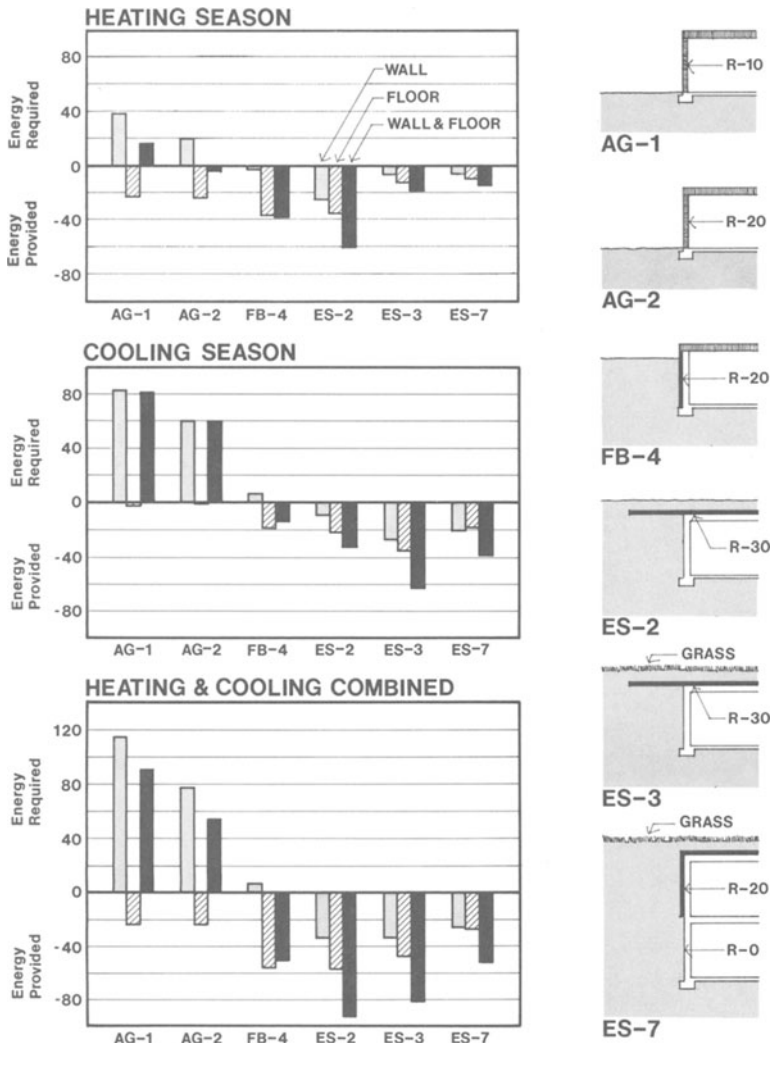


Fig. 6.11: Wall and floor energy transfer—Tucson, Ariz. Units are Btu × 10^8 per lineal foot (8 ft wall and 10 ft floor).

The energy required to offset winter losses and summer gains through the above grade wall corresponds directly to the severity of the exterior conditions with the greatest heat loss in January and the greatest heat gain in July. Because the finite difference program does not include the effect of solar radiation on above grade walls, the summer heat gain for case AG-1 shown in Fig. 6.12 is lower than it would be in reality. The pattern of the earth contact wall is to provide cooling in all five months of the cooling season, but with greater amounts at the beginning of the season gradually diminishing through September as the ground temperatures increase. A similar pattern is evident in the winter months as the greatest amount

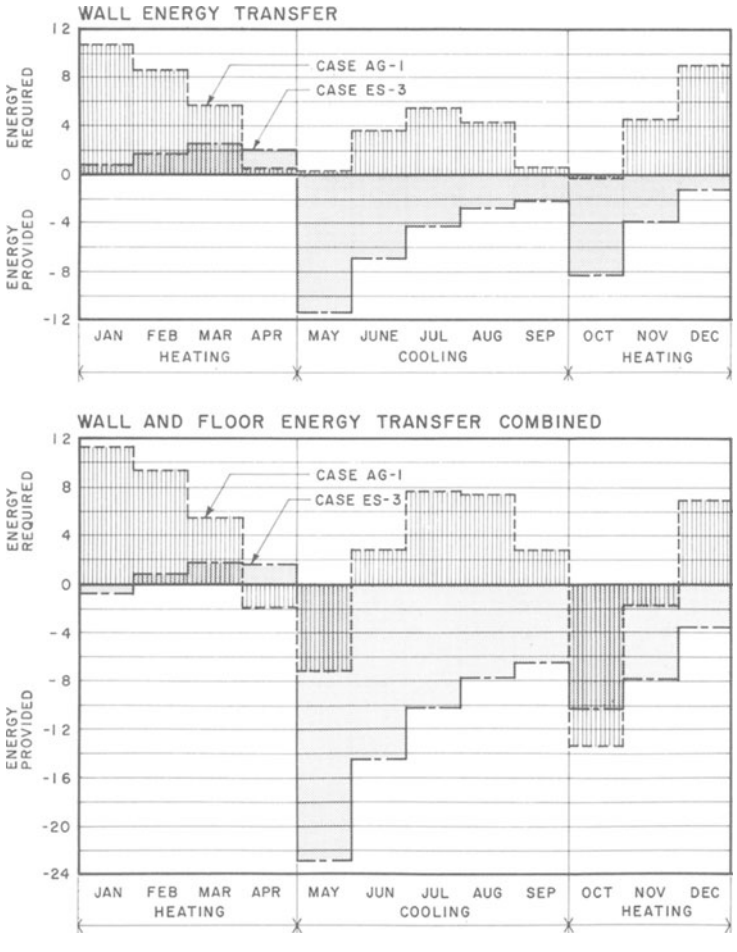


Fig. 6.12: Monthly energy transfer—Tucson, Ariz. Units are MBtu per lineal foot (8 ft wall and 10 ft floor).

of heat is provided through the earth contact walls in the autumn as outside temperatures begin to drop. During the middle and later winter months, either small gains or losses occur through the walls representing a significant improvement over the above grade walls.

When the energy transfer through walls and floors in combined, the monthly distribution patterns are similar to those for the walls alone. For the earth contact case (ES-3), the total cooling provided is more than doubled in each cooling month with the majority of the cooling provided in the first part of the summer. Winter heating is also increased in October through January, but the floor's contribution in the late winter months is negligible.

The slab-on-grade floor in case AG-1 has some of the characteristics of the deeper earth contact walls and floor of case ES-3. Thus, the floor does not increase greatly the winter heat loss or summer heat gain in the most severe months. In addition, the slab-on-grade floor provides some cooling in the first summer month and some heating in the first two winter months (mainly October). In fact, the slab-on-grade floor in case AG-1 actually produces more heat in October than the earth contact wall and floor. The magnitude of these benefits on an annual basis, however, is considerably less than those provided by the earth contact case and virtually no benefit is provided during the peak of the cooling season in the above grade case.

6.14.2 Columbus, Ohio

The six cases for Columbus, Ohio, are illustrated and their wall and floor energy transfer is compared in Fig. 6.13. In this climate, the heating season is considered to be eight months (October through May) and the cooling season four months (June through September). The severity of the winter in Columbus is between the other two climates—5670 heating degree days (HDD) as opposed to 1700 HDD in Tucson and 8250 HDD in Minneapolis. The cooling requirement is 30% to 40% greater than that of Minneapolis, but still only half as great as that of Tucson. With regard to ground temperatures and the related effect of earth contact walls, the Columbus climate would be similar to that of Kansas City, Missouri, or Washington, D.C., although both have a slightly smaller need for heating and a slightly greater need for cooling than Columbus.

6.14.2.1 Above Grade Cases. The base case (AG-1) for Columbus is identical to that for the other climates—a conventional above grade wall insulated to a value of R-10 with an uninsulated slab-on-grade floor. The ground surface in all cases in Columbus is grass. Winter heat loss through the wall and floor in case AG-1 is approximately equal. In the summer the floor contributes a moderate amount of cooling which more than offsets heat gain through the wall. By increasing the wall insulation to R-20 and placing R-10 insulation to a depth of four feet around the foundation (case AG-4), the winter heat loss is reduced by 40% while the summer cooling is affected to a minor degree.

6.14.2.2 Fully Bermed Case. The wall and floor are placed in the earth to a depth of seven feet in the third case (FB-4). The concrete wall is insulated on the outside to a value of R-20 so that the earth contact is the only difference between cases AG-4 and FB-4. For the fully bermed structure, winter heat loss is reduced somewhat in both the wall and floor, while increased cooling is provided in summer through both components. The net effect of both heating and cooling season improvements is that the total annual energy transfer for case FB-4 is about 50% less than the energy transfer for the well insulated, above grade case (AG-4).

6.14.2.3 Earth Covered Case with a Typical Wall Insulation. In case ES-2, two feet of earth is placed on the roof with the floor level eleven feet below grade. The wall insulation remains at a value of R-20, making it comparable to the previous cases. Due to the greater depth of the wall and floor in this case, additional

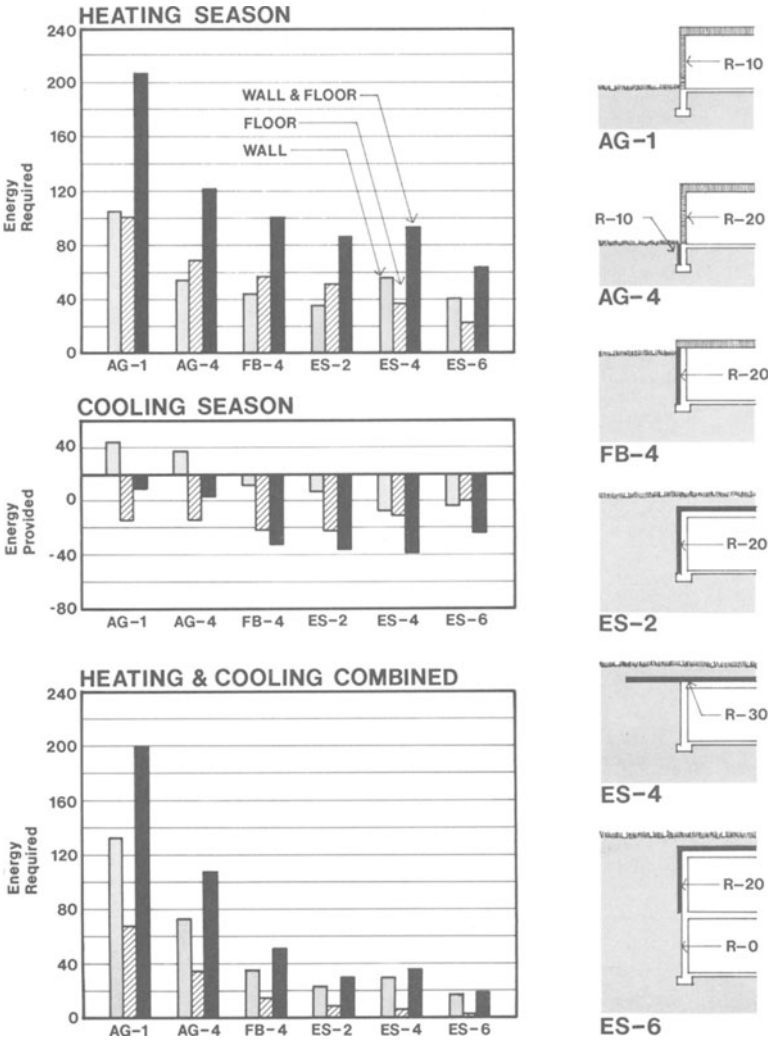


Fig. 6.13: Wall and floor energy transfer—Columbus, Ohio. Units are Btu $\times 10^5$ per lineal foot (8 ft wall and 10 ft floor).

reduction in winter heat loss occurs while sumnmer cooling is increased compared to the previous fully bermed case. The net annual energy transfer for the wall and floor in case ES-2 is further reduced to a point where it is approximately 70% less than the well insulated, above grade wall (case AG-4).

6.14.2.4 Earth Covered Case with Extended Roof Insulation. In order to evaluate the potential for increased earth contact cooling, the wall is uninsulated in case ES-4 with the roof insulation (R-30) extended horizontally eight feet beyond the walls. As might be expected, the cooling provided through the wall is more

than doubled compared to the wall in case ES-2. This increase in cooling is more than offset by the greater heat loss through the wall in winter. It might also be expected that the energy transfer through the floor would be unchanged in cases ES-2 and ES-4 since the floor is uninsulated and at the same depth in both configurations. This is not the case, however. Both heat loss through the floor in winter and cooling provided through the floor in summer are reduced in case ES-4. Since there is greater energy transfer through the wall in both summer and winter, the ground temperature is changed around the building which affects the energy transfer through the floor. It is apparent from this interrelationship between the wall and floor that energy transfer through these earth contact components cannot be examined separately. While both the wall and floor energy transfers differ significantly in cases ES-2 and ES-4, the net effect of combining the heating and cooling seasons for both components results in only a small total difference.

6.14.2.5 Two Story Earth Covered Case. Identical to the two story case in Tucson, the structure in case ES-6 is covered with two feet of earth resulting in a lower floor level at twenty feet below the surface. The upper nine feet of the wall is covered with insulation at a value of R-20, with the lower portion of the wall remaining uninsulated. In case ES-6, the energy transfer numbers for a one-foot-wide section of both wall and floor are divided in half for comparison to the one level cases. Since this is the most compact geometry, winter heat loss is reduced considerably. The deeper wall and floor provide a moderate amount of summer cooling, but not as much as the one story cases. The net annual energy transfer for this case is the best of all cases shown here, over 80% less than the above grade, well-insulated structure with foundation insulation (AG-4). Unlike Tucson, where the compact geometry of the two story case is a detriment to providing cooling, the same configuration in Columbus is desirable because the winter heat loss reductions outweigh any compromise in cooling.

6.14.3 Minneapolis, Minnesota

For Minneapolis, Minnesota, six cases are illustrated and their wall and floor energy transfer is compared in Fig. 6.14. In this climate, the heating season is considered to be nine months (September through May) and the cooling season three months (June through August).

6.14.3.1 Above Grade Cases. While the walls in the two above grade cases (AG-1 and AG-4) are identical to those in Tucson and Columbus with insulating values of R-10 and R-20, the foundation or floor insulation is increased in Minneapolis to reflect more typical protection from the severe winter. In fact, most houses in this climate are built with basements instead of slab-on-grade, however, this complicates the comparative analysis. In the first case (AG-1), insulation at a value of R-5 is extended over the foundation to a depth of four feet. In case AG-4, which has increased wall insulation (R-20), insulation is placed under the entire floor slab at a value of R-10. The ground surface in all cases in Minneapolis is grass.

As expected, the increased wall insulation and insulation under the entire floor slab in case AG-4 reduce the energy transfer significantly in winter compared

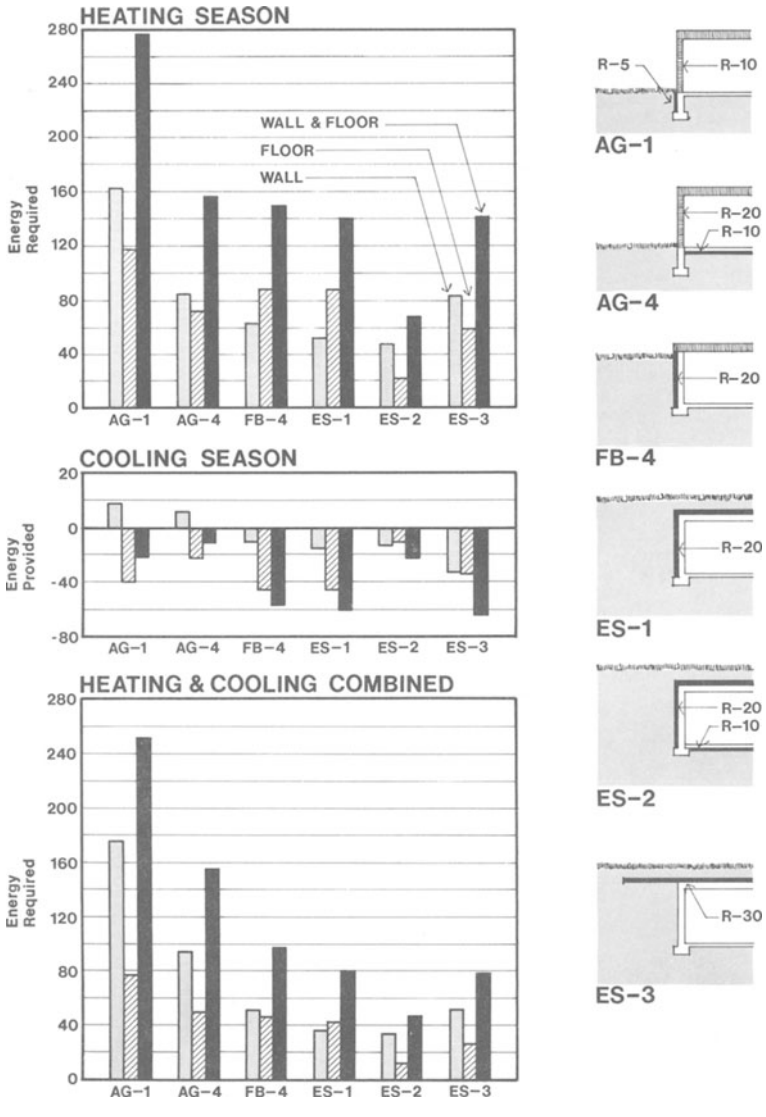


Fig. 6.14: Wall and floor energy transfer—Minneapolis, Minn. Units are Btu $\times 10^5$ per lineal foot (8 ft wall and 10 ft floor).

to case AG-1. The uninsulated floor slab in case AG-1 provides a moderate amount of cooling in summer that is reduced by the insulation in case AG-4. The net annual energy transfer for both above grade cases is dominated by the large heating season requirements.

6.14.3.2 Fully Bermed Case. In the fully bermed case (FB-4), the wall and floor are placed at a depth seven feet below the surface. The entire wall is covered

on the outside by insulation at a value of R-20, but the floor is uninsulated. The wall in this case shows reduced heat loss in winter and some increase in summer cooling compared to the equally well-insulated wall above grade. As expected, the uninsulated floor in case FB-4 has a greater heat loss in winter than the insulated slab in case AG-4, but more summer cooling is provided. The net energy transfer is lower than the above grade cases due to the improved wall performance and the provision of additional summer cooling.

6.14.3.3 Earth Covered Case. Two feet of earth is placed on the roof in case ES-1, resulting in a deeper placement of both wall and floor (eleven feet from surface to floor level). Compared to the previous fully bermed case, the energy transfer through the wall is reduced in winter while additional cooling is provided in summer. The energy transfer through the uninsulated floor varies little between cases FB-4 and ES-1. The net effect is annual energy transfer through the combined components of case ES-1 that is 40% less than the well-insulated, above grade case (AG-4) and almost 20% less than the fully bermed wall.

6.14.3.4 Earth Covered Case with Floor Insulation. In case ES-2, all conditions are identical to the previous case except that insulation at a value of R-10 is placed under the floor slab. In terms of insulation value on wall and floor components, this case is most directly comparable to the well-insulated, above grade case (AG-4). There is little change in the energy transfer through the wall comparing cases ES-1 and ES-2. The floor insulation, however, accounts for a significant reduction in energy transfer in the winter with a loss in summer cooling as well. Since the winter season is dominant, the net annual effect is to reduce energy transfer for the combined components to a level that is approximately one-third of the above grade case with similar insulation levels (AG-4). The desirability of reducing winter energy requirements at the expense of a moderate amount of passive cooling appears likely from this analysis. However, this trade-off should only be made in the context of the total building loads combined with other sources of passive heating or cooling.

6.14.3.5 Earth Covered Case with Extended Roof Insulation. In a configuration identical to the one modeled in Tucson and Columbus, the wall is uninsulated in case ES-3 with the roof insulation (R-30) extended horizontally eight feet beyond the exterior walls. In a manner similar to the same cases in Columbus, the heating season performance of the wall is poorer than a well-insulated, below grade wall (case ES-1), but the cooling through the wall is improved. In spite of the apparent similarity of the floors in cases ES-1 and ES-3, there is less energy transfer in the latter case in both winter and summer. For the extended roof insulation case (ES-3), the net annual energy transfer for wall and floor components combined is very similar to the fully insulated earth covered case (ES-1), even though the separate components are quite different. Case ES-3 still represents a slight improvement over the well-insulated, above grade case (AG-4) in winter and a significant increase in providing summer cooling.

6.14.3.6 Two Story Case. Although no two story earth covered cases are shown for Minneapolis in Fig. 6.14, it is quite likely that such a case would have the best annual performance, as it did in Columbus. Due to the more severe winter and

limited need for cooling, the more compact two story geometry would result in a lower total energy transfer than any of the single story cases.

6.15 LIMITATIONS OF PARAMETRIC STUDIES

The two-dimensional finite difference model used in this analysis does not reflect three-dimensional effects in earth contact walls and floors which would generally result in slightly greater energy transfer in both summer and winter. Also, the model is based on daily average temperatures and does not reflect hourly variations in exterior climate. In addition, the interior temperature is held constant in the model and not allowed to fluctuate. Thus, the benefits of a high-mass structure with fluctuating temperatures is not reflected in this analysis.

Analyzing energy transfer for individual components in the parametric studies reflects the potential performance, but must be put into the context of total loads and other load-reduction strategies. This is beyond the scope of this chapter. However, a few points related to other energy-related effects of earth contact systems should be made. In analyzing total loads, earth contact structures have some inherent characteristics that are not represented in this analysis of conduction through the envelope. The first is reduced infiltration. While infiltration can be reduced considerably in properly built above grade structures, berming and covering structures with earth—particularly on the sides toward prevailing winter winds—can result in very low infiltration without any extraordinary sealing techniques. Secondly, the concrete or masonry structures usually used in earth contact buildings provide a great deal of mass that is available to be coupled with intermittant sources of energy, that is, direct gain solar or wood-burning stoves, or both. Finally, peak loads are reduced in these more massive structures resulting in lower costs for mechanical equipment.

6.16 PRELIMINARY CONCLUSIONS

A broader range of parametric studies coupled with total load calculations is necessary to develop complete and useful design guidelines for earth contact structures.[44] It is possible, however, to make some preliminary observations based on the sample of parametric studies presented here.

(1) In a predominantly cooling climate such as Tucson, Arizona, earth contact walls and floors can provide reasonable amounts of passive cooling while providing a moderate amount of passive heating in the winter. Above grade components cannot compete with this performance. Maximizing exterior surface area in contact with earth is the recommended strategy in this climate. A key to increasing cooling benefits is the ground surface treatment—a surface such as gravel or grass—can reduce heat gain and keep ground temperatures lower.

(2) In a climate with moderate heating and cooling requirements such as Columbus, Ohio, various earth contact wall and floor designs represent improvements over above grade components, assuming there is an equivalent amount of insulation. Increased insulation above grade can equal the performance of the

below grade components in winter, but cannot provide the summer cooling. Since winter heating is somewhat more predominant than summer cooling in Columbus, more compact designs minimizing exterior surface area are recommended. In similar climates that have a slightly smaller winter heating requirement and slightly greater summer cooling requirement—Kansas City, Missouri, or Washington, D.C., for example—it may be reasonable to optimize earth contact cooling with a more extended configuration.

(3) In a predominantly heating climate such as Minneapolis, Minnesota, earth contact walls and floors represent improvements in both summer and winter compared to above grade components with equivalent amounts of insulation. As with the more moderate Columbus climate, increased insulation in above grade walls and floors can equal the performance below grade in winter but cannot provide summer cooling at the same time. Since winter heat loss is more important than providing a limited amount of passive cooling, the most compact geometry possible is desirable in this climate.

6.16.1 Building Configuration Considerations

The extensive use of earth contact as a conservation or passive cooling strategy definitely places constraints on design which result in a few key prototypical layouts on which most variations are based. The most common is the elevational type with a single exposed facade and earth placed on the remaining three sides. Either bermed on flat land or set into a hillside on one or two levels, this type of design inherently encompasses a number of conservation and passive strategies particularly appropriate to temperate and colder climates. The geometry of the house is usually compact, window area is reduced, all glass faces south, the house includes substantial thermal mass to store solar gain, wind is diverted, and infiltration is reduced. These effects, combined with the improved characteristics of the earth contact envelope, make this design especially appropriate for energy efficient housing.

The second prototypical design, exhibited in much indigenous architecture as well as modern variations, is the atrium or courtyard type of house. This design is particularly appropriate in warmer climates. In a typical atrium design, usually on one level, the exterior surface area is increased compared to an elevational structure and the entire perimeter wall can be placed in contact with the earth. Maximizing the earth contact in this manner is ideal for providing passive cooling in summer as well as passive heating in the winter from moderately high ground temperatures. In addition, the courtyard design reduces window area, lends itself to various shading techniques, and provides a space that can collect cool night air.

6.16.2 Interior Surface Considerations

The energy transfer for walls and floors indicated in the parametric studies are based on solid concrete walls and floors that are exposed to the interior space. Any coverings on these surfaces—insulating the walls or carpeting the floor—would reduce the energy transfer. This might be desirable in winter but would inhibit the passive cooling in summer. Interior finishing and design must be coordinated with any earth contact cooling strategies. For example, in some cases, insulating

the walls may be acceptable if most of the cooling is provided by the floor. In other cases the reverse may be true. If floors are to be relied upon for cooling, an exposed concrete finish or ceramic tiles may be the best solution. In some cases, it may be possible to maximize for both heating and cooling simply by placing area rugs over masonry floors in winter and removing them in summer.

6.17 REFERENCES

1. Sterling, Carmody, and Farnan, *Earth Sheltered Residential Design Manual* (Van Nostrand Reinhold Co., New York, 1982).
2. Carmody and Sterling, *Underground Building Design: Commercial and Institutional Structures* (Van Nostrand Reinhold Co., New York, to be published late 1983).
3. Sterling, Carmody, and Elnicky, *Earth Sheltered Community Design: Energy-Efficient Residential Development* (Van Nostrand Reinhold Co., New York, 1981)
4. M. G. Swenson, "An economic test for earth-sheltered design," *Underground Space* **7**, 105–109.
5. R. K. Maxwell, "Temperature measurements and the calculated heat flux in the soil." M. S. thesis, University of Minnesota, 1964.
6. T. P. Bligh, "Energy conservation by building underground," *Underground Space* **1**, 19–23.
7. D. Carter, "Terraset School," *Underground Space* **1**, 317–323.
8. Setter, Leach and Lindstrom, Inc. and Underground Space Center, *Energy, Cost and Design Evaluation of Nonresidential Buildings*. Report to the Naval Facilities Engineering Command, May 1981.
9. Booz, Allen and Hamilton, Inc., "Case study profile: Williamson Hall," *Design and Performance Trends for Energy Efficient Commercial Buildings*. A study for the U. S. Department of Energy, 1982, pp B-35 through B-48.
10. K. Labs, "Direct-coupled ground cooling: Issues and opportunities," *Passive Cooling*. Proceedings of the International Passive and Hybrid Cooling Conference, ASES. Miami Beach, Fla., 1981.
11. B. Givoni, "Modifying the ambient temperature of underground buildings," *Earth Covered Buildings: Technical Notes*, Vol. 1, Conference Proceedings from The Use of Earth Covered Settlements, edited by F. Moreland. NTIS, Springfield, Va., 1979.
12. Underground Space Center, *Earth Sheltered Housing Design*, (Van Nostrand Reinhold Co., New York, 1978).
13. K. Labs, *Regional Analysis of Ground and Above-Ground Climate*, ORNL Report No. Sub-81/40451/1, Oak Ridge National Laboratories, Dec. 1981.
14. T. Kusuda, *Earth Temperatures Beneath Five Different Surfaces*, National Bureau of Standards Report 10373, Feb. 1981.
15. P. H. Shipp, E. Pfender, and T. P. Bligh, "Thermal characteristics of a large underground building—Parts I and II," *Underground Space* **6**, 54–64.

16. P. H. Shipp, Personal communication. Aug. 1982.

17. J. Mostaghimi Tehrani, "Measurement of thermal conductivity of Soils," M. S. thesis, University of Minnesota, July 1978.

18. L. A. Salomone, *Thermal Behavior of Fine-Grained Soils*, NBS Building Science Series BSS 149. U. S. Department of Commerce National Bureau of Standards, Washington, D.C., 1982.

19. F. Wang, "Mathematical modeling and computer simulations of insulated systems in below grade applications." ASHRAE/DOE Conference on Thermal Performance of the Exterior Envelopes of Buildings. Orlando, Fla., Dec. 1979.

20. G. D. Meixel, P. H. Shipp, and T. P. Bligh, "The impact of insulation placement on the seasonal heat loss through basement and earth-sheltered walls," *Underground Space* **5**, 41–47, (1980).

21. G. D. Meixel, "Effects of insulation placement on two-dimensional heat transfer through basement and earth-sheltered walls." ASCE Conference, Minneapolis, Minn., Sept. 1982.

22. P. H. Shipp, "Natural convection within masonry block basement walls," *ASHRAE Transactions* **89** (1), (1983).

23. T. Kusuda and P. R. Achenbach, "Numerical analysis of the thermal environment of occupied underground spaces with finite cover using a digital computer," *ASHRAE Transactions* **69**, (1963).

24. P. D. Metz, "Design, construction, and operation of the solar assisted heat pump ground coupled storage experiments of Brookhaven National Laboratory," *Fourth Annual Heat Pump Technology Conference*, Stillwater, Okla., 1979.

25. Control Data Corporation, MITAS, CYBERNET Publications Department, HQWOSF, P. O. Box O, Minneapolis, Minn., 55440, Publication Number 86615000.

26. M. F. McBride et al., "Measurement of subgrade temperatures for prediction of heat loss in basements," *ASHRAE Transactions* **85** (1), (1979).

27. J. J. Speltz, "A numerical simulation of transient heat flow in earth-sheltered buildings for seven selected U.S. cities," M. S. thesis, Trinity University, 1980.

28. P. H. Shipp, "Thermal characteristics of large earth-sheltered structures." Ph.D. thesis, University of Minnesota, 1979.

29. R. F. Szydlowski, "Analysis of heat loss in earth-sheltered structures." M.S. thesis, Iowa State University, 1980.

30. G. D. Meixel et al., "Earth contact systems: Detailed model development, response factor subroutines and preliminary design guidelines," *Proceedings of the Passive Solar Update Conference*, Washington, D. C., Sept. 1982.

31. K. J. Bathe, "ADINAT—A finite element program for automatic dynamic incremental nonlinear analysis of temperature," Massachusetts Institute of Technology Report No. 82448-5, Acoustic and Vibration Laboratory, December 1978.

32. C. Voss, "TEMPFEM Manual: Finit elementprogram for tvadimensionella varmelednignsbergakningar," VBB Vattenbyggnadsbyran, Stockholm, Sweden, 1980.

33. R. L. Taylor, "*HEAT*, A finite element program for heat conduction analysis," Report 75-1 Civil Engineering Laboratory, Naval Construction Battalion Center, Port Hueneme, Calif., May 1975.

34. H. T. Ceylan and G. E. Myers, "Long-time solutions to heat conduction transients with time-dependent inputs," *Journal of Heat Transfer* **102**, 115 (1980).

35. F. C. Houghton et al., "Heat loss through basement walls and floors," Paper No. 1213, ASHVE *Transactions* **48**, 369–384, 1942.

36. J. M. Elliot and M. Baker, "Heat loss from a heated basement," Paper No. 1724, ASHRAE *Transactions* **66**, 400–413, 1960.

37. G. G. Boileau and J. K. Latta, *Calculation of Basement Heat Losses*, Division of Building Research Report 10477, National Research Council of Canada, Ottawa, 1968.

38. K. B. Labs, "Building underground: A tempered climate, earth as insulation and the surface-undersurface interface, in *Energy Efficient Buildings*, edited by W. Wagner, (Architectural Record Books/McGraw-Hill Book Co., New York, 1980), pp. 82–90.

39. P. H. Shipp and T. B. Broderick, "Comparison of annual heating loads for various basement wall insulation strategies using transient and steady state models," *Proceedings of the Conference: Thermal Insulation, Materials, and Systems for the 80s.* DOE/ORNL/ASTM, Clearwater Beach, Fla., Dec. 1981.

40. W. C. Brown, *Mark XI Energy Research Project Comparison of Standard and Upgraded Houses*, Division of Building Research BR Note Number 160, National Research Council of Canada, Ottawa, 1980.

41. G. P. Mitalas, "Basement heat loss studies at DBR/NRC," National Research Council of Canada, Division of Building Research, NRCC 20416, 1982.

42. J. M. Akridge, "A decremented average ground temperature method for estimating the thermal performance of underground houses," *International Passive and Hybrid Cooling Conference*, ASES, edited by A. Bowen, G. Clark, and K. Labs, Miami Beach, Fla., 1981.

43. James Poulos, "Thermal performance of underground structures: Development of the decremented average ground temperature method for predicting the thermal performance of underground walls." M. S. thesis, College of Architecture, Georgia Institute of Technology, 1982.

44. W. D. Turner, D. C. Elrod, and I. I. Siman-Tov, *HEATING V*, "IBM 360 heat conduction program," ORNL/CSD/TM-15, Oak Ridge National Laboratory, Oak Ridge, Tenn., 1977.

45. T. Kusuda, M. Mizimo, and J. W. Bean, *Seasonal Heat Loss Calculation for Slab-On-Grade Floors*, National Bureau of Standards Report NBSIR 81-2420, March 1982.

46. A. H. Lachenbruch, *Three Dimensional Heat Conduction in Permafrost Beneath Heated Buildings*, Geological Survey Bulletin 1052-B, Government Printing Office, Washington, D. C., 1982.

47. T. P. Bligh et al., "Earth contact systems: Soil temperature and thermal conductivity data, heat flux data and meter calibration," *Proceedings of the Passive Solar Update Conference*, U. S. Department of Energy, Washington, D. C., Sept. 1982.

48. T. P. Bligh and B. H. Knoth, "Data from one, two and three dimensional temperature fields in the soil surrounding an earth sheltered house," ASHRAE *Transactions* **89**, (1) (1983).

49. B. H. Knoth, "A thermal study of an earth-sheltered residence: Instrumental, data processing techniques, soil temperature and heat flux data," M. S. thesis, Massachusetts Institute of Technology, 1982.

50. M. F. McBride et al., "Measurement of subgrade temperature for prediction of heat loss in basements, ASHRAE *Transactions* **85** (1) (1979).

51. R. F. Szydlowski, "An earth sheltered building research facility at Ames Laboratory," *Earth Shelter Performance and Evaluation*, Conference Proceedings, Oklahoma State University, edited by L. Boyer, 1981.

52. G. D. Meixel et al., "Earth contact systems: Detailed model development, response factor subroutines and preliminary design guidelines," *Proceedings of the Passive Solar Update Conference*, Washington, D. C., Sept. 1981.

53. Final report on Task 1.4-Develop Detailed Earth Contact Model to the U. S. Department of Energy, Passive Cooling Program, San Francisco Operations Office under contract DE-ACO3-80SF11508, Feb. 1983.

54. S. V. Patankar, *Numerical Heat Transfer and Fluid Flow*, (McGraw-Hill, 1980).

55. W. D. Sellers, *Physical Climatology* (University of Chicago Press, Chicago, 1965).

56. G. Clark and C. G. Allan, "The estimation of atmospheric radiation for clear and cloudy skies," *Proceedings of the 2nd National Passive Solar Conference*, Vol. 2, edited by Donald Prowler, Philadelphia, Pa., 1978, 676–679.

57. F. Kreith and W. D. Sellers, "General principles of natural evaporation," *Heat and Mass Transfer in the Biosphere, Part I: Transfer Processes in the Plant Environment*, (Scripta Book Company, 1975).

58. M. E. Jensen (ed.), *Consumptive Use of Water and Irrigation Waste Requirements*. (American Society of Civil Engineers, New York, 1973).

59. NAHB Research Foundation, Inc. *Insulation Manual* Second edition (Rockville, Maryland, 1979).

60. Underground Space Center, *Preliminary Design Guidelines for Earth Contact Buildings*, a report from the earth contact systems portion of Passive Cooling Research sponsored by the U. S. Department of Energy under contract DE-ACO3-80SF11508, University of Minnesota, 1983.

CHAPTER 7

TESTING SOLAR COLLECTORS

James E. Hill
Byard D. Wood
Kent A. Reed

7.1 ABSTRACT

The state of the art in testing solar collectors to determine their thermal performance is reviewed here. A brief history of the significant work done to establish the procedures for testing is given, including the current proposed and adopted worldwide procedures. In the second section, the basic mathematical equations describing the thermal performance of solar collectors are given as the basis for showing what information can be obtained from such tests. In the third section, a detailed description is given of the testing techniques used to determine collector thermal performance under clear-sky, full-irradiance conditions, as typically called for in most of the currently adopted standards. This section will also include a discussion of typical instrumentation, its accuracies and limitations, the results of major round-robin programs conducted to date showing the comparability of data between laboratories when the tests are conducted outside under real-sun conditions, and special techniques required to test concentrating collectors. In the fourth section, a technique for obtaining collector efficiency indoors under zero-irradiance will be discussed. In the fifth section, special considerations when testing air collectors will be discussed. In the final section, concepts for using test data to calculate all-day collector performance will be presented.

7.2 INTRODUCTION

The successful use of solar collectors depends on their thermal efficiency or thermal performance as well as upon other factors such as cost, durability and reliability, and safety characteristics. Waksman et al.[1] describe several kinds of tests that could be used for determining some of these characteristics, such as 30-day no-flow, thermal shock, rain, hail, thermal cycling, pressure, leakage, fire, and various

structural loading tests. However, these tests will not be described or discussed in this chapter.

Although thousands of solar domestic water heating systems were used in Southern California in the 1920s and 1930s, in Florida in the 1940s and 1950s, and in other places around the world during the same period, the number of collector manufacturers was small and there is no record of attempts to standardize test procedures. The earliest record of quantitative data related to performance of flat-plate collectors was published by Brooks in the United States in 1936.[2] Experimental investigation and verification of the theory related to the performance of flat-plate collectors was commenced by Hottel and Woertz at the Massachusetts Institute of Technology in the 1940s.[3] The application of these studies led to the development of a standard test proposed by Whillier and Richards of South Africa in 1961.[4] The proposed test involved determining the efficiency of a liquid-heating, flat-plate collector outdoors under full-irradiance, near-normal-incidence conditions over a range of inlet water temperatures. The data were then used to determine the basic thermal parameters in the governing equation describing collector efficiency (to be described in the next section) so collector efficiency could be calculated under any conditions. This basic approach, with modifications, is still used today, as will be shown in the remainder of this chapter.

Over the past 20 years, considerable research was done in Israel, Australia, selected countries in Europe, and the United States to refine the basic techniques as proposed by Whillier and Richards. The Israelis adopted a standard test procedure for liquid-heating, flat-plate collectors in 1966 very similar to the one proposed by Whillier and Richards.[5] In 1976, Tabor of Israel outlined how four collectors could be connected in series during a test and considerably more data obtained to determine the collector parameters during one exposure.[6] The Australians have used procedures similar to those of the South Africans and Israelis for testing liquid-heating solar collectors since the early 1970s.[7-11] They have given considerable emphasis to collecting a large amount of data under a variety of environmental conditions and then reducing the data in such a way as to determine the nonlinear dependence of collector efficiency on environmental conditions.

Following the oil embargo of 1973-74 in the United States, a major research program for the use of solar energy led to considerable testing of solar collectors, primarily flat-plate, liquid-heating ones. The major testing has been done at the Los Alamos National Laboratory (LANL),[12] the Lewis Research Center of the National Aeronautics and Space Administration (NASA),[13-15] the Honeywell Corporation,[16] the Argonne National Laboratory,[17] Sandia Laboratories,[18] the National Bureau of Standards (NBS),[19] and at a number of commercial testing laboratories where the results were proprietary and remain largely unpublished. With the exception of the work done by Hill, Jenkins, and Jones,[19] the emphasis in the U.S. has been on the development of collectors, not the development of the testing procedures. In addition, considerable use has been made of solar simulators.[13-16]

In 1974, NBS proposed a standard procedure for testing solar collectors involving a series of outdoor, steady-state tests to determine their near-normal-incidence efficiencies over a range of temperature conditions.[20-22] Again, the proposal was

similar in concept to the Whillier and Richards method, but with considerable detail added on the test set-up, instrumentation, and testing technique. The American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) adopted Standard 93-77 in 1977.[23] It is similar to the original NBS proposal, but calls for additional tests to determine the transient response characteristics of the collector as well as how its near-normal-incidence efficiency varies as the incidence angle changes. The ASHRAE procedure was written for both liquid-heating and air-heating collectors, flat-plate and concentrating collectors, and for use in outdoor testing or indoors with a solar irradiance simulator. ASHRAE also adopted Standard 96-1980 in 1980 for testing unglazed liquid-heating, flat-plate collectors, which is virtually identical to Standard 93-77 except for one or two small changes in experimental technique necessary for unglazed collectors.[24]

Mention has already been made of the Israeli standard for liquid-heating, flat-plate collectors adopted in 1966.[5] There have been other notable standards proposed or adopted to date. In December 1977, the Association Francaise de Normalisation (AFNOR) of France adopted Standard 50-501, which is very similar to ASHRAE Standard 93-77 except it is only for liquid-heating collectors and calls for additional tests to determine the back and side losses when the collector aperture is covered and insulated.[25] In May 1978, Bundesverband Solarenergie (BSE) of Germany proposed a procedure calling for a combined indoor (zero-irradiance) and outdoor test for flat-plate, liquid-heating and air-heating collectors to improve the repeatability of test data and reduce the dependence of testing on favorable, full-irradiance, outdoor conditions seldom achieved in Europe.[26] This procedure is much different than the others mentioned in this section and will be described below.

In January 1980, the Commission of the European Communities (CEC) published its "Recommendations for European Solar Collector Test Methods (Liquid-Heating Collectors)."[27] It is for flat-plate, liquid-heating collectors and essentially contains the three procedures already mentioned (ASHRAE, AFNOR, and BSE), except it excludes the use of a solar irradiance simulator. The document suggests that any of the methods can be used and the particular one chosen will depend primarily on the environmental conditions of the locality tested. In 1982, the British Standards Institution published a draft of a standard "for development," DD77:1982.[28] It is for flat-plate, liquid-heating collectors and specifies either a steady-state test under a solar irradiance simulator (consistent with the CEC procedures) or a technique for conducting a transient test outdoors. The latter technique allows data to be taken when the incident solar irradiance radiation is varying with time (and consequently the outlet temperature for the collector). An elaborate data reduction process is used to produce the necessary information to establish the efficiency curve. The Canadian Standards Association (CSA) adopted a performance standard entitled "Solar Collectors," F378-M1982, in 1982 which specifies, in part, that the thermal performance of the collectors be determined outdoors in accordance with ASHRAE Standard 93-77 (with a few minor variations noted in the CSA Standard).[29] In 1982, the Standards Association of Australia adopted Standard 2535-1982, "Glazed Flat-Plate Solar Collectors with Water as the Heat-Transfer Fluid—Method for Testing Thermal Performance."[30] It specifies that the test be done outdoors and, except for not requiring a test to be run to determine the collectors' time constant, it produces

the same results as ASHRAE Standard 93-77. The specific experimental techniques are based on those used at the Commonwealth and Industrial Research Organization (CSIRO), and the standard requires more data to be taken and a more elaborate data reduction process than is specified by ASHRAE.[8] ASHRAE Standard 93-77 is currently in the process of being revised. In addition, the International Standards Organization (ISO) has recently formed Technical Committee 180, which has on its agenda the development of an ISO standard for testing solar collectors.

Below, a detailed description will be given of how to test solar collectors under full-irradiance conditions outdoors. This will be largely in accordance with the current ASHRAE Standards. However, attempts will be made to point out, as appropriate, modificaitons to the procedure that are likely to be adopted in the near future and other alternative ways of achieving the same objective that have been adopted or used elsewhere. Also, a description of how to test solar collectors under zero-irradiance conditions, as proposed by BSE, will be made.[26]

7.3 BASIC EQUATIONS GOVERNING THE THERMAL PERFORMANCE OF SOLAR COLLECTORS

7.3.1 Thermal Efficiency

When a solar collector is operating under steady-state conditions, the relationship that governs its thermal performance is [31]

The rate of energy extracted from the collector	=	The rate of energy absorbed in the collector by the absorber	−	The rate of energy loss from the collector by conduction, convection, and radiation	=	The rate of energy carried away by the transfer fluid

In equation form,

$$\frac{\dot{Q}_u}{A_a} = F_R I_t (\tau\alpha)_e - F_R U_L (t_{f,i} - t_a) = \frac{\dot{m} c_p}{A_a} (t_{f,e} - t_{f,i}) \tag{7.1}$$

This equation is often referred to as the Hottel-Whillier equation. If solar collector efficiency is defined by

$$\eta_a = \frac{\dot{Q}_u / A_a}{I_t} \tag{7.2}$$

then the efficiency can be written as

$$\eta_a = F_R I_t (\tau\alpha)_e - F_R U_L \frac{(t_{f,i} - t_a)}{I_t} \tag{7.3}$$

or

$$\eta_a = \frac{\dot{m}c_p(t_{f,i} - t_a)}{A_a I_t} \tag{7.4}$$

Eq. (7.3) implies that if the efficiency is plotted against $(t_{f,i} - t_a)/I_t$ and if F_R, U_L, and $(\tau\alpha)_e$ are constant. a straight line will result where the slope is $-F_R U_L$ and the y intercept is $F_R(\tau\alpha)_e$. In reality, $F_R U_L$ is not a constant but rather a function of the operating temperature of the collector and of the ambient weather conditions. In addition, $F_R(\tau\alpha)_e$ varies with the incident angle between the outward-drawn normal to the collector aperture and the direct solar beam. Nevertheless, efficiency data taken near normal incidence. when $(\tau\alpha)_e = (\tau\alpha)_{e,n} \approx$ constant. for most collectors operating at temperatures <80°C (<175°F) can be correlated reasonably well with Eq. (7.3). This will be discussed in more detail in the next section.

7.3.2 Time Constant

Whenever transient conditions exist, the equality indicated by Eq. (7.1) does not govern the thermal performance of the collector since part of the energy being absorbed is used for heating up the collector and its components. The corresponding relationship for transient conditions is

The rate of change of the energy of the collector and its components	=	The rate of energy absorbed in the collector by the absorber	−	The rate of energy loss from the collector by conduction, convection, and radiation	=	The rate of energy carried away by the transfer fluid

In equation form,[31,32]

$$\frac{C_A}{A_a}\frac{d\bar{t}_f}{d\theta} = F_R I_t(\tau\alpha)_e - F_R U_L(t_{f,i} - t_a) - \frac{\dot{m}c_p}{A_a}(t_{f,e} - t_{f,i}) \tag{7.5}$$

Eq. (7.5) can be solved for the exit temperature of the transfer fluid, $t_{f,e}$, as a function of time, θ, after making the following assumptions:

(1) The exit temperature of the transfer fluid is related to the average temperature by

$$\frac{d\bar{t}_f}{d\theta} = K\frac{dt_{f,e}}{d\theta} \tag{7.6}$$

where[14]

$$K = \frac{\dot{m}c_p}{F'U_L A_a}\left(\frac{F'}{F_R} - 1\right) \tag{7.7}$$

(2) $I_t, (\tau\alpha)_e, U_L, t_a, \dot{m}, c_p$, and $t_{f,i}$ are all constant for the period covered by the transient solution.

The solution to Eq. (7.5) is then

$$\frac{F_R I_t(\tau\alpha)_e - F_R U_L(t_{f,i} - t_a) - \frac{\dot{m}c_p}{A_a}(t_{f,e} - t_{f,i})}{F_R I_t(\tau\alpha)_e - F_R U_L(t_{f,i} - t_a) - \frac{\dot{m}c_p}{A_a}(t_{f,e,\text{initial}} - t_{f,i})} = \exp - \left(\frac{mc_p}{KC_A}\right)\theta \qquad (7.8)$$

It is conventional practice in the physical sciences that whenever an equation of the form of Eq. (7.8) describes the response of a particular system, the quantity in the exponential divided into time shall be called the system's time constant. The physical interpretation of the time constant is the time required for the quantity on the left side of Eq. (7.8) to change from 1.0 to 0.368, where $0.368 = 1/e$. It is felt that this concept should be maintained in order to be consistent with the accepted interpretation of time constant. Unfortunately, this has not always been done in the solar collector field. Simon published the value of "time constant" for nine different water-heating collectors using the 0-to-99% change time.[14] Wijeysundera computed the "response time" of typical one-, two-, and three-cover air heaters using the 0-to-99% change time.[33]

7.3.3 Incident Angle Modifier

Souka and Safwat[34] have shown that the effective transmittance-absorptance product, $(\tau\alpha)_e$, of a flat-plate solar collector can be described by

$$(\tau\alpha)_e = K_{\alpha\tau}\ (\tau\alpha)_{e,n} \qquad (7.9)$$

where $K\alpha\tau$ is a function of incident angle only. The relationship was derived assuming that the optical properties of the cover plate or plates were the only factors affecting the change in absorbed solar radiation as incident angle changes. In testing solar collectors, it has been popular to assume that a very similar relationship, suggested by Simon and Buyco,[35] applies

$$F_R(\tau\alpha)_e = K_{\alpha\tau}F_R(\tau\alpha)_{e,n} \qquad (7.10)$$

and to develop a testing procedure to determine $K_{\alpha\tau}$ as a function of its independent variables whatever they may be for the particular collector being tested. The incident angle modifier (IAM), $K_{\alpha\tau}$, simply describes how the optical efficiency of the collector changes as a function of incident angle and other independent variables and becomes an essential factor in determining the all-day performance of most stationary and single-axis tracking collectors. It is a simple correction factor which is applied to the efficiency curve.

For any incident angle, the collector thermal efficiency is given by

$$\eta_a = F_R\ K_{\alpha\tau}\ (\tau\alpha)_{e,n} - F_R U_L \frac{(t_{f,i} - t_a)}{I_t} \qquad (7.11)$$

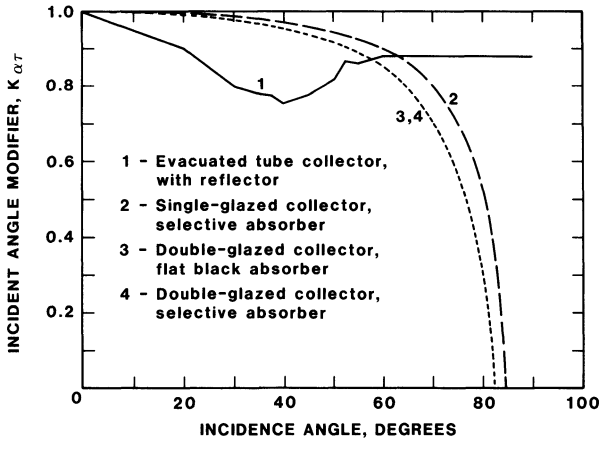


Fig. 7.1: Incident angle modifier as a function of the incident angle for four stationary solar collectors.

It is important to note that $K_{\alpha\tau}$ influences only the y intercept and not the slope of the efficiency curve. Figure 7.1 shows how $K_{\alpha\tau}$ varies with incident angle for four collectors. For many flat-plate collectors, $K_{\alpha\tau}$ can be correlated with the following equations, as shown in Fig. 7.2:

$$K_{\alpha\tau} = 1 - b_o\left(\frac{1}{\cos\theta} - 1\right) \tag{7.12}$$

It is recognized that there are collectors for which $K_{\alpha\tau}$ is not symmetrical with the normal to the aperture plane. Mather and Beekley suggested that the optical characteristics of these collectors can be described using a biaxial IAM.[36] For parabolic trough. concentrating collectors and evacuated tubular collectors, there are two mutually perpendicular directions of symmetry; one parallel to the longitudinal absorber axis and the other perpendicular to the axis. They suggested $K_{\alpha\tau}$ can be expressed as the product of two separate IAM values, each determined in the mutually perpendicular directors of symmetry

$$K_{\alpha\tau} = K_1 K_2 \tag{7.13}$$

This concept has been used in the Solar Rating and Certification Corporation (SRCC) standard for rating solar collectors.[37]

7.4 TESTING SOLAR COLLECTORS UNDER CLEAR-SKY, FULL-IRRADIANCE CONDITIONS

7.4.1 ASHRAE Standards

Solar collectors can be tested outdoors under natural solar irradiance or they can be tested indoors under artificial sunlight, that is, a solar irradiance simulator.

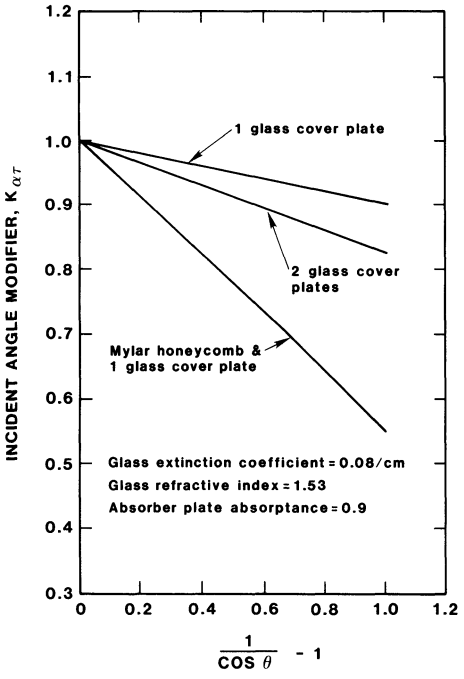


Fig. 7.2: Incident angle modifier for three flat-plate solar collectors with nonselective coatings on the absorber.

This section is concerned with outdoor tests.

It is important to recognize that it is not possible to control a number of the test parameters during the outdoor test. The solar/meteorological parameters, such as irradiance, ambient temperature, and wind speed, cannot be controlled. One has to wait for the desired test conditions or live with what is available. This situation leads to probably the strongest argument for using solar simulators, providing the opportunity to conduct a well-controlled test.

In testing solar collectors, the ASHRAE Standards specify that the collectors be isolated so that they have one inlet and one outlet. The energy of the fluid entering and leaving the collector is determined by making appropriate measurements in the fluid streams. These quantities are then used in conjunction with the energy incident upon the collector (also determined by measurement) in order to calculate the pertinent performance factors for the collector. The fluid can be either a liquid or a gas, but not a combination of the two. Test work has been done on collectors in which a phase change occurs;[38,39] however, no standard procedures have been adopted to date.

As part of the procedure, an apparatus is specified both for when a liquid (Figs. 7.3, 7.4, or 7.5) or air (Fig. 7.6) is to be the heat-transfer fluid. The detailed requirements of the apparatus are given along with specifications for instrumentation to be used in measuring incident solar radiation, temperature, temperature

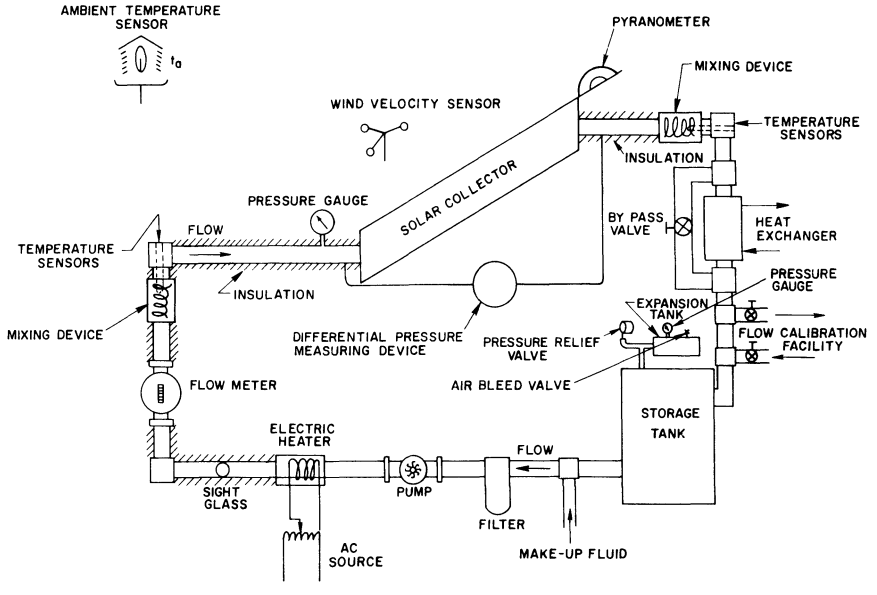


Fig. 7.3: Closed-loop testing configuration for a solar collector when the transfer fluid is a liquid.[23]

difference, liquid flow rate, air flow rate, pressure, pressure drop, time, and weight. A discussion about the instrumentation, what is typically used, and its accuracy limitations will be given later in this section.

The test procedure is carried out in four steps:

(1) The collector is required to undergo a preconditioning test prior to the start of the thermal tests. The collector must be exposed for three cumulative days with no fluid passing through it and with the mean incident solar radiation measured in the plane of the collector aperture exceeding 17,000 kJ/m^2 d (1500 Btu/ft^2 d). The intentions of this test are to ensure that the collector component properties have stabilized prior to the start of the thermal tests. It is likely that this test will be eliminated in the next revision of the standards.

(2) The time constant test is done.

(3) The efficiency tests are conducted.

(4) A series of tests are conducted to determine the collector's incident angle modifier.

7.4.1.1 Time Constant Test. The time constant for a collector could be theoretically calculated using Eq. (7.8). However, there is a large uncertainty in knowing the value of the effective heat capacity, C_A. Duffie anad Beckman have shown that the upper limit for this heat capacity is the summation of the mass times the specific heat for each of the components that make up the collector (absorber, glass, insulation, and so on), plus that of the heat-transfer fluid required to fill the collec-

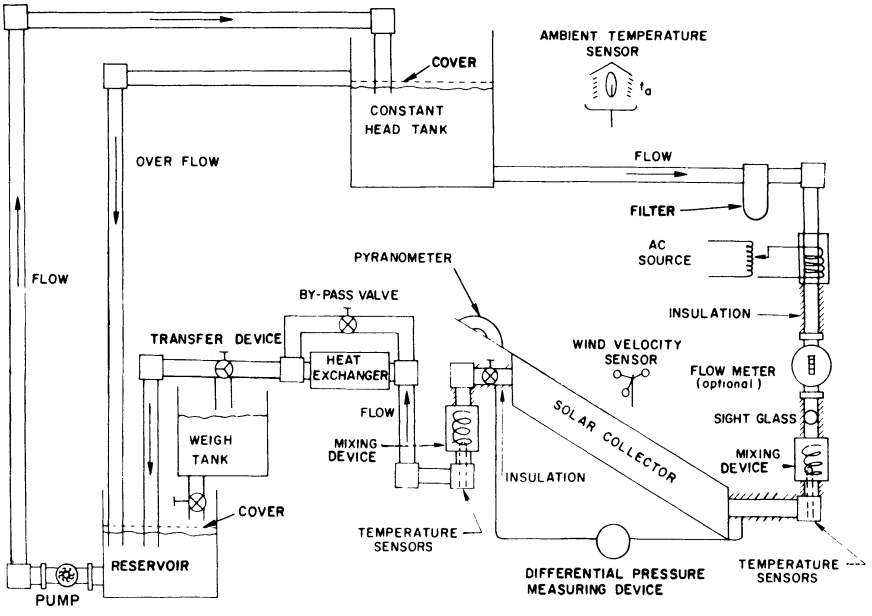


Fig. 7.4: Open-loop testing configuration for a solar collector when the transfer fluid is a liquid.[23]

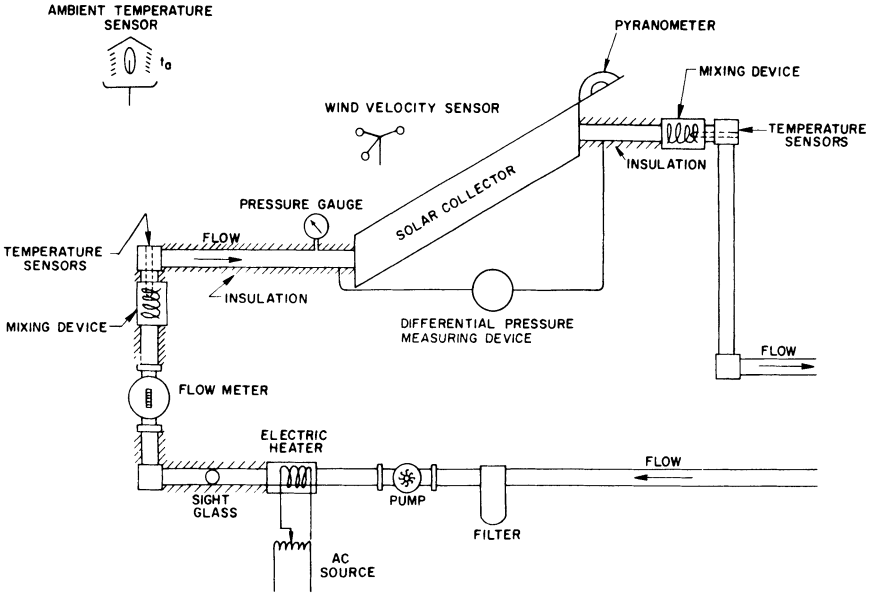


Fig. 7.5: Open-loop testing configuration for a solar collector when the transfer fluid is a liquid and is supplied continuously.[23]

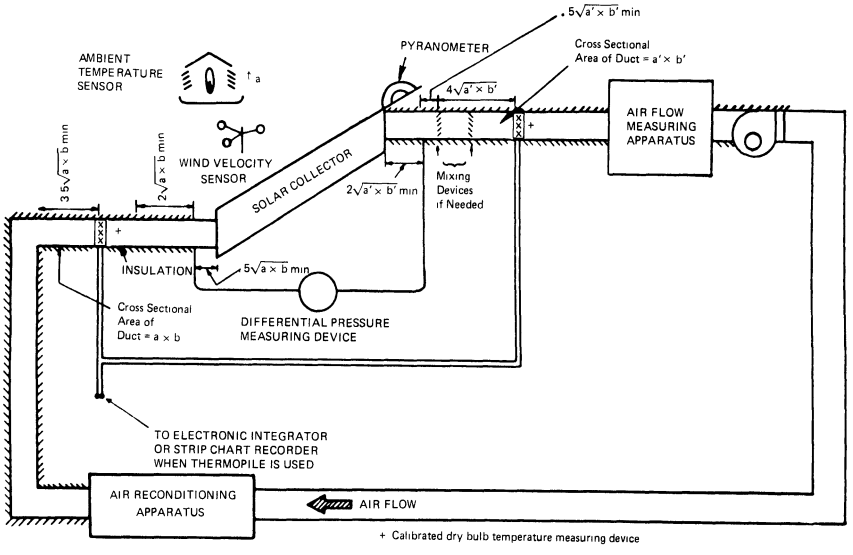


Fig. 7.6: Testing configuration for a solar collector when the transfer fluid is air.[28]

tor.[31] However, if this were used for C_A, the time constant would be overestimated, because the temperature of some materials in the collector changes by only a fraction of the amount that the fluid temperature changes in such a transient process. Consequently, the time constant is required to be determined experimentally in the ASHRAE Standards.

The actual test can be carried out in one of two ways. The most straightforward technique is to expose the collector to the solar radiation and, after entering and exiting fluid temperatures have stabilized, suddenly shield the collector from the sun and monitor the exit fluid temperature on a strip-chart recorder. The incident radiation initially must be above 790 W/m^2(250 Btu/ft^2 h). The entering fluid should be within ± 1°C (± 2°F) of the ambient temperature, if possible, for the duration of the test. This latter requirement simplifies the data-reduction process using the left side of Eq. (7.9).

A second technique that can be used in accordance with the ASHRAE Standards is to shield the collector from the sun altogether (conduct the test inside, for example). The inlet fluid temperature is adjusted to 30°C (54°F) above the ambient and, after the exit temperature has stabilized, the inlet temperature is suddenly decreased to within ± 1°C (± 2°F) of the ambient.

The results using the two techniques will give slightly different results since the heat transfer processes are different. In the CEC recommended procedures,[27] the first technique is specified to be used, but in reverse; that is, the collector shield is suddenly removed in bright sunlight ($I > 750$ W/m^2) and the collector is allowed to heat up. A separate test is also done to determine the effective thermal capacity of the collector. The experiment is done in the laboratory in a manner similar to the second technique described above but, again, in reverse. The inlet temperature

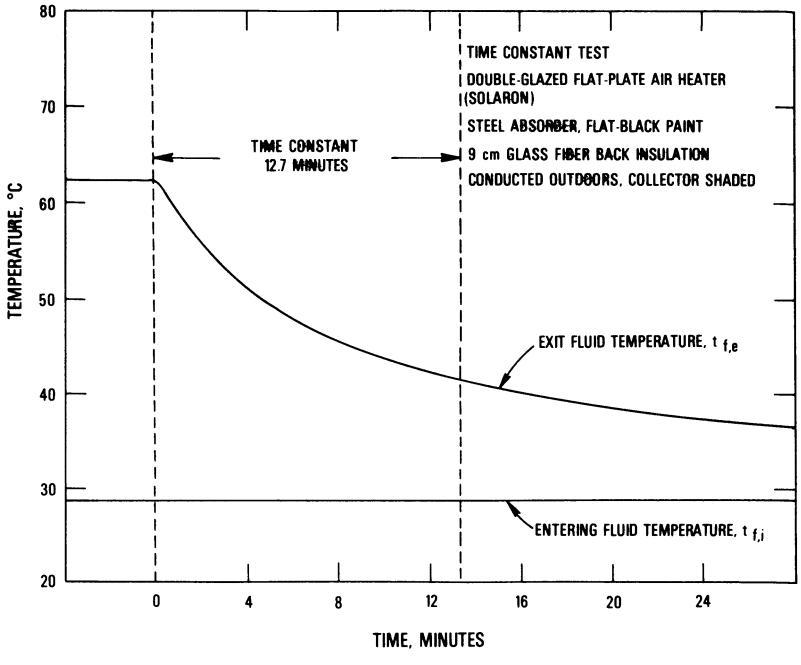


Fig. 7.7: Experimental results for the time constant test for an air-heating solar collector.[19]

is suddenly increased 10°C (18°F) above the ambient temperature used to stabilize the collector. Data are taken during the transient period and essentially used to evaluate the integral of all terms in an equation similar to Eq. (7.5), recognizing that $I_t = 0$ and assuming C_A to be constant for the process.

Figure 7.7 shows typical test results of the time constant test done on a double-glazed, flat-plate air heater.[19] The test was conducted outside under a clear, sunny sky by shading the collector after steady-state conditions were reached. Initially, the fluid entered the collector at 28.5°C (83.3°F) and exited at approximately 62°C (144°F), with a flow rate of 0.01 $m^3/(m^2s)$(2 cfm/ft^2). As can be seen, the time constant was found to be approximately 12.7 min. This is much larger than for flat-plate collectors which use a liquid as the heat-transfer fluid (1-2 min), but less than that for many liquid-heating, evacuated-tubular collectors ($\sim$ 20 min).[19]

7.4.1.2 Thermal Efficiency Test. The efficiency test series consists of determining the average efficiency (integrating the energy quantities) over a range of temperature differences between the average fluid temperature and the ambient air. The efficiency is calculated by

$$\eta_g = \frac{\int_0^\tau \dot{m} c_p (t_{f,e} - t_{f,i}) d\theta}{A_g \int_0^\tau I_t d\theta} \tag{7.14}$$

The time period, τ, to be used for the integration of the energy quantities is

either 5 min or one time constant, whichever is larger. In the CEC recommended procedures, this time is specified to be 30 min. In the Australian Standard, it is specified to be three times the residence time,* or 5 min plus two residence times.[30] Also note that the efficiency is computed based on gross collector area. The flow rate is required to be steady and vary by less than ± 1.0% for the duration of each test. In addition, the heat-transfer fluid shall have a known specific heat which varies by less than 0.5% over the temperature range of the fluid during a particular test period. Consequently, the efficiency can be determined by

$$\eta_g = \frac{\dot{m}c_p \int_0^\tau (t_{f,e} - t_{f,i})d\theta}{A_g \int_0^\tau I_t d\theta} \tag{7.15}$$

The test apparatus specified by ASHRAE were designed so that the temperature of the fluid entering the collector can be controlled to selected values. This feature is used to obtain the data over the temperature range desired. At least 16 data points are required for a complete test series, and they must be taken symmetrically with respect to solar noon (to prevent biased results due to possible transient effects).

During each test period, the incident solar radiation must be quasi-steady, as indicated in Fig. 7.8 (in constrast to days in which cloud cover can cause a time distribution such as shown in Fig. 7.9). The CEC recommended procedures include a quantitative specification of quasi-steady conditions. That is, for each test point the following must be satisfied:

(1) I_t values vary by less than 50 W/m^2(15 Btu/(ft^{2}h)),

(2) $t_{f,i}$ values vary by less than 0.1°C (0.2°F),

(3) t_a values vary by less than 0.1°C (0.2°F),

(4) $\dot{m}$ values vary by less than ± 1%, and

(5) $(t_{f,e} - t_{f,i})$ values vary by less than 0.1°C (0.2°F).

Other requirements that must be satisfied for each data point are that the average irradiance be greater than 630 W/m^2 (200 Btu/ft^{2}h) (600 W/m^2 in the CEC recommended procedure) and the incident angle between the sun and the outward drawn normal from the collector be less than 30° (40° in the CEC recommendations). In addition, the range of ambient temperatures for the entire test series must be less than 30°C (54°F).

The measurements made and the calculated efficiency for each data point are reported in tabular and graphical form such as shown in Fig. 7.10. The ordinate is the efficiency and the abscissa is the measured temperature difference (between fluid inlet and ambient) divided by the irradiance.

Figure 7.10 shows typical results for a double-glazed, flat-plate, water-heating

* Residence time is defined as the time for an element of heat transfer fluid to travel from the collector inlet to the collector outlet.

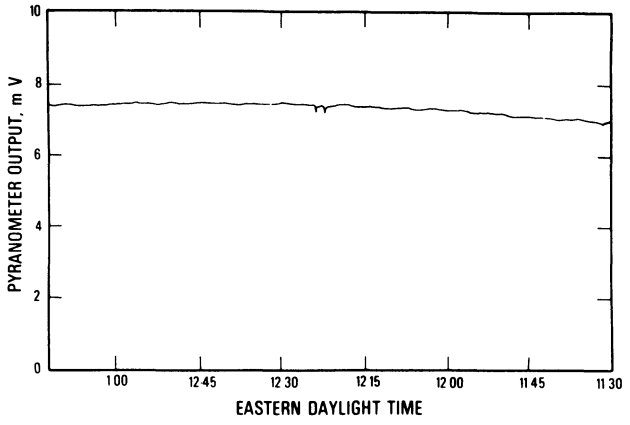


Fig. 7.8: Incident solar radiation on a horizontal surface in Gaithersburg, Maryland, March 13, 1974.[23]

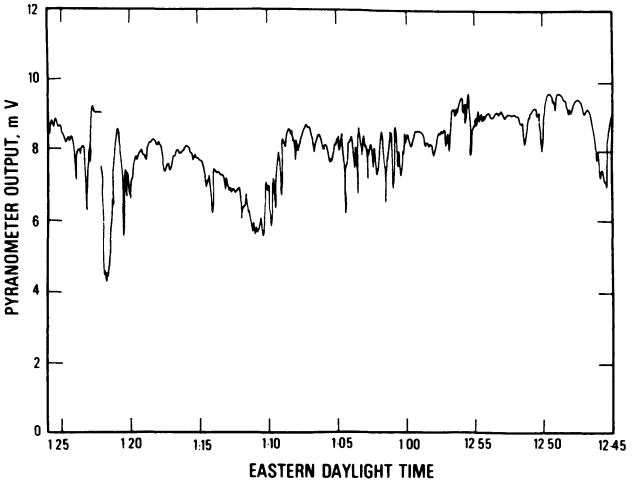


Fig. 7.9: Incident solar radiation on a horizontal surface in Gaithersburg, Maryland, March 11, 1974.[23]

collector.[19] The efficiency values are based on the aperture of the collector, 1.40 m^2. Normally, test data are obtained over the entire range at $\Delta t/I$ values primarily by changing the operating temperature of the collector, $t_{f,i}$. This is because the ambient temperature and solar irradiance are generally fixed (within limits) for the location and time of year in which the test is being conducted. However, Fig. 7.10 shows that it was possible to establish part of the efficiency curve by using a solar screen to reduce the irradiance level to a low but steady value.

7.4.1.3 Incident Angle Modifier Test. As with the time constant test, there are two ways in which the incident angle modifier test can be done in accordance

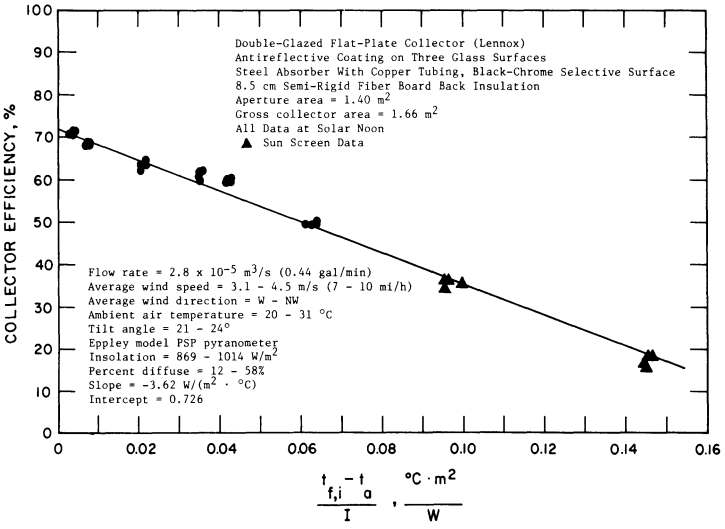


Fig. 7.10: Experimental results for the near-normal-incidence efficiency test for a water-heating solar collector.[19]

with the ASHRAE Standards. The first technique is applicable when the incident angle can be arbitrarily adjusted, such as with a solar irradiance simulator or with an outdoor movable test rack. In this case, a thermal efficiency test is conducted in accordance with all the requirements of the procedure used to get the near-normal-incidence efficiency curve, except that:

(1) The inlet fluid temperature is held within $\pm 1°C$ ($\pm 2°F$) of the ambient temperature, and

(2) The test is done for incident angles of 0°, 30°, 45°, and 60°.

By comparing the test results here with those obtained in establishing the efficiency curve at incident angles of less than 30°, values of $K_{\alpha\tau}$ can be computed as a function of incident angle (out to 60°).

The second technique is applicable for testing outside with a permanent test rack where the collector orientation cannot be arbitrarily adjusted with respect to the direction of the incident solar radiation (except for perhaps adjustments in tilt). The collector is tested for a complete day with the inlet fluid temperature held constant as above. The efficiency values are computed continuously and pairs of values are selected, one from the morning and one from the afternoon, which correspond to values of 30, 40, and 60°for the incident angle. The two efficiency values in each pair are averaged to compensate for transient effects and then used as above to compute the incident angle modifier.

Figure 7.11 shows typical test results for the double-glazed collector whose efficiency was shown in Fig. 7.10.[19] Both techniques were used for this collector and, as can be seen, gave comparable results. The second technique involved orienting a test stand so that the collector faced north with the tilt angle adjusted in order

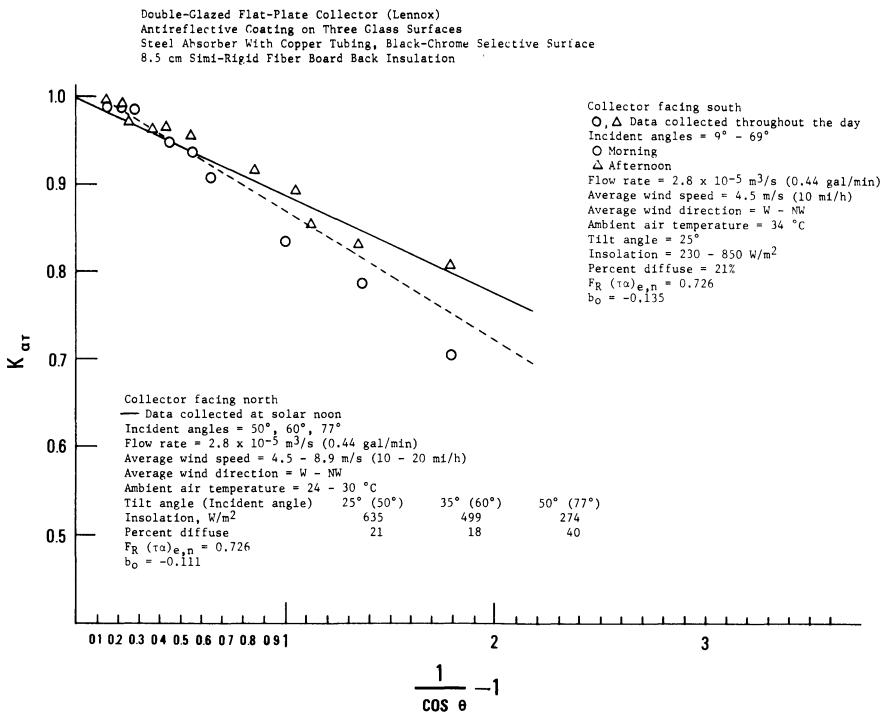


Fig. 7.11: Experimental results for the incident angle modifier test for a water-heating solar collectors.[19]

to obtain a specific incident angle across solar noon and was used to determine the solid line in the figure. The first technique was completed in one day, the data points indicated by circles taken in the morning and those by triangles in the afternoon. The dashed line is a curve fit through the average of these points. As can be seen by the fact that the circles and triangles are relatively close together, there was only a small thermal lag effect with this collector. This is consistent with its small time constant of 1.6 min. As can also be seen in Fig. 7.11, there was essentially a linear relationship between $K_{\alpha\tau}$ and the term $(1/\cos\theta) - 1$, which is consistent with the theoretical model for optical efficiency of a stationary flat-plate collector.[34]

It should be noted that the efficiency of the collector at incidence angles greater than zero can be used to characterize the performance of the collector in other ways than through the use of an incidence angle modifier as defined by Eq. (7.10). In the Australian Standard, a complete efficiency curve is established for incidence angles of 30°, 45°, and 60° in the same way the efficiency curve at normal incidence is determined.

7.4.1.4 Instrumentation. This subsection reviews the instrumentation commonly used to measure the primary physical variables during a collector test, discusses measurement techniques and difficulties encountered in using the instrumentation, and examines both typical and unexpected uncertainties contributed by the

instrumentation.*

The rate of energy collection defined by the numerator of Eq. (7.15) is commonly determined by measurements of the $\dot{m}, c_p$, and of the temperature difference. In practice, it is usually the volumetric flow rate of the heat-transfer fluid that is measured, so that the specific heat and the density of the fluid must be known. Water is well characterized as a heat-transfer fluid, but serious difficulties are encountered when aqueous glycol solutions, organic oils, or silicone fluids are used in order to reach higher operating temperatures. Their thermophysical properties depend on the specific composition of the fluid in the test loop, and may deviate from tabulated generic values by considerably more than the ± 0.5% uncertainty allowed in the ASHRAE Standards. When volumetric flow rate is measured, the deviations in the fluid density values alone can account for the ± 1.0% uncertainty specified for mass flow rate measurements.

The allowable uncertainty in temperature difference measurement is given in terms of temperature, reflecting the nature of the sensors themselves. Since the temperature difference can vary continuously to zero, a fixed measurement uncertainty can lead to a very large relative uncertainty. In ASHRAE Standard 93-77, the allowable uncertainty of the temperature difference measurement is ±0.1°C (±0.2°F). In ASHRAE Standard 96-1980 for unglazed solar collectors, where much smaller temperature differences occur, the requirement was tightened to ±0.01°C (±0.02°F).

The allowable uncertainty for irradiance measurements was set by fiat in the ASHRAE Standards. The expectation was that the overall uncertainty would be ± 2-3% of the measured value.

When the above factors are combined assuming statistical independence, the result is an expected ± 2-3% measurement uncertainty in the thermal delivery rate. Combined with the expected irradiance measurement uncertainty, this leads to a predicted uncertainty in the measured thermal efficiency of ± 4-5%.

7.4.1.4.1 Fluid Flow. A recent survey identified and compared 18 different generic types of flowmeters.[40] Despite the great variety available, turbine flowmeters continue to be the instrument commonly used in testing solar collectors using liquid heat-transfer fluids. Turbine flowmeters are available for a variety of flow ranges and offer excellent performance in terms of repeatability (± 0.1% of flow), absolute accuracy (± 0.25% of flow), and linearity (± 0.5% of flow in the linear range).[41] Turbine flowmeters are used as transfer standards because of this performance. Other attractive features include extended range (10 to 1 or more), digital output, rapid response, and retention of calibration. Fluid passing through a turbine flowmeter causes a magnetized rotor to spin on its bearings. Electrical pulses induced in an external pick-off coil by the spinning rotor are detected by an electronic discriminator. Since the rotation rate of the rotor is proportional to the

* This section addresses the instrumentation used in testing of liquid-heating collectors, although most of the information is also applicable to the testing of air collectors. Later, some precautions to be taken when testing air collectors will be described.

fluid velocity, the turbine flowmeter is inherently a volumetric flow device. A turbine flowmeter calibration is expressed in terms of a curve of calibration coefficient K (cycles/m^3) as a function of frequency. For a typical instrument, the curve is flat above a transition "hump" to within ± 0.5% over at least a decade in frequency, while the calibration coefficient falls below the hump, showing the nonlinear (but very reproducible) behavior of the instrument in the laminar flow regime.

Good as they are, turbine flowmeters are not perfect instruments. They must be calibrated for the working fluid, since their nonlinearity in the laminar and transition flow regimes makes their calibration viscosity-dependent. They are also particularly sensitive to flow patterns such as flow profiles, swirl, pulsation, and turbulence. Most instruments now include an internal flow straightener, but these are rarely sufficient protection. It has been shown that mounting a turbine flowmeter just downstream of a pipe bend creates shifts of 0.5 to 1% in calibration.[42] Sudden changes in pipe size, due to fittings, for example, are also troublesome. A typical rule of thumb calls for mounting the turbine flowmeter in a package incorporating ten or more diameters of uniform pipe on the upstream side with a flow conditioner at the inlet end, and five diameters of similar pipe downstream.[42] Wear or fouling of the bearings or rotor is a potential hazard that depends on the fluid, any fluid-borne particulate matter, the flowmeter design, and materials used in construction. In one instance, the observed 2.5% shift in one year of the calibration of a turbine flowmeter in a solar collector test loop was attributed to bearing wear.[19] An in-line filter is essential. Unsuspected two-component flow is another hazard. A recent study, noting that small amounts of gas are often present in liquid flow even when every attempt is made to eradicate them, went on to show that as little as 2% of air by volume could cause as much as 30% overestimation of the liquid flow rate in an extreme case.[43] Problems also arise in the signal conditioning. The electrical pulses induced in the pick-off coil decrease both in frequency and in magnitude as the flow rate decreases. An improperly adjusted disciminator or the pickup of AC noise can lead to substantial bias in the indicated flow. Large electrical currents through nearby resistance heaters can induce just this effect.

Flowmeters that measure the actual mass flow rate eliminate the need to accurately know the density of the heat transfer fluid. Both inertial and thermal instruments have been available for some years.[44] Gyroscopic mass flowmeters that depend on the Coriolis effect are gaining popularity.[45] The flow to be monitored is passed through a U-tube that is vibrated at its natural frequency. A flow-induced torque causes an angular deflection of the tube that is proportional to mass flow rate. Optical sensors measure the deflection and electronic circuitry processes the output. These flowmeters are available in a number of sizes and typically have a stated accuracy of ± 0.4% of mass flow with excellent linearity. Because they are mass flow transducers, their calibration is not susceptible to the two-component flow problem or the viscosity-dependent, nonlinearity mentioned above for turbine flowmeters. They also do not experience bearing or rotor wear. However, there has not been enough independent experience with these levels of accuracy, repeatability, and linearity, as with turbine flowmeters. One problem already known is that mechanical vibrations transmitted through the case will lead to degradation in performance. Specifications that these meters be mounted rigidly

to a firm foundation or structural member may be difficult to meet in many field installations.

Flowmeters that measure the $\dot{m}c_p$ product of the heat transfer fluid directly would be even more desirable than mass flowmeters for this application. There is a class of thermal flowmeters that accomplishes this,[44] although in the past, ironically, the readings were corrected for specific heat in order to give mass flow rate. These bulk heating flowmeters apply a constant thermal flux to the fluid and determine the $\dot{m}c_p$ product from the measured temperature rise. Thomas described the use of commercial instruments of this type to measure the flow rate of gas in a pipe in 1911.[46] These thermal flowmeters, also called reference heat sources[47] and calorimeters.[48,49] are currently the subject of investigations in the U.S., Canada, the U.K., and Sweden as part of an International Energy Agency (IEA) Solar Heating and Cooling Research Program.[50] They show promise for giving the performance of more conventional methods when using well-characterized heat transfer fluids and for giving unmatched performance when using fluids whose thermophysical properties are only approximately known. This level of performance will be achieved only if the instruments are designed, calibrated, and used carefully. It is clear that heat losses must be kept to a minimum and the thermal efficiency be well characterized for the measurement applications.

Although the thermal flowmeter takes into account the specific heat of the working fluid, the temperature dependence of the specific heat is still an issue. Other considerations become apparent when the thermal flowmeter is viewed as a substitute for an in-line reference solar collector. Steady-state operation is imperative. Provision must be made to accommodate the no-flow (stagnation) condition. At low-flow rates, the flowmeter may be operating in the transition or laminar flow regime. At high flow rates, the applied thermal flux must be kept high to maintain a reasonable temperature rise. The objectives of the IEA program are to explore these issues, to establish the levels of accuracy and precision attainable with these devices, and to establish a recommended practice for their design, construction, and use.

Whatever the choice of flowmeter in collector testing, measurement assurance is enhanced when tandem instruments are used to provide redundant measurements. While agreement between readings does not guarantee correct measurement, differences help flag flowmeter problems. When thermal flowmeters are used, it is essential to monitor the flow conditions independently using some kind of mechanical flowmeter.

7.4.1.4.2 Temperature Difference. Thermopiles or pairs of precision resistance thermometers (including thermistors) are the available choices for measuring the small temperature differences found in collector testing. The measurement practice using these sensors is described in ASHRAE Standard 41.1-74.[51] A thermopile generates an output voltage that is proportional to the temperature difference between multiple thermocouple junctions. The number of junctions selected to make up a thermopile is dictated by the measurement accuracy, precision requirements, and the quality of associated readout. Typically, six to ten is the minimum acceptable number of type-T thermocouples in a thermopile intended to provide $\pm 0.1°C$

($\pm 0.2°F$) to $\pm 0.01°C$ ($\pm 0.02°F$) accuracy. These tight accuracy requirements can be met only with careful construction using selected thermocouple wire for the junctions and the extension wires, and careful calibration using controlled temperature baths.

Platinum resistance thermometers remain the predominant choice in resistance thermometers. Because of their accuracy and repeatability, they are capable of being secondary standards. The temperature coefficient of the electrical resistance of very pure platinum wire is well known. The change in resistance of a platinum resistance thermometer can thus be related precisely to a change in temperature. To provide the required accuracy, however, platinum resistance thermometers must be calibrated in controlled baths with their associated cabling, signal conditions, and readout devices in circuit. Although the temperature coefficient of the resistance is well known, the base-reference resistance varies slightly from probe to probe due to normal manufacturing tolerances. Specifying matched pairs of probes can reduce the magnitude of the resulting temperature measurement offset, but, in general, does not eliminate it. It is possible to calibrate precision resistance thermometers in the laboratory, with millidegree accuracy.

Matched thermistors are emerging as an alternate choice in resistance thermometers. Made of semiconducting materials, thermistors offer potentially high precision in temperature difference measurements due to their very large temperature coefficient. It has been difficult to maintain the desired accuracy over an extended temperature range, however, and long-term stability has been a problem in the past. Nevertheless, new sensors are becoming available that exceed the capabilities of the old.

Using any of these devices, the major difficulty lies in making accurate measurements in situ rather than in controlled temperature baths. The problems that arise are commonly deficiencies in practice rather than in the specific instrumentation, but the uncertainty introduced in the temperature difference measurement can be large, $\pm 0.5°C$ ($\pm 0.9°F$) and greater. The temperature probes are typically inserted into the fluid loop through compression fittings, and it is easy to incur measurement offsets due to insufficient immersion of the probes into the heat-transfer fluid. Cooling of the probe by thermal conduction along the connecting leads, and, in the case of resistance thermometers, self-heating of the probe by its excitation current can result in significant measurement errors. Shifts in calibration due to mechanical strain, spurious readings due to electrical noise, inadequate fluid mixing to ensure bulk temperature measurements, and inadequate insulation around the temperature measurement sections leading to nonrepresentative measured temperature differences are additional hazards often encountered. As in the case of flow measurements, using redundant temperature instrumentation adds to the measurement assurance by aiding the detection of measurement problems.

7.4.1.4.3 Irradiance. Geist observed a number of years ago that "This field (radiometry) has the dubious distinction of being an area of measurement science that is characterized by surprisingly low accuracy. One percent is rarely achieved ... Typically, five percent radiometry, if not pushing the state-of-the-art, is at least very near it."[21] Losing sight of this fact has led to innumerable disagree-

ments in collector testing. The recent wide availability of self-calibrating absolute radiometers has greatly improved the interlaboratory comparability in making measurement of the direct solar irradiance, but the situation remains essentially unchanged in making measurements of the global solar irradiance using pyranometers with thermopile sensors. In these instruments, solar radiation falling on a blackened sensor warms it relative to the instrument case, and the resulting temperature difference is sensed by a thermopile. They are, in fact, miniature solar collectors and share all the characteristics of their big brethren. Hence, in general, their output depends on a number of factors besides the intensity of the incident radiation, including the ambient temperature. the inclination and orientation of the instrument, and the angular and spectral distributions of the incident radiation.

In ASHRAE Standards 93-77 and 96-1980, these dependencies are acknowledged, but the allowable measurement uncertainty was set by fiat.

The Standards specify that the pyranometers used in collector testing must meet or exceed the World Meteorological Organization (WMO) specifications for a first class instrument.[52] These specifications result in an expected irradiance uncertainty of ± 2 to 3%. Despite the fact that only selected thermopile pyranometers have been so classified, it has been common practice in collector testing to use off-the-shelf instruments that are considered as a class to meet the WMO specifications de facto. Presumably, this practice has persisted because few laboratories have the instrumentation and test facilities required to calibrate and check their own pyranometers. In addition, the standard calibration provided with pyranometers is one that is appropriate for meteorological work in which the instruments are labeled and used to make long-term solar energy availability measurements. Meteorological calibrations ensure that the uncertainty of this long-term measurement is minimized by averaging over angle of incidence and irradiance level. The end-use conditions are quite different in testing solar collectors, where the irradiance level is always high, the angles of incidence are specified, and the instrument is tilted far from horizontal.

This situation has been reviewed in a recent summary of some round-robin pyranometer calibration experiments.[53] This paper showed that the uncertainty introduced in solar collector testing due to various pyranometer-dependent effects and the use of meteorological calibrations reach ± 4% and more, but that this uncertainty could be reduced markedly through appropriate calibration procedures that take into account the end use of the instrument.[54] The proper characterization of pyranometers is a continuing research activity in the International Energy Agency Solar Heating and Cooling Research Program.[55] Since the irradiance measurement uncertainty dominates the overall measurement uncertainty in determining the thermal performance of solar collectors even when the WMO specifications are met, it is imperative that the pyranometers be characterized to at least that level of accuracy.

7.4.2 Shortcomings of the Assumed Collector Model

In describing the performance of a solar collector by Eqs. (7.3) and (7.5), it has already been mentioned that the heat loss coefficient, U_L, is not a constant. It is a function of the operating temperature of the collector (that is, the plate

temperature or fluid temperature), the tilt angle of the collector (effect on the free convection loss component), and the ambient weather conditions (that is, environmental temperature and wind velocity and direction). In specifying standard methods of testing collectors outdoors, the tilt angle can be specified. In addition, the tests are generally conducted under very low wind conditions. Standardizing some of the conditions in this way helps to make the test results comparable; however, estimating the performance of the collector at other tilt angles or wind conditions, or both, must be done using other sources of information. The other test conditions can only be required to be within prescribed ranges; otherwise, it would be virtually impossible to complete a test series outdoors. Unfortunately, this necessary variation can cause results on the same collector obtained at different times or different test sites to be different. The extent to which this happens for flat-plate collectors will be described in the next section.

The variation in the value of the top loss coefficient* with operating conditions for typical flat-plate collectors is given in Fig. 6.4.4 to 6.4.6 of Duffie and Beckman[31]. In addition, Klein developed an empirical relationship for the coefficient that includes the four major variables (in addition to material properties) affecting the coefficient: (1) tilt angle, (2) ambient temperature, (3) mean plate temperature, and (4) wind heat transfer coefficient at the outer surface of the exterior glazing.[56] Because, in general, there is both radiation and free convection occurring in the heat-loss process, the top-loss coefficient is a function of both the average of the mean plate and ambient temperature and the difference between them.

In order to account for other than a constant U_L in testing collectors and correlation of test results, a wide range of options has been used and is available. The large majority of the techniques has been developed for flat-plate collectors which operate at relatively low temperatures for solar domestic water heating and space heating applications. Tabor[6] states that U_L can be described by

$$U_L = C_1(\bar{t}_f - t_a)^{C_2} \tag{7.16}$$

Dunkle and Cooper[9] and Symons and Cooper[10] proposed †

$$F'U_L = C_3 + C_4(\bar{t}_f - t_e) \tag{7.17}$$

Shewan and Hollands[57] proposed

$$U_L = C_5\left(\frac{\bar{t}_p + t_a}{2}\right)^{C_6} \tag{7.18}$$

for flat-plate collectors designed to substantially reduce the free convection compo-

* The heat-loss coefficient, U_L, differs from the top-loss coefficient only by the amount of losses occurring through the edges and back of the collector, which are generally small.

† In the correlation of data, the Australians use an environmental temperature, t_e, instead of t_a, where t_e is assumed to be 3°C lower than t_a.

nent of the heat loss or

$$U_L = C_5\left(\frac{\bar{t}_p + t_a}{2}\right)^{C_6} + C_7(\bar{t}_p - t_a) \tag{7.19}$$

when both radiation and free convection are substantial. Proctor[58] has proposed

$$F'U_L = (C_8 + C_9 v)\frac{\bar{t}_f - t_e}{I_t} + (C_{10} + C_{11}v)\frac{(\bar{t}_f - t_e)^2}{I_t} + (C_{12} + C_{13}\bar{t}_f)\frac{(\overline{T}_f^4 - \overline{T}_s^4)}{I_t} \tag{7.20}$$

which obviously accounts for both radiation and convection heat-loss phenomena in a rather rigorous manner.

In ASHRAE Standards 93-77 and 96-1980, the possibility that test data when plotted will not result in a linear relationship as indicated by Eq. (7.3) (and hence variable U_L) is acknowledged by an allowance to use a second-order fit to the data with the independent variable being $(t_{f,i} - t_a)/I_t$. Unfortunately, if one curve is plotted, this results in an assumed dependence of U_L on irradiance that is in violation of the known physical phenomena described above.[59,60] If the data were correlated by

$$\eta_a = C_{14} + C_{15}\left(\frac{t_{f,i} - t_a}{I_t}\right) + C_{16}\left(\frac{t_{f,i} - t_a}{I_t}\right)^2 \tag{7.21}$$

then by comparing Eqs. (7.3) and (7.21),

$$U_L = -\frac{C_{15}}{F_R} - \frac{C_{16}}{F_R}\left(\frac{t_{f,i} - t_a}{I_t}\right) \tag{7.22}$$

If a second- or higher-order fit is to be used with test data to account for variations in U_L, there are two choices in properly correlating the data. This first is to use a correlation that is consistent with the assumed dependence of U_L on independent variables. For example, in accordance with the Australian Standard,[30] the data are correlated by

$$\eta_a = C_{17} + C_{18}\left(\frac{\bar{t}_f - t_e}{I_t}\right) + C_{19}\left(\frac{\bar{t}_f - t_e}{I_t}\right)^2 \tag{7.23}$$

which is consistent with the assumptions about U_L in Eq. (7.17). A second choice is to continue to use the kind of plot that has become accepted through the use of ASHRAE Standard 93-77 with $(t_{f,i} - t_a)/I_t$ as the independent variable, but to plot the data on several curves to account for the correct dependence of U_L. For example, if for low temperature applications of flat-plate collectors Eq. (7.17) is assumed to be correct, Eq. (7.23) could have been written

$$\eta_a = C_{17} + C_{18}\left(\frac{\bar{t}_f - t_e}{I_t}\right) + I_t C_{19}\left(\frac{\bar{t}_f - t_e}{I_t}\right)^2 \tag{7.24}$$

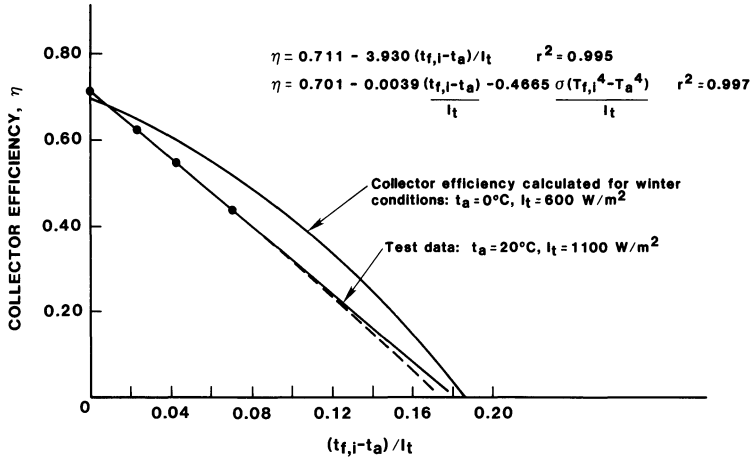


Fig. 7.12: Illustration of potential errors in extrapolation of solar collector test data using multiple correlation parameters.

and the type of plot called for in the ASHRAE Standards used provided separate curves are given for different levels of irradiance.

A priori, it can be argued that for the higher operating temperatures radiation heat losses become significant. A correlation that has been used to account for this is [61]

$$\eta_a = C_{20} + C_{21}\left(\frac{t_{f,i} - t_a}{I_t}\right) + C_{22}\sigma\left(\frac{T_{f,i}^4 - T_a^4}{I_t}\right) \tag{7.25}$$

In a recent discussion of the performance of high-temperature, flat-plate collectors having relatively low heat-loss coefficients, Gani, Proctor, and Symons recognized that the loss by free convection is very low and that U_L is primarily a function of the average fluid temperature.[62] As the result of an analysis similar to the one above, they recommended using the standard plot, providing different curves for different values of fluid temperature.

For a given set of test data, Eqs. (7.23), (7.24), and (7.25) may fit the data better than Eqs. (7.3) and (7.18). However, Fig. 7.12 illustrates what may happen when the equations are used to predict collector performance for conditions other than the test conditions. Multiple correlation parameters should be used only with great care.

It should be noted that design procedures commonly used for predicting the performance of solar energy systems require a linear model for the solar collector as indicated by Eqs. (7.3) and (7.5).[63,64] Consequently, if the results of a collector test warrant the use of a higher-order fit, as discussed above, care must be taken in linearizing the results prior to use in the systems models. This is discussed in detail by Gani, Proctor, and Symons,[62] Cooper and Dunkle,[65] and Proctor.[66]

7.4.3 Comparability of Results from Outdoor Tests

Since the mid-1970s, there have been at least four major solar collector programs involving the testing of one or more collectors at several outdoor test sites where the results have been compared. In some cases, quite a lot of analysis was done to explain the differences in results.

In 1976 and 1977, a round-robin program was organized by the National Bureau of Standards to test two liquid-heating, flat-plate collectors at 21 test facilities located across the United States.[67,68] The program was initiated after NBS had made its recommendations for a standard procedure, but prior to the adoption of ASHRAE Standard 93-77.[20-22] The participants included three government laboratories, eleven universities, and seven industrial organizations covering a wide range of geographic locations and climatic conditions. Two commercially available collectors were selected from firms employing established manufacturing techniques. Collector #1 had two tempered glass cover plates over an aluminum absorber with integral flow passages and a nonselective black coating backed with glass fiber insulation and contained in a sheet metal enclosure. Collector #2 had a single tempered glass cover over a stitch-welded and pressure-expanded absorber made from two steel sheets. The absorber was coated with a selective surface and also backed with glass fiber insulation.

The tests were conducted during a time when many organizations were just beginning to establish their facilities and modifications were being made to meet the requirements of the new ASHRAE Standard, then still under consideration. Consequently, the variation in test results from all the organizations was large and not indicative of the variation that would occur if the program were repeated today. The coefficient of variation (standard deviation expressed as a percent of the mean value) for the y-intercept of efficiency curve for collectors #1 and #2 were 7.7% and 4.7%, respectively, and for the slopes assuming a linear plot were 15.8% and 24.9%, respectively.

One of the most interesting aspects of this program was the analysis done to determine what effect the difference in environmental conditions at the different test sites had on the spread of results. A detailed, computerized thermal model was made of each of the two collectors and used to calculate the change in efficiency that would result from a change in test conditions. The data from all the test laboratories were then screened to determine which results were completely documented with respect to test conditions. They were then further reviewed to determine which met the requirements of ASHRAE Standard 93-77.

Figure 7.13 shows, collectively, the efficiency of collector #2 as reported by ten test sites reporting complete data. The efficiency values, as reported (Fig. 7.13a), show considerable scatter. Figure 7.13b shows the theoretical efficiencies, from the analytical mode, for the actual test conditions reported and, consequently, the allowable scatter in the absence of experimental error. Figure 7.13c shows the results of correcting the measured efficiencies to a set of reference test conditions (chosen to be the mean of all actual test conditions to minimize the overall adjustment). The mean value of the square of the distance from the points to the curve (mean square)

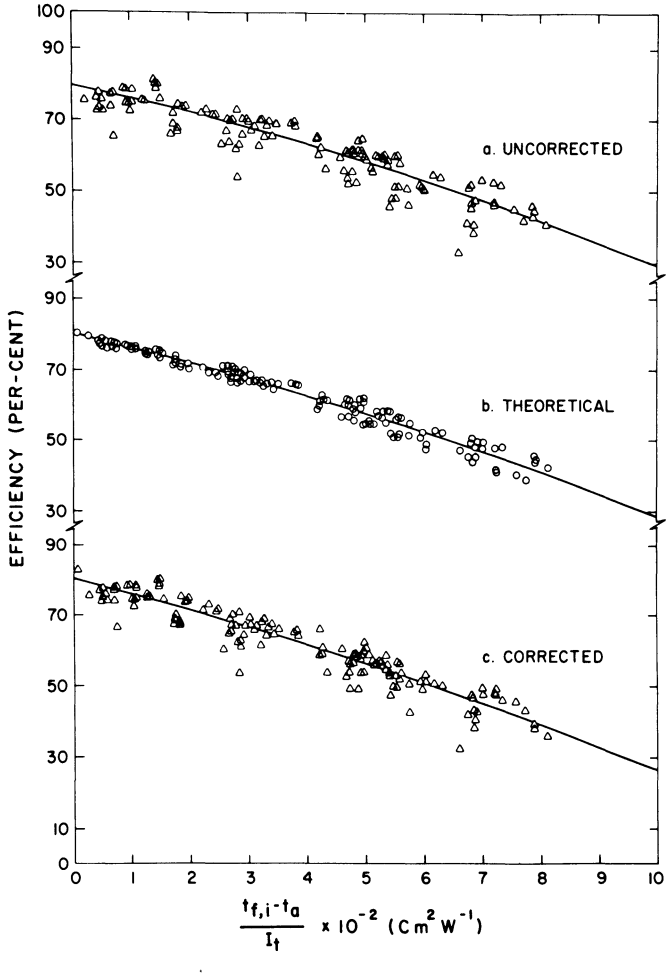


Fig. 7.13: Collector #2 efficiency curves based on measurements from 10 test sites (160 values) in the NBS Round-Robin program.[68]

was used in the analysis to quantify the closeness of correlations. A mean square value of 17.5 in Fig. 7.13a was reduced to 13.5 in Fig. 7.13c. Figure 7.13b shows the scatter theoretically justified in Fig. 7.13a by the variations in the environmental test conditions, a mean square value of only 3.7. Consequently, the scatter remaining in Fig. 7.13c was caused by reasons other than the different test environments.

Figure 7.14 shows the results for collector #2 for the subset of tests which reportedly met the requirements of ASHRAE Standard 93-77. The mean square value was substantially reduced to 4.6 in Fig. 7.14a, and Fig. 7.14b had a mean square value of 3.7. Applying the correction procedure, the mean square value was only slightly reduced to a value of 4.0 in Fig. 7.14c. The substantial improvement in results when considering those tests which reportedly met the requirements of Standard 93-77 was probably due as much to the improved experimental procedures

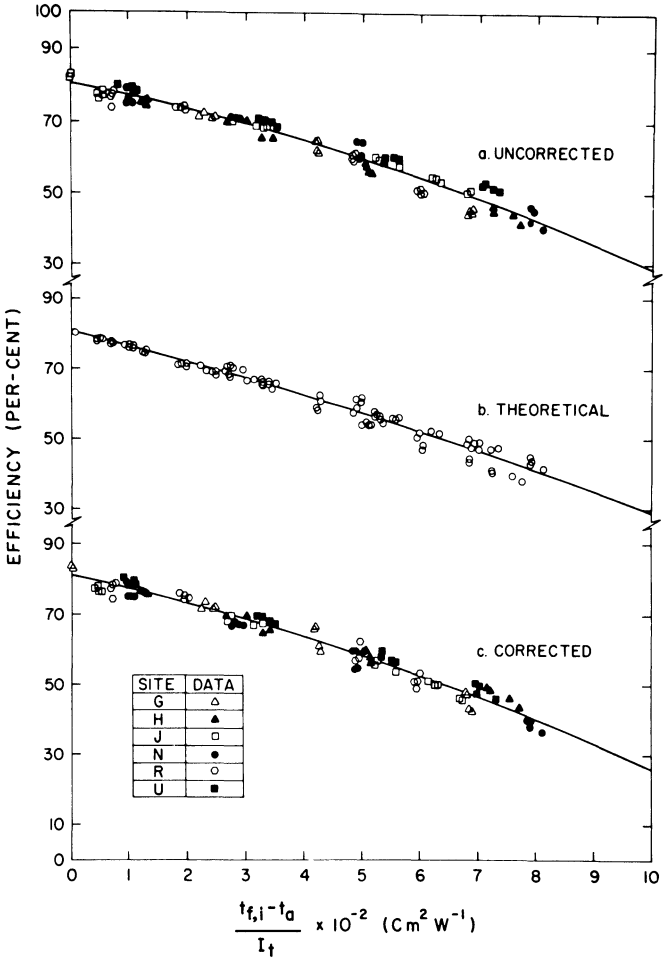


Fig. 7.14: Collector #2 efficiency curves based on measurements from six test sites (96 values) meeting ASHRAE Standard 93-77 requirements in the NBS Round-Robin program.[68]

being used by those participants as to the more restrictive requirements.

When the data of Fig. 7.14 were examined closely, it was observed that test results from individual sites exhibited relatively small scatter. (This was substantiated when the data of Fig. 7.13 were identified with individual test sites.) The relatively large differences between measured efficiencies at different sites suggested systematic instrumentation errors or, at least, systematic differences between facilities.

Taking into consideration the characteristics of collector #2 and the allowable uncertainty in the measurements of temperature, temperature difference, mass flow rate, and solar radiation as specified or estimated from ASHRAE Standard 93-

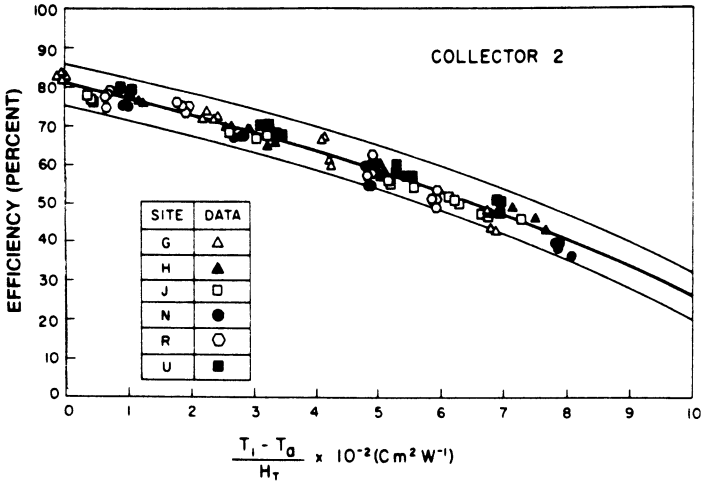


Fig. 7.15: Collector #2 efficiency data based on measurements from six test sites (96 values) meeting ASHRAE Standard 93-77 requirements with 'maximum' error band in the NBS Round-Robin program.[68]

77, a maximum error band was computed for typical test conditions. It is shown in Fig. 7.15 along with the data from all participants meeting the requirements of Standard 93-77 after the data were corrected to reference conditions. Initially it was felt this figure showed that the scatter was entirely due to the allowable uncertainty in instrumentation; however, further analysis indicated that a more probable error band based on random occurrence of the instrumentation inaccuracy fell inside the data spread of Fig. 7.15.[67,68] Thus, systematic differences from facility to facility undoubtedly did exist.

The important conclusion drawn from this round-robin program was that the primary reason for the differences in measured collector efficiency between test sites resulted from experimental error or systematic differences from facility to facility. Less than one-third of the scatter could be attributed to different test conditions.

As part of the International Energy Agency tasks on solar heating and cooling, a round-robin program similar to the one described above was initiated and tests completed in 1977 and 1978.[69] Two liquid-heating, flat-plate collectors (one of which was identical to collector #2 in the NBS round-robin program) were distributed to 16 laboratories situated in 12 different countries in the Northern Hemisphere. Tests were completed in ten of the laboratories on one of the collectors and in 11 on the other in accordance with the ASHRAE recommended procedure. No analysis was done to account for the effect of environmental test conditions as in the previous program since the data available were not complete. However, it was shown that all the reported data fell within a band bracketed by theoretical efficiency curves established, assuming that extremes of allowable environmental test conditions occurred simultaneously with allowable measurement errors. In addition, there was not a prominent grouping of data around the mean values as would be expected and therefore systematic differences between test facilities were suspected.

A project was initiated at NBS in 1978 to investigate accelerated and real-time test methods for predicting collector durability and reliability.[70] Representative, commercially available, flat-plate collectors and their materials were exposed outdoors in four different U.S. climatic regions. As part of the collector characterization, initial thermal performance tests were made on multiple samples of each collector at each test site. These tests consisted of efficiency and incident angle modifier tests on eight collector types at each of the four sites, with two to four collectors being tested at each site. Therefore, the data indicated not only the reproducibility of results between test sites as in the previous two programs, but also the repeatability of results within a single test site.[71,72]

In contrast to the relatively large spread in y-intercept and slope (assuming a linear fit to the data) experienced in the original NBS round-robin program, the average coefficient of variation for the y-intercept for all the eight collector types tested* was 2.4%, and the slope 8.4%. This improvement compared to the results of the first NBS round-robin program (4.7% to 7.7% in y-intercept and 15.8% to 24.9% in slope), indicates improvement in experimental procedures in collector testing over the intervening two to three years. When the within-test-site variation was analyzed separately, the average coefficient of variation in the y-intercept and slope only reduced to 2.1% and 5.9%, respectively.

The most interesting aspect of this test program was the analysis done on the incident angle modifier (IAM) results,[73] tests not conducted in either of the two previous round-robin programs. Test data on four of the eight collector types tested at three of the four laboratories were analyzed.

In describing how the near-normal-incidence efficiency of the collector changes with changes in incident angle, the concept of the incident angle modifier defined by Eq. (7.9) was originally conceived as accounting for the change in material radiative properties of the glazing system and absorber of a flat-plate collector and, consequently, only a function of θ.[34] In analyzing the data from the durability/reliability program, Thomas et al. showed that the effect of shading the absorber by the ends and sides of the four collectors considered was the same order of importance as the change in transmitting properites of the cover-absorber assembly as the incident angle changes.[73] Consequently, flat-plate collectors may exhibit a bidirectional, angular response.

Figure 7.16 shows the results for the four collector types analyzed.[73] The spread in results and experimental curves shown was much larger than had been expected. The IAM calculated from theory would lie somewhere within the area bracketed by the dotted lines, depending on the orientation of the collector (effect of shading mentioned above). As can be seen, the calculated IAM values would lie generally within the spread of experimental curves. The mean IAM curve based on all measurements corrresponded closely to curves 8, 7, 5, and 6 for collectors A, B, D, and E, respectively. The relative agreement between calculated and measured results was about the same for the collectors with flat-black absorbers (B and E)

* There were four collectors of each type tested at two sites with two collectors of each type tested at the other two sites, for a total of 96 collectors tested.

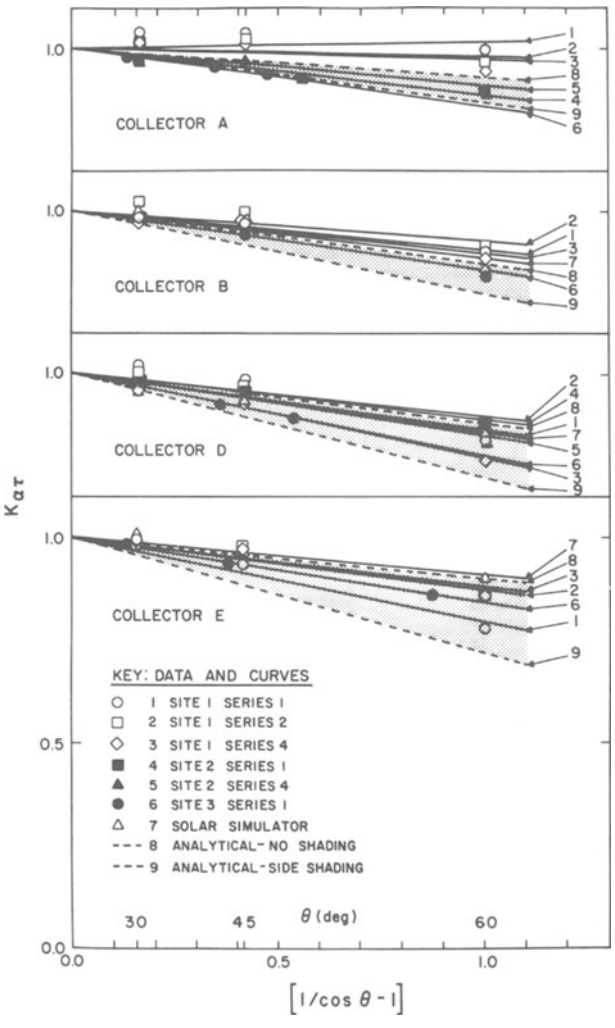


Fig. 7.16: Incident angle modifier data, correlations, and calculated curves for four collectors.[73]

and those with selective absorbers (A and D).

The probable uncertainties in results due only to the uncertainties in measurements made were calculated* and are shown in Fig. 7.17 as the dashed lines. Also shown are the actual data and resulting mean experimental curves. By comparing Figs. 7.16 and 7.17, it is seen that the spreads in experimental IAM curves (Fig. 7.16) are about the same as the ranges due to measurement uncertainty (Fig. 7.17).

* Assuming the uncertainties in the individual measurements were equally probable, this is the square root of sum of the squares of the individual uncertainties.[74]

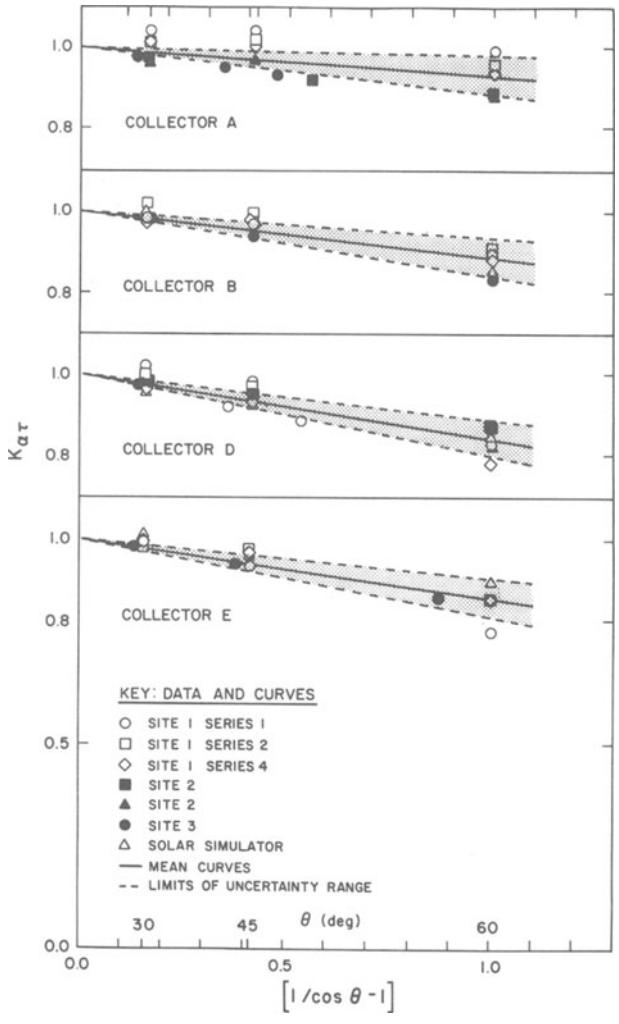


Fig. 7.17: Correlations and uncertainty ranges for incident angle modifiers based on measurements for four collectors.[73]

The large uncertainty in test results for the IAM of conventional flat-plate collectors was found to have only a minor effect on the uncertainty in predicted seasonal performance of solar space heating and domestic hot water systems with relatively efficient collectors and met a reasonable load. For example, the uncertainty in calculated seasonal solar fraction for a combined system in Madison, Wisconsin, was only 0.05 as a result of the uncertainty for collectors D and E indicated in Fig. 7.16.

As a result of the above analysis, Thomas et al. concluded that for flat-plate collectors, the uncertainty in determining the IAM experimentally is the same order

of magnitude as the efficiency reduction from the normal to the maximum incident angle. Therefore, a simplified calculation procedure was recommended in lieu of the experimental procedure. In addition, it was recommended that the experimental procedure be used for collectors with a strong thermal performance dependence on incident angle, that is, with complex geometries on concentrating features. However, for these collectors the IAM is expected to depend strongly on the scattered fraction of solar radiation which is not accounted for in the present procedure.

The fourth and most recent major collector program was the DOE Solar Collector Testing Program, funded by the U.S. Department of Energy. It was a comprehensive project to provide uniform and comparable thermal performance test data on a significant portion of the solar collectors commercially available in the U.S. in the late 1970s. This was a joint effort of the federal government and the U.S. solar industry to accelerate the development and implementation of a nationally recognized certification, rating, and labeling program for solar collectors. Kirkpatrick summarized the results for the 116 solar collectors tested.[75]

Several laboratories tested samples of the same collector model. A summary of the test results for the flat-plate, liquid-heating collectors is shown in Table 7.1. The comparability of results between the laboratories is given in Table 2. Note that the variations in the measured y-intercept are within 5% as one would expect from an uncertainty analysis. However, the variations for the slope are larger than one would expect. There were also large differences in the measured uncertainty of the incident angle modifier as Thomas et al. found in the NBS durability/reliability program.

7.4.4 Testing Concentrating Collectors

The testing procedures described in this section were developed primarily for flat-plate collectors. However, the procedures contain the essential elements required to test a concentrating collector and can be applied if the individual tests are adapted to fit the particular type of concentrator under consideration.

Figures 7.3 through 7.5 are typical of test loops used for collectors operating at temperatures below approximately 100°C (212°F) and at only moderate pressures. Harrison et al.[76] describe a test facility of similar concept, but built to allow testing to 315°C (600°F) and 0.51 MPa (75 psig). Because of the high temperatures and pressures, heat transfer oils are frequently used as the working fluid and reference heat sources or calorimeters are employed to accurately determine the collector thermal output.[47-49]

It is unrealistic to require tracking collectors having high concentration ratios to be exposed to any preconditioning test, such as the three-day test specified in ASHRAE Standard 93-77 because of the extremely high temperatures likely to occur. The normal control strategy with such collectors is to defocus the concentrator whenever the absorber temperature reaches a preset temperature.

The time constant test can be applied as described previously in this section. Wood et al. conducted the full range of tests specified in Standard 93-77 for a single-axis, tracking, linear, parabolic, trough-concentrating collector having a geometrical

Table 7.1

Results of the DOE solar collector testing program for liquid heating collectors[75]

	Lab	First Order Equations (Btu) Baseline	Post-Stag.	IAM	Time Constant	Pressure Drop	A_g	A_a
1*	C	0.675–1.047 P	0.662–1.129 P	0.998–0.079 P	2.46	0.15/0.6	33.18	29.93
	B	0.736–1.023 P	0.739–1.050 P	0.98 –0.095 P	2.67	0.21/0.95	33.17	29.99
	E	0.658–0.975 P	0.650–1.112 P	1.01 –0.150 P	3.34	0.15/0.9	32.71	29.93
	A	0.618–0.897 P	0.646–0.960 P	1.005–0.075 P	2.62	0.20/0.88	31.66	30.05
	G		0.726–1.111 P					
	I	0.681–1.285 P	0.707–1.318 P					
2*	A	0.753–0.634 P	0.716–0.582 P	0.998–0.084 P	2.10	No data	21.21	19.16
	E	0.746–0.762 P	0.704–0.712 P	1.02 –0.120 P	2.43	0.9/0.6	21.21	19.18
	I	0.798–0.873 P	0.757–0.835 P	1.00 –0.125 P	1.83		21.30	19.26
3*	G		0.643–1.437 P					
	D	0.662–1.289 P	0.652–1.201 P	Very scattered points; no equation	1.9	0.20/0.6	24.4	24.51
	A	0.674–1.249 P	0.675–1.233 P	1.009–0.122 P	1.3	0.40/0.74	24.49	22.76
	I	0.707–1.179 P	0.701–1.179 P	1.00 –0.080 P	1.0	No data	24.43	22.92
	E	0.666–1.154 P	0.621–1.128 P	1.02 –0.120 P	2.21	0.62/0.69	24.46	23.09
	B	0.712–1.098 P	0.697–1.138 P	0.98 –0.100 P	1.50	0.20–0.38/0.75	24.81	23.00
4*	A	0.597–0.702 P	0.577–0.781 P	0.996–0.109 P	1.47	No data	24.18	23.21
	E	0.618–0.832 P	0.565–0.750 P	1.01 –0.220 P	2.01	0.60/0.7	25.69	23.19
	F	0.579–0.921 P	0.605–0.947 P	1.00 –0.180 P	1.27	0.179/0.7	25.64	23.78
	B	0.66 –0.773 P	0.647–0.850 P	0.98 –0.175 P	1.30	0.10/0.75	25.76	23.82
	I	0.609–0.910 P	0.573–1.071 P					
5*	E	0.676–1.016 P	0.657–1.0073 P	1.02 –0.13 P	2.57	0.12/1.2 0.25/2.2	48.46	45.11
	D	0.686–1.216 P	0.657–1.166 P	Data no good	2.60	0.013/0.8	48.58	46.01
	A	0.683–1.059 P	0.649–1.080 P	1.005–0.163 P	2.33	1.5/2.3	48.46	45.17
	B	0.726–1.175 P	0.708–1.163 P	1.00 –0.100 P	2.95	0.4/1.8	48.33	46.08
	G	0.646–0.978 P	0.699–1.259 P	1.00 –0.090 P	2.66	0.24/1.5	48.33	46.02
6*	C	0.685–1.033 P	0.679–1.304 P	0.010–0.088 P	1.91	0.08/0.6	21.62	18.84
	A	0.681–0.916 P	0.679–1.196 P	1.007–0.118 P	1.80	0.16/0.56	20.20	18.88
	H	0.716–1.126 P	0.699–1.171 P	1.00 –0.116 P	2.0	—	21.62	18.92
7*	I	0.787–1.390 P	0.798–1.420 P	1.00 –0.090 P	1.00	—	22.98	20.40
	H	0.801–1.481 P	0.792–1.355 P	1.00 –0.131 P			22.92	21.31
	G		0.78 –1.719 P					
	C	0.775–1.425 P	0.762–1.373 P	1.02 –0.138 P			23.12	21.33
		$P = t_i - t_a/I_t$	$P = t_i - t_a/I_t$	$P = 1/\cos\theta$	min	lb/in² at a flow rate of gal/min	ft²	ft²

*Collector number.

concentration ratio of 64:1, an operative area of 16.38 m² (176.3 ft²), and a focal length of 91.4 cm (3 ft).[77] The receiver was a 2.5-cm (1-in.) diameter copper tube coated with a selective surface. The time constant test was performed by maintaining the inlet temperature to the collector while in the defocused position and suddenly moving it to the focused position and vice versa. The time constant was found to be 1.4 min and 1.8 min, respectively.

The testing and determination of the peak noon efficiency of five different concentrating solar collectors having geometric concentration ratios up to 67:1 have

Table 7.2

Comparability of Results of the DOE Solar Collector Testing Program for Liquid–Heating Collectors[75]

	Baseline						Post-Stagnation								
	Intercept			Slope			Intercept			Slope			IAM		
Collector	Mean	Std. Dev.	%	Mean	Std. Dev.	%	Mean	Std. Dev.	%	Mean	Std. Dev.	%	Mean	Std. Dev.	%
1	0.658	0.028	4.3	−1.051	0.168	16.0	0.678	0.036	5.3	−1.126	0.127	11.3	−0.100	0.035	35
2	0.766	0.028	3.7	−0.756	0.120	15.9	0.726	0.028	3.9	−0.710	0.127	17.9	−0.110	0.022	20
3	0.677	0.021	3.1	−1.218	0.062	5.1	0.658	0.031	4.7	−1.236	0.119	9.6	−0.107	0.024	22
4	0.601	0.017	2.8	−0.841	0.101	12.0	0.58	0.017	2.9	−0.887	0.150	16.9	−0.171	0.046	27
5	0.673	0.018	2.7	−1.067	0.105	9.8	0.666	0.023	3.4	−1.128	0.109	9.7	−0.121	0.033	27
6	0.694	0.019	2.7	−1.025	0.105	10.2	0.687	0.012	1.7	−1.224	0.071	5.8	−0.107	0.017	16
7	0.788	0.013	1.6	−1.432	0.046	3.2	0.783	0.016	2.0	−1.467	0.170	11.6	−0.120	0.026	22
Average Variance (σ)		0.021	2.98		0.101	10.3		0.023	3.4		0.125	11.8		0.029	24

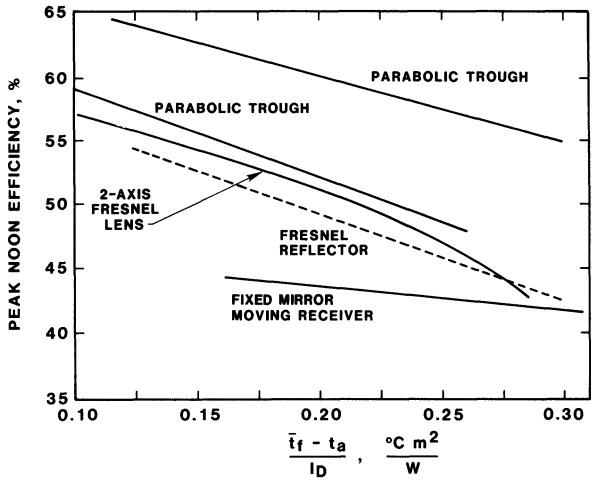


Fig. 7.18: Comparison of peak solar noon efficiency for five concentrating solar collectors.[16]

been reported by Dudley and Workhoven.[18] The results of their tests are shown in Fig. 7.18 (with the average fluid temperature in the axis of the plot rather than the inlet fluid temperature normally used) and are in agreement with the results of Wood et al.[77] The tests in Standard 93-77 were modified to meet the specific characteristics of the collectors. Whereas the optical efficiency of a simple flat-plate collector is the same at an incident angle of 30° as 0°, it can be quite different at those two incident angles for a concentrator. Therefore, the near-normal-incidence efficiency tests should be conducted well within the acceptance angle of the collector and not just anywhere in the range 0° to 30°. The tests by Dudley and Workhoven[18] and Wood et al.[77] were conducted at times close to solar noon with the collectors in the tracking mode.

Standard 93-77 does require that, if concentrators are tested, both a pyrheliometer and pyranometer be used so that both the total incident radiation and the direct component can be determined. Two efficiency curves are then required: one in which the global incident radiation is used for the calculation of efficiency and in the x-axis of the plot and a second in which only the direct component is used. Figure 7.18 is a plot where the direct component has been used and is the only one of importance for the collectors tested by Dudley and Workhoven,[18] since they all had high concentration ratios and, for all practical purposes, only used the direct normal irradiance.

The use of a single curve to describe the performance of collectors, such as those in Fig. 7.18, is even less appropriate for concentrating collectors than for flat-plate collectors because of the change in heat-loss coefficient with operating conditions as discussed previously. Following the recommendation of Gani et al.,[62] the heat loss from high temperature collectors is predominantly a function of operating temperature and, therefore, curves like those of Fig. 7.18 for different operating temperatures would give a more complete characterization of the collectors' perfor-

mance. Dudley and Workhoven conducted tests on each of the five concentrating collectors while in the defocused position and determined the heat loss as a function of the difference between receiver temperature and ambient.[18] The data could have been used for developing additional efficiency curves, such as those recommended by Gani et al.

The most significant deviation from the specifications of a procedure such as Standard 93-77 to adequately test concentrators is required in the conduct of the incident angle modifier test. With flat-plate collectors, the incident angle modifier changes in a smooth manner with incident angle and, hence, the test can be conducted for specific incident angles of 30°, 45°, and 60°.* For concentrators, the incident angle modifier test should be conducted based on the collector design and intended use.

Wood et al.[77] analyzed the characteristics of the single-axis tracking linear concentrator they tested and concluded that the incident angle modifier could be described by

$$K_{\alpha\tau} = K_o K_E K_{AP} \tag{7.26}$$

where K_o is optical loss coefficient to account for losses caused by off-normal reflection or refraction from the concentrator and incident angle effects for the receiver; K_E is end loss coefficient to account for the fact that part of the receiver is not illuminated during off-normal operation; and K_{AP} is cos *theta* to account for the change in projected aperture area normal to the direct solar beam.

Since the collector had an east-west axis and tracked north-south, tests were conducted in accordance with Standard 93-77 for various incident angles in the east-west direction measured relative to the outward drawn normal. The results are shown in Fig. 7.19. Since the last two terms in Eq. (7.26) could be calculated knowing the incident angle, θ, and the geometry of the collector, the optical loss coefficient was calculated from the test results and is shown in Fig. 7.20 along with the other two coefficients. Note that the end losses and projected aperture losses were greater than the optical losses for incident angles up to 45°. Beyond 45°, the optical losses were dominant.

If the concentrator is a two-axis tracking collector, the incident angle modifier should be an identical one except for tracking error. At present, there is no procedure incorporated in any of the specific standard testing procedures for determining tracking error.

7.5 TESTING SOLAR COLLECTORS UNDER ZERO-IRRADIANCE CONDITIONS

As mentioned above, a German Bundesverband Solarenergie (BSE) working group recommended in May 1978 a standard test procedure for solar collectors that is different from the other standard procedures proposed and adopted. The other procedures generally require complete testing under full solar irradiance; the BSE

* As established by Thomas et al., orientation of the collector should also be accounted for because of shading of the absorber.[78]

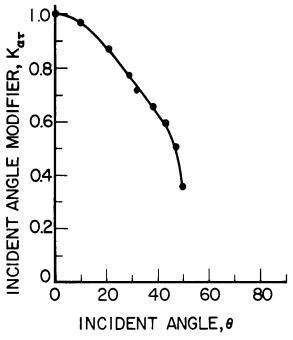


Fig. 7.19: ASHRAE 93-77 incident angle modifier test results for a concentrator.[77]

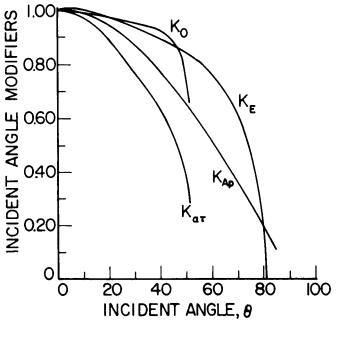


Fig. 7.20: Loss coefficients which make up the incident angle modifier for a concentrator.[77]

procedure requires the collector optical efficiency and thermal loss characteristics to be determined independently through outdoor and indoor tests, respectively. The optical efficiency is determined during outdoor testing in which the operating conditions are regulated so that the collector experiences negligible heat loss, whereas the thermal loss characteristics are determined under indoor laboratory conditions with zero solar irradiance and by circulating the working fluid through the collector over a range of operating temperatures above ambient air temperature. The two separately determined properties are then used to generate the collector near-normal-incidence instantaneous efficiency curve as a function of various operating conditions.

Prior to the publication by BSE, research on the technique had been completed by Symons,[78] Christie,[79] Smith and Weiss,[80] Whillier,[81] and Reed.[17] In addition, staff of the National Bureau of Standards conducted an experimental evaluation of the procedure after its publication by testing five single- and double-glazed, flat-plate, water-heating collectors[82,83] and two unglazed, flat-plate, water-heating collectors,[84,85] using both the BSE and ASHRAE procedures, and compared the results. The studies showed that (1) the difference in efficiency between the two procedures was less than the experimental uncertainty of the efficiency determined

in accordance with the ASHRAE procedures, (2) the repeatability of test results was improved by using the BSE procedure due to keeping the amount of outdoor testing to a minimum and using indoor testing under controlled laboratory conditions, and (3) the total test time associated with the use of the BSE procedure was reduced by minimizing the dependence of testing upon favorable weather conditions. The procedure was found particularly advantageous for the unglazed collectors since their efficiency is very sensitive to environmental conditions and it was possible to determine how it varied over a whole range of wind speeds and directions.

Despite the experiences and recent experimental evaluations, there has been a reluctance to adopt the concept of an indoor heat loss test due to differences that exist in collector thermal losses determined indoors relative to those experienced outdoors. Losses determined indoors will normally be lower than those experienced outdoors because (1) the environmental and operating conditions are generally more favorable and (2) for the same environmental and operating conditions, the fundamental heat loss mechanisms are different; consequently, the collector efficiencies will be higher when calculated based on indoor heat-loss tests. However, the NBS experimental studies have shown how the environmental and operating conditions can be controlled during the indoor laboratory tests to closely match those occurring outdoors.[82-85] In addition, recent analytical studies have shown that collector differences can be predicted and then used to correct the results after the mixed indoor/outdoor tests are completed. As a result, there may be increased use of the indoor heat-loss test in the future since it offers a method of determining the fundamental dependence of U_L on operating conditions.

Analytical studies completed recently by Svendson,[86] Gillett,[87] and Jenkins and Bushby[88] have all addressed the differences in heat loss likely to occur between indoor and outdoor tests. Svendson modeled several flat-plate collectors and predicted differences in efficiencies as a function of operating conditions, while Gillett analyzed the various mechanisms causing the efficiency differences for flat-plate collectors. Both investigators' results were in agreement with the NBS experimental studies.[82-85] Jenkins and Bushby developed an equation for calculating the expected difference in collector efficiency between outdoor and indoor testing, assuming the environmental and operating conditions are identical and a procedure for correcting the results once the mixed indoor/outdoor tests are completed (summarized below).[88]

Figure 7.21 shows. conceptually, the difference in efficiency, $\Delta\eta$, that occurs assuming environmental and operating conditions are identical. Figure 7.21(a) is a typical result from an outdoor test under full-irradiance conditions.* An identical outdoor test is conducted to determine the efficiency curve y-intercept, η_o, when the mixed indoor/outdoor test procedure is used as shown in Fig. 7.21(b). The collector thermal losses are determined in the indoor heat loss test as shown in Fig. 7.21(c) and then combined with η_o to generate an efficiency curve for a selected irradiance level as shown in Fig. 7.21(b). If the results of the mixed indoor/outdoor tests are plotted in a manner consistent with the results of the outdoor tests under full-irradiance

* Average fluid temperature, $\bar{t}_f$, was used in the x-axis instead of inlet fluid temperature, $t_{f,i}$, for convenience of analysis by Jenkins and Bushby.[88]

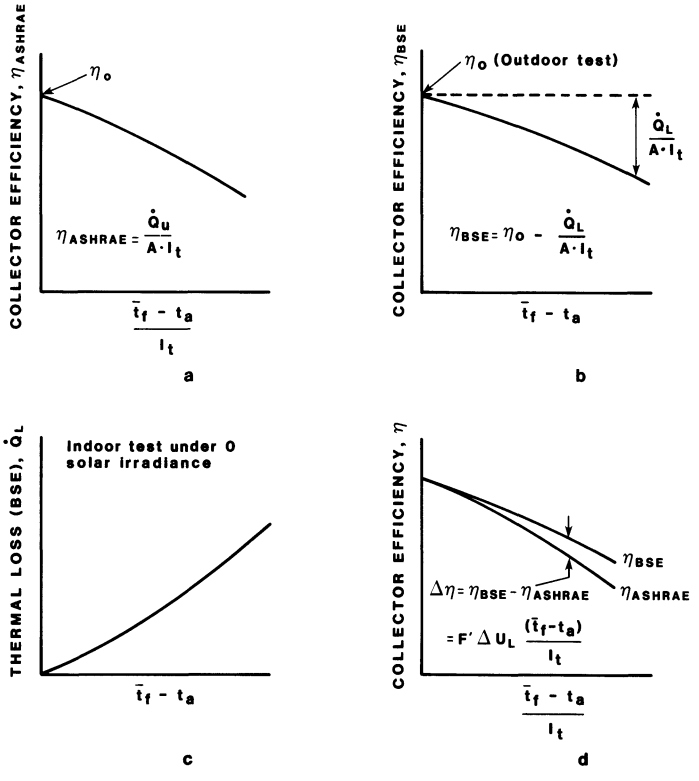


Fig. 7.21: a) ASHRAE collector efficiency curve. b) BSE collector efficiency curve. c) BSE collector thermal losses. d) Comparison of BSE and ASHRAE efficiency curves.

conditions (Fig. 7.21(a)), the curves in Fig. 7.21(d) are obtained. Note that the curve intercepts are equal due to the identical procedure for determining η_o using either procedure. The magnitude of $\Delta\eta$ is directly related to differences between characteristic slopes of the two efficiency curves $\Delta(F'U_L)$. However, because the collector efficiency factor, F', is essentially constant for a specific collector and set of operating conditions, $\Delta\eta$ is fundamentally dependent upon the difference between the collector loss coefficients ΔU_L for the same operating conditions.

As shown in Fig. 7.22, the collector heat-transfer loss coefficient, U_L, can be represented as a linear function of the absorber plate temperature, t_p. The slope and magnitude of the U_L versus t_p curve is unique for a specific collector and set of environmental conditions. Looking at Fig. 7.22, ΔU_L, and consequently, $\Delta\eta$, are dependent upon factors which contribute to differences between the mean absorber plate temperature, $\Delta\bar{t}_p$, during indoor testing, $\bar{t}_{p,i}$, and outdoor testing, $\bar{t}_{p,o}$, for the same fluid temperature. The fluid temperature is the temperature of reference in characterizing the collector efficiency. The primary factor which contributes to $\Delta\bar{t}_p$ is inverse fluid-absorber temperature profiles between indoor and outdoor testing.

Reverse modes of heat transport between indoor and outdoor testing are

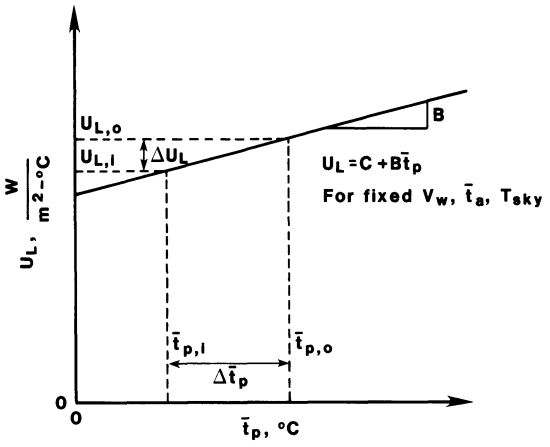


Fig. 7.22: Collector loss coefficient, U_L, as a function of the mean absorber plate temperature, $\bar{t}_p$.

responsible for the inverse fluid-absorber temperature profiles. During outdoor testing, the useful thermal energy is transported from the absorber to the working field. As a result, the mean absorber plate temperature, $\bar{t}_{p,o}$, is higher than the mean working fluid temperature, $\bar{t}_f$. Conversely, during indoor testing under zero irradiance conditions, the thermal losses from the working fluid are transported to the absorber and then from the collector into the environment. Consequently, the mean absorber plate temperataure, $\bar{t}_{p,i}$, is less than $\bar{t}_f$. Because of these reverse modes of heat transport, the absorber plate temperatures are reversed such that $\bar{t}_{p,o} \geq \bar{t}_f \geq \bar{t}_{p,i}$ for a given value of t_f. Therefore, the magnitude of Δt_p depends upon the absorber-fluid thermal resistance and the magnitude of heat transfer occurring in the collector.

Jenkins and Bushby developed an equation for $\Delta\eta$ containing four terms and verified its accuracy by doing detailed thermal modeling of the performance of selected single- and double-glazed, flat-plate collectors operating both outdoors and indoors.[88] They found that the equation could accurately predict $\Delta\eta$ to within ± 2.5 efficiency points and was valid for any type of collector whose performance could be expressed in terms of the Hottel-Whillier equation. More importantly, they developed a short procedure whereby all the terms in their equation for $\Delta\eta$ can be calculated from the results of the mixed indoor/outdoor tests.

7.6 CONSIDERATIONS IN TESTING AIR COLLECTORS

7.6.1 Air Leakage

Air leakage in air heaters is a troublesome problem that can affect test results, as well as the actual performance of an installed system. During testing, the test loop can be sealed sufficiently well with duct tape, caulking, and so on. Leakage in and around the collector may occur as a result of the collector design and recommended installation practice. This type of leakage of course should not be eliminated.

However, measurements should be made in such a way as to determine the true output of the collector array. In addition, the array should be tested such that the air leakage occurring during the tests will be indicative of the leakage that will occur in an actual installation.

Close and Yusoff have recently published the results of an analysis of air leakage in air-heating collectors.[89] They assumed a constant leakage rate along the length of the collector and determined its effect on the efficiency measurements for all combinations of:

(a) operating the collector under negative pressure (air leaking in) and positive pressure (air leaking out), and

(b) measuring the air flow rate before (upstream) the collector or after (downstream) the collector.

The results are shown in Fig. 7.23. The abscissa is the leakage rate divided by the measured flow rate and the ordinate is the ratio of the actual collector efficiency to the measured efficiency. As can be seen, for three of the testing configurations, the discrepancy between measured and actual efficiency is a direct function of the leakage rate. In addition, for the case where the collector is operating under negative pressure and air is leaking in, the difference in efficiency depends on the difference in temperature between the ambient air and the entering air stream to the collector. Therefore, the error in efficiency for this case is larger for the data points at the higher inlet fluid temperatures relative to ambient. Also note that when the air leaks in, the actual efficiency is larger than indicated by the measurements, whereas just the opposite is true when air leaks out. This is consistent with what one would intuitively expect. When air leaks into the collector at a cooler temperature than is measured at the collector inlet, the collector is heating up the air over a larger temperature difference than is indicated by the measurements. In addition, if the air flow measurement is upstream of the collector, the collector is heating a larger quantity of air than is indicated by the measurements. When air leaks out of the collector, and the flow rate is measured upstream, the quantity of useful heated air is less than indicated by the measurements. However, if air leaks out and the air flow measurement is made downstream of the collector, the quantity of useful heated air is precisely what is measured and, as a result, there is no difference between actual and measured collector efficiency. Data for this case are not shown in Fig. 7.23 since they would preduce a horizontal straight line with an ordinate value of 1.

The results of the above analysis have a direct implication for the testing of air-heating collectors:

(1) If the collector is normally operated under positive pressure, it should be tested while operating under positive pressure and the air flow rate measured downstream of the collector.

(2) If the collector is normally operated under negative pressure, it should be tested while operating under negative pressure and the air flow rate measured both upstream and downstream of the collector in order to quantify the leakage rate. In this manner, an estimate of the actual collector efficiency could be made by the user of the collector with the data of Fig. 7.23, even

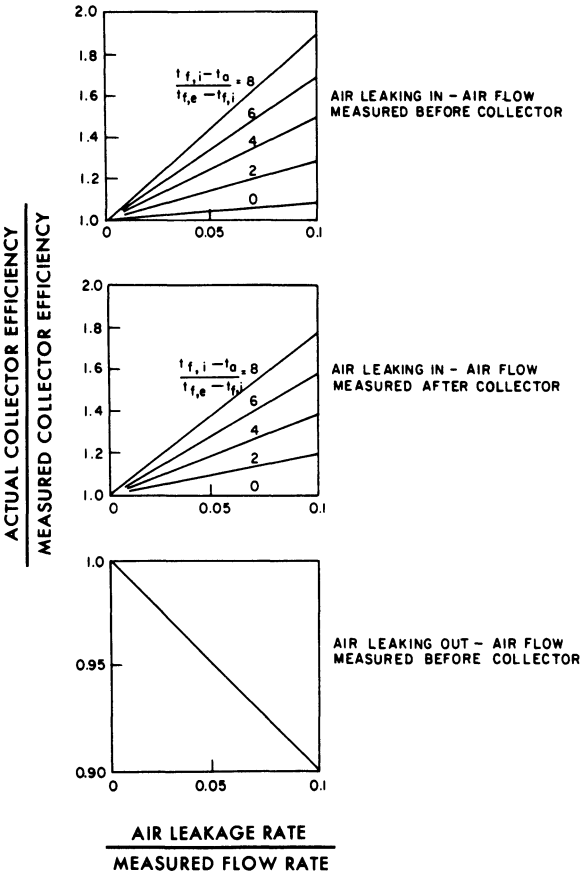


Fig. 7.23: Relationship between actual efficiency and measured efficiency as a function of air leakage rate for air-heating solar collectors.[89]

though the correction might not be made and published as part of the test results.

7.6.2 Predicting Collector Array Performance from Tests on Modules

As indicated by Eq. (7.1), the useful heat output of the collector is directly proportional to the collector heat removal factor, F_R. If the collector test results are to be used without correction to predict the performance of the collector in an actual field installation, then the value of F_R which occurs during the test should be approximately the same as that which will occur in the field. The expression for F_R is [31]

$$F_R = \frac{\dot{m}c_p}{A_a U_L}\left(1 - \exp - \frac{A_a U_L F'}{\dot{m}c_p}\right) \tag{7.27}$$

Therefore, the conditions during testing and in actual field operation should be such that $\dot{m}/A_a$, c_p, U_L, andF' are approximately the same in both cases. For liquid-heating collectors, this is easy to accomplish.* The flow rate per unit collector area, $\dot{m}/A_a$, the type of fluid and, hence, specific heat, c_p, and the collector tilt angle, temperature, incident solar radiation, and ambient conditions are such that the heat loss coefficient, U_L, is approximately the same. In addition, the collector efficiency factor, F', is primarily a function of the geometry of the absorber (material, thickness, distance between tubes, and so on) and, of course, is the same for the tested and installed collectors at any flow rate.

The situation is slightly different for air heaters. Although $\dot{m}/A_a$, c_p, and U_L can be made the same in a similar fashion as with the liquid-heating collectors, assuring the same value of F' is more difficult. For air heaters, F' is primarily a function of the convection heat transfer coefficient between the absorber and the air stream. Since the air flow is nearly always in the turbulent flow range (to maximize the heat transfer), the heat transfer coefficient is determined primarily by the value of the Reynolds number in the collector.

Quite often, air collectors are designed to be installed with at least two modules in series. A procedure has been developed so that the test results on a single module can be corrected to predict the performance of the collectors after installation in the array.[19,92] The single module must be tested at a flow rate that will result in the same flow velocity (and hence Reynolds number) that will occur in the array of collector modules in series. The collector output for the single module is then multiplied by a correction factor to obtain the output for the array. The correction procedure was experimentally verified by tests on one and then two modules of a flat-plate air heater connected in series.[19]

7.7 CALCULATING ALL-DAY COLLECTOR PERFORMANCE

7.7.1 Calculation Including Diffuse Solar Irradiance

The standard test methods described previously were developed to provide sufficient data for calculating the useful daily energy collected by the solar collector. This is essential for design purposes and, since the performance of collectors can vary so much, it becomes necessary to consider their daily output when making a comparison between any two of them.

The ASHRAE Standard 93-77 and 96-1980 data can be used to calculate the collector all-day performance, as illustrated in Fig. 7.24. Zerlaut et al.,[93] Hill et al.,[19] Wood et al.,[77] and Beach[94] have all obtained excellent agreement between calculated and measured all-day performance for clear sky conditions. The agreement in thermal energy collected was typically ± 5% for a wide range of collectors. Farber et al.[95] compared the calculated and measured all-day performance of a flat-plate

* However, if a different fluid is used in the test than in the actual field installation, substantial differences in collector efficiency can result. Youngblood et al.[90] and Thomas[91] give test results for typical flat-plate collectors using water, mixtures of water and a commercial ethylene glycol-base antifreeze, and a mineral-base heat transfer oil.

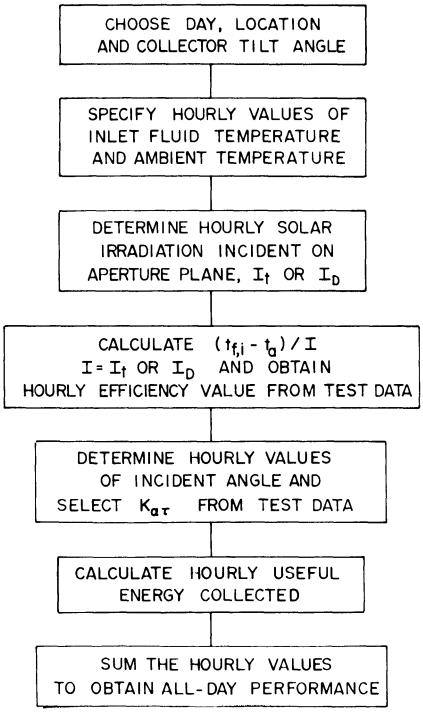


Fig. 7.24: Flow chart for calculating all-day energy collection using ASHRAE 93-77 test data.

collector under various sky conditions. Even for cloudy days, the agreement was ± 6%.

The all-day performance of a collector is, a priori, dependent on the intensity-weighted directionality of the incoming solar irradiance. Consequently, on a clear day the majority of solar irradiance strikes the aperture plane of the collector at incident angles defined by the sun's path through the sky (beam component). On partially diffuse or diffuse days, a large portion of the incident solar irradiance may come at incident angles significantly different than that experienced on a clear day.

The ASHRAE Standard test methods are prescribed for clear skies with low levels of diffuse solar irradiance. Consequently, the equations above do not explicitly consider the response of the collector to diffuse solar irradiance. However, the required test data does include both the hemispherical and direct normal irradiance from which the global diffuse irradiance can be determined. Eq. (7.11) can be modified to explicitly consider both the beam and diffuse components of the solar irradiance*

* Eq. (7.11) has also been modified by the inclusion of A_a/A_g since the efficiency as determined in the ASHRAE Standards is based on gross collector area.

$$\eta_g = \frac{A_a}{A_g} F_R(\tau\alpha)_{e,n} \frac{K_{\alpha\tau} I_{DN} \cos\theta + K_d I_d}{I_t} - \frac{A_a}{A_g} F_R U_L \frac{t_{f,i} - t_a}{I_t} \tag{7.28}$$

The term K_d is defined as the diffuse incident angle modifier[35] such that

$$K_d = \frac{\int_o^{2\pi} I_d(\omega) \cos\theta d\omega}{\int_o^{2\pi} I_d(\omega) \cos\theta d\omega} \tag{7.29}$$

For the case of isotropically distributed diffuse irradiance and for collectors whose $K_{\alpha\tau}$ can be described by Eq. (7.12), K_d becomes

$$K_d = (1 - b_o) \tag{7.30}$$

Note that $(1 - b_o)$ is the value of $K_{\alpha\tau}$ at an incident angle of 60°. This means that beam irradiance at an incident angle of 60° has the same transmittance-absorptance as all of the isotropic diffuse irradiance, that is, the effective incident angle for isotropic diffuse irradiance is 60°. Brandemuehl and Beckman have calculated the effective beam angle for solar collectors which "see" both the sky and the ground.[96]

7.7.2 SRCC Rating Calculation Methods

The Solar Rating and Certification Corporation (SRCC) has developed a consensus standard for calculating the all-day thermal performance of solar collectors using ASHRAE Standard test data.[37] The rating itself is a calculated set of numbers representing the all-day energy output of the solar collector under prescribed rating conditions. The SRCC method includes a diffuse solar irradiance term very similar to that in Eq. (7.28) and it assumes isotropic diffuse irradiance.

Two kinds of incident angle modifiers are used by SRCC. First, for most flat-plate collectors, $K_{\alpha\tau}$ is assumed to be symmetrical with respect to the normal direction from the aperture plane, that is, $K_{\alpha\tau}$ is a function of the magnitude of θ only, and not the direction of the incident rays. The diffuse incident angle modifier is given by Eq. (7.29), which, for isotropic diffuse irradiation, becomes

$$K_d = \int_o^{\pi/2} K_{\alpha\tau}(\theta) \sin 2\theta d\theta \tag{7.31}$$

This integral can be evaluated either numerically or analytically, if an expression for $K_{\alpha\tau}(\theta)$ is available.

Second, the SRCC method considers the response of the collectors when $K_{\alpha\tau}$ is a function of both the angle of incidence and the direction of the solar irradiance. SRCC uses the biaxial incident angle modifier described by Eq. (7.13).[36] Two separate incident angle modifiers are experimentally determined as a function of the off-altitude angle, and the off-azimuth angle (see Fig. 7.25). Mather suggested that the angle ψ in Fig. 7.25, rather than the angle Γ, was better to use in the

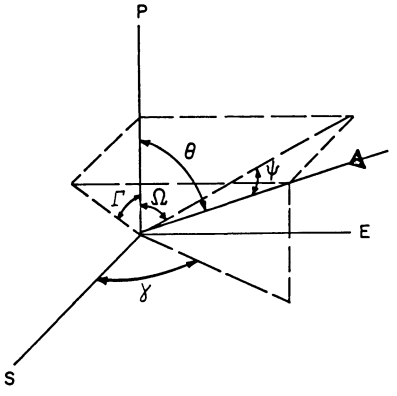


Fig. 7.25: Diagram of off-altitude and off-azimuth angles used in describing the incident angle modifier for evacuated tube collectors.

correlation for evacuated tubes.[97] Hence,

$$K_{\alpha\tau} = K_1(\Omega)K_2(\psi) \tag{7.32}$$

He argued that $K_2(\psi)$ was a better measure of the Fresnel reflection effects that occur when the solar rays are not perpendicular to the tube axis. Using ψ allows Eq. (7.29) to be simplified to

$$K_d = \frac{4}{\pi}\int_o^{\pi/2} K_1(\Omega)\cos\Omega d\Omega \int_o^{\pi/2} K_2(\psi)\cos^2(\psi)d\Gamma \tag{7.33}$$

7.7.3 Effects of Diffuse Irradiance on Calculations

To validate the SRCC method for calculating the all-day energy collection using the diffuse incident angle modifier, the Solar Energy Research Institute funded DSET Laboratories, Inc. to measure the performance of flat-plate and stationary, concentrating collectors and to compare the calculations using ASHRAE 93-77 test data and several models with all-day measurements. This subsection is a summary of the work reported by Putman et al.[98]

An extensive literature survey showed that the effect of diffuse irradiance on the thermal performance of solar collectors could be taken into account for the majority of commercially available solar collectors as long as their directional dependence could be established. The major obstacle was the inherent inability to characterize the solar irradiance in a meaningful way.

The thermal performance of three generic collector types (flat-plate, single-glazed, nonselective absorber; flat-plate, single-glazed, selective absorber; and evacuated tube, stationary, concentrating reflector) were determined in accordance with ASHRAE Standard 93-77, including extensive tests to determine the incident angle

modifiers. Subsequently, the all-day, optical performance of each collector was determined (that is, the collector inlet fluid temperature was maintained at ambient temperature during each test day) under three distinctly different sky conditions: clear, partly cloudy, and overcast. The collectors were oriented due south at a 34° tilt angle. The directional intensity of the incoming diffuse radiation during the Standard 93-77 and all-day tests was measured with a Directional Diffuse Radiation Flux Mapper, an instrument designed specifically for this project. These data were then used to compare the predicted performance using the various techniques to the measured all-day performance. The uniqueness of this research was the ability to measure the actual direction and intensity of the radiation during testing and, thereby, evaluating unproven sky and collector models.

The measured sky radiation data using the Diffuse Radiation Flux Mapper showed that the diffuse radiation for the overcast sky was isotropic, that is, uniformly distributed over the sky. Similar data for the partly cloudy day and the clear day data showed that the diffuse radiation was anisotropic. It was concluded that the clear sky diffuse irradiance could be modeled as a "pseudo-solar beam" and a "constant or isotropic background." Due to the random distribution of discrete clouds, there is no generalized model for diffuse irradiance for a partly cloudy day.

As was expected, the optical characteristics of the evacuated tubular collector were not symmetrical when comparing off-altitude and off-azimuth incident angle modifier data. However, the optical characteristics of the other two collectors were found to be symmetrical.

Measurements of the time-dependent variables were available as averages for each 15-min interval over each of the three different test days. Using these measurements, the all-day performances were calculated using five different calculation techniques. Each of the calculation techniques treated the diffuse solar irradiance differently. The five methods were as follows:

BEST method used the beam irradiance as measured with the Directional Diffuse Radiation Flux Mapper and the measured incident angle modifier for each view angle defined by the flux mapper. The *BEST* calculations were the only ones which included the actual measured distribution of the sky diffuse radiation. Four different ways were used to calculate the incident angle modifier. In *BEST 1* and *BEST 2*, a biaxial incident angle modifier was defined as the product of the incident angle modifiers from two mutually perpendicular directions. A full two-dimensional mapping of the incident angle modifier was completed for the evacuated tube collector and the all-day calculations were designated the *BEST 4* method. The *BEST 3* method applied to collectors with axisymmetric properties, such as the flat-plate collectors. In *BEST 3*, the incident angle modifier was dependent only on the incident angle and not on the off-altitude and off-azimuth angles.

SRCC-ASYM method used the beam irradiance and assumed the measured diffuse irradiance to be isotropic. The incident angle modifiers were biaxial and defined as the product of the off-altitude and off-azimuth incident angle modifiers as in *BEST 1* and *BEST 2*. The all-day calculations were designated as *SRCC-ASYM 1* and *SRCC-ASYM 2*. *SRCC-ASYM 1* method followed the

Table 7.3

Percent Difference Between Calculated and Measured All-Day Performance for Three Collectors Using Various Calculation Methods[98]

	BEST				SRCC					Measured All-day
	1	2	3	4	ASYMI	ASYM2	SYM	SYMFIT	TOT	Efficiency
COLLECTOR 1										
Clear day	4.8	4.8	4.5		3.5	3.0	2.6	1.3	3.9	53.8
Partly cloudy day	5.1	5.1	4.5		4.0	2.5	2.1	0.2	5.3	53.0
Diffuse day	0.9	0.9	1.1		0	−2.6	−2.4	−4.4	3.9	54.0
COLLECTOR 2										
Clear day	−1.0	−1.0	−0.7		−2.1	−2.6	−2.5	−2.9	−1.0	68.0
Partly cloudy day	−1.0	−1.0	−1.5		−2.2	−3.9	−4.0	−4.9	−0.3	67.3
Diffuse day	−3.4	−3.4	−2.8		−5.2	−7.9	−7.0	−8.2	0.7	67.2
COLLECTOR 3										
Clear day	3.4	3.4	4.6	2.7	1.6	1.4	3.0	6.2	7.0	43.9
Partly cloudy day	3.5	3.5	3.5	0.3	2.8	0.9	1.9	5.6	5.3	43.1
Diffuse day	0.2	0.2	1.2	−0.2	0	−3.2	−0.9	2.6	0.7	43.1

Solar Rating and Certification Corporation (SRCC) method for anisotropic collectors.

SRCC-SYM method used the beam irradiance and assumed the measured diffuse irradiance to be isotropic. The incident angle modifiers were symmetrical with respect to the normal direction from the collector aperture plane. This method was the same as the SRCC isotropic collector method.

SYMFIT method used the beam irradiance and an assumed isotropic diffuse irradiance. The incident angle modifier was determined from a single correlation for all the measured incident angle data, that is, Eq. (7.12). The effective incident angle for the diffuse irradiance was taken as 60°.

TOT method treated the global irradiance as beam irradiance with all incident angle modifier data "fit" to get b_o in Eq. (7.12).

All of these methods gave surprisingly similar results for all-day performance. In fact, treating the global irradiance as if it were all beam irradiance, an assumption which greatly simplified the calculations, gave very good results for all collectors on clear, partly cloudy, and even totally cloudy days. The agreement between the measured all-day performance and the *SRCC-SYM* or *SRCC-ASYM* methods was improved by substituting an anisotropic diffuse sky model for the isotropic diffuse model. Using a biaxial incident angle modifier as done in the SRCC method appears to be reasonable. The results of Putman et al. are summarized in Table 7.3.[98]

It is interesting to note that the all-day collector optical efficiencies for each collector were essentially the same for each of the three solar days. This further suggests that the different diffuse irradiance distributions had little effect on the all-day collector performance.

7.8 NOMENCLATURE

A_a	aperture area of the collector, m^2
A_g	gross area of the collector, m^2
b_o	defined by Eq. (7.11)
$C_1 \ldots C_{22}$	constants in Eqs. (7.16) through (7.25)
C_A	effective heat capacity of the collector, its components, and the transfer fluid in the collector, J/°C
c_p	specific heat of the transfer fluid, J/kg s
F'	collector efficiency factor
F_R	collector heat-removal factor
I_t	total solar energy incident upon the plane of the collector per unit time per unit area, W/m^2
I_D	direct component of solar energy incident upon the plane of the collector per unit time per unit area, W/m^2
I_h	total solar energy incident upon a horizontal surface per unit time per unit area, W/m^2
K	defined by Eq. (7.7)
$K_{\alpha\tau}$	incident angle modifier
K_d	diffuse incident angle modifier
K_o	optical loss coefficient
K_E	end loss coefficient
K_{A_p}	$\cos\theta$
K_1	off-azimuth incident angle modifier
K_2	off-altitude incident angle modifier
$\dot{m}$	mass flow rate of the heat transfer fluid through the collector, kg/s m^2
$\dot{Q}_L$	rate of heat loss from the collector, W
$\dot{Q}_u$	rate of useful energy extracted from the collector, W
t_a	ambient air temperature, °C
T_e	effective environmental temperature, °C
$\overline{T}_f$	average temperature of the heat-transfer fluid in the collector, K
$\overline{t}_f$	average temperature of the heat-transfer fluid in the collector, °C
$t_{f,e}$	temperature of the heat-transfer fluid leaving the collector, °C
$t_{f,i}$	temperature of the heat-transfer fluid entering the collector, °C

$\bar{t}_p$	average temperature of collector absorber, °C
$\bar{t}_{p,i}$	average temperature of collector absorber during indoor testing, °C
$\bar{t}_{p,o}$	average temperature of collector absorber during outdoor testing, °C
$\overline{T}_s$	average sink temperature for radiation loss, K
U_L	heat-transfer loss coefficient for the collector, W/m^2 °C
$U_{L,i}$	heat-transfer loss coefficient for the collector during indoor testing, W/m^2 °C
$U_{L,o}$	heat-transfer loss coefficient for the collector during outdoor testing, W/m^2 °C
v	wind speed, m/s
η_a	collector efficiency based on collector aperture area
η_g	collector efficiency based on collector gross area
η_o	collector optical efficiency
η_{ASHRAE}	collector efficiency determined under full-irradiance conditions in accordance with ASHRAE procedure
η_{BSE}	collector efficiency determined by a combined indoor/outdoor test in accordance with the BSE procedure
$\Delta\eta$	$\eta_{\text{BSE}} - \eta_{\text{ASHRAE}}$
θ	time, incident angle between the direct solar beam and the outward-drawn normal to the plane of the collector
$(\tau\alpha)_e$	effective transmittance-absorptance product for the collector
$(\tau\alpha)_{e,n}$	effective transmittance-absorptance product for the collector at normal incidence

REFERENCES

1. D. Waksman et al., "Provisional flat-plate solar collector testing procedures, first revision," NBS Report NBSIR 77-1305A, June 1978.
2. F. A. Brooks, "Solar energy and its use for heating water in California," University of California Agricultural Experiment Station Bulletin No. 602, November 1936.
3. H. C. Hottel and B. B. Woertz, "The performance of flat-plate solar heat collectors," *ASME Transactions* **64**, 91 (1942).
4. A. Whillier and S. J. Richards, "A standard test for solar water heaters," *Proceedings of the Conference on New Sources of Energy*, Paper S/97, Rome, August 21-31, 1961.
5. "Solar water heaters: Test methods," The Standards Institution of Israel, Tel Aviv, Israel Standard, U. D. C. 602.93, S.I 609, May 1966.

6. H. Tabor, "Testing of solar collectors," *Solar Energy* **20**, 293-303 (1978).

7. P. I. Cooper, "A method of testing flat-plate solar water heaters to determine performance characteristics," *Proceedings of the 1975 Australian and New Zealand Section Meeting of the International Solar Energy Society* (Melbourne, July 2, 1975).

8. P. Pott and P. I. Cooper, "An experimental facility to test flat-plate solar collectors outdoors," CSIRO Division of Mechanical Engineering Report No. TR 9, 1976.

9. R. V. Dunkle and P. I. Cooper, "A proposed method for the evaluation of performance parameters of flat-plate solar collectors," *Proceedings of the 1975 Solar Energy Congress, Solar Use Now—A Resource for People* (Los Angeles, July 28-August 1, 1975).

10. J. G. Symons and P. I. Cooper, "Thermal performance testing of flat-plate solar collectors," *Proceedings of the 1978 International Solar Energy Society Meeting, Sun: Mankind's Future Source of Energy*, (New Dehli, January 1978), Vol. 2, pp. 1004-1008.

11. D. Proctor, "A generalized method for testing all classes of solar collectors, part I—Attainable accuracy," *Solar Energy*, submitted for publication, 1982.

12. S. W. Moore, J. D. Balcomb, and J. C. Hedstrom, "Design and testing of a structurally-integrated steel solar collector unit based on expanded flat-metal plates," *Proceedings of the 1974 United States Section Meeting of the International Solar Energy Society Meeting*, (Fort Collins, Colo., August 19–23, 1974).

13. R. W. Vernon and F. F. Simon, "Flat-plate collector performance determined experimentally with a solar simulator," NASA TM X-71602, *Proceedings of the 1974 United States Section Meeting of the International Solar Energy Society Meeting* (Fort Collins, Colo., August 19–23, 1974).

14. F. F. Simon, "Flat-plate solar-collector performance evaluation with a solar simulator as a basis for collector selection and performance prediction," NASA TM X-71793, *Solar Energy* **18**, 451–466 (1976).

15. S. M. Johnson and F. F. Simon, "Evaluation of flat-plate collector efficiency under controlled conditions in a solar simulator," NASA TM X-73520, *Proceedings of the 1976 International Solar Energy Society Conference* (Winnipeg, Canada, August 15–20, 1976).

16. J. W. Ramsey, J. T. Borzon, and T. H. Holland, "Development of flat-plate collectors for heating and cooling of buildings," NASA CR-13804, June 1975.

17. K. A. Reed, "Interim solar collector test plan," *Proceedings of the ERDA Concentrating Solar Collector Conference* (Atlanta, Sept. 26–28, 1977).

18. V. E. Dudley and R. M. Workhoven, "Summary report: Concentrating solar collector test results, collector module test facility," Sandia Report SAND 780815, Sandia Laboratories, May 1978.

19. J. E. Hill, J. P. Jenkins, and D. E. Jones, "Experimental verification of a standard test procedure for solar collectors," NBS Building Science Series 117, January 1979.

20. J. E. Hill and T. Kusuda, "Methods of testing for rating solar collectors based on thermal performance," NBSIR 74-635, Dec. 1974.

21. J. E. Hill et al., "Development of proposed standards for testing solar collectors and thermal storage devices," NBS Technical Note 899, Feb. 1976.

22. J. E. Hill and E. R. Streed, "A method of testing for rating solar collectors based on thermal performance," *Solar Energy* **18**, 421–429 (1976).

23. "Methods of testing to determine the thermal performance of solar collectors," ASHRAE Standard 93-77 (ASHRAE, Atlanta, 1977).

24. "Methods of testing to determine the thermal performance of unglazed flat-plate liquid-type solar collectors," ASHRAE Standard 96-1980 (ASHRAE, Atlanta, 1980).

25. "Solar collectors, measurement of thermal performance," AFNOR Standard 50-501 (AFNOR, France, Dec. 1977).

26. "Guidelines and directions for determining the usability of solar collectors, a solar collector efficiency test," (Bundesverband Solarenergie, Kruppstrasse 5, 4300 Essen 1, Federal Republic of Germany, May 1978).

27. "Recommendation for European solar collector test methods (liquid heating collectors)," (Commission of the European Communities (CEC), Jan. 1980).

28. "Draft for development, methods of test for thermal performance of solar collectors," Standard DD 77:1982 (British Standards Institution (BSI), 1982).

29. "Solar collectors," Canadian Standards Association (CSA) Standard F378-M1982 (Canadian Standards Association, 178 Rexdale Boulevard, Rexdale, Ontario M9W 1R3, Aug. 1982).

30. "Glazed flat-plate solar collectors with water as the heat-transfer fluid—Method for testing thermal performance," Australian Standard 2535-1982 (Standards Association of Australia, Standards House, 80 Arthur St., North Sydney, NSW, 1982).

31. J. A. Duffie and W. A. Beckman, *Solar Engineering Thermal Process* (Wiley, New York, 1980).

32. S. A. Klein, J. A. Duffie, and W. A. Beckman, "Transient considerations of flat-plate solar collectors," *ASME Journal of Engineering Power* **96A**, 109 (1974).

33. N. W. Wijeysundera, "Response time of solar collectors," *Solar Energy* **18**, 65–68 (1976).

34. A. F. Souka and H. H. Safwat, "Determination of optimum orientations for the double-exposure, flat-plate collector and its reflectors," *Solar energy* **10**(4), 170–174 (1966).

35. F. F. Simon and E. H. Buyco, "Outdoor flat-plate collector performance prediction from solar simulator test data," NASA TM X-7107, *Proceedings of the 10th AIAA Thermal Physics Conference* (Denver, Colo., May 27–29, 1976).

36. G. R. Mather and D. C. Beekley, "Performance of an evacuated tubular collector using non-imaging reflectors," *Proceedings of the 1976 International Solar Energy Society Conference*, Vol. 2, p. 74, (Winnipeg, Canada, Aug. 15–20, 1976).

37. "Methodology for determining the thermal performance rating for solar collectors," SRCC Standard RM-1 (Solar Rating and Certification Corporation, Washington, D.C., April 1981).

38. B. Doron, "A technical note on testing of collectors," *Solar Energy* **18**, 405 (1976).

39. R. S. Soin et al., "Performance of flat-plate solar collectors with fluid undergoing phase change," *Solar Energy* **23**, 69–73 (1979).

40. J. Hall, "Choosing a flow monitoring device," *Instruments and Control Systems*, 51–59 (June 1981).

41. W. O. Strohmeier, "Turbine flowmeters, past, present, and future," in *Flow: Its Measurement and Control in Science and Industry*, Vol. 1, edited by R. B. Dowdell (Instrument Society of America, Pittsburgh, 1974), pp. 687–693.

42. C. A. E. Clay, C. Griffiths, and E. A. Spencer, "Improving the confidence in hydraulic laboratory calibrations," in *Flow: Its Measurement and Control in Science and Industry*, Vol. 2, edited by W. W. Durgin (Instrument Society of America, St. Louis, 1981), pp. 789–807.

43. F. C. Kinghorn and A. McHugh, "The performance of turbine meters in two-component gas/liquid flow," in *Flow: Its Measurement and Control in Science and Industry*, Vol. 2, pp. 471–480.

44. M. P. Wilson, "A survey of mass flowmeters," in *Flow: Its Measurement and Control in Science and Industry*, Vol. 1, pp. 864–879.

45. K. O. Plache, "Coriolis/Gyroscopic flow meter," *Mechanical Engineering* 36–41 (March 1979).

46. C. C. Thomas, "The measurement of gases," *Journal of the Franklin Institute* **172**, 411–460 (1911).

47. J. P. Jenkins, "The design and evaluation of a reference heat source to be used in conjunction with solar collector testing," Letter Report, National Bureau of Standards, May 1979.

48. K. A. Reed and J. W. Allen, "Thermal performance testing of solar collectors: The calorimetric ratio technique," *Proceedings of the 1978 Annual Meeting of the American Section of the International Solar Energy Society*, (American Solar Energy Society, Boulder, Colo., 1978), pp. 345–346.

49. M. Collares-Pereia et al., "A calorimeter for solar thermal collector testing," *Solar Energy* **27**(6), 581–582 (1981).

50. H. D. Talarek, "Task II: Performance testing of solar collectors, annual progress report, IEA-program to develop and test solar heating and cooling systems," (Kernforschungsanlage, Julich, Federal Republic of Germany, January 1981).

51. "Standard measurements guide: Section of temperature measurements," *ASHRAE Standard 41.1-74* (ASHRAE, Atlanta, 1973).

52. "Guide to meteorological instrument and observing practices," WMO-no. 8, TP.3, Supplement No. 5 (World Meteorological Organization, Davos, Switzerland, Aug. 1965).

53. G. A. Zerlaut, "What standard pyranometer calibrations are inappropriate for solar collector testing." *Proceedings of the 1981 Annual Meeting of the American Section of the International Solar Energy Society* (ASES, Boulder, Colo., 1981), pp. 1536–1540.

54. "Standard method for calibration of reference pyranometers with axis tilted by the shading method," Draft ASTM Standard: Document No. 141R3 (American Society for Testing and Materials, Philadelphia, Jan. 1982).

55. M. R. Riches, T. L. Stoffel, and C. V. Wells, "International Energy Agency Conference on pyranometer measurements." Draft SERI *Technical Report TR-642-1156* (Solar Energy Research Institute, Golden, Colo., April 1981).

56. S. A. Klein, "Calculation of flat-plate loss coefficients," *Solar Energy* **17**, 79 (1975).

57. E. C. Shewan and K. G. T. Hollands, "Equations for representing the U_L dependence in collector test procedures," *Proceedings of the 1979 Meeting of the International Solar Energy Society* (ISES, Melbourne, 1979), pp. 360–364.

58. D. Proctor, "A generalized method for testing all classes of solar collectors, part II—Evaluation of collector thermal constants," *Solar Energy*, submitted for publication, 1982.

59. H. Tabor, "Letter to the Editor," *Solar Energy* **24**, 113–115 (1980).

60. J. M. Gordon, "On non-linear effects in flat-plate collector efficiency curves," *Solar Energy* **26**, 265–266 (1981).

61. "Product certification standard 1-79," (Solar Energy Industries Association, Washington, D. C., 1979).

62. R. Gani, D. Proctor, and J. G. Symons, "Linear efficiency characterization for high temperature flat-plate collectors," *Solar Energy* **26**, 271–273 (1981).

63. W. A. Beckman, S. A. Klein, and J. A. Duffie, *Solar Heating Design by the F-Chart Method* (Wiley, New York, 1977).

64. S. A. Klein and W. A. Beckman, "A general design method for closed loop solar energy sytems," *Solar Energy* **22**, 269 (1979).

65. P. I. Cooper and R. V. Dunkle, "A non-linear flat-plate collector model," *Solar Energy* **26**, 133–140 (1981).

66. D. Proctor, "A generalized method for testing all classes of solar collectors, part III—Linearized efficiency equations," *Solar Energy*, submitted for publication, 1982.

67. E. R. Streed et al., "Results and analysis of a round-robin test program for liquid-heating flat-plate solar collectors," *Solar Energy* **22**, 235–249 (1979).

68. E. R. Streed et al., "Results and analysis of a round-robin test program for liquid-heating flat-plate solar collectors," *Solar Energy* **22**, 235–249 (1979).

69. H. D. Talarek, "Task III: Performance testing of solar collectors, results and analysis of IEA round-robin testing," (Kernforschungsanlage, Julich, Federal Republic of Germany, Dec. 1979).

70. D. Waksman, E. R. Streed, and J. Seiler, "Solar collector durability/reliability test program plan," NBS Technical Note 1136, Jan. 1981.

71. E. R. Streed and D. Waksman, "Uncertainty in determining thermal performance of liquid-heating flat-plate solar collectors," NBS Technical Note 1140, April 1981.

72. E. R. Streed and D. Waksman, "Uncertainty in determining thermal performance of liquid-heating flat-plate solar collectors," *ASME Journal of Solar Energy Engineering* **103**, 126–134 (May 1981).

73. W. C. Thomas et al., "Incident angle modifiers for flat-plate solar collectors: Analysis of measurements and calculation procedures," *ASME Journal of Solar Energy Engineering* **104**, 349 (Nov. 1982).

74. S. J. Kline and F. A. McClintock, "Describing uncertainties in single-sample experiments," *Mechanical Engineering*, p. 3 (Jan. 1953).

75. D. L. Kirkpatrick, "Solar collector data manual," Final Report for P. O. No. B-O-5254-1-M1 (Solar Energy Research Institute, Golden, Colo., July 1981).

76. T. D. Harrison, W. D. Dworzok, and C. A. Folkner, Jr., "Solar collector module test facility, instrumentation fluid loop number one," Sandia Report SAND 76-0425 (Sandia Laboratories, Jan. 1977).

77. B. D. Wood, P. J. Fiore, and C. R. Christopherson, "Application of ASHRAE Standard 93-77 for testing concentrating collectors for the purpose of predicting all-day performance," *Proceedings of the 1979 Meeting of the International Solar Energy Society* (ISES, Melbourne, 1979), Vol. 1, pp. 487–491.

78. J. G. Symons, "The direct measurement of heat loss from flat-plate solar collectors on an indoor testing facility," CSIRO Division of Mechanical Engineering Report No. TR 7, 1976.

79. E. A. Christie, "A method of measuring the performance characteristics of flat-plate solar collectors," CSIRO Division of Mechanical Engineering Report No. TR 6, 1976.

80. C. C. Smith and T. A. Weiss, "Design application of the Hottel-Whillier-Bliss Equation," *Solar Energy* **19**(2), 109-114 (1977).

81. A. Whillier, "The thermal performance of solar water heaters," *Solar Energy* **9**(1) (1965).

82. J. P. Jenkins and J. E Hill, "Testing flat-plate water-heating solar collectors in accordance with the BSE and ASHRAE procedures," NBSIR 80-2087, Aug. 1980.

83. J. P. Jenkins and J. E Hill, "A comparison of test results for flat-plate water-heating solar collectors using the BSE and ASHRAE procedures," *ASME Journal of Solar Energy Engineering* **102**, 2–15 (Feb. 1980).

84. J. P. Jenkins and K. A Reed, "A comparison of unglazed flat-plate liquid solar collector thermal performance using the ASHRAE Standard 96-1980 and modified BSE test procedures," NBSIR 82-2522, May 1982.

85. J. P. Jenkins, "Testing unglazed flat-plate liquid-heating collectors according to ASHRAE Standard 96-1980 and BSE procedures," *ASME Journal of Solar Energy Engineering*, submitted for publication, 1982.

86. S. Svendson, "Theoretical investigation of the methodical errors of the BSE-procedure for testing solar collectors," University of Denmark Report No. 78-27, 1978.

87. W. B. Gillett, "The equivalence of outdoor and mixed indoor/outdoor solar collector testing," *Solar Energy* **25**, 543–548 (1981).

88. J. P. Jenkins and S. T. Bushby, "Differences in determining collector thermal performance between the BSE and ASHRAE collector test procedures," *ASME Journal of Solar Energy Engineering*, submitted for publication.

89. D. J. Close and M. B. Yusoff, "The effects of air leaks on solar air collector behavior," *Solar Energy* **20**(6), 459–463 (1978).

90. W. W. Youngblood, W. Schultz, and R. Barber, "Solar collector fluid parameter study," NBS GCR 79-184, Oct. 1979.

91. W. C. Thomas, "Effects of test fluid composition and flow rates on the thermal efficiency of solar collectors," NBS GCR 80-254, Aug. 1980.

92. R. L. Oonk, D. E. Jones, and B. E. Cole-appel, "Calculation of the performance of N collectors in series from test data on a single collector," *Solar Energy* **23**, 535–536 (1979).

93. G. A. Zerlaut, W. T. Dokos, R. F. Heiskell, "The use of ASHRAE Standard 93-77 in predicting all-day performance of flat-plate collectors," *Proceedings of the 1977 Flat-Plate Solar Collector Conference*, Florida Solar Energy Center, 1977.

94. C. D. Beach, "A solar collector testing program, part 2: All-day test results," Final Report for DOE Grant EG-77-G-05-5561, Report FSEC-RD-79-2, Florida Solar Energy Center, May 1979.

95. E. A. Farber, R. W. Dixon, and S. D. Wix, "Investigation of the accuracy of predicted day-long solar collector performance from the ASHRAE 93-77 thermal performance and angle modifier curves," Unpublished Report (University of Florida, Gainesville, Fla., Jan. 1981).

96. M. J. Brandemuehl and W. A. Beckman, "Transmission of diffuse radiation through CPC and flat-plate collector glazings," *Solar Energy* **24**, 511 (1980).

97. G. R. Mather, "ASHRAE 93-77 instantaneous and all-day tests of the Sunpak evacuated-tube collector," *ASME Journal of Solar Energy Engineering* **102**, 294 (1980).

98. W. J. Putman, D. E. Evans, and B. D. Wood, "The effect of different sky conditions on the optical performance of flat-plate and stationary concentrating collectors," Final Report for SERI Contract XX-1-1178-1, Dec. 1982.

CHAPTER 8

CONCENTRATING SOLAR COLLECTORS

A. Rabl

8.1 ABSTRACT

This chapter provides an introduction to concentrating solar collectors. The optical and thermal characteristics are described in relatively simple terms, and copious references to the more technical literature are given. A unified framework is used for analyzing the performance of all solar collector types; it involves optical efficiency, U-value, and heat transfer factor. Two measures of performance are of particular interest: the instantaneous efficiency under peak insolation (which is easy to measure), and the annual energy delivered by the collector (which is difficult to measure but is crucial for the economic evaluation). A simple but accurate graphical procedure is presented for obtaining the annual energy from the instantaneous efficiency curve. The chapter includes a discussion of practical aspects, covering choice of materials as well as problems of cleaning and tracking. Data for cost and performance of current collectors are provided.

8.2 INTRODUCTION

There are basically two motives for considering concentrating solar collectors: the achievement of high temperatures and the hope for cost reduction. High temperatures require that the collector have low heat losses. Since heat losses are more or less proportional to the absorber area, one can obviously reduce the heat loss per aperture area by concentrating the radiation incident on the aperture onto a smaller absorber. The other motive is clear in principle: if concentrators are less expensive per unit area than absorbers, then one may be able to reduce the cost of a collector by replacing expensive absorber area with less expensive concentrator area.* Photovoltaic cells, for example, are expensive at the present time, while

* As for terminology, the term collector designates the complete assembly consisting of concentrator (reflector or lens) and receiver (absorber plus associated enclosure).

reflector material is relatively cheap. In practice, however, this cost advantage of concentrators is less clear because concentrators can also become quite costly when one has to pay for support structure, tracking mechanism, and so on.

For the collection of solar radiation, concentrators seem to have two obvious disadvantages. They miss part or all of the diffuse radiation, and if the concentration is sufficiently high they require tracking. On the other hand, tracking can enhance the annual energy collection by keeping the aperture normal to the sun at all times. Also, some low concentration collectors do not need any tracking.

To clarify these issues, it is useful to define two critical properties of concentrating collectors: the concentration ratio and the acceptance angle. The geometrical concentration* C is the ratio of aperture A and absorber surface area A_{abs}

$$C = \frac{A}{A_{abs}} \tag{8.1}$$

The acceptance angle, or field of view, is defined as the angular range over which all or almost all† of the incident rays are accepted without moving all or part of the collector. The acceptance angle is one of the most important characteristics of a solar concentrator because it determines the tracking requirement. By considering phase space conservation[1] or reciprocity relations for radiation shape factors,[2] one can show that the second law of thermodynamics imposes an upper limit on the concentration ratio achievable by any optical system with nonzero acceptance angle; this is sometimes called the thermodynamic (or ideal) limit of concentration. There must be a connection between optics and the second law of thermodynamics, because if solar radiation could be concentrated onto an arbitrarily small receiver, the receiver temperature could exceed the surface temperature of the sun. This would obviously be a violation of the second law, which states that heat cannot flow from a cold surface to a hot surface without an external source of work. The maximum possible concentration for a given acceptance half-angle, θ_a, for a two-dimensional (trough-like, line focus) concentrator is

$$C_{ideal,\ 2D} = \frac{1}{\sin \theta_a} \tag{8.2}$$

and for three-dimensional ones (cones, dishes, pyramids, point focus) it is

$$C_{ideal,\ 3D} = \frac{1}{\sin^2 \theta_a} \tag{8.3}$$

Since the angular radius of the sun is $\Delta_s \simeq 1/4^\circ \simeq 5$ mrad, this limit implies a maximum of 200 for the concentration of a single axis tracking solar concentrator,

* This quantity depends only on the geometry. Sometimes a flux concentration ratio has been used, defined as intensity ratio at aperture and absorber; it depends on absorption effects in addition to geometry. In the present chapter, the geometrical concentration ratio will be used exclusively.

† In practice, most collectors are designed to accept about 95% to 100% of the rays from the solar disk.

whereas for a point focus collector geometry it is 40,000. However, the concentration achievable in practical systems is reduced by a number of factors:

(1) Most conventional concentrators, in particular, line or point focus types, are based on optical designs which fall short of the thermodynamic limit by a factor of 2 to 4.

(2) Receiver misalignment, tracking errors, and errors in mirror surface and contour necessitate design acceptance angles considerably larger than the angular diameter of the sun.

(3) No lens or mirror material is perfectly specular; therefore, the acceptance angle must be enlarged further.

(4) Due to atmospheric scattering, a significant portion of the solar radiation may come from directions other than the solar disk itself.

The concentration ratio achievable with nontracking collectors is determined by the magnitude of the angular motion of the sun during the day and the year.[2] For fixed collectors, the highest practical concentration ratio is about 2. This value can be increased to about 3 if one permits a summer-to-winter adjustment of the collector tilt. With daily tilt adjustments, the concentration limit for nontracking collectors is about 10. These limits can be reached only by nonfocusing collectors of the CPC (compound parabolic concentrator) class.[3]

Solar concentrators which require little or no tracking must have a fairly large acceptance angle and, therefore, can collect a significant amount of diffuse radiation. A precise calculation of this effect would require detailed information about the angular distribution of diffuse sky radiation. Since, at the present time very little data on this distribution are available, one usually assumes that the hemispherical (or total) insolation, I_h, is the sum of the beam (or direct) component, I_b, and an isotropic background of diffuse insolation, I_d,

$$I_h = I_b + I_d \tag{8.4}$$

There is simple proof, in terms of radiation shape factors, that a concentrator of concentration C accepts $1/C$ of the diffuse radiation if the latter is isotropic.[2] Due to the predominance of near forward scattering in the atmosphere, the sky radiation tends to be centered around the sun, and, therefore, the isotropic model is a slight underestimate of the actual acceptance for diffuse radiation.

The choice of an optimal collector for a given application depends on many factors, and it is unlikely that a single concentrator type will be desirable for all applications. Accordingly, we present a review of the most promising concentrator types and describe a simple model for predicting how much energy they can deliver during a year.

8.3 NONTRACKING CONCENTRATORS

The principal nontracking collector types are flat plates, flat plates enhanced by side reflectors or V-troughs, tubular collectors, and compound parabolic con-

centrators (CPCs). For a given relation between aperture and radiation source, the highest possible concentration is achieved by a CPC. Hence, it is instructive to begin this section with a fairly detailed description of the large class of CPC configurations. This is followed by a discussion of reflector configurations for evacuated tubes. After that, we address the important practical case of V-troughs and side reflectors.

8.3.1 Compound Parabolic Concentrators

Concentrators that reach the thermodynamic limit of concentration, Eqs. (8.2) and (8.3), have been called ideal concentrators because of their optical properties. In the solar energy literature, names such as compound parabolic concentrators and nonimaging concentrators have also been used. We shall refer to all concentrators of this class as CPCs, even though some of them are not even parabolic in shape.

Ideal concentrators are a surprisingly recent discovery. The first example of a CPC, shown in Fig. 8.1, was found independently in the United States by Hinterberger and Winston,[4] in Germany by Ploke,[5] and in the USSR by Baranov and Melnikov.[6] It consists of parabolic reflectors that funnel the radiation from aperture to absorber. The right and left halves belong to different parabolas, as expressed by the name CPC. The axis of the right branch, for instance, makes an angle, θ_a, with the collector midplane, and its focus is at A. At the end points C and D, the slope is parallel to the collector midplane. Tracing a few sample rays reveals that this device has the following angular acceptance characteristic: all rays incident on the aperture within the acceptance angle, that is, with $|\theta| < \theta_a$, will reach the absorber, whereas all the rays with $|\theta| > \theta_a$ will bounce back and forth between the reflector sides and eventually reemerge through the aperture. This property, plotted schematically by the solid line in Fig. 8.2, implies that the concentration is equal to the thermodynamic limit.

Subsequent to the discovery of the basic CPC, Fig. 8.1, several generalizations of the ideal concentrator have been described which are relevant to special applications. These generalizations concern the following:

(1) The use of arbitrary receiver shapes,[2,7] for example, fins and tubes (see Fig. 8.3). The latter are important because of their ability to carry a heat transfer fluid.

(2) The restriction of exit angles, θ_{out}, at the receiver[8] to values $\theta_{out} < \theta_2 < \pi/2$. It is important because some receivers have poor absorptance at large angles of incidence. It is also needed for the design of CPCs that function entirely by total internal reflection.[9]

(3) The asymmetric orientation of source and aperture, see Fig. 8.4 (for the design of collectors with seasonally varying outputs).[2,10]

(4) The matching of a CPC to a finite source of radiation, see Fig. 8.5 (useful as second-stage concentrators to collect radiation from a first stage which is a finite distance away).

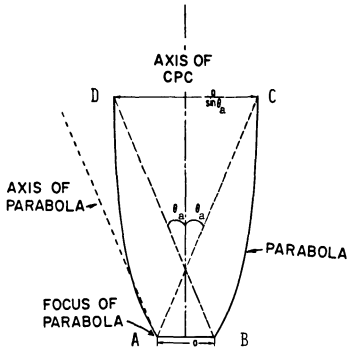


Fig. 8.1: Compound parabolic concentrators (CPC).

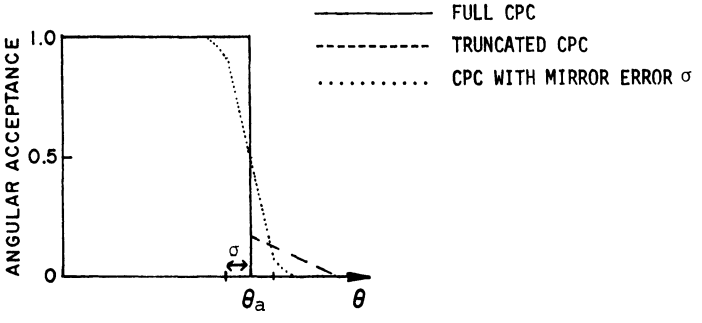


Fig. 8.2: Fraction of the radiation incident on aperture at angle θ that reaches absorber, for CPC with acceptance half-angle θ_a, assuming reflectivity $\rho = 1$. Full line: untruncated CPC with perfect reflectors; dashed line: truncated CPC with perfect reflectors; dotted line: untruncated CPC with surface errors σ.[2]

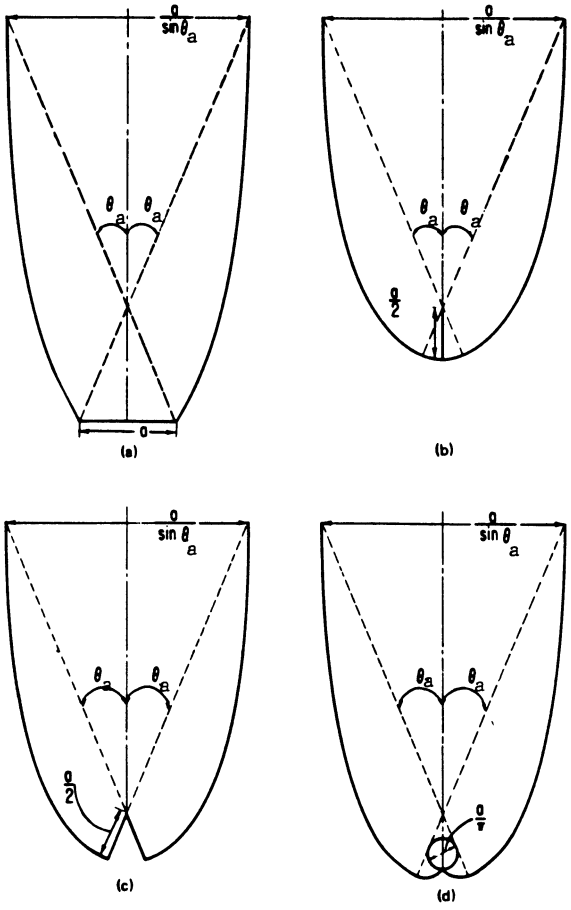


Fig. 8.3: CPCs with different absorber shapes. All have the same absorber perimeter, a, and acceptance half-angle θ_a.

The design goal is to concentrate radiation maximally, subject to any of the subsidiary conditions (1) through (4) that may have been specified. The design of two-dimensional concentrators is determined uniquely by the extreme rays—extreme rays being defined as rays coming from the edge of the source. In three dimensions (geometry of cones and pyramids), the design is in general overdetermined, but at least for flat receivers, a good compromise is achieved by choosing a surface of revolution whose cross section has been determined by the two-dimensional solution.[3]

To describe the extreme ray design procedure more fully, let us consider concentrators that use only reflective elements and let us demand that the average number of reflections be as small as possible. Then the solution is to maximize the slope of the profile curve of the mirror, subject to the condition that extreme rays illuminate the absorber within the prescribed angular limits $\pm\theta_2$. This implies that rays originating at angle θ_2 from a convex absorber (flat absorbers are treated as

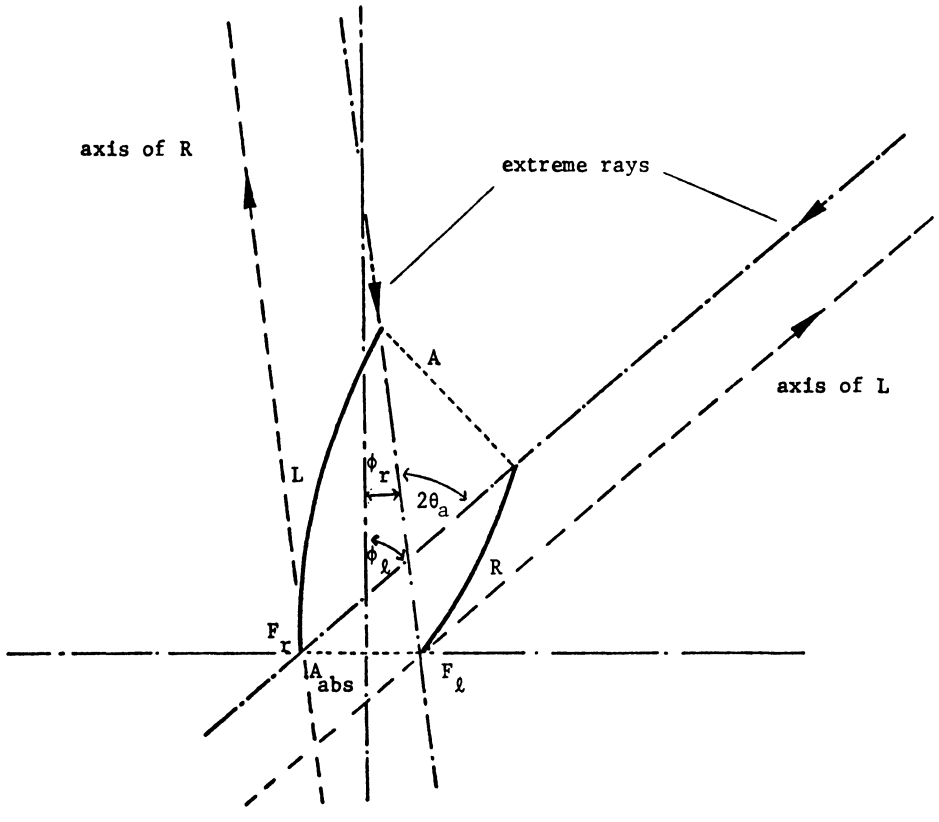


Fig. 8.4: Asymmetric CPC with acceptance angle $2\theta_a = \phi_l + \phi_r$ and geometric concentration $C = 1/\sin\theta_a$. The effective concentration varies with angle of incidence. A = aperture, A_{abs} = absorber, R = right parabola. L = left parabola, F_r = focus of R, F_l = focus of L.[2]

limiting cases of convex absorbers and, in nonconvex absorbers, cavities are replaced by flat chords stretched across cavity openings) emerge from the concentrator as extreme rays after undergoing at most one reflection. The procedure is illustrated in Fig. 8.5. Radiation emanating from the diffuse source S, and entering the aperture BB' is concentrated onto the absorber R. The mirror curve must have such a slope that a ray emitted from any point P of the absorber at an angle $+\theta_2(-\theta_2)$ may be reflected toward the edge $A'(A)$ of the source. The mirror starts at the edge $C(C')$ of the absorber and it terminates at the intersection $B(B')$, with the limiting rays $AC'(A'C)$ to avoid shading the absorber. Speaking mathematically, the concentrator slope is uniquely determined by a first-order differential equation with one boundary condition (the reflector curve must pass through the edge of the absorber). Explicit solutions and equations for the reflector shapes can be found in the individual references for each CPC type.

As for the choice between different CPC types, the configurations with fin or tube absorbers, Figs. 8.3(b) and 8.3(d), are preferable for most solar applications.

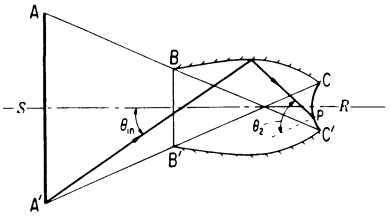


Fig. 8.5: Exemplary design of an ideal concentrator for finite source. Radiant energy from diffuse source (S) enters aperture BB' and is maximally concentrated onto the absorber (R). The angle of incidence on R is restricted to $|\theta_{out}| \leq \theta_2$. A typical extreme ray originates from A', impinges on the aperture with angle of incidence θ_{in}, and, after one reflection, is directed to point P on the absorber with angle of incidence θ_2.

Not only is the absorber material used more efficiently than in other designs, but heat losses through the back are low. This may be quite an important advantage, because it may not be cost-effective to reduce the effective U-value of the back of a collector much below 0.5 W/m^2 K. Compared to a frontal U-value on the order of 3 W/m^2 K for threefold[11] and 1.4 W/m^2 K for tenfold concentration, the losses through the back are indeed significant. Thus, the reduction in back losses possible with the configurations of Figs. 8.3(b) and 8.3(d) will more than compensate for the slightly higher optical losses (the average number of reflections for the configurations of Figs. 8.3(b) and 8.3(d) is approximately 0.5 higher than for the configuration of Fig. 8.3(a)).

In their optical properties, all CPC types are exactly or almost exactly alike. Above all, the same relation, Eq. (8.5), exists between their concentration and angular acceptance, with the sharp cutoff implied by Fig. 8.2. The flux distribution at the absorber depends on the angle of incidence and on absorber shape, and must be determined by detailed ray tracing. However, the following important statement can be made about all CPCs, without any need for ray tracing: if the radiation incident on the aperture is spread uniformly over the entire acceptance angle, then it will be isotropic when it reaches the absorber—unless the design was chosen to restrict the exit angles to values below $\theta_2 < \pi/2$, in which case the radiation at the absorber will uniformly fill the angular range from $-\theta_2 + \theta_2$. This consideration of uniform illumination is very important because it gives a simple and reliable estimate of the average performance of a CPC solar collector. When beam insolation is incident at certain angles, hot spots of high flux concentration (up to about 40) may appear on the absorber.

To analyze the optical performance of a CPC, it is convenient to think in terms of the number of reflections.

The number of reflections varies both with angle of incidence and with point of incidence on the aperture. A very good estimate of the transmission losses is

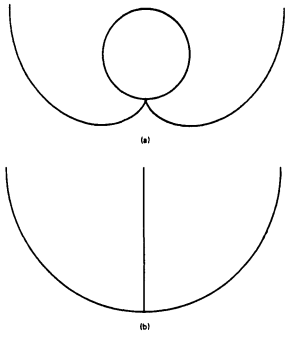


Fig. 8.6: Involute with unit concentration ratio, for tubular absorber (a) and for flat two-sided absorber (b).

given by the simple formula

$$\tau = \rho^{<n>} \tag{8.5}$$

where $< n >$ is the average number of reflections; the validity of this approximation has been demonstrated by Rabl.[12] This reference also shows how $< n >$ can be calculated in closed form for a large number of concentrators and radiation passages.

CPCs have large reflector areas. Fortunately, this disadvantage can be alleviated by truncation: the top portion of a CPC can be cut off with little loss in concentration since it does not intercept much radiation. Detailed graphs showing the effects of truncation on height, reflector area, number of reflections, and concentration ratio have been published by Rabl[2,13] and by McIntire.[14]

Regarding sensitivity to mirror surface errors, the analysis is equally simple for all CPCs because their geometry implies that all rays incident near the cutoff angle, that is, with $|\theta| \lesssim \theta_a$, undergo exactly one reflection on their way to the absorber. In almost all practical applications, the acceptance half-angle, θ_a, will be larger than 5°, and the mirror surface errors, σ, will usually be small compared to θ_a. Therefore, all the rays with $|\theta| < \theta_a - 2\sigma$ and none of the rays with $|\theta| > \theta_a + 2\sigma$ will reach the absorber, whereas in the transition region, $\theta_a - 2\sigma < |\theta| < \theta_a + 2\sigma$, some rays are accepted and some are rejected. The resulting angular acceptance is shown schematically by the dotted line in Fig. 8.2. Further details on optical and geometric properties of CPCs can be found in Welford and Winston[3] and in Rabl.[15]

The involute with unit concentration ($C=1$) is included in this chapter as a special case of the CPC. The acceptance half-angle is 90°. The examples in Fig. 8.6 distribute incident radiation over the surface of a tubular or fin absorber, and the average number of reflections is $\pi/4$ for isotropic radiation.[12]

Recently, some new concentrator types have been discovered that reach or closely approach the thermodynamic limit, even though they are quite different from the CPC. Welford and Winston discovered a generalized design principle that yields not only the CPC, but also a novel type of second-stage concentrator, called "trumpet" because of its shape.[16] An interesting prismatic reflector with total internal reflection has been described by Mills and Giutronich.[17] It closely approaches

the thermodynamic limit and is well suited as a second-stage concentrator for line focus photovoltaic collectors.[18]

We note in passing that the optical designs described here have many applications beyond solar energy. They are useful wherever one wants to redistribute radiation in a particular manner. The reflector of a photographic flash tube, for instance, is to distribute the light from the tube in such a way as to uniformly illuminate the field of view of the camera. The appropriate design is a CPC, or a CPC-like lens-mirror combination. (The flashlight industry seems to have discovered these principles independently.) Obvious energy-conserving applications lie ahead in the lighting industry when one wants to maximize the transmission of light from fluorescent tubes at the ceiling to the room below.

8.3.2 Reflectors for Evacuated Tubes

By and large, the cost per unit area of reflectors is lower than the cost of evacuated tubes. For this reason, most of the evacuated collectors sold today use some kind of reflector enhancement. Because of imperfect reflectivity, the use of a reflector incurs some optical losses. At the same time, heat losses are reduced because there is less absorber area per aperture area. Thus, the choice of a reflector involves trade-offs between optical performance, thermal performance, and cost. In general, one wants a large acceptance angle to minimize the need for tilt adjustments. From an optical point of view, the CPC is the best reflector, but in some cases one might, for practical reasons, choose a different reflector, such as the V-trough.

The choice of the reflector depends on the shape of the absorber. Figure 8.7 shows several reflector arrangements that have been used with tubular absorbers. The diffuse reflector in Fig. 8.7(a) is just a plain white surface behind the tubes. It has the lowest cost and the lowest performance since much of the reflected radiation misses the tubes. V-groove and circular cylindrical reflectors, shown in Figs. 8.7(b) and 8.7(c), are also quite easy to fabricate, and in small quantities their cost can be significantly lower than the cost of CPCs. V-trough reflectors are fairly well matched to flat absorbers, but with tubular absorbers they do not utilize the back of the tubes very well. The best optical performance is achieved with a CPC reflector, as in Fig. 8.7(d). With mass production technologies, the shape of the reflector has little influence on the cost, and it becomes advantageous to use the best shape. One way to describe the superior optical performance of the CPC is that the entire aperture area is effective in directing radiation to the absorber, over the entire range of incidence angles within the acceptance angle. This is illustrated by Fig. 8.8. With the absorber tube in the design position of a CPC reflector trough, the entire aperture appears black if viewed from any angle within the acceptance angle. An analogous photograph taken of the V-groove or circular reflector arrangements would reveal large shining portions in the aperture where incident radiation misses the absorber. At particular incidence angles, the entire aperture may be active even with these reflectors, but only with a CPC is the entire aperture active over the entire range of incidence angles that has been specified.

It is worth pointing out that in some cases, one may want to design for a geometric concentration ratio below unity. Some absorber coatings of evacuated

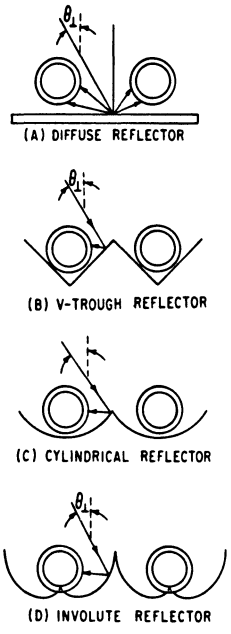


Fig. 8.7: Several reflector configurations for evacuated tubes.[19] **a)** diffuse reflector; **b)** V-trough reflector; **c)** circular reflector; **d)** involute reflector (CPC).

tubes have a rather low absorptivity. Together with reflections off the glass envelope, these tubes may reflect 20% to 30% of the incident radiation. If the reflected rays have another chance to hit a tube, the overall effective absorptivity of the collector is improved. This kind of absorption enhancement is possible if the geometric concentration ratio is less than unity. In effect, the absorber tubes plus their reflectors can act like a radiation cavity whose absorptivity exceeds the absorptivity of the absorber surface itself. This question has been studied by O'Gallagher et al.,[19] and the involute (CPC with concentration less than unity) was shown to give the greatest possible absorption enhancement.

An important design problem for thermal CPC collectors arises from the need for a gap between reflector and absorber.* The design principles for CPC reflectors demand that the reflector extend all the way to the absorber. However, this is not possible with evacuated tubes because of the finite thickness of the glass and a finite spacing between absorber surface and glass envelope. Several modifications of the basic CPC design are possible to deal with this gap problem. One could, for example, leave the reflector intact and reduce the absorber size. Or one could leave the absorber as it is and truncate the adjacent reflector. These and several other solutions have been investigated and the following can be recommended as the best in terms of minimizing the optical loss.

* A gap is also needed in nonevacuated CPC collectors if the reflector has a high conductivity, otherwise the reflector may turn out to act as a cooling fin for the absorber.

Fig. 8.8: a) Photograph of CPC troughs with 35° acceptance half-angle. Evacuated absorber tube held in design position makes the entire aperture look black if viewed from anywhere within the acceptance angle (Countesy of Argonne National Laboratory). **b)** Cross section of evacuated 1.5X CPC with shaped glass tube.

For flat absorbers the optical loss of the gap is minimized if the reflector is truncated adjacent to the absorber, as sketched in Fig. 8.9. As shown in Rabl et al.,[20] the optical loss for uniform illumination within the acceptance angle can be calculated in closed form using radiation shape factors. This is an excellent approximation for the yearly average loss in actual operation. For small gaps in the configuration of Fig. 8.9, the fraction of the incident radiation that is lost is in the

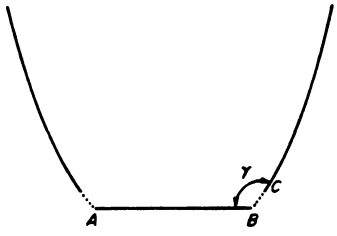


Fig. 8.9: Truncation of CPC to accommodate gap between flat one-sided absorber and reflector.[20]

range of 0.3 g/a to 0.5 g/a (depending on CPC shape) where

$$g = \overline{BC} = \text{gap width} \tag{8.6}$$

and

$$a = \overline{AB} = \text{absorber width} \tag{8.7}$$

For tubular absorbers, one could also truncate the reflector, as suggested in Fig. 8.7(d). The associated gap loss has been calculated by Rabl et al.[20] and turns out to be quite small. But with tubular absorbers one can do even better by avoiding gap losses altogether. This possibility was suggested by McIntire[21] and explored systematically by Winston.[22] This design replaces the cusp by a grooved cavity as shown in Fig. 8.10. The number and dimensions of the grooves depend on the gap width. With a single groove, the gap can be as large as 0.27 times the tube radius. A design with two grooves permits a lossless solution up to $g \simeq 0.4r$. With a larger number of grooves, the lossless design can be extended up to a maximum gap equal to the radius. The lossless design necessarily entails a certain sacrifice of concentration ratio. For instance, the design of Fig. 8.10 achieves only a concentration of 1.0 even though the thermodynamic limit for this acceptance angle is $1/\sin 60^\circ = 1.15$. In most solar applications, high optical efficiency is more important than the attainment of the highest possible concentration ratio, hence one will usually choose the lossless solution.

For a detailed optical analysis of evacuated tubes with reflectors, the reader is referred to articles by McIntire[23] and by Window and Basset.[24]

8.3.3 V-Troughs

V-troughs are a classic concentrator design for applications with large acceptance angle; see, for example, Tabor[25] and Hollands.[26] To analyze the multiple reflections in V-troughs, the method of images is convenient. The angular acceptance of the V-trough of Fig. 8.11(a) is shown schematically by the solid line in Fig. 8.11(b). Compared to the CPC of the same concentration $C + 1/\sin(\phi + \theta_1)$ (see dotted line of Fig. 8.11(b)), the useful acceptance angle θ_1 of a V-trough is significantly smaller. Only when the trough angle ϕ approaches zero does the acceptance angle approach that of the CPC; however, in this limit, the trough becomes too deep and the reflection losses become excessive.

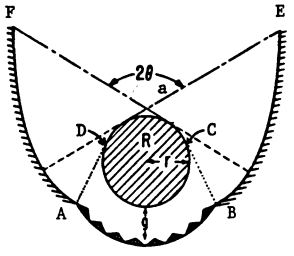


Fig. 8.10: Modification of CPC for tubular absorber with finite gap, g, and zero optical loss. For the example chosen, the gap is $g = 0.8r$, the groove angle $2\psi = 118°$, the acceptance half-angle $\theta_a = 60°$, and the concentration ratio $C = 1.0$.[22]

The higher the desired concentration, the greater the relative advantage of the CPC over the V-trough. The upper limit of concentration for a practical V-trough is about 3 (as a nontracking collector with daily tilt adjustments). With summer/winter adjustments only, the V-trough is limited to concentration values below 2 and, for a completely fixed collector, a V-trough gives almost no concentration. As for absorber shapes, the V-trough is limited to flat one-sided absorbers. An interesting asymmetric V-trough that requires summer-to-winter adjustment has been described by Selcuk.[27]

8.3.4 Side Reflectors

Side reflectors can be a useful means of increasing the output of a flat plate collector. Especially in situations where a suitable surface is available right next to a collector, it can be quite cost-effective to add a reflector, for instance a sheet of anodized aluminum. If a tilted collector is mounted on a flat horizontal roof, one might place a reflector in front. In large installations, several rows of collectors may be mounted one behind another; in this case, one may put reflectors behind the collector rows creating a saw-tooth pattern of alternating collector and reflector surfaces. If a separate support structure is needed for a side reflector, the cost may be too high to be practical.

Flat plates with a side reflector tend to produce a rather nonuniform output over the course of the year, just like an extreme asymmetric (or one-sided) CPC[2] to which it is, so-to-speak, a straightline approximation. Whether such nonuniform output is desirable depends on the load distribution. Depending on the geometry, the reflector may at some times of the year cast a shadow on the collector and actually reduce the performance. There is one case where a reflector is guaranteed to improve the collector output. A single tilted collector always receives some radiation from the ground during daylight hours; hence, a horizontal reflector in front of the collector can only help. Whether it is cost-effective is, of course, another question. As for the choice between a specular and a diffuse reflector, a diffuse side reflector has very little effect as shown by Grassie and Sheridan.[28] The side reflector should be reasonably specular and its reflectivity should be as high as possible. To evaluate

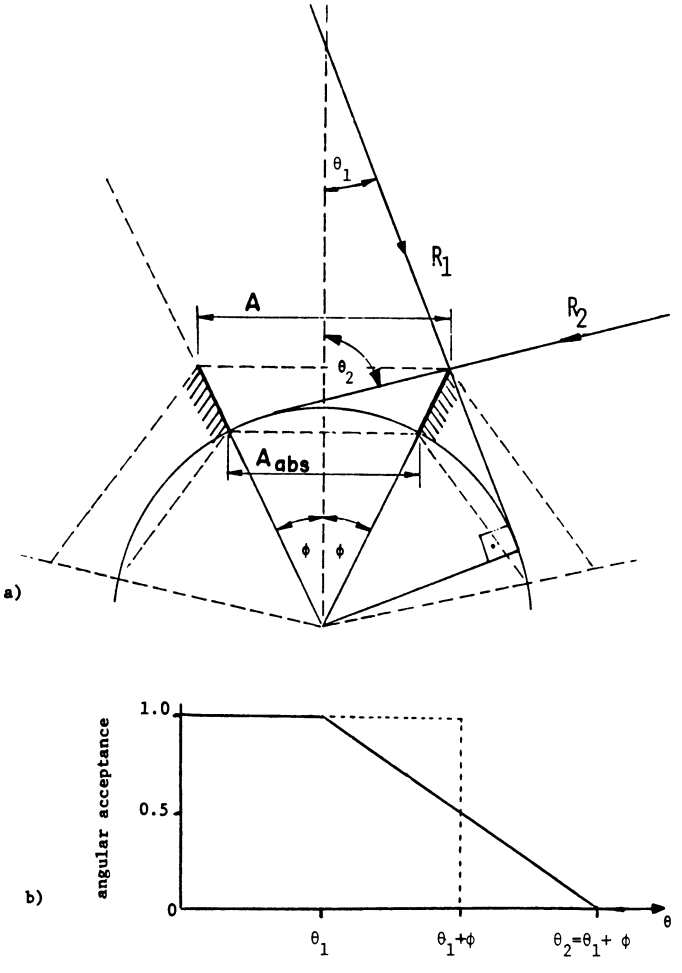


Fig. 8.11: **a)** V-trough concentrator with aperture, A, and absorber, A_{abs}. Mirror images and reference circle are also shown. The rays P_1 and P_2 have angle of incidence θ_1 and θ_2, respectively; they pass through the edge of the absorber and are tangential to the reference circle. **b)** Angular acceptance of V trough (schematic, neglecting difference between polygon and circle in **a**).[2]

the benefit of side reflectors, one must look at system performance, not just the optical enhancement. Since the heat loss is the same with and without reflector, the increase of the collector output can be significantly higher than the optical enhancement factor.

Apart from these simple facts, it is difficult to make general statements about side reflectors. A detailed evaluation is needed in each case because the benefit of side reflectors depends on a large number of variables, not only the geometric/optical parameters, but also the system configuration and load. Even the optical analysis of the reflector/collector combination is fairly complicated. The number of variables

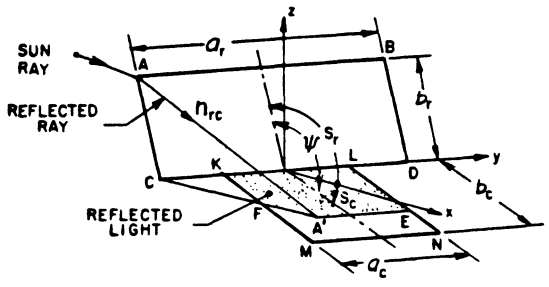


Fig. 8.12: Flat-plate collector (KLMN) with side reflector (ABCD).[26]

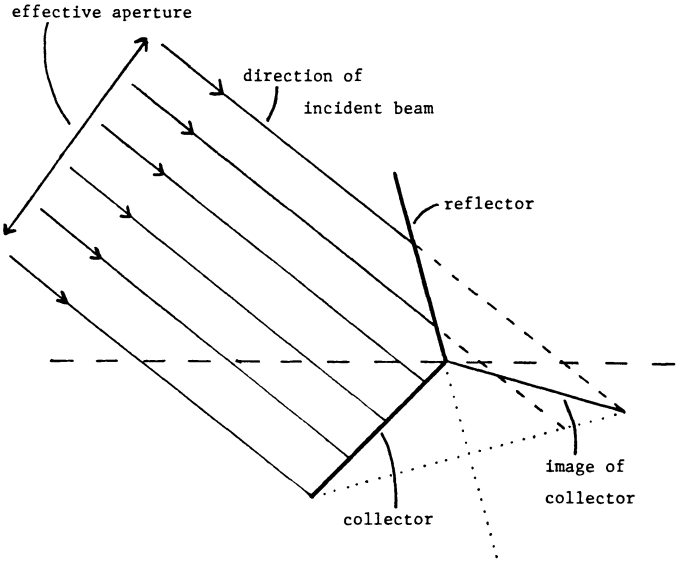


Fig. 8.13: Image and effective aperture for reflector/collector combination.

involved is indicated in Fig. 8.12. Both reflector and collector have finite length and, for the incident direction of beam insolation in this figure, the portion ($KLEA'F$) of the collector receives reflected radiation. For certain incidence angles, the reflector may cast shade: if the shadow of the corner A falls on A', then the entire portion $KLEA'F$ is shaded. The reflector in Fig. 8.12 is above the collector; one could also place it below.

The analysis is much simpler if the reflector is long compared to the collector because then the problem reduces to two dimensions. It is instructive to look at the two-dimensional case and add the image of the collector in the reflector, as shown in Fig. 8.13. For each incidence angle, the effective aperture is the projected aperture of the collector plus the projected aperture of the image.

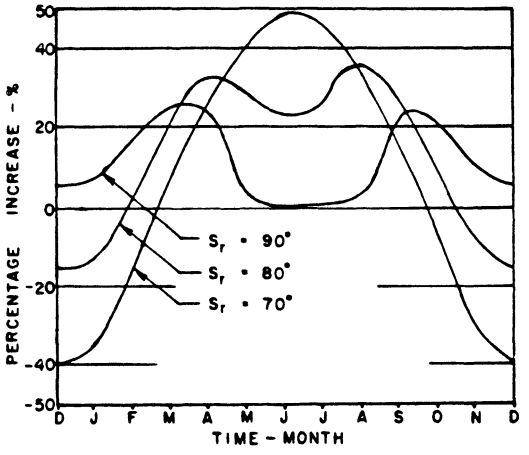


Fig. 8.14: Percentage increase of energy delivered by a hot water system due to the addition of a reflector above the collector, at tilt angle, S_r, from horizontal.[28]

How the addition of a side reflector might effect the system performance is illustrated in Fig. 8.14. In this particular example, a domestic hot water system uses a flat plate collector at tilt 35° due north in Brisbane, Australia, at latitude 27.5° south. Figure 8.14 displays the monthly performance change of the system due to the addition of a side reflector above the collector. Three different curves are shown, corresponding to three different reflector angles, S_r. For $S_r = 90°$, the delivered energy is always increased. For $S_r = 70°$, the output is highly peaked in summer while it drops to 40% below the no-reflector case in winter. For more specific results, in particular for the optical performance, the reader is referred to detailed studies in the literature.[29-32]

8.4 TRACKING CONCENTRATORS

In principle, any collector, even a flat plate, could be made to track the sun, but, in practice, one will usually resort to tracking only with collectors of fairly high concentration ratio. Hence, the acceptance angle of tracking collectors is relatively small, and care must be exercised in the optical design. Since even direct solar radiation is not perfectly collimated, but comes from a range of directions with further spread due to optical errors, the receiver must have a minimum size if it is to intercept most of the incident radiation. The choice of the optimal absorber size involves a compromise between optical and thermal performance. If the absorber is too large, most of the incident solar radiation will be intercepted, but the heat losses are excessive. On the other hand, a very small absorber has low heat losses, but it will also miss too much of the available solar radiation. A proper analysis of these effects needs to take into account the angular distribution of direct solar radiation and of optical errors, and, hence, we begin with a discussion of these distributions.

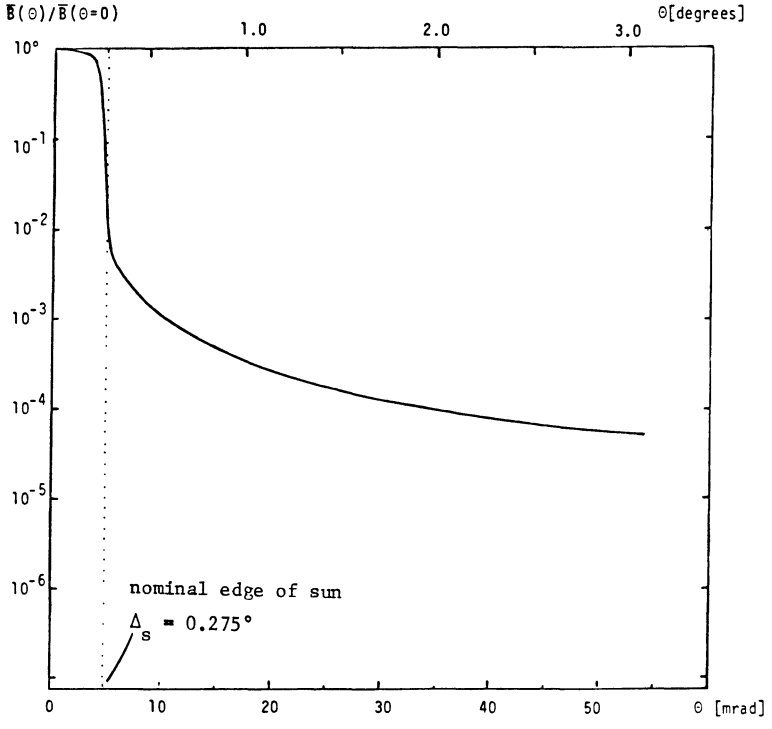


Fig. 8.15: Brightness distribution (dimensionless) of standard solar and circumsolar scan.[33]

8.4.1 Image Spread Due to Finite Width of the Sun and Optical Errors

When viewed from the earth, the sun appears as a disk of angular radius

$$\Delta_s = 4.7 \text{ mrad} \tag{8.8}$$

Seasonal variations in the Δ_s due to the eccentricity of the Earth's orbit are so small ($\pm 1.7\%$) that they can be neglected. For most applications, one need not worry about the detailed brightness distribution of the solar disk. Only collectors with very high concentration are sensitive to details of the angular brightness distribution of the sun. A typical solar brightness distribution is plotted in Fig. 8.15. This distribution is the simple average of all solar and circumsolar data obtained by the Lawrence Berkeley Laboratory circumsolar telescope.[33]

The distribution in Fig. 8.15 is seen to be fairly flat over the center of the sun, but it decreases near the edge, a phenomenon known as limb darkeninig. At the edge, the brightness drops sharply by several orders of magnitude. The region beyond 4.7 mrad and out to 50 mrad is called circumsolar region. The limb darkening of the solar disk is caused by absorption and scattering in the photosphere of the sun and the atmosphere of the Earth. If the sun radiated isotropically and if there were

no narrow angle scattering in the atmosphere, then the brightness of the solar disk would be perfectly uniform.

Actual solar brightness distributions vary a great deal with atmospheric conditions: under hazy sky, the radiation from the solar disk is less, the circumsolar radiation more than under a clear sky. Nevertheless, the standard circumsolar scan of Fig. 8.15 (with separate weather-dependent weighting factors for the solar and the circumsolar portions) is very useful because it permits a simple yet accurate calculation of the long-term average effect of circumsolar radiation on the performance of highly concentrating collectors, as proved by Rabl and Bendt.[33]

For a statistical analysis of the radiation intercepted by a collector, it is convenient to use the rms width of the sun rather than its radius. For point focus geometry (e.g., parabolic dish), the rms width, $\sigma_{\text{sun,point}}$, is given by*

$$\sigma^2_{\text{sun,point}} = \frac{\int_0^\infty d\theta\, \theta^3\, B(\theta)}{\int_0^\infty d\theta\, \theta\, B(\theta)} \tag{8.9}$$

For line focus geometry (e.g., parabolic troughs), one has to distinguish the angular variables, $\theta_\parallel$ and $\theta_\perp$, parallel and perpendicular to the direction of the absorber. They are indicated in Fig. 8.16. Only the direction perpendicular to the absorber needs to be considered for line focus collectors because in the parallel direction all rays reach the absorber, no matter how large their angle, $\theta_\parallel$. Therefore, the relevant rms width of the sun for the line focus geometry is given by

$$\sigma^2_{\text{sun,line}} = \sigma^2_\perp = \frac{\int_{-\infty}^\infty \int d\theta_\parallel\, d\theta_\perp\, \theta_\perp^2\, B(\theta)}{\int_{-\infty}^\infty \int d\theta_\parallel\, d\theta_\perp\, B(\theta)} \tag{8.10}$$

Since $\theta^2 = \theta_\parallel^2 + \theta_\perp^2$, it is clear from Eq. (8.10) that

$$\sigma_{\text{sun,point}} = \sigma_\parallel^2 + \sigma_\perp^2 = 2\sigma_\perp^2 \tag{8.11}$$

hence, the rms width of the sun for point focus geometry is $\sqrt{2}$ times the rms width for line focus geometry

$$\sigma_{\text{sun,point}} = \sqrt{2}\sigma_{\text{sun,line}} \tag{8.12}$$

An intuitive explanation for this difference between line and point focus geometry is given in Fig. 8.16. The solid circle represents the source while the dotted lines show the field of view of a collector of high concentration with point focus (a) and line focus (b). As drawn in this figure, the field of view characterized by an acceptance angle, $2\theta_a$, is smaller than the width, $2\Delta_s$ of the source. Thus, the radiation from the shaded portion of the source is not collected. For the same value of θ_a, the portion that is lost is seen to be smaller for a line focus collector than for a

* For simplicity, the limit of integration has been extended to infinity because in practice, $B(\theta)$ can be assumed to be negligible beyond a few degrees.

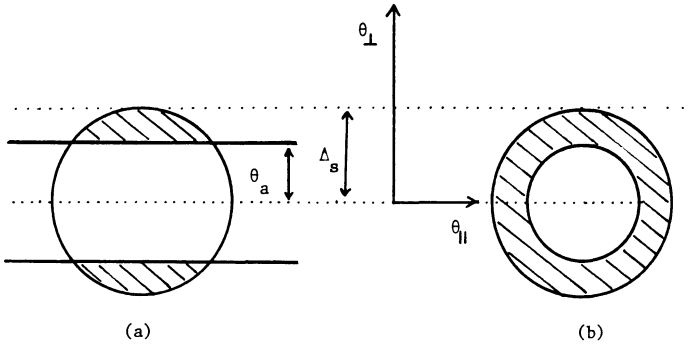


Fig. 8.16: Field of view of line focus **a)** and point focus **b)** collector aimed at solar disk. Shaded region indicates radiation that is not accepted by collectors. The orientation of the angular coordinates $\theta_{\parallel}$ and $\theta_{\perp}$ parallel and perpendicular to the absorber in part **a)** is also shown.

point focus collector. In this sense, the sun appears smaller to a line focus collector than to a point focus collector.

If the sun were a uniform disk, the rms width would be

$$\sigma_{\text{disk,point}} = \left(\frac{\int_0^{\Delta_s} d\theta\ \theta^3}{\int_0^{\Delta_s} d\theta\ \theta} \right)^{1/2} = \frac{\Delta_s}{\sqrt{2}} \tag{8.13}$$

For real solar brightness distributions, the rms width varies slightly with atmospheric conditions (due to the contributions of circumsolar radiation). Under very clear sky (average clear sky), the width is

$$\sigma_{\text{sun,point}} \approx 3.6 \text{ mrad} (5.6 \text{ mrad}) \tag{8.14}$$

and

$$\sigma_{\text{sun,line}} \approx 2.7 \text{ mrad} (4.0 \text{ mrad}) \tag{8.15}$$

("Very clear sky" corresponds to circumsolar scan #1, "average clear sky" to the average over the rms widths for circumsolar scans #1 through #10 of Grether and Hunt[34]; their I_b values range from 639 to 954 W/m^2. A description of these scans can also be found in Biggs and Vittitoe[35] and in Rabl and Bendt.[33])

Snell's law of reflection is an idealization based on perfectly smooth surfaces. Real surfaces tend to have all kinds of irregularities, from microscopic roughness to macroscopic undulations. Furthermore, real reflector surfaces will not conform exactly to the design shape, and tracking collectors may have alignment errors. All these effects contribute to enlarging the focal zone of a solar concentrator. For solar energy calculations, a statistical analysis of these effects is adequate, because one is only interested in overall or averaged performance features, quite apart from the

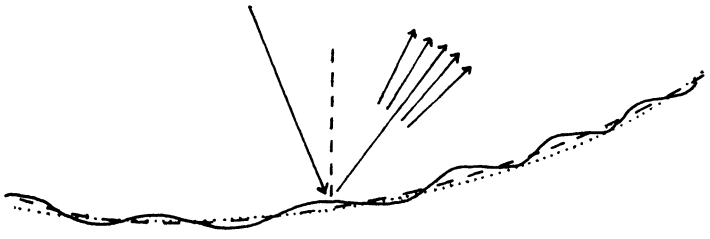


Fig. 8.17: Schematic sketch of large-scale, medium-scale, and microscale errors of reflector surface. Dotted line shows design surface, dashed line shows average of actual surface, solid line shows actual surface. Microroughness of surface causes scattering of reflected radiation over a range of angles.

fact that it would be practically impossible to measure and monitor the reflector surface to its last detail.

In discussing these features of reflector surfaces, it is conceptually helpful to distinguish three scales, as illustrated in Fig. 8.17. The dotted line shows the design shape, typically as a parabola. The solid line shows the actual reflector surface with its waviness. The dashed line averages over the undulations. The difference between the design shape (dotted line) and the average shape (dashed line) represents large-scale optical errors, caused by gravity, wind, materials stress, or manufacturing errors. The waviness, with typical wavelengths on the order of centimeters to decimeters represents medium-scale errors. Finally, there is the small scale: microscopic surface roughness which causes scattering away from the specular direction.

This classification is not always clear-cut and in some cases there could be a continuous transition from small- to large-scale errors. Nonetheless, it seems to be useful for solar concentrators, and it corresponds to different measuring methods. Small-scale errors are a property of the reflector surface itself, whereas medium- and large-scale errors are due to the substrate and the support structure. Thus, small-scale errors are determined by measuring the spread of a collimated beam of light after it has been reflected from a small sample of the material. Medium- and large-scale errors can be determined by mapping the contour of the reflector using mechanical or laser ray tracing.

Each of these optical errors can be characterized by a statistical distribution function. Only angular variations matter. Since image preservation is irrelevant, displacements in position need to be considered only to the extent that they cause deviations in angle. Ideally, the distributions should be derived from measurements. From the available data it appears that the surface error distributions are usually well approximated by a gaussian or normal distribution.[36] Fortunately, in many cases the details of the distributions do not matter even if they are not gaussian. This follows from the central limit theorem of statistics which says that the convolution of a large number of independent distributions approaches a gaussian, even if the

individual distributions are not gaussian.*[37,38] Thus, only the standard deviations of the individual distributions are needed, provided they have zero mean, as they usually do in solar applications. Because of the central limit theorem one can frequently (with the exception of collectors with very high concentrations and very small errors) approximate the brightness distribution of the sun by a gaussian. If a single gaussian does not give an accurate enough representation of the sun, one can employ the formulation developed by Vittitoe and Biggs[39] which treats the sun as a superposition of six gaussian terms.

The lack of specularity of solar reflector materials has been investigated extensively by Pettit.[40] Pettit found that for many reflector materials, the scattering of radiation about the specular direction can be described by a gaussian

$$R(\theta) \propto R_1 \exp\left(-\frac{\theta^2}{2\sigma_1^2}\right) \tag{8.16}$$

where θ is the angular deviation of a particular ray from the specular direction and $R(\theta)$ is the intensity of radiation reflected into the direction θ. R_1 is the total (i.e., hemispherical reflectance) and σ_1 is the width of the distribution. Some materials, for example Alzac, a rolled and polished aluminum sheet, exhibit different surface roughness in different directions. For these Pettit found it necessary to fit the data with a superposition of two gaussians

$$R(\theta) \propto R_1 \exp\left(\frac{-\theta^2}{2\sigma_1^2}\right) + R_2 \exp\left(\frac{-\theta^2}{2\sigma_2^2}\right) \tag{8.17}$$

The rms width σ_{optical} for the total optical error is obtained by adding the squares of the individual widths:

$$\sigma_{\text{optical}}^2 = 4\sigma_{\text{contour}}^2 + \sigma_{\text{specular}}^2 + \sigma_{\text{displacement}}^2 + \sigma_{\text{tracking}}^2 \tag{8.18}$$

$\sigma_{\text{contour}}^2$ is multiplied by two because of Snell's law; in Fresnel reflectors, σ_{tracking} must also be multiplied by two. The total beam width, σ_{tot}, is obtained by adding the rms width of the sun according to

$$\sigma_{\text{tot}}^2 = \sigma_{\text{optical}}^2 + \sigma_{\text{sun}}^2 \tag{8.19}$$

8.4.2 Parabolic Reflectors

One of the best-known solar concentrators is the parabolic reflector; it can be built either as a trough or as a dish. The absorber can take a variety of shapes, the most common being flat or round.

The parabola is the unique reflector shape that focuses a collimated beam of radiation into a single point. Perfect focusing is possible only for rays that are

* Unless the distributions are dominated strongly by a nongaussian term.

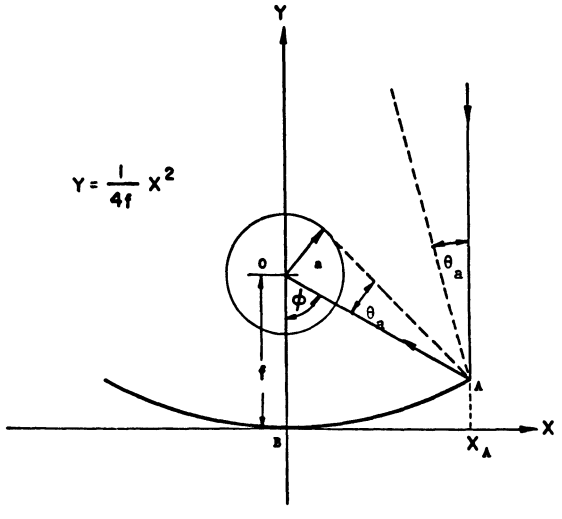


Fig. 8.18: Absorber radius a and acceptance half-angle θ_a for a focusing parabola.

incident parallel to the optical axis (symmetry axis) of the parabola. A collimated beam coming from other directions will not only miss the focus, but, due to off-axis aberrations of the parabola, it will not even converge into a single point.

If the absorber is to intercept most or all rays, it must be made sufficiently large. The example of the parabolic trough reflector with cylindrical absorber tube in Fig. 8.18 serves to illustrate this problem for focusing collectors. The absorber tube is placed concentrically around the focal line. If the ray with the largest deviation, θ_a, is to reach the absorber just barely, as shown by the dashed line in Fig. 8.18, then the concentration must be

$$C_{2D,\text{parab., cyl. abs.}} = \frac{2x_A}{2\pi a} = \frac{\sin\phi}{\pi \sin\theta_a} = \frac{\sin\phi}{\pi} C_{\text{ideal}} \tag{8.20}$$

where ϕ is the rim angle $\measuredangle AOB$. The maximum occurs for $\phi = 90°$ and falls a factor π short of the ideal limit. This is typical of all single stage focusing concentrators; that is, they reach only one-fourth to one-half of the thermodynamic concentration limit. Eq. (8.20) determines how large the absorber must be to intercept all rays incident within the acceptance angle, i.e., within $\pm\theta_a$ from the optical axis. But this is not the complete answer since the distribution of incident rays does not have a sharp cut off; also in a nonideal concentrator, some rays from outside the nominal acceptance angle are accepted. An accurate analysis is quite involved. As a simple short cut, one frequently takes the rule of choosing θ_a equal to twice the rms width, σ_{tot}, of the radiation distribution incident on the absorber, given by Eq. (8.19). This rule corresponds to intercepting 95% of the radiation for a gaussian distribution, and it seems to be a reasonable approximation.

A more accurate optimization of the concentration ratio involves weighing in-

cremental optical gains against incremental thermal losses. Usually this problem has been treated with computer ray trace programs. Recent analytical or semianalytical solutions have been published by Bendt et al.[41] for parabolic troughs and by Rabl and Bendt[33] for parabolic dishes.* The crucial optical quantity to be calculated is the intercept factor, γ, defined as the fraction of the rays incident on the aperture that is intercepted by the absorber. With this definition γ depends only on geometry and radiation distribution. Reflectivity, ρ, of the reflector and absorptivity, α, of the absorber are treated separately, so that the optical efficiency, η_o, is given by

$$\eta_o = \rho\gamma\alpha \tag{8.21}$$

If the absorber is covered by glazing of transmissivity, τ, then Eq. (8.21) contains an additional factor, τ, unless one wants to worry about multiple reflections between absorber and glazing, in which case an additional correction factor is needed.

If one makes the approximation that the distribution of radiation at the absorber is a gaussian of width σ_{tot} then the intercept factor depends only on rim angle ϕ, on absorber shape, and on the quantity $\sigma_{\text{tot}}C$ for troughs or $\sigma_{\text{tot}}^2 C$ for dishes. Thus, it can be presented readily in graphic form. Figure 8.19 shows γ for a parabolic trough, with cylindrical receiver (a) and with flat one-sided receiver (b); for the parabolic dish with flat one-sided receiver, γ is shown in Fig. 8.20.

8.4.3 Fresnel Reflectors

The smooth optical surface of a reflector or lens can be broken into segments, a trick invented by Fresnel. An example of a point focus Fresnel mirror system is the central receiver in Fig. 8.21. Fresnel reflectors can be very useful for solar energy, especially in large installations. As one increases the size of parabolic collectors, the mechanical problems of weight and wind loading increase. In practice it appears to be uneconomical to build parabolic collectors with aperture areas much larger than 100 m^2. When one needs larger areas, one would therefore have to install several collector modules, each with its own receiver. Collecting the heat from the individual receiver to a central point of use can be quite expensive. Therefore, it may be better to collect the solar energy optically by means of a central receiver. The receiver is mounted on top of a tower, surrounded by a field of so-called heliostats, mirrors that track the sun individually. Linear Fresnel mirror systems can also be built, with linear receiver and one-axis tracking reflectors (see Fig. 8.22).

The optical analysis of Fresnel reflectors is fairly complicated. Since the location of the reflectors is fixed,† the angle of incidence of the sun on the reflector varies with time. This causes shading and blocking. Shading occurs if direct sunlight fails to reach a mirror because the mirror is in the shadow of another mirror; blocking occurs if light reflected by a mirror fails to reach the absorber because

* Bendt et al.[41] also show how three-dimensional effects need to be accounted for if a parabolic trough is analyzed in two dimensions.

† Usually the reflectors are mounted on the ground (e.g., for the central receiver) or in a tilted plane (e.g., for small line focus Fresnel collectors).

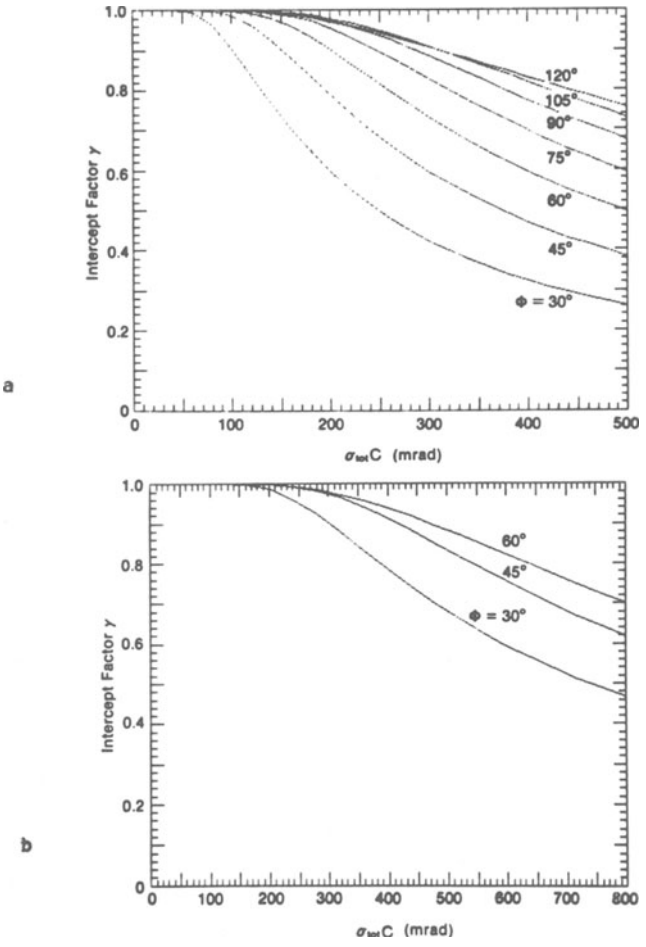


Fig. 8.19: Intercept factor γ versus $\sigma_{tot}C$ for parabolic trough with different rim angles ϕ (Gaussian approximation) for **a)** a cylindrical receiver and **b)** a flat receiver.[41]

it is intercepted by the backside of another mirror. In the interest of efficient mirror utilization, the reflector segments should be spaced far enough apart to avoid excessive shading and blocking. The optimal spacing depends on average incidence angles and on the relative cost of reflector area and rest of system. The ratio of heliostat area over total ground area is called ground cover factor ψ. Values of ψ between 0.3 and 0.6 are practical for Fresnel reflectors that track around the north-south axis or about two axes. For linear Fresnel reflectors with east-west tracking axes, the shading and blocking effects are less severe and ψ can be larger.

There is a further complication caused by varying incidence angles of the sun. The size and shape of the solar "images" from the individual reflectors vary with time of day and year. This effect is particularly complicated if the individual mirrors

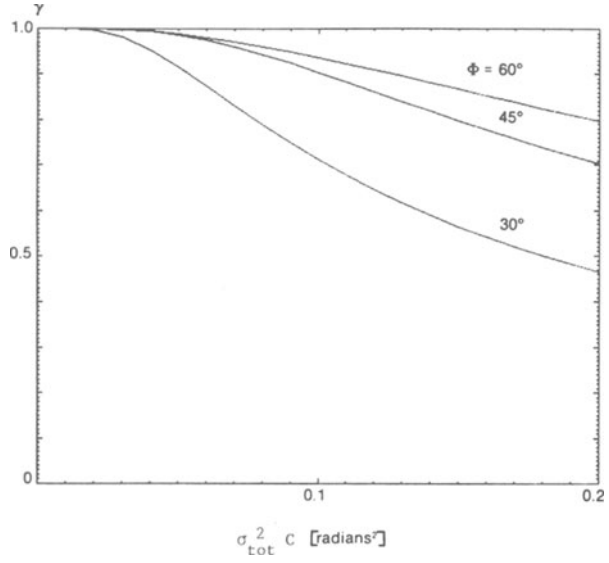


Fig. 8.20: Intercept factor γ for parabolic dish with flat receiver geometric concentration C, and rim angle ϕ, if source is Gaussian with width σ_{tot}.[33]

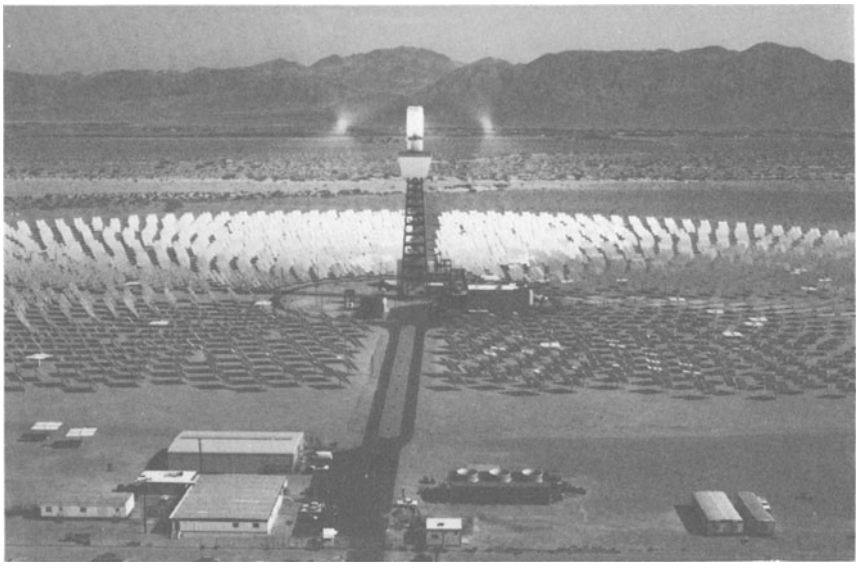


Fig. 8.21: Central receiver power plant at Barstow, California (Courtesy of Sandia Laboratories).

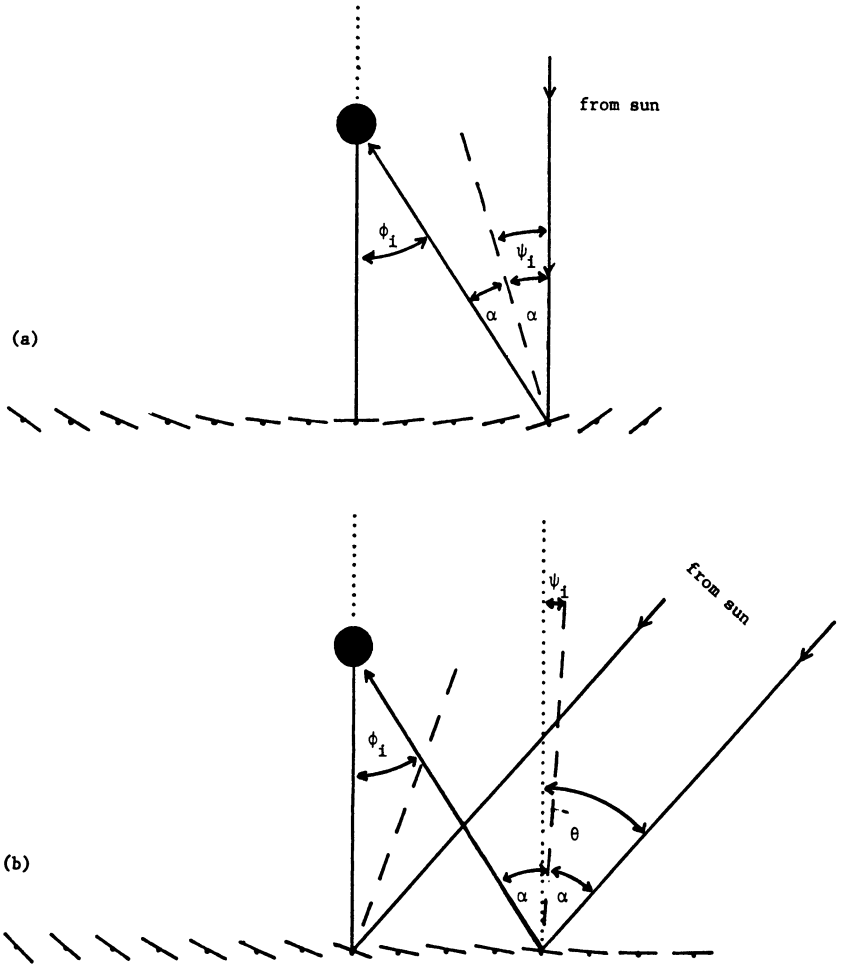


Fig. 8.22: Tracking motion for line focus Fresnel concentrator. **a)** Sun overhead, **b)** sun at incidence angle θ. ψ_i is tracking angle of i'th reflector.

have some curvature. If the number of mirrors is large, the gain in concentration achievable by curving the reflectors is small. When the number of mirrors exceeds on the order of 100 for linear or 10,000 for point focus Fresnel systems, one gets diminishing returns from the extra expense* of curving the mirrors and one may be better off with flat mirrors. On the other hand, when the number of mirrors is small, one can increase the concentration significantly by curving the mirrors.

* The cost of curving mirrors varies greatly with manufcturing technique. It is very easy to impart cylindrical curvature to glass mirrors. Spherical curvature in glass mirrors is more difficult and causes stress. One technique that produces low-cost reflectors of nearly spherical curvature is to stretch a reflective plastic film over a circular frame and apply negative pressure. For an analysis of curved heliostats, the reader is referred to Biggs and Vittitoe[35] or Rabl.[15]

In view of the complexity of the problem it is not surprising that no closed form analytical formulas are available for calculating the optical performance of central receivers. Only some partial solutions have been attempted. For example, Riaz[42] has derived equations for the effects of shading and blocking. Various fast algorithms have been developed for carrying out the computations,[43] but the complete analysis seems to require a computer program. Several computer programs have been developed for this purpose. The HELIOS program[35] is a very accurate and flexible program that can model such details as circumsolar radiation, arbitrary reflector shapes, and second stage concentrators; however, a program like HELIOS is slow and expensive to run. Other examples of computer programs for central receivers are MIRVAL,[44] DELSOL,[45] and SCRAM.[46] The latter make various approximations and are much faster than HELIOS, but they are less accurate and less (or far less) flexible. For more analytical approaches to calculating the optical behavior of central receiver systems, the reader is referred to the work of Riaz[42] and of Vant-Hull and his co-workers (Walzel et al.,[47] Lipps et al.,[48] and Vant-Hull[43]).

A particularly interesting result of these calculations is shown in Tables 8.1a and 8.1b.[49] These two tables show the field efficiency (essentially the incidence angle modifier), $K(\theta_z, \phi_s)$, for particular central receiver designs. For any values of the solar zenith angle, θ_z, and the solar azimuth angle, ϕ_s, the product of beam irradiance and $K(\theta_z, \phi_s)$ yields the radiation intercepted by the receiver. The field efficiency as listed in Tables 8.1a and 8.1b includes (and is limited to) the effects of incidence angle cosine, shading, blocking, mirror reflectivity, tower shadowing, atmospheric attenuation between mirror and receiver, and intercept factor (i.e., loss of rays that miss the receiver). The atmospheric attenuation depends on the distance from mirror to receiver, and is approximately 10% per kilometer. For an accurate model of this attenuation, the reader is referred to Pitman and Vant-Hull.[50] The reflectivity of the mirrors is assumed to be $\rho = 0.90$. The absorptivity of the receiver is not included in these tables. The need for two different tables arises from the difference in mirror field design for small and for large systems. For small systems at intermediate latitudes the heliostats will be deployed only to the north of the tower, whereas in large systems the mirror field will surround the tower (albeit with more mirrors to the north than to the south); this design difference reflects different economies of scale in tower and heliostat costs. Table 8.1a is for a small field of 2.4 MW_t capacity, while Table 8.1b is for a large field of 195 MW_t capacity. Both are optimized for a geographic latitude of 35° and for a specified ratio of tower and heliostat costs. The reader should understand that different designs, as appropriate for different latitudes, cost ratios or system sizes, may have quite different incidence angle modifiers.

A 5 MW_t central receiver test facility has been operated in Albuquerque, New Mexico, since 1978, and a 10 MW_e central receiver electric power plant (using a steam turbine) in Barstow, California has been producing power since 1982. Measurements at the Albuquerque test facility have confirmed the calculations of the HELIOS computer program, and optical errors as small as a few mrad have been achieved.[35]

Table 8.1

Field efficiency for two central receiver designs. (Includes losses due to cosine, tower shadowing, blocking and shading, atmospheric attenuation, intercept factor, and a value $\rho = 0.90$ for the heliostat reflectivity)[49]

Elevation (degrees, horizon = 0°)	Azimuth (degrees, south = 0) 0	30	60	75	90	110	130
5	.384	.404	.366	.330	.300	.240	.212
15	.701	.687	.576	.495	.429	.367	.315
25	.789	.772	.662	.584	.521	.445	.391
45	.814	.814	.757	.708	.661	.603	.544
65	.811	.806	.754	.753	.724	.689	.642
89.5	.723	.729	.748	.726	.730	.736	.736

(a) Small central receiver ($2.4MW_t$)

Elevation (degrees, horizon = 0°)	Azimuth (degrees, south = 0) 0	30	60	75	90	110	130
5	.216	.215	.206	.204	.199	.194	.192
15	.446	.448	.425	.423	.405	.392	.385
25	.560	.558	.537	.522	.516	.498	.491
45	.719	.640	.626	.618	.605	.594	.599
65	.684	.670	.671	.668	.660	.655	.641
89.5	.683	.683	.686	.672	.682	.687	.681

(b) Large central receiver ($195MW_t$)

8.4.4 Fresnel Lenses

Use of an ordinary lens, Fig. 8.23(a), is impractical in most solar applications because the lens would have to be too thick. The mass per aperture ratio is proportional to aperture width and for widths larger than a few centimeters the

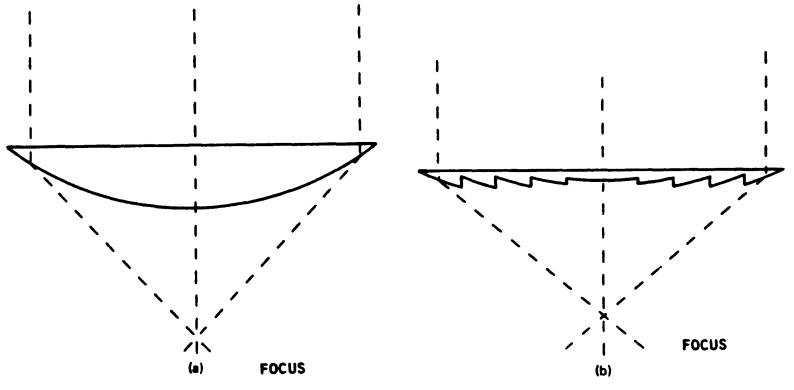


Fig. 8.23: a) Ordinary lens. b) Fresnel lens.

mass and weight become excessive. As an alternative one can use a Fresnel lens, Fig. 8.23(b). The optical performance of the Fresnel lens is almost equivalent* to an ordinary lens.

In principle both sides of a Fresnel lens could be grooved, but in practice one will use only lenses with one smooth surface. Furthermore, the grooves should face down to minimize dirt accumulation. Flat Fresnel lenses with smooth surface towards the sun have poor transmission because of large incidence angles at the outer prisms. Also they have large off-axis aberrations.

Starting with the extreme ray principle for ideal concentrators, Kritchman, Friesem, and Yekutieli[51] found an elegant lens design that avoids these problems, and that comes very close to the thermodynamic limit of concentration. An example of such a lens with an acceptance half-angle, $\theta_a = 10°$, is shown in Fig. 8.24. Included in this figure are the ray paths for two incident directions, $\theta = 0°$ and $\theta = 10°$; of course, the rays coming from $\theta = \theta_a = 10°$ all hit the edge of the absorber. The geometrical concentration of these lenses is at least 75% of the thermodynamic limit.† Like the CPC, they can be truncated with little loss in concentration. That is, the outer portions of the lens do not intercept much radiation, but require a large amount of material per aperture area. In practice it will probably be reasonable to truncate the lens to about half of the full depth; for the lens shown in Fig. 8.24, such a truncation reduces the concentration from $0.75/\sin 10°$ to $0.67/\sin 10°$. If designed with an acceptance angle equal to the solar size, i.e., $\theta_a = \Delta_s = 5$ mrad,

* The principal difference lies in the fact that the optical path length for different rays is different in a Fresnel lens, whereas in a normal lens the optical path from object to image is the same for all rays. This difference is inconsequential for solar energy since imaging and coherence properties are irrelevant.

† It is easy to see why this lens does not quite reach the thermodynamic limit, despite the extreme ray principle used for its design. Some of the rays emitted from the absorber to the lens hit the wrong prism facets and are refracted to leave the lens outside the acceptance angle. In general, a Fresnel lens cannot reach the thermodynamic limit because of the discontinuity in optical path length from one prism facet to another.

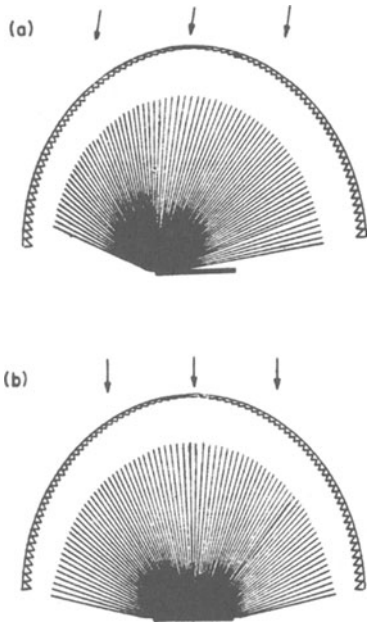


Fig. 8.24: Shape of Fresenel lens with maximal concentration for acceptance half-angle 10°. From Ref. 52.

such a lens reaches a concentration ratio of $0.745/\sin\,\theta_a = 149$ in line focus geometry. Truncated to half of its full depth, its concentration is still equal to 133.

Compared to other Fresnel lenses, the design of Kritchman, Friesem, and Yekutieli has the following advantages:

(1) it reaches the highest geometric concentration possible with a Fresnel lens;

(2) reflection losses are small because the angle between incident ray and first surface is close to the angle between exiting ray and prism surface;

(3) the concave shape adds mechanical strength to the lens; and

(4) the loss of performance for nonzero elevation angles (i.e., when the sun is not in plane of the paper of Fig. 8.23) is less severe than for other linear Fresnel lenses. By the same token it is less sensitive to chromatic aberrations.

The same lens design can be used for point focus systems by rotating the contour of the line focus lens about the optical axis.

Several other Fresnel lens designs have been published recently that come fairly close in performance to the optimal design of Kritchman, Friesem, and Yekutieli. Collares-Pereira[52] approximated the curved lens profile by two straight portions and used a second stage CPC reflector in the focal zone. O'Neill[53] took as a starting point the criterion for minimal reflection loss: equal angles between light rays both at front and back of the lens. Together with the requirement for a smooth front surface, this determines the lens uniquely. Since the design of Kritchman,

Friesem, and Yekutieli yields almost equal angles at front and back between light rays and prism surfaces, it is not surprising that O'Neill's design approaches the design for maximal geometric concentration very closely.

Compared to reflective optics, Fresnel lenses offer several advantages:

(1) Compared to a reflector, a Fresnel lens has dramatically greater tolerance to surface contour errors as shown by O'Neill.*[53]

(2) A Fresnel lens can do double duty as collector cover, protecting the inside not only from heat loss by forced convection, but also from dirt. In a well sealed collector, light rays have to penetrate only once through a layer of dirt. In an exposed reflector, on the other hand, light rays have to penetrate the dirt layer twice, first on the way to the reflecting surface and then again after reflection, because any reflector for outdoor exposure has a sufficiently thick protective coating in front of the aluminum or silver surface itself that incident and reflected rays will be affected by different dust particles.

Despite these advantages, Fresnel lenses do have problems of their own. Above all, the focal length of linear Fresnel lenses shrinks as the elevation of the incident ray from the plane of the paper increases. This effect is so severe that Fresnel lenses with horizontal east-west tracking axes are impractical. Even with polar mount which minimizes the elevation changes to $\pm 23.5°$, use of a Fresnel lens is marginal. In fact, most linear Fresnel lenses that have been built have been intended for installations with seasonal adjustment of the tracking axis or even two-axis tracking. If the tracking axis of a linear Fresnel lens collector is to be fixed, a geometric concentration ratio of about fifteen is the upper limit.[52,54]

Ultimately, the choice between reflective and refractive concentrators will be made in terms of cost per delivered energy. This will depend critically upon cost and durability of mirror and lens materials. Glass is very durable, but as lens material it is ill-suited at least for the grooved surface because the high surface tension and the high softening temperature of glass prevent accurate molding of small prisms. Plastics, especially acrylic, are easy to mold, but their durability is questionable, the main problem being resistance to scratching and to UV degradation. As for the optical transmission through Fresnel lenses, one can expect the following losses for a well designed lens made of acrylic (index of refraction is $n = 1.491$).[53]

* This can be made plausible by considering rays passing through the center of the lens. If the slope of the facet differs from the design direction by an angle δ, then the refracted ray differs from its design direction by $(n-1)\delta$ where n is the index of refraction. In a reflector, by contrast, a contour error δ causes a deviation 2δ of the reflected ray. For typical lens materials the index of refraction is $n \simeq 1.5$ and the sensitivity to contour errors near the center of the lens is a factor four smaller than for reflectors. (The sensitivity to tracking errors is, of course, the same as for reflectors because a tracking error δ is equivalent to an enlargement of the source by δ.)

Loss Type	Fraction of Incident Radiation Lost (%)
Front surface reflection	4.3
Absorption and scattering in acrylic	2.0
Back surface reflection	4.0

In addition, there may be losses up to 5% from the intercept factor because the receiver will usually be made small enough that some rays miss the receiver.

8.4.5 Fixed Reflectors with Tracking Receivers

As an alternative to moving a large reflector, one can design systems where the reflector is fixed and only a small receiver needs to track the sun. Three such systems are known: the hemispherical reflector, the circular cylindrical reflector, and the reflector slats on circular cylindrical mount.

8.4.5.1 Spherical Reflectors. Collimated radiation incident on a spherical reflector will, after one or several reflections, cross the line that extends through the center of this sphere in the direction of the incident direction. Rays close to this line reach the focus which is on this line, half-way between reflector and center of the sphere, as shown in Fig. 8.25(a). Other rays intersect this line between the focus F and the apex A. Therefore, a receiver that extends from focus F to apex A will intercept all rays. Because of the spherical symmetry of the reflector, only the receiver needs to track if the radiation source moves. As a solar collector, such a system can attain geometric concentration ratios up to about 270.[55] The flux concentration along the receiver is quite nonuniform.[56] Even though the focusing is independent of incidence angle, the effective aperture is not. This is illustrated in Fig. 8.25(b) where the reflector portion from R to S is inactive for the incidence angle shown.

8.4.5.2 Circular Cylindrical Reflector with Tracking Receiver. One could use the design principle of the hemispherical reflector for a two-dimensional arrangement as well (i.e., a tracking receiver plate placed along a circular cylindrical reflector trough). However, the concentration ratio of such an arrangement is less than 2 and too low to be practical. However, as the ray trace diagram of Fig. 8.26 shows, a sufficiently large tubular absorber placed slightly below the focal line (i.e., the line half-way between apex and center of the reflector) can intercept most of the reflected rays. This arrangement was investigated by Tabor and Zeimer,[57] who designed and tested a concentrator with geometric concentration ratio 2.5 and a fairly large acceptance angle, $2\theta_a = 17°$. More recently, a group at Lawrence Livermore Laboratory[58] built and tested a similar collector type with inflated cylindrical reflector that needs approximately 12 tilt adjustments per year and that is expected to be suitable for temperatures up to 175°C.

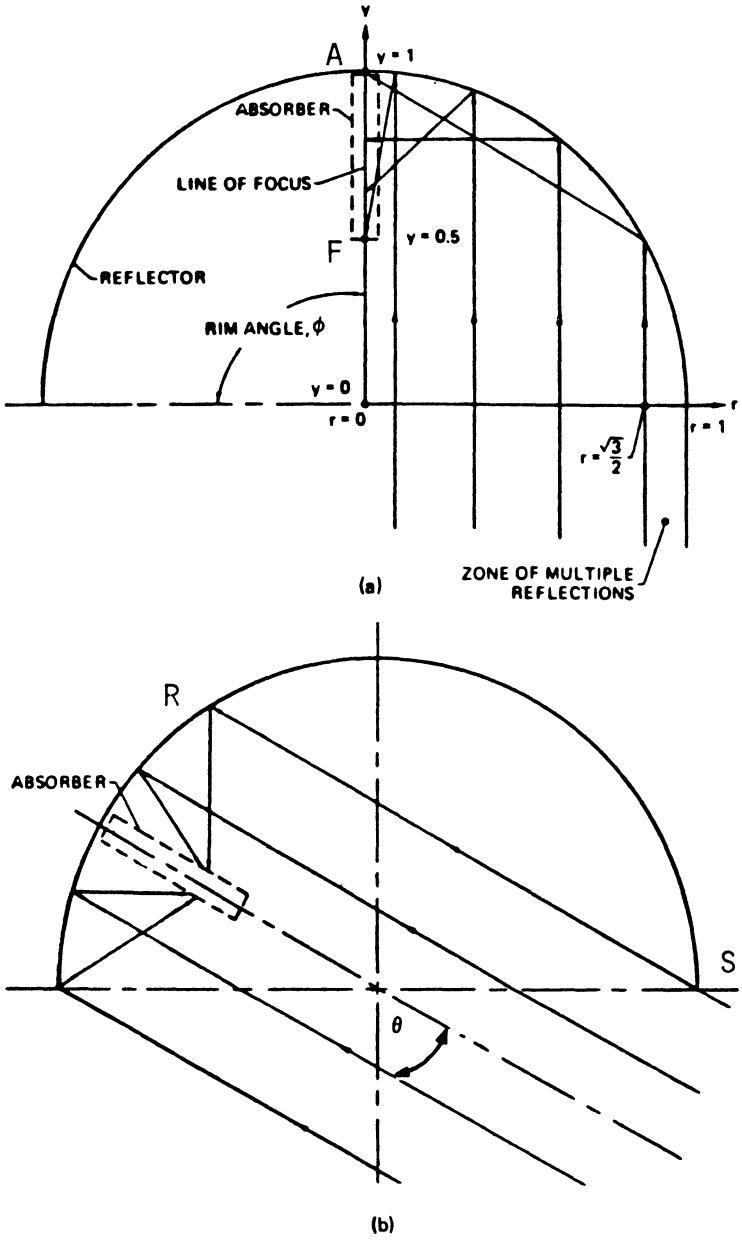


Fig. 8.25: Fixed spherical reflector with tracking receiver. **a)** Normal incidence. **b)** Cosine loss at off-normal incidence.[55]

An interesting improvement of this concept has recently been designed and tested by Edmonds et al.[59] It uses a prismatic reflector[17] as second stage to boost the overall concentration ratio close to the thermodynamic limit; its main application appears to be for photovoltaic cells.

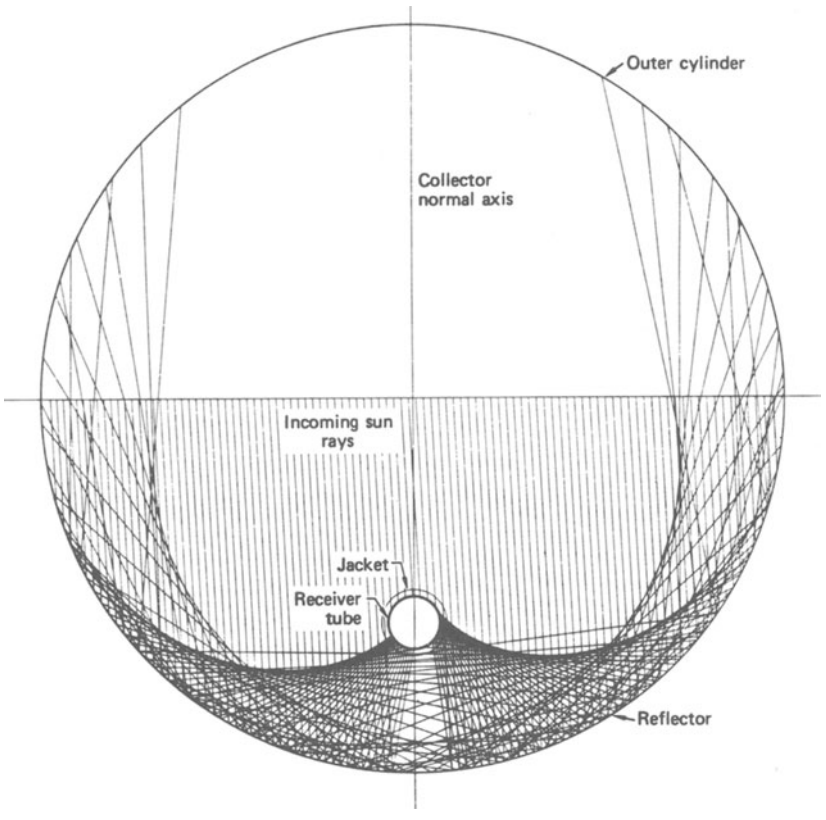


Fig. 8.26: Raytrace diagram for circular reflector with cylindrical absorber. In this example, the incident beam is 3^c off axis (line absorber center to reflector center).

8.4.5.3 Reflector Slats on Circular Cylindrical Mount. The third fixed reflector scheme consists of narrow flat reflector slats that are mounted on a circular cylindrical surface. The cross section is shown in Fig. 8.27. If the mirror slats have the correct slope, this system will always produce a perfect line focus (in the limit of infinitely narrow slats), regardless of incidence angle. The focus is on the same circle as the reflector slats.

Figure 8.27 explains the basic geometry. The solar radiation as shown happens to be incident at an angle of 30° from the vertical, and the rim angles 75° to the north and 45° to the south of vertical (for east-west trough axis) are approximately optimal for deployment at 35° latitude. A particular ray is shown incident on a mirror facet located at an angle, β, from the line to the sun. Simple geometry determines the mirror facet angle at this point to be $\beta/4$ below the horizontal. The angles of the other facets are determined in the same way. Now consider what happens as the sun moves by an angle δ to the right. Then all reflected rays move by an angle δ to the left. The included angle between two rays reflected from two

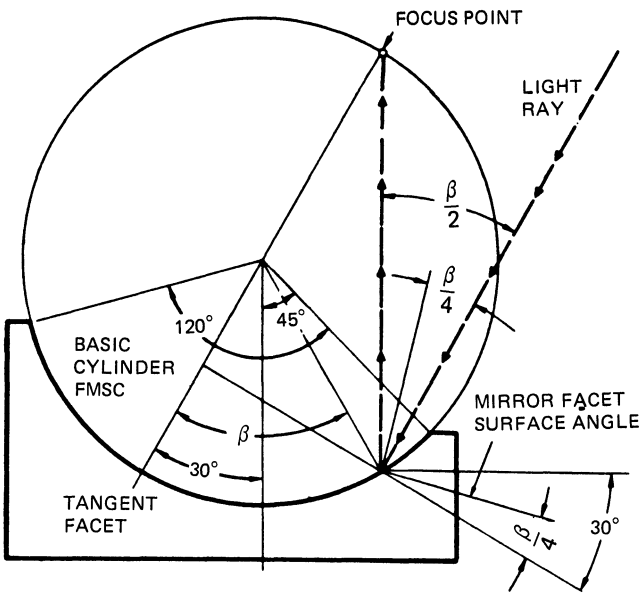


Fig. 8.27: Fixed mirror slats on cylindrical mount with tracking receiver. Basic geometry.

different facets remains constant. According to a well-known theorem,* the focus moves, therefore, on the same circle on which the mirror facets are positioned. In the limit of infinitely narrow slats, the focus is perfect. There are, however, losses due to shading and blocking.[60,61]

8.4.6 Concentrator Configurations for Low Cost Manufacture

Some manufacturing processes naturally lend themselves to the reproduction or approximation of certain reflector geometries. Flat float glass mirrors, for example, can be used to approximate the profile of a CPC. Reflector surfaces of near spherical shape (for example, Brantley[62]) can be obtained by stretching a metallized plastic film over a circular frame and applying air pressure to one side. The profile is not exactly spherical because the fixed frame prevents the film from stretching uniformly in all directions. A very slight air pressure difference, less than 1% of atmospheric pressure, is adequate for maintaining the proper shape, and the surface is close enough to a true paraboloid to yield concentration values of several hundred with two-axis tracking. Air-inflated reflectors may also be suited for some of the fixed reflector systems described in the previous subsection; see for example, Tabor and Zeimer[57] and Gerich.[63]

The catenary or hyperbolic cosine profile can be obtained by suspending metallized plastic film from a frame. A parabolic trough reflector can also be

* For three points A, B, and C on a circle the angle ABC remains constant if the end points A and C are fixed while the apex B moves along the circle.

approximated by applying an appropriate bending moment to the edges of an elastic reflector sheet.[64]

An intriguing technique for producing a low-cost point focus concentrator has been invented by Steenblick at Georgia Institute of Technology.[65] A properly designed spiral pattern is cut from a flat sheet of reflector material. By slightly winding the spiral like a spring and mounting it on a planar surface with appropriate support points and tilt angles, one obtains a point focus reflector. A full-scale model with concentration ratio 500 has been successfully demonstrated. This is really a Fresnel reflector approximation of a parabolic dish. As an alternative manufacturing technique for point focus Fresnel reflectors, one can press a fine circular pattern of grooves into a transparent plastic sheet and cover it with a reflective coating. The possibility of mounting a point focus reflector on a flat support structure is attractive because one can take advantage of the mass production potential of heliostats for central receiver systems and use the basic structure of a heliostat to make point focus collectors that are much smaller than a central receiver system.

A particularly interesting design encloses the concentrator inside an inflated transparent bubble. This approach has been investigated by both Boeing and General Electric. It is suitable for heliostats, as well as for parabolic dishes. This approach has the main advantage of eliminating the windload on the concentrator itself, therefore permitting light-weight, low-cost fabrication. A heliostat inside a bubble could, for instance, be made by stretching an aluminized plastic film over a circular frame. Since it is not practical to make the bubble large enough (i.e., larger than several tens of meters in diameter) to enclose an entire central receiver plant, only individual heliostats will be enclosed and light has to traverse the enclosures twice, at fairly high optical loss. The main difficulty of the bubble approach seems to lie in the stringent requirements for the bubble material. The material must have a high specular transmittance, it must have sufficient mechanical strength, and it must have a long life when exposed to sunlight—all that at very low cost.

Straight reflector cones can be made quite easily by appropriate bending of a flat reflector sheet. The absorber is a narrow cone extending along the symmetry axis. Despite the need for two-axis tracking, the achievable concentration is relatively low, in the range of a few hundred.[66]

8.4.7 Second-Stage Concentrators

In collectors with rim angle less than 90°, the radiation incident on the absorber has a limited angular spread and can, therefore, be further concentrated.

To illustrate the design of second-stage concentrators, we consider the line-focus parabola in Fig. 8.28 with rim angle ϕ (angle between BA and optical axis). Without a second stage, the receiver would be the flat surface from B to B'. A second-stage CPC with acceptance half-angle ϕ boosts the concentration by a factor $1/\sin\ \phi$.

For small rim angles the concentration of the two-stage system approaches the thermodynamic limit, and even for ϕ as large as 30°, it is still within 15%. For large rim angles, the second stage becomes less effective. In a central receiver the

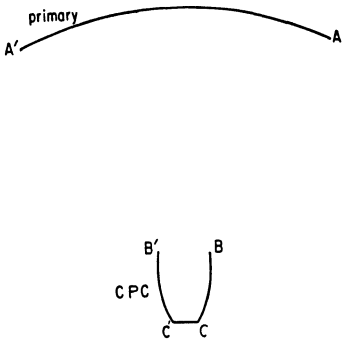


Fig. 8.28: Second-stage CPC concentrator for focusing parabola. Absorber = CC'.

rim angles may be fairly large, about up to 60°, and a single second-stage CPC could not accomplish much; in that case, one would, however, divide the heliostat field and the receiver into several separate sections, each of which has in effect a small rim angle. As for the precise design of the second stage, it should be based on the design of Fig. 8.5 because the source of radiation (i.e., the first stage) is a finite distance away. For a flat absorber the ideal shape turns out to be compound elliptic rather than compound parabolic.

An intriguing alternative second-stage concentrator, called "trumpet" because of its shape, has recently been discovered by Winston and Welford[16] and evaluated in great detail by Kritchman.[67,68] For a flat absorber the trumpet has a hyperbolic profile as shown in Fig. 8.29. The full profile extends all the way to the primary concentrator, and for a parabolic primary reflector that would cause complete shading. In practice, a trumpet would be highly truncated, as suggested by the diference between the dashed line (complete profile) and the solid line (actual profile). For point-focus geometry, the shading loss for a reasonably truncated trumpet is only a few percent of the total incident radiation, while the gains in concentration can be on the order of 2. In choosing between a trumpet and a CPC as second stage, one should consider many factors besides geometric concentration, for example, average number of reflections and absorption loss, problems of cooling the second stage, rejection of skew rays, shading, flux distribution at aperture, size and weights, etc. For a detailed analysis the reader is referred to the papers by Kritchman.[67-69]

Whether one wants to use a second-stage concentrator at all depends also on many factors, in particular on the relation between desired concentration ratio and optical accuracy.

The fact that for a given acceptance angle the concentration reached with a second stage is two to four times as high as for a system without a second-stage concentrator is an obvious advantage for the design of ultrahigh-temperature power plants (using power cycles such as high-temperature gas turbines, magnetohydrodynamics, or thermionic conversion). But it can be just as important for solar collectors of low or intermediate temperature, because the acceptance half-angle,

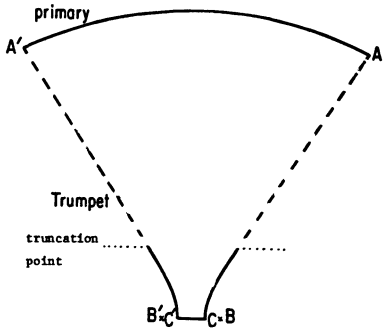


Fig. 8.29: Second-stage trumpet concentrator for focusing parabola. Absorber = CC'.

θ_a, can be doubled for a specified concentration, thus allowing a very significant relaxation of the mirror accuracy. For example, a power tower with effective mirror and tracking error, $\Delta_m = \Delta_s = 4.7$ mrad in its heliostats and with CPC second stage achieves as high a concentration as a power tower without CPC but with perfect mirrors. Contour errors of the second-stage CPC are insignificant as long as they are small compared to θ_a; this is a crucial difference between an imaging Cassegrain system and the nonimaging CPC system.

8.5 PERFORMANCE OF CONCENTRATING COLLECTORS

The value of a solar energy system is determined by its long-term average performance, that is, the useful energy delivered over its lifetime. The long-term performance depends not only on the properties of the collector, but on the characteristics of a specific application (e.g., system configuration, load distribution, operating temperatures, location, and climate). Even the efficiency of the collector itself is not constant; it can vary with collector temperature, ambient temperature, beam and diffuse insolation, incidence angle, and so on. In order to characterize a solar collector in a simple and unambiguous manner, one uses the instantaneous efficiency under standard conditions (clear sky near solar noon, with standard angles of incidence). This is described below. Later, we will present a set of correlations by which one can translate the instantaneous efficiency into the long-term average performance.

8.5.1 Instantaneous Efficiency

The instantaneous efficiency, η, is defined as ratio of the output, $\dot{q}$, over the solar irradiance, I, on the collector aperture (both in units of W/m^2 of aperture area)

$$\eta = \frac{\dot{q}}{I} \tag{8.22}$$

While for flat plates one usually takes I to be the hemispherical irradiance, for collectors with high concentration the beam irradiance is the appropriate choice; it is measured by a pyrheliometer. For nontracking concentrators the contribution of

diffuse insolation can be significant and, hence, it is most appropriate to base I on the irradiance, $I_b + I_d/C$ within the acceptance angle. It does not really matter which convention one chooses for I, as long as one is consistent and states clearly what has been assumed.

The efficiency of photovoltaic collectors is essentially the product of the optical efficiency of the concentrator and the electrical efficiency of the photovoltaic cells. The output, $\dot{q}$ (in W)

$$\dot{Q} = Aq \tag{8.23}$$

of a thermal collector of aperture area, A, is related to the temperature rise, $(T_{out} - T_{in})$, of the heat transfer fluid flowing through the collector

$$\dot{Q} = \dot{m}\, c(T_{out} - T_{in}) \tag{8.24}$$

where $\dot{m}$ is the flow rate (in kg/sec) and c is the specific heat (in kJ/kg °C). If the optical efficiency of the collector is η_0 and the heat loss is q_{loss} (in W/m^2 of aperture area), then the output $\dot{q}$ is

$$\dot{q} = \eta_0 I - \dot{q}_{loss} \tag{8.25}$$

It is convenient to define a heat loss coefficient or U-value by dividing $\dot{q}_{loss}$ by the temperature difference between absorber and ambient

$$U = \frac{\dot{q}_{loss}}{T_{abs} - T_{amb}} \tag{8.26}$$

where T_{abs} is the mean temperature of the absorber surface and T_{amb} is the ambient temperature. The U-value depends only weakly on temperature, and in many cases one gets a good approximation if one chooses a constant U-value corresponding to typical operating conditions. The resulting linearized efficiency equation

$$\eta = \eta_0 - U\left(\frac{T_{abs} - T_{amb}}{I}\right) \tag{8.27}$$

is extremely useful because it characterizes the performance of a solar collector in terms of only two parameters: η_0 and U. If the efficiency is plotted against the variable $(T_{abs} - T_{amb})/I$, one obtains a straight line with intercept η_0 and slope $-U$. Strictly speaking, the U-value is not constant, but depends on absorber temperature, air temperature, sky temperature, wind velocity, and collector tilt. In practice, most of these variables have only a small effect. However, the variation of U with absorber temperature can be quite important, especially in high temperature collectors. In that case the efficiency depends on T_{abs}, T_{amb}, and I separately. To account for the temperature dependence of U, it is usually sufficient to expand with two constant coefficients, U_0 and U_1, as

$$U = U_0 + U_1(T_{abs} - T_{amb}) \tag{8.28}$$

Cooper and Dunkle[70] have examined this approach in full detail and provided a procedure for approximating test results which are stated in terms of Eq. (8.28) by

a linearized efficiency equation, Eq. (8.27), which is the best approximation for a specified set of operating conditions; the linearized equation has the advantage of being far more convenient for system performance calculations.

The above equations are not directly suitable for practical applications because, in practice, one cares about the temperature(s) of the heat transfer fluid, not the absorber temperature. Fortunately, only a simple multiplicative factor is needed to bridge this gap. By a well-known argument (see, for example, Kreith and Kreider[71]) one can show that in terms of the mean fluid temperature

$$T_m = \frac{T_{in} + T_{out}}{2} \tag{8.29}$$

the efficiency equation takes the form

$$\eta = F_m \left(\eta_0 - U \left(\frac{T_m - T_{amb}}{I} \right) \right) \tag{8.30}$$

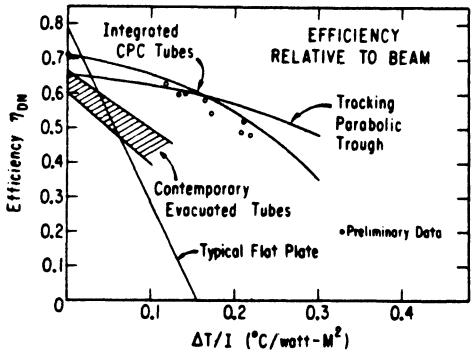
F_m is a collector parameter which accounts for the heat transfer from the absorber surface to the fluid. F_m has been called collector efficiency factor or heat transfer factor and is also known by the symbol F'. It depends on the collector construction, but is nearly independent of operating conditions. For flat plate collectors, typical values are in the range of 0.8 to 0.9 for nonevacuated air collectors, 0.9 to 0.95 for nonevacuated liquid collectors, and 0.95 to 1 for evacuated tube collectors. There is no need to determine F_m separately since only the products $(F_m \eta_0)$ and $(F_m U)$ are ever needed in practice.

Efficiency curves for several collectors are plotted in Fig. 8.30: a parabolic trough, an evacuated fixed tilt CPC of concentration 1.5 X (the design of Fig. 8.8(b) with shaped glass tube), a typical range of commercially available evacuated tubular collectors and a typical flat plate.

One fairly important item that has not yet been mentioned is the incidence angle modifier. As the incidence angle increases, the transmissivity of glazings and the absorptivity of absorber coatings decrease. In some collectors the intercept factor also decreases with incidence angle. All of this lowers the optical efficiency. The easiest way to account for this effect in practice is to include an incidence angle modifier, $K(\theta)$, as multiplicative factor with the optical efficiency, where θ is the angle of incidence. In tubular collectors one has to distinguish two directions for the incidence angle modifier: parallel and perpendicular to the tubes. Incidence angle modifiers for typical flat plates and evacuated tubes are shown in Fig. 8.31. Data for incidence angle modifiers of several parabolic troughs (in direction parallel to absorber) are plotted in Fig. 8.32.

For some applications it is more convenient to specify the fluid inlet temperature T_{in} than T_m. In terms of T_{in}, the efficiency equation reads

$$\eta = F_{in} \left(\eta_0 - U \left(\frac{T_{in} - T_{amb}}{I} \right) \right) \tag{8.31}$$



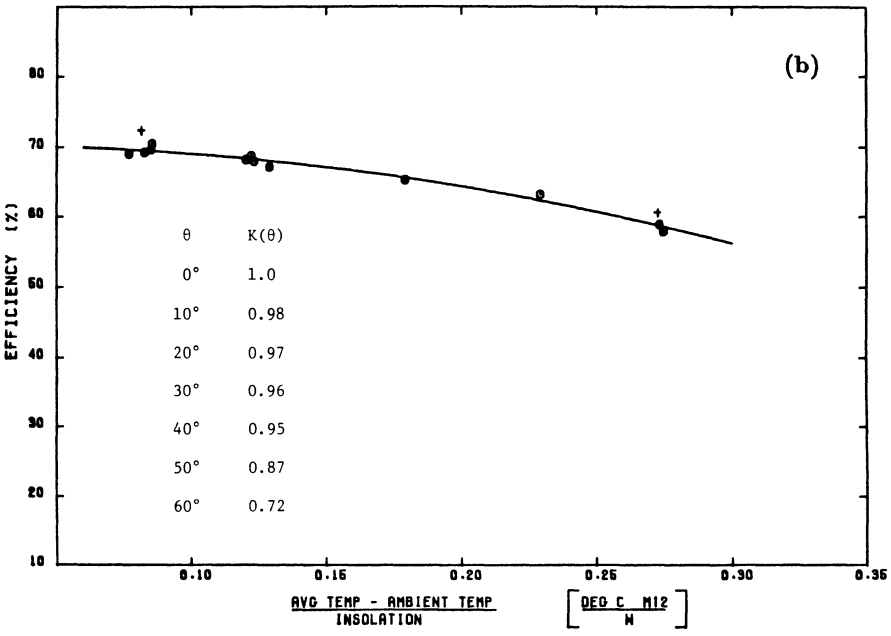


Fig. 8.30: a) Instantaneous efficiency curve for several collectors, plotted versus beam irradiance I_b. "Integrated CPC Tubes" refers to nontracking fixed-tilt collector with evacuated glass tubes shaped as 1.5X CPC. Data are measured with 0.94 m long prototype, curve is projection for 1.83 m long commercial module.[72] b) Efficiency and incidence angle modifier of advanced parabolic trough.[73]

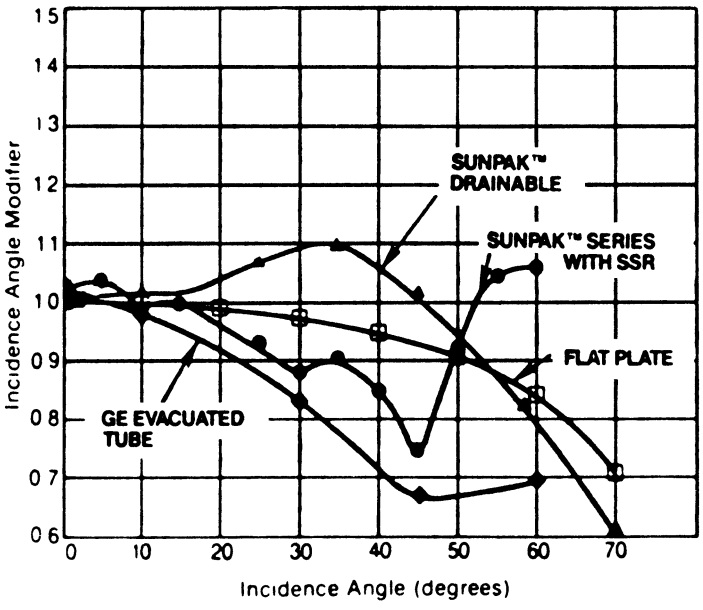


Fig. 8.31: Incidence angle modifier for several evacuated-tubular collectors (incidence angle in direction perpendicular to tubes).[74]

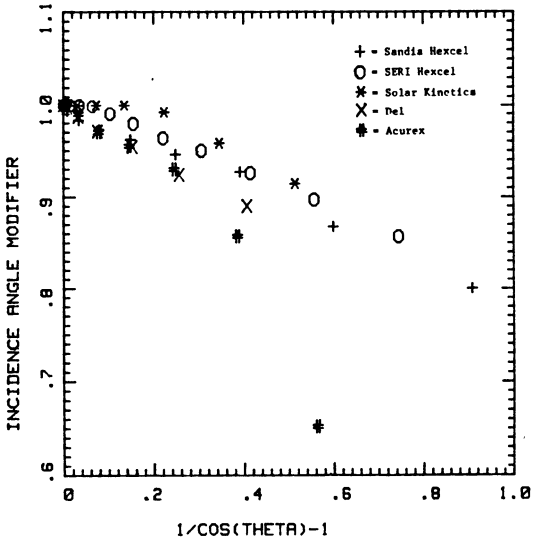


Fig. 8.32: Incidence angle modifier for several parabolic troughs, versus $1/\cos\,\theta - 1$.[75]

where F_{in} is called the heat removal factor, defined by the equation

$$F_{in} = \frac{\dot{m}c}{UA}\left(1 - \exp\left(-\frac{UAF_m}{\dot{m}c}\right)\right) \tag{8.32}$$

Standard thermal test procedures, such as ASHRAE No. 99-77,[76] determine only the instantaneous or average thermal efficiency of a single collector module. In practical applications one needs to evaluate the performance of a collector field or a collector-heat exchanger system. Losses from transfer lines between individual collectors and between the field and the point of end use can be accounted for by modifying η_0 and U, as shown by Beckman.[77] The effect of a heat exchanger can be treated by a similar method, as shown by de Winter.[78]

8.5.2 Long-Term Average Performance

Prediction of long-term average system performance can be quite complicated, and frequently computer simulations are considered necessary. For the most important standard applications, such as water heating, space heating, and industrial process heat, shorthand procedures are also available that are simple enough to permit calculation of long-term average performance by hand or programmable calculator. The correlations presented in this section predict the long-term average performance for any system where the average collector temperature is known. This procedure requires only the reading of a single graph and multiplication of the coordinates by the optical and thermal parameters of the collector field. Alternatively, one can evaluate the corresponding polynomial curve fit. This simple procedure reproduces the results of hourly computer calculations with an accuracy (rms errors) on the order of 2% for flat plates and 2% to 4% for concentrators. Only three variables are needed: the operating threshold (for thermal collectors, this is the average heat loss divided by the optical efficiency), the geographical latitude, and the yearly average direct normal insolation. This remarkable simplification is possible because, to a solar collector, all climates appear very much alike and differ only in their average normal insolation.

The procedure is based on a correlation between yearly energy, Q, delivered by a collector and the threshold, X.[79] For thermal collectors, the threshold is the ratio of heat loss and optical efficiency

$$X = \frac{\dot{q}_{loss}}{\eta_0} \tag{8.33}$$

and Q is given in the form $Q/(AF\ \eta_0)$ where F is equal to unity, to F_m, or to F_{in}, depending on the temperature base (T_{abs}, T_m, or T_{in}) chosen for the collector efficiency equation. The correlation is

$$\frac{Q}{AF\eta_0} = q(x) = q_0 - q_1\ X + q_2\ X^2 \tag{8.34}$$

where q_0, q_1, and q_2 are second-order polynomials in latitude and, in the daytime-average direct normal solar irradiation, $\bar{I}_b$. This correlation is plotted as a function of X in Figs. 8.33 to 8.39 for each of the principal collector types and deployment modes. For photovoltaic systems, one uses these correlations with $X = 0$ and $F = 1$. For convenience, the results of $X = 0$ have been plottetd as q_0 versus average direct normal irradiance, $\bar{I}_b$, in Fig. 8.40. On the upper border of this graph, some typical locations for these insolation values are indicated. Two brief examples will illustrate this method.

First consider a parabolic trough reflector with photovoltaic cell as an absorber with the following conditions:

reflectivity of reflector $\rho = 0.90$

intercept factor $\gamma = 0.95$

electric efficiency of cell $\eta_c = 0.14$

tracking axis horizontal east–west
location Albuquerque with $\bar{I}_b = 0.60 \text{kW/m}^2$

Solution: the optical efficiency is

$$\eta_0 = \rho\gamma = 0.855 \tag{8.35}$$

From Fig. 8.40 we find $q_0 = 6.8$ GJ/m^2 for the east-west trough in Albquerque. Hence, the annual electricity production is

$$q_e = \eta_e \, \eta_0 \, q_0 = 0.81 \text{ GJ/m2} \tag{8.36}$$

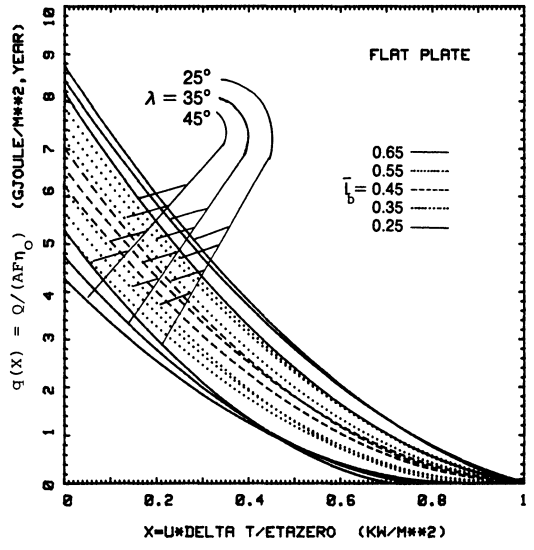
To illustrate the correlations for thermal collectors, consider a parabolic trough with the parameters

$$F_m\eta_0 = 0.65 \text{ and } F_mU = 0.67 \text{ W/m}^2 \text{ °C} \tag{8.37}$$

mounted with horizontal east-west axis and operating in Albuquerque at a fluid temperature $T_m = 300°$ C while the annual average ambient temperature is $T_{amb} = 13°$ C. Solution: the threshold is

$$X = 0.67\left(\frac{300 - 13}{0.65 \text{ W/m}^2}\right) = 0.294 \text{ kW/m}^2 \tag{8.38}$$

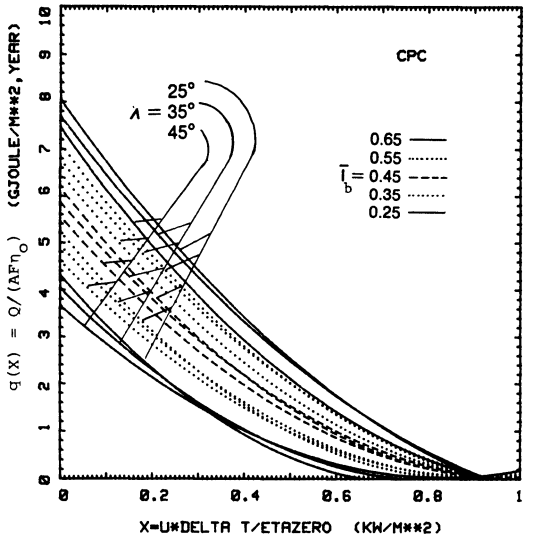
Interpolation between the $\bar{I}_b = 0.55$ and 0.65 curves in Fig. 8.35 yields $Q/(AF_m\eta_0) = 3.65$ GJ/m^2. Hence, this collector can deliver $Q/A = 2.37$ GJ/m^2 per year. To illustrate the effect of dirt and degradation, let us suppose that the optical efficiency is reduced by 10% to $F_m \, \eta_0 = 0.585$, averaged over the lifetime of the collector, while the U-value remains constant. Then the threshold increases to 0.327 kW/m^2 and the yearly output is only 1.96 GJ/m^2.



q(X) = Q/(AFη$_o$) =

[(+5.215 +6.973*$\bar{I}_b$)+(-5.412 +4.293*$\bar{I}_b$)*L+(+1.403 -.899*$\bar{I}_b$)*L**2]
+[(-18.596 -5.931*$\bar{I}_b$)+(+15.468 +18.845*$\bar{I}_b$)*L+(-.164 -35.510*$\bar{I}_b$)*L**2]*X
+[(+14.601 -3.570*$\bar{I}_b$)+(-13.675 -15.549*$\bar{I}_b$)*L+(+1.620 +30.564*$\bar{I}_b$)*L**2]*X**2

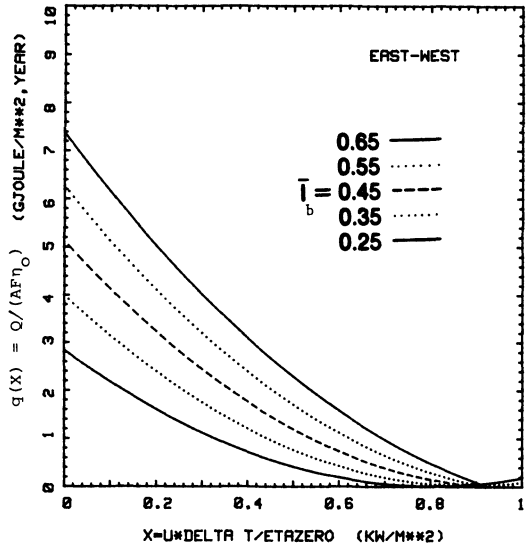
Fig. 8.33: Yearly collectible energy for flat plate at tilt equal latitude; shown as graph and as curve fit for $Q/(AF\eta_0)$, as function of threshold X (in kW/m^2), of average direct normal irradiance $\bar{I}_b$ (in kW/m^2), and of latitude (λ in degrees, L in radians).



q(X) = Q/(AFη$_o$) =

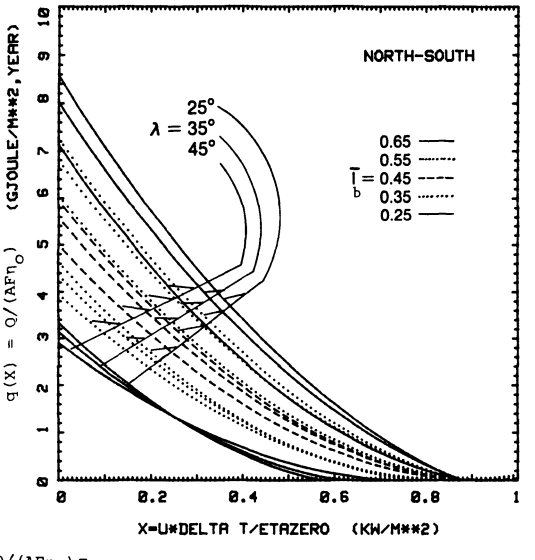
[(+1.738 +11.758*$\bar{I}_b$)+(+1.990 -8.875*$\bar{I}_b$)*L+(-3.236 +7.617*$\bar{I}_b$)*L**2]
+[(-13.240 -14.688*$\bar{I}_b$)+(+3.979 +43.653*$\bar{I}_b$)*L+(+7.345 -52.556*$\bar{I}_b$)*L**2]*X
+[(+14.015 -1.437*$\bar{I}_b$)+(-11.884 -25.852*$\bar{I}_b$)*L+(-.079 +39.538*$\bar{I}_b$)*L**2]*X**2

Fig. 8.34: Yearly collectible energy for CPC, shown as graph and as curve fit for $Q/(AF\eta_0)$, as function of threshold X (in kW/m^2), of average direct normal irradiance $\bar{I}_b$ (in kW/m^2), and of latitude (λ in degrees, L in radians).



$$q(X) = Q/(AF\eta_o) = \quad (\quad -.098 \quad +11.944*\bar{I}_b \quad -.657*\bar{I}_b{}^{**2})$$
$$+ (\quad -.599 \quad -30.363*\bar{I}_b \quad +17.788*\bar{I}_b{}^{**2})*X$$
$$+ (\quad +1.093 \quad +17.606*\bar{I}_b \quad -17.290*\bar{I}_b{}^{**2})*X**2$$

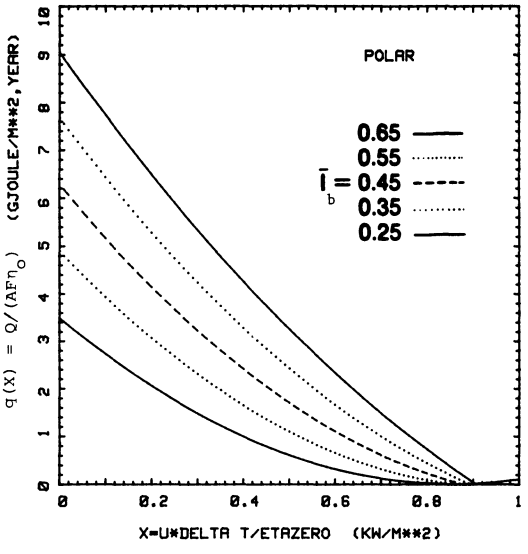
Fig. 8.35: Yearly collectible energy for concentrator tracking about east-west axis, shown as graph and as curve fit for $Q/(AF\eta_0)$ as function of threshold X (in kW/m^2), and of average direct normal irradiance $\bar{I}_b$ (in kW/m^2).



$$q(X) = Q/(AF\eta_o) =$$

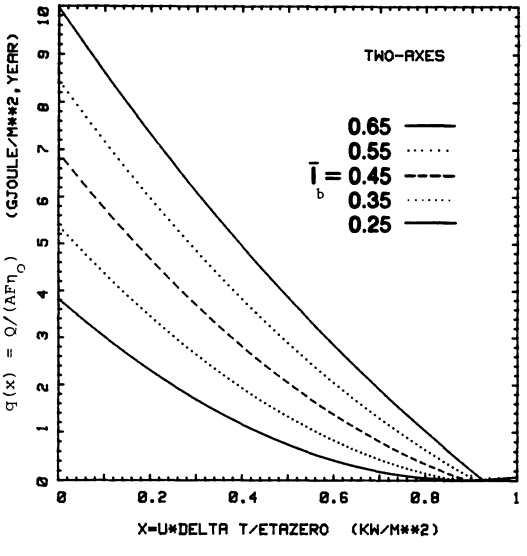
$$[(\quad +.640 \quad +11.981*\bar{I}_b)+(\quad -2.365 \quad +7.979*\bar{I}_b)*L+(\quad +2.380 \quad -12.409*\bar{I}_b)*L**2]$$
$$+[(\quad -6.021 \quad -19.086*\bar{I}_b)+(\quad -4.592 \quad +20.298*\bar{I}_b)*L+(\quad +10.570 \quad -22.978*\bar{I}_b)*L**2]*X$$
$$+[(\quad +6.440 \quad +5.219*\bar{I}_b)+(\quad +6.986 \quad -30.500*\bar{I}_b)*L+(\quad -14.095 \quad +40.089*\bar{I}_b)*L**2]*X**2$$

Fig. 8.36: Yearly collectible energy for concentrator tracking about horizontal north-south axis, shown as graph and as curve fit for $Q/(AF\eta_0)$ as function of threshold X (in kW/m^2), of average direct normal irradiance $\bar{I}_b$ (in kW/m^2), and of latitude (λ in degrees, L in radians).



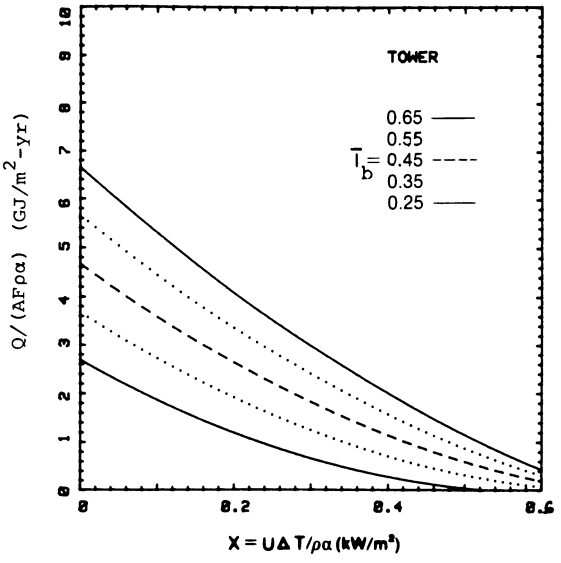
$$q(X) = Q/(AF\eta_o) = (-.075 + 14.432*\bar{I}_b - .592*\bar{I}_b^{**2}) + (-.780 - 35.670*\bar{I}_b + 24.520*\bar{I}_b^{**2})*X + (+1.373 + 19.604*\bar{I}_b - 23.965*\bar{I}_b^{**2})*X^{**2}$$

Fig. 8.37: Yearly collectible energy for concentrator tracking about polar axis, shown as graph and as curve fit for $Q/(AF\eta_0)$, as function of threshold X (in kW/m^2) and of average direct normal irradiance $\bar{I}_b$ (in kW/m^2).



$$q(X) = Q/(AF\eta_o) = (-.147 + 16.084*\bar{I}_b - .792*\bar{I}_b^{**2}) + (-.886 - 37.659*\bar{I}_b + 26.983*\bar{I}_b^{**2})*X + (+1.700 + 18.883*\bar{I}_b - 24.887*\bar{I}_b^{**2})*X^{**2}$$

Fig. 8.38: Yearly collectible energy for concentrator with two-axis tracking, shown as graph and as curve fit for $Q/(AF\eta_0)$, as function of threshold X (in kW/m^2), and of average direct normal irradiance $\bar{I}_b$ (in kW/m^2).



$$q(X) = Q/(AF\rho\alpha) = 0.206 + 9.916\,\bar{I}_b + (-5.715 - 12.968\,\bar{I}_b)X + (7.956 - 2.585\,\bar{I}_b)X^2$$

Fig. 8.39: Yearly collectible energy for central receiver, shown as graph and as curve fit for $Q/(AF\rho\alpha)$, as a function of threshold X (in kW/m^2), and of average direct normal irradiance $\bar{I}_b$ (in kW/m^2). This correlation is based on the particular design with incidence angle modifier of Table 15b (195 MW$_b$ plant at 35° latitude). For the incidence angle modifier of Table 15a, the collectible energy would be significantly higher.[79]

For a direct comparison between several typical collectors, Fig. 8.41 shows the yearly average energy delivery, Q/A, versus operating temperature (as ΔT = difference between mean fluid and ambient temperatures) at a latitude of 35° for two climates: (a) a sunny climate with average beam irradiance $\bar{I}_b = 0.6$ kW/m^2 and (b) a fairly cloudy one with $\bar{I}_b = 0.3$ kW/m^2. The collector parameters, listed in Table 8.2, are typical values based on standard outdoor tests. One striking conclusion from this comparison concerns the crossover between flat-plate and concentrating collectors. At temperatures more than approximately 25° C above ambient, all of the concentrating collectors in Fig. 8.41 surpass the flat plate in performance. This conclusion holds even for relatively cloudy climates where concentrating collectors are at a disadvantage because they miss most of the diffuse insolation. But the high heat loss of flat-plate collectors turns out to be more important; at times when the insolation is above the threshold for a flat-plate collector (with $\Delta T \gtrsim 25$°C), there is enough direct radiation to run a concentrating collector. Of course, energy delivery is only the denominator of the cost per energy ratio. The best choice of a collector depends as much on its capital and operating costs as on its thermal performance.

To achieve a compact presentation of the results, it has been necessary to

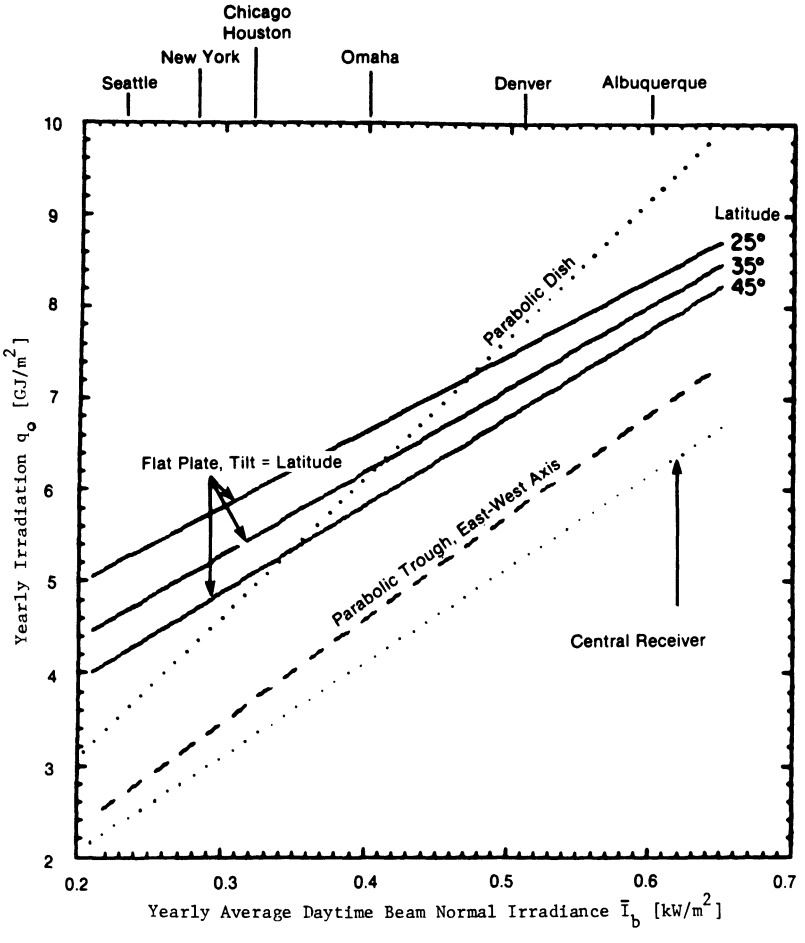


Fig. 8.40: Yearly irradiation q_o (GJ/m^2) incident on aperture of principle collector types. These correlations include angle modifiers. (Curve for central receiver is based on one specific design at 35° latitude only and includes intercept factor.) This graph shows only radiation availability, not system output.

make certain standardizing assumptions:[79]

- the collector uses all solar radiation above a specified threshold, X, and the portion of the solar radiation that is above X is used with constant efficiency.
- Transient effects are neglected. This is of little consequence if the time constant of the collector field is significantly less than one hour.
- Flat plate and CPC are deployed at tilt equal to latitude.
- Standard spacing between concentrator modules is assumed to account for shading. (No shading for flat plate and CPC; ground cover ratio is equal to 0.5 for one-axis and 0.25 for two-axis tracking.)
- Standard incidence angle modifiers are assumed.

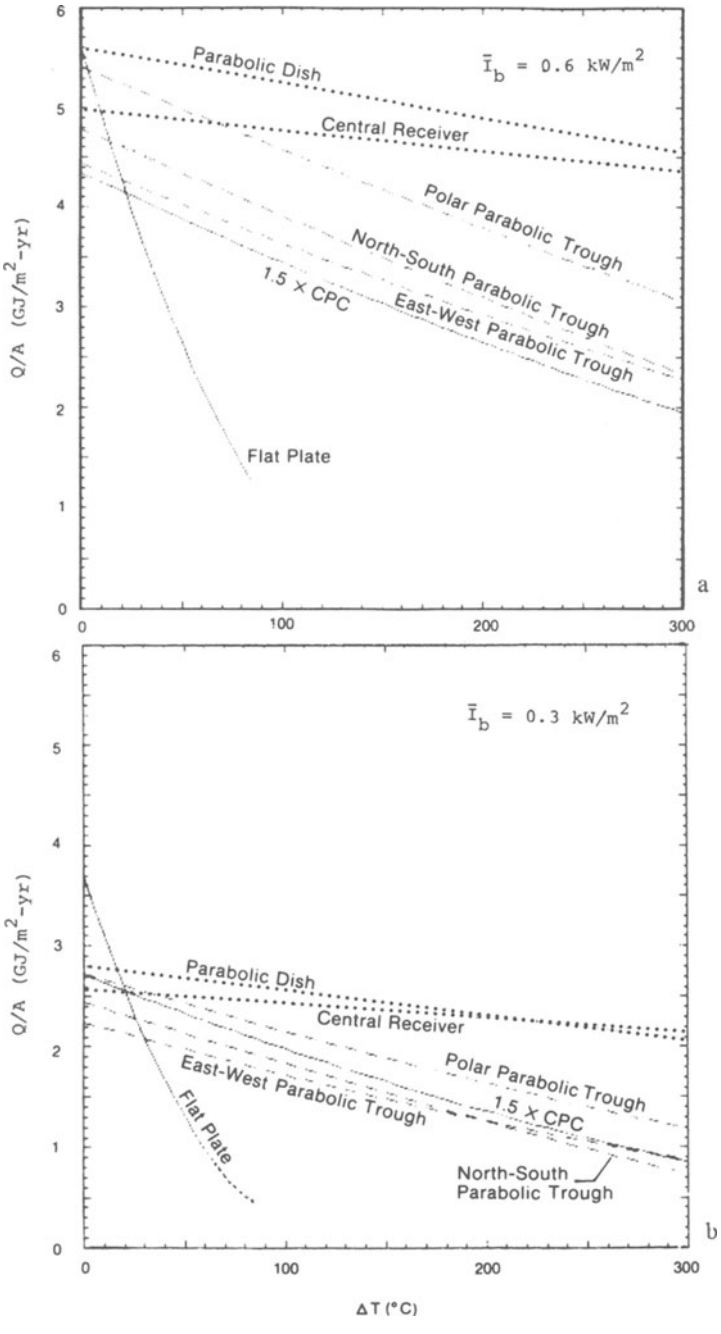


Fig. 8.41: Yearly collectible energy versus average difference between fluid temperature and ambient for typical collectors (for collector parameters see Table 8.2):
a) $\bar{I}_b = 0.6$ kW/m^2. **b)** $\bar{I}_b = 0.3$ kW/m^2.

Table 8.2

Typical collector parameters assumed[40]

Collector Type	$F_m\eta_o$	F_mU[W/m^2°C]
Flat plate, double glazed[a]	0.70[b]	5.0
CPC 1.5X with evacuated tubes	0.60[c]	0.75[c]
Parabolic trough[d]	0.65[e]	0.67
Parabolic dish[d]	0.61[e]	0.27
Power tower[f]	$\rho\alpha = 0.81$[g]	0.16

[a] Average of 4 best double-glazed collectors, Florida Solar Energy Center, Summary Test Package, issued 15 Feb. 1978, and Supplements 1 (April 1978) and 2 (June 1978).

[b] With respect to radiation within acceptance angle.

[c] With respect to radiation within acceptance angle.

[d] Typical test results from Dudley and Workhoven.[80]

[e] With respect to pyrheliometer.

[f] Projected performance.

[g] ρ = heliostat reflectance.
α = receiver absorptance.

These assumptions are realistic for a wide range of applications. This method can be applied when the collector system operates year-round in such a way that no collected energy is discarded and, in the case of thermal collectors, where the average threshold is known. This includes

- photovoltaic systems;
- solar-augmented industrial process heat systems; and
- solar thermal power systems.

In addition, the method is useful for rating collectors of different types or different manufacturers on the basis of yearly average performance. Such a rating provides a convenient criterion for selecting the most cost-effective collector. The method can also serve for evaluating the effects of collector degradation, the benefits of collector cleaning, and the gains from collector improvements due to measures such as enhanced optical efficiency or decreased heat loss per absorber surface. For many of these applications, the method is accurate enough to replace a detailed system simulation.

8.6 PRACTICAL CONSIDERATIONS

8.6.1 Absorber Coatings

Several absorber coatings are listed in Table 8.3. The performance benefit of selective coatings is greatest in evacuated collectors; in fact, if one goes to the trouble of evacuating a collector, then one should also use a selective coating with a low emissivity. In nonevacuated collectors, convection begins to dominate if ε is reduced below a certain value (0.2 to 0.5 depending on temperatures and absorber design).

The optimal choice of an absorber coating depends not only on collector efficiency, but also on cost and durability. Generally, paints are cheaper and easier to apply than coatings deposited by chemical or electric processes. As Table 8.3 shows, there are quite a few coatings that can be used up to about 200°C, but at high temperatures durability becomes a problem. Black chrome, a favorite because of performance and durability, begins to degrade about 350°C. For temperature stability, the Solkote coating looks particularly impressive: it has $\alpha = 0.95$, $\varepsilon = 0.45$, and a temperature limit of 870°C. The higher the operating temperature, the more difficult it is to achieve selectivity. Not only does material degradation become a more serious problem at high temperatures, but the emissivity increases; furthermore, the overlap between solar in thermal radiation becomes greater. Some sophisticated multilayer coatings have been developed that maintain high selectivity up to 500°C, in particular, a coating by Seraphin and Carver[83] with $\alpha = 0.91$ and $\varepsilon = 0.11$ at 500°C. Above 500°C no coatings with high selectivity are available.

8.6.2 Glazing

Many concentrating collectors have a transparent cover in front of the absorber in order to reduce heat losses. In practice, the ideal glazing has high transmittance for solar radiation, low transmittance for thermal radiation from the absorber, high durability, and low cost. The properties of the most common glazing materials are listed in Table 8.4.

Optical losses of the glazing are due to reflection at the glazing surfaces and due to absorption inside the glazing material. The reflection loss depends on incidence angle and on index of refraction and is given by Fresnel's equation. Almost all glazing materials for solar collectors have an index of approximately 1.5 and the corresponding reflective loss at normal incidence is 4% per surface or 8% per glazing. Teflon has the lowest index of refraction, 1.34, and the highest solar transmittance, 96%.

Absorption losses depend on extinction coefficient and thickness according to Bouguer's law. Ordinary window glass has undesirably high absorption losses, because of its iron content. Low-iron glass has almost no absorption loss, but it is more expensive. Actually, it is not so much the iron content itself that matters as the nature of the iron bond since Fe^{+2} ions absorb much more strongly than Fe^{+3} ions. Hence, one can cut absorption losses by reducing the concentration of Fe^{+2} ions rather than eliminating iron altogether. For example, Corning 0317

Table 8.3

Properties of absorber coatings

Material	Type	Supplier/ Developer	Absorptance	Emittance	Temperature Stability (°C)	Comments
Pyromark	paint	Tempil	0.95	0.85(500°C)	750	(1)
Enersorb	paint	Desoto	0.96	0.92	not available	(1)
PbS pigment with Silicon	paint	experimental	0.90	0.37	350	Expected life less than 2 yrs.; may require repainting of local hot spots (2, 3)
Copper–chromium oxide pigment, ethylene–propylene diene monomer binder	paint	Harshaw Chemical (pigment) Exxon (binder)	0.92–0.94	0.30–0.45	not available	Durability expected to be better than acrylic paint; estimated life less than 5 yrs. (2)
Solkote Hi/Sorb–I	paint	Solar Energy Corporation	0.95	0.45	870	Manufacturer's data for lab application on Al
Black chrome	electro-deposited	many	0.94–0.96	0.20–0.25(300°C)	300	Estimated life greater than 30 yrs. (1,2)
Black nickel (Tabor black)	electro-deposited	Miromit	0.91	0.14	not	(1)
Black iron	chemically deposited	Ethone, Inc.	0.84–0.90	0.07–0.35	300–370	Estimated life greater than 10 yrs.; sensitive to humidity (2)
Black copper	chemically deposited	Ethone, Inc.	0.88–0.91	0.12–0.20	190	Guaranteed to be 80% effective after 5 yrs.; sensitive to humidity (2)
Alcoa black	chemically deposited	Alcoa	0.90	0.30–0.40	175	Estimated life greater than 10 yrs.; coating destroyed by water (2)
Silicon nitride over molybdenum	chemical vapor deposited	Seraphin and Carver	0.91	0.11	500	Withstood 1000 hours at 500°C, 1 torr vacuum (4)

(1)Call[97] (2)Vresk[79] (3)Kreith and Kreider[72] (4)Seraphin and Carver[77]

Table 8.4

Properties of glazing materials adapted from Butler and Claassen[98] and Jorgenson[60]

Material	Index of Refraction	Short-wave Transmittance (0.4–2.5μ)	Long-wave Transmittance (2.5–4.0μ)	Density (kg/m^3)	Thickness (mm)	Maximum Temperature (°C)	Cost (in 1979 \$, \$/m^2)
Soda Lime Float Glass	1.52	0.84	0.01	2500	3.175	232	5.4
Water White Low-Iron Glass	1.52	0.90	0.01	2473	3.175	204	10.6
Fiberglass Reinforced Polyester (Sunlite)	1.54	0.87	0.076	1399	0.635	NA	NA
Acrylic (Plexiglass)	1.49	0.90	0.02	1189	3.175	93	9.0–13.0
Polycarbonate (Lexan)	1.586	0.84	0.02	1199	3.175	121	24.0–31.0
Polytetrafluoroethylene (Teflon)	1.343	0.96	0.256	2148	.0508	NA	NA
Polyvinyl Fluoride (Tedlar)	1.46	0.92	0.207	1379	.1016	66	0.60
Polyester (Mylar)	1.64	0.87	0.178	1394	.1270	NA	NA
Polyvinylidene Fluoride (Kynar)	1.413	0.93	0.230	1770	.1016	NA	NA
Polyethylene (Marlex)	1.500	0.92	0.810	910	.1016	NA	NA

fusion glass has good transmittance despite high iron content.[86] Many plastics have some absorptivity for the infrared portion of the solar spectrum. Borosilicate glass is excellent: it has high transmissivity, strength, and temperature stability. It is best suited for use in focusing collectors where its higher cost does not matter as much.

Two comments must be added to the solar transmittance values in Table 8.4. First, these values refer to normal incidence; at larger incidence angles the transmittance is smaller. Second, these transmittance values are for total, as opposed to specular, transmittance. In a flat-plate collector only the total transmittance is of interest since the scattered rays also reach the absorber. Even textured glass can be used. On the other hand, for a cover of a focusing collector (e.g., an inflated plastic dome enclosure for lightweight heliostats or parabolic dishes) the specular transmittance is relevant, and it may be significantly lower than the values in Table 8.4. Surface scratching has little effect on total transmittance, but it can seriously reduce the specular transmittance.

The infrared transmittance affects radiative heat losses from the absorber. Glass and thick layers of plastic transmit little or no thermal infrared. The thin plastic films are far less effective in blocking infrared. The importance of that point depends on collector design and application, in particular on the emissivity of the absorber. If the absorber has a low emissivity, then radiative losses are small anyway and it does not matter very much whether this radiation goes to cover or to ambient. Table 8.4 also lists density, thickness, maximum temperature, and cost of the glazing materials.

Of crucial concern is the durability. In this regard, glass is unsurpassed, with the possible exception of breakage from thermal shock, hail, and vandalism. But even the risk of breakage can be greatly reduced by tempering.[82] Glass is hardly affected by ultraviolet radiation and is very scratch-resistant. The durability of glass as a collector cover is highlighted by Jorgensen[85] who found that samples of 20- to 40-year-old glass from south-facing windows showed less than 2% decrease in transmittance compared to unexposed glass from the same windows.

At the present time there is still much uncertainty about the use of plastic glazings in solar collectors. The most reliable plastic, acrylic, offers no cost advantage over glass. The thin films can be significantly cheaper in first cost, but they may have to be replaced after several years. Thus, it is not surprising that so far almost all commercial collectors have used glass covers. This situation may well change with further research and development in the chemical industry. This process takes time because of the need for long-term outdoor testing of collector materials. But if plastic films with high transmittance and long life are developed, the cost reductions of solar collectors could be very significant.[87]

Another area for progress lies in coatings. Two kinds of coatings are of interest for collector glazing: antireflection (AR) coatings and infrared reflecting coatings, also known as heat mirrors. Coatings of tin and/or indium oxide are examples of the latter. They can be added to glass and have been tested in solar collectors. In their effect on collector performance, they are equivalent to selective absorber coatings. The main drawback of the heat mirror/coatings known so far is their low solar transmittance, caused by a high index of refraction. A typical heat mirror

with a nonselective absorber will be equivalent to a selective absorber with ε around 0.1 and α around 0.8 to 0.85. Selective absorbers offer better performance at lower cost. This might change as a result of further research, for example, by combining AR coatings with a heat mirror. The principal advantage of a heat mirror is that the absorber can get very hot without degradation of the selectivity.

AR coatings are produced by adding a layer with an index of refraction intermediate between air and glazing material. Ideally, the intermediate layer should have an index equal to the square root of the index of the glazing, and its thickness should be one-quarter wavelength. Several low-cost AR coatings are available for solar collectors. Etching by hydrofluoric acid can produce a gradual transition in the refractive index at the surface of glass. Reflection losses can easily be reduced below 1% per glass/air interface.[88] This type of etching is microscopically fine and does not increase the sensitivity to dirt or dust.[82] Thin films of polymers with appropriate index of refraction can also provide low-cost AR coatings. A reduction of reflection losses from 6% to 0.50% at 550 nm with a coating of this type has been reported by Beauchamp.[89]

8.6.3 Reflector Materials

Reflectors for solar collectors should have as high a reflectivity as possible. Furthermore, the reflectors should be highly specular since scattered radiation is likely to miss the absorber. Only collectors with very low concentration (side reflectors, V-troughs, and CPCs) can use reflectors with relatively low specularity, provided the total reflectance is high. Apart from total internal reflection, there are only two materials with sufficiently high specular reflectivity to be suitable for solar reflectors: aluminum and silver. Uses for total internal reflection in solar collectors are limited; some possible designs have been analyzed by Winston,[9] Rabl,[90] and Smestad and Hamill.[91] A pure polished aluminum surface has a solar reflectance of 0.90 while polished silver has 0.96. Both silver and aluminum must be protected from the environment, otherwise the reflectance would rapidly deteriorate to unacceptable levels. Unprotected aluminum tarnishes less rapidly, because, upon contact with air, aluminum immediately develops a thin transparent oxide coating that inhibits further oxidation; however, this coating is not enough to protect against the moisture and pollutants found in most environments.

To protect the reflector surface, one places the aluminum or silver layer behind a transparent cover of glass or plastic. The back of a second-surface solar reflector must also be protected, but that is relatively easy because one can use opaque coatings. First-surface aluminum or silver reflectors can be used in a vacuum, for example, in the shaped tube CPC collector of Fig. 8.8(b). The protective oxide layer of an aluminum sheet can be enhanced by an anodizing process to the point where aluminum can be used as first-surface reflector outdoors. Aluminum sheets of this type are sold under the trade names of Alzac or Kinglux and are frequently used for lighting and interior decorating. Outdoors they survive quite well in dry, unpolluted environments, but in other areas they may tarnish within years or even months.

Most reflectors that are exposed to the elements will probably be second-

surface reflectors. Glass provides the best protection, but it is rigid and fragile. Most concentrators need curved reflectors, and curving glass can be a problem. One could shape the glass while it is hot, but such a process is very capital intensive and justifiable only by a large production volume. The problems of shaping cold glass mirrors increase with curvature; the stresses induced in the glass make it more susceptible to breakage. In this regard a plastic film like acrylic looks much better. In fact, quite a few manufacturers offer solar reflector sheets consisting of aluminized plastic films with adhesive backing. They are easy to handle and attach to a shaped substrate. The main problem with these plastic reflector foils is durability. The plastic surface scratches easily and loses specular reflectivity. Also, pollutants may reach the metal surface through scratches or pinholes, and corrosion will set in. Silver reflectors with a plastic coating seem to be particularly vulnerable: so far. none has been developed that holds promise of surviving outdoors for 20 years.

Absorption in the cover material causes another problem with second-surface reflectors. In mirrors with ordinary window glass, the absorption is too large to be practical (note that the path of a light ray through the glass is twice the thickness, divided by the cosine of the incidence angle in the glass). Hence, one needs either low-iron glass or very thin glass, a millimeter or less in thickness. The latter, also called microsheet, looks most promising for the future. Since microsheet by itself is extremely fragile, one needs to bond it to a thin plastic substrate. Such a sandwich of microsheet glass, silver, and plastic substrates could be rolled and unrolled easily. It combines the highest possible reflectance with durability and can readily be bonded to curved reflector substrates. Unfortunately, the development costs of this material are high and require a large market. At the present time, microsheet reflectors are not yet commercially available, but they do appear to be the solar reflector material of the future.

Data for some typical solar reflectors are shown in Table 8.5. For the optical analysis of focusing collectors, one needs not only the total reflectivity, but also the distribution of reflected rays about the specular direction. Table 8.5 provides this information according to the formation explained above. $R_s(2\pi)$ is the total or hemispherical reflectivity. The distribution of reflected rays is modeled either as a single gaussian of width σ_1 or as a sum of two gaussians of widths σ_1 and σ_2. The coefficients R_1 and R_2 give the relative weights of the two gaussians. For collectors of high concentration, one needs high values of $R_s(2\pi)$ and low values of σ. If σ is small compared to the angular radius of the sun, then scattering is no problem. Silvered glass of several millimeters thickness is excellent in this regard. With microsheet, the beam spread depends very much on how the microsheet is attached to the substrate. Aluminum sheet like Alzac or Kinglux has rolling marks caused by the production process, and the beam spread is different in the directions parallel and perpendicular to these roll marks. In any case, the beam spread is so large, with σ_2 on the order of 10 mrad, as to put this material at a severe disadvantage as far as collectors with high concentration are concerned. For example, if the acceptance half-angle of the collector is 10 mrad, then only 0.71 of the light incident on Kinglux is reflected to the absorber even though the hemispherical reflectance is 0.85.

As for costs, metallized plastic reflector films sell for about 8 to 35 $\$/m^2$ and silvered glass for about 10 to 15 $\$/m^2$. (The cost of the silver itself makes only a

Table 8.5

Properties of reflector materials. $R_s(2\pi)$ is hemispherical reflectivity, σ_1 and σ_2 indicate beam spread, adapted from Pettit[40] and Butler and Claassen[84]

Material	Measurement Wavelength (nm)	R_1	σ_1 (mrad)	R_2	σ_2 (mrad)	$R_s(2\pi)$
Second-surface glass						
Laminated glass (Carolina Mirror Co.)	500	0.92	0.15			0.83
Laminated glass (Gardner Mirror Co.)						
perpendicular to streaks	600	0.92	0.4			0.90
	500	0.92	0.4			
parallel to streaks	800	0.88	<0.05			
	500	0.92	<0.05			
Corning microsheet (vacuum chuck)	550	0.77	1.1	0.18	6.2	0.95
Corning 0317 glass (1.5mm)						0.95[84]
Metallized plastic films						
3M Scotchcal 5400	500	0.86	1.9			0.85
	600	0.86	2.0			
	700	0.82	2.1			
	900	0.84	1.9			
3M FEK-163	500	0.86	0.90			0.85
	600	0.86	0.78			
	700	0.82	0.86			
	900	0.84	0.86			
Sheldahl aluminized Teflon	400	0.73	1.4	0.15	12.1	0.87
	500	0.80	1.3	0.07	30.9	
	700	0.80	1.6	0.04	39.8	
	900	0.81	1.4	0.03	31.4	
Polished bulk aluminum						
Alcoa Alzak						
perpendicular to	670	0.66	0.39	0.21	9.7	0.85
rolling marks	505	0.56	0.42	0.33	10.1	
	407.5	0.45	0.53	0.42	9.8	
parallel to	670	0.70	0.24	0.17	7.7	
rolling marks	505	0.62	0.29	0.27	7.1	
	407.5	0.58	0.46	0.29	9.0	
Kingston Ind. Kinglux						
perpendicular to						
rolling marks	498	0.65	0.37	0.23	16.1	0.85
parallel to						
rolling marks	498	0.67	0.43	0.21	18.5	
Metal Fabrications						
bright aluminum	550	0.44	1.4	0.43	10.3	0.84

small contribution because the layer can be very thin.) Polished aluminum sheet like Alzac or Kinglux can be bought for about 25 $\$/m^2$. Of course the cost of a collector is not determined by the reflector surface alone. For example, an aluminum sheet has some rigidity by itself and may not need the extra expense of a shaped substrate.

8.6.4 Other Materials

Collector enclosures, collector support, insulation, seals, pipes, and heat transfer fluids are vital parts of a collection system. Insulating materials should have low thermal conductivities and low cost. They must also withstand the highest expected operating temperatures and thermal cycling. Moisture can cause severe problems. For example, if fiberglass or open cell foam become wet, they can act like a heat pipe: water evaporates near the absorber and condenses near the enclosure, creating a thermal short circuit from absorber to ambient. Hence, it is crucial to protect the insulation from external or internal leaks. Closed cell insulation is preferable because it avoids this problem. However, Freon-blown insulation releases corrosive hydrofluoric or hydrochloric vapors if heated above 180° C. In fact, many insulating materials outgas at elevated temperatures, and these gases may condense on the inside of the glazing, thereby reducing the transmittance even if they are noncorrosive. Most conductivities increase rapidly with temperature. Therefore, heat losses must be calculated with the conductivity at operating temperature.

Enclosures and supports must provide enough rigidity to keep the collector components in place. The rigidity requirement is particularly severe for collectors with high concentration. The enclosure should keep the inside dry and clean. It should allow rain and snow to run off. Common materials for collector enclosures or support are galvanized or painted steel, aluminum, fiberglass, plastics, wood and wood products, and concrete. Compatibility of materials is important. For instance, different metals in close contact with each other can cause severe electrolytic corrosion. Also, the possibility of differential thermal expansion must be taken into account to prevent warping or breakage of glazing or absorber. Wood is a risky choice as collector enclosure; it may outgas, warp, rot, or ignite.

Last, but not least, one needs to select the heat transfer fluid. Table 8.6 lists the properties of the most important candidates for solar application. Air costs nothing, is noncorrosive and stable at all temperatures, but its heat transfer coefficient is low and the pumping power requirement high. Air is practical only for applications where hot air itself is the desired product.

Water is the ideal heat transfer fluid as long as there is no risk of freezing or boiling. It offers excellent heat transfer, and pumping requirements are low. The temperature limits of water can be extended by adding antifreeze. Typically, one chooses a 50/50 mixture of water and glycol; this prevents freezing down to almost $-40°$C and permits operation up to about 120° C. At higher temperatures, glycols may decompose into corrosive by-products. Hence, it is advisable to check the condition of the antifreeze occasionally, in particular after an accidental overheating. Glycols are somewhat toxic, and U.S. building codes require a double wall heat exchanger between a drinking water supply and a collector with glycol.

Table 8.6

Properties of heat transfer fluids[1]

Commercial Name (Manufacturer)	Chemical Composition	Freezing Point (°C)	Boiling Point (°C)	Thermal Conductivity (cal/s cm^2°C/ cm × 10^4)	Specific Heat (Kcal/kg °C) at °C	Viscosity (Centistokes) at °C	Volume Expansion (°C^{-1} × 10^4)	Specific Gravity at °C
Propylene glycol								
Dowfrost (Dow)[1] 59 weight % aqueous solution	propylene glycol K_2HPO_4 one %	−31°	102°	9.3	0.83, 27°	3,38° 70, −18°	7.3	1.033, 27°
Sunsol 60 (Sunworks)[2] 60 weight % aqueous solution	propylene gylcol inhibitor	−48°	109°	8.3	0.81, 25°	4,38° 140, −20°	7.5	1.055, 25°
Ethylene glycol								
Dowtherm SR-1 (Dow)[3] 59 weight aqueous solution	ethylene glycol K_2HPO_4 two %	−37°	110°	10.0	0.82,27°	2.6, 38° 22, −18°	6.3	1.074, 20°
Polyglycols								
UCON 500 (Union Carbide)	polyglycol 90 % inhibitor	−37°	289° rec max temp	3.7	0.53, 100°	11.5, 100° 61, 38°	7.9	0.98, 100°
Paraffinic hydrocarbon oils								
Caloria HT 43 (Exxon)[4]	petroleum distillate moderate molecular weight, low aromatic content	−9.5°	311° rec max temp	3.1	0.5, 100°	5,100°	10.0	0.80, 100°

Table 8.6

Properties of heat transfer fluids[1](cont'd)

Commercial Name (Manufacturer)	Chemical Composition	Freezing Point (°C)	Boiling Point (°C)	Thermal Conductivity (cal/s cm^2°C/ cm × 10^4)	Specific Heat (Kcal/kg °C) at °C	Viscosity (Centistokes) at °C	Volume Expansion (°C^{-1} × 10^4)	Specific Gravity at °C
Synthetic Hydrocarbon oils								
Brayco 888 HF (Bray Oil Co.)[5]	polymerized 1-decene	−85°	227° flash point	3.0	0.54, 25° 0.61, 100°	4.5, 100° 22.5, 38°	4.8	0.83, 15°
Silicone oils								
Syltherm 444 (Dow Corning)[6]	polydimethy-siloxane	−46°	>315°	3.4	0.39, 100°	70.0, 100° 20.25°	10.7	0.95, 25°
Water		0°	100°	14.3	1.01, 93°	0.3, 93° 1.8, 0°	2.1	1.00, 4°

Notes: Similar physical properties exist for: 1) UCAR 35, 50 percent solution (Union Carbide). Sunsafe 230, 50 percent solution (NPD Energy Systems)

2) Solar Winter Ban (CAMCO), Corona Solar Fluid (A.O. Smith)

3) UCAR 17, 50 percent solution (Union Carbide)

4) Mobiltherm 603 (Mobil). Thermia C (Shell). Dowtherm HP (Dow)

5) PAO-20E (Uniroyal). H-30 (Mark Enterprises)

6)SF-96 (General Electric)

Information for the following products was taken from manufacturers' data sheets: Sunsol 60, Brayco 888, Syttherm 444, Sunsafe 230, UCAR 35, Solar Winter Ban, Corona Solar Fuild PAO-20E and H-30

For operation in the range of 100° C and 300°C, one can use water only if one is willing to pay the price of high pressure pipes. Otherwise one will resort to heat transfer oil. In their heat transfer characteristics, oils are inferior to water, but certainly far better than air. Hydrocarbon oils are flammable and require precautions against fire. Also, they tend to become very viscous or even freeze at ambient temperatures, thereby posing problems for system startup. Silicone oils avoid the flammability and viscosity problems of hydrocarbon oils, but they are more expensive and even more leak prone. For applications above 300° C, the principal heat transfer fluids are air, steam, helium, molten salt, and liquid metal.

8.6.5 Collector Orientation

When collectors are placed together in arrays, there will be shading from one collector row to the next. Here a compromise must be made between losses due to shading on one hand and cost of land and longer piping runs on the other. The collector spacing can be expressed as ground cover ratio defined as ratio of total aperture and total ground area of a collector field. The optimal ground cover ratio depends on latitude and collector type; typical values for midlatitudes range from 0.2 to 0.4 for point focus collectors and 0.4 to 0.6 for line focus and nontracking collectors.

The collector orientation should be chosen to maximize useful solar energy. Nontracking collectors should face due south, although some deviation from the rule may be acceptable; however, one must be careful that the sun will be within the acceptance angle most of the time. Close to the equator the tolerance for azimuth deviations is greater because the collector tilt is smaller. The optimal tilt for a fixed aperture is approximately equal to the latitude for year-round loads and approximately equal to latitude +15° in the Northern Hemisphere for space heating. For parabolic troughs and linear Fresnel reflectors, the basic goal is to keep the incidence angle small on the average. The optimal orientation for year-round loads is the polar mount (i.e., tracking axis north-south with tilt equal to latitude), but outside the equatorial regions, this may cause problems with collector supports, wind loading, and long piping connections. Hence, one may prefer to keep the tracking axis horizontal. In this case, the usual choices are to align the tracking axis either in the east-west or the north-south direction. The north-south direction yields more yearly energy, but is highly nonuniform over the year; the optimal choice depends on the application. In any case, the precise orientation of the tracking axis does not matter; just as with fixed collectors there is some tolerance. In fact, a sloping ground facilitates draining of the collector. Also, a south-facing slope (in the northern latitude) enhances energy collection. Linear Fresnel lenses are more demanding. If they are designed for a fixed tracking axis at all, then it must be the polar axis. If the concentration of a linear Fresnel lens is sufficiently high, then seasonal adjustments of the tracking axis, or even full two-axis tracking, are required.

8.6.6 Cleaning

The need for collector cleaning varies widely with collector type and location.

To understand the effect of dirt on different collector types, let us take a closer look at the interaction of light with dust particles. Dust causes both absorption and scattering. Most dust particles are very small and they affect the passage of light as much by scattering as by absorption.[93] Since the scattered radiation is distributed over a wide range of angles, it will miss the receiver of a focusing collector. In a flat-plate collector, by contrast, even the radiation scattered by dust on the cover (at least the radiation scattered into the forward hemisphere) will reach the absorber. For that reason, flat-plate collectors are quite insensitive to dust; they may function surprisingly well even when they look terrible. Besides, most flat-plate collectors are deployed at a tilt where rain provides adequate cleaning. In residential applications, people rarely bother to clean flat-plate collectors. Industrial systems are larger and economies of scale may make cleaning of flat-plate collectors cost-effective, at least in dusty environments. Concentrating collectors with large acceptance angle, e.g., CPCs, lose some of the scattered radiation and are intermediate between flat-plate and focusing collectors in their sensitivity to dust.

The effect of dust depends on optical design in yet another way. On an exposed second-surface reflector, a light ray has to penetrate the dust layer twice on its way to the receiver. The thickness of the protective coating in front of the reflective metal layer itself is large compared to the size of dust particles. Therefore, the scattering and absorption on the way to the metal layer is uncorrelated with the scattering and absorption on the way out from the metal layer. Hence, the loss of radiation with an exposed second-surface reflector is twice as large as it would be if light had to pass the dust layer only once. The latter is the case with Fresnel lenses (assuming the back of the lens stays clean) and with parabolic dish collectors that are completely enclosed inside inflated transparent bubbles.

Several factors determine the accumulation of dust on collectors. The deposition of dust depends on wind velocity, surface shape, and particle size. The bonds that dust or dirt particles form with a surface may be chemical or physical. There is a tendency for the bond to become stronger with time. Atmospheric moisture plays an important role in this bonding process, because moisture brings a dust particle closer to the surface, thereby increasing the forces of adhesion; the increased bond persists even after the moisture has evaporated again.[93] Dust accumulation is most severe at night when dew condenses on airborne dust and makes it settle on a collector, with the water acting like a glue. A light rain in the desert causes similar problems if the precipitation is only enough to collect dust from the air but not enough to rinse off the collector. For this reason it is important to stow reflectors so they do not face upwards at night or during dust storms. For heliostats, the vertical position, facing out of the wind, is also acceptable. Fixed reflector systems, i.e., the hemispherical bowl and the cylindrical slats, suffer in this regard and require extra cleaning.

For a more quantitative discussion of the effect of dirt on reflectors, we cite a study by Freese.[94] The specular reflectances of mirrors were monitored through nine months of outdoor exposure and different cleaning frequencies. Differences in reflector elevation and mounting angle (other than inverted) were found to have no significant effect on the rate of dirt and dust accumulation. Figures 8.42(a), (b), and (c) show reflectance as a function of time for mirrors cleaned every two, six, and

twelve days. The sawtooth pattern indicates that reflectance deteriorates rapidly after cleaning but is restored to a fairly constant maximum with each cleaning. Figure 8.42(d) depicts the same relationship for a mirror never cleaned except by natural weather conditions. It was found that specular reflectance decreased most rapidly on just-cleaned mirrors, that light rain with windy, dusty conditions could decrease reflectance up to 15% to 20%, and that rain on a newly cleaned mirror slightly decreased its reflectance, but rain and melting snow on dirty mirror surfaces cleaned them well. In fact, snow conditions could increase reflectance to within 0.01 reflectance units of the value obtained after ultrasonically cleaning the mirrors in a laboratory.

Ultrasonic cleaning, of course, is not a practical solution to problems of dirt and dust accumulation on reflectors. Realistic cleaning methods include high pressure water sprays, sprays of commercial detergents, and wiping with soap and water, all followed by a water rinse. Wiping may not be feasible if access to reflectors is blocked by supports, receivers, or guy wires. Cleaning large reflector arrays with detergent may not be environmentally sound and it does not appear to be significantly more effective than cleaning with plain water.[36] High pressure plain water sprays of 500 psi are recommended; they can recover about 95% of the reflectance lost due to particle accumulation.[93] Either an automatic sprinkler system or a spray truck with water tanks can supply the water jets. To minimize particle accumulation between cleanings, reflectors should be stored facing vertically or downward when possible and be rinsed after cleanings with an antistatic solution.[95]

While dust and dirt on an exposed surface are relatively easy to clean off, the absorber and the inside of the glazing are difficult or impossible to clean. Therefore, these surfaces should be adequately protected. In a nonevacuated collector, hermetic sealing seems impractical because of the thermal expansion of air inside the collector; instead vent holes should be covered with dust filters.

8.6.7 Tracking

The performance of tracking collectors depends on the tracking accuracy that can be maintained in actual operation. Many of the tracking collectors installed during the 1970s fared rather poorly in this regard.[96] A frequent problem was caused by the inability of the first generation of tracking sensors to function well under variable insolation. They could be adjusted to work correctly for a fixed insolation level, but with changing beam or diffuse insolation they would be come erratic. Sometimes these trackers would chase the silver lining of a cloud instead of the sun. Another common problem with the early trackers had to do with misalignment between sensor and collector. Some systems tried to bypass the problems caused by a sun sensor and used instead a computer to calculate the position of the sun at any moment. However, this latter approach requires very careful determination of the orientation of the collector, otherwise there will be a serious sytematic tracking error.

Apparently these early problems have been more or less solved. When measuring the accuracy of recent commercial trackers for parabolic troughs, Gee found that correct tracking could be maintained within about 1 mrad. With heliostats,

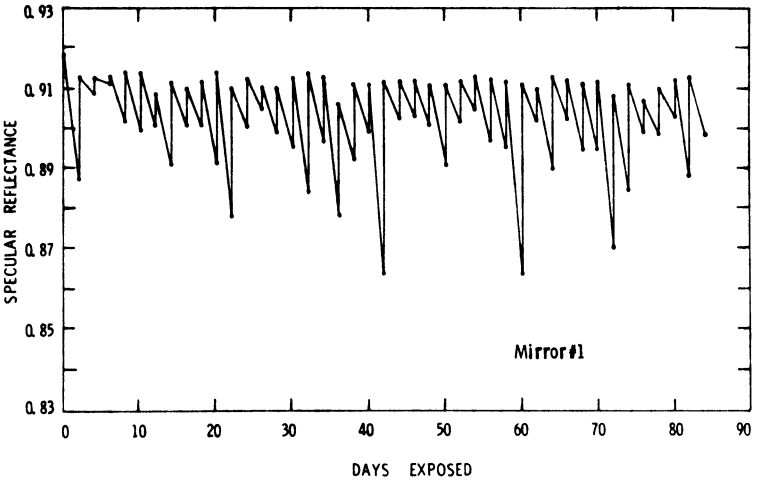


Fig. 8.42: Specular reflectance versus time for a mirror under several different cleaning cycles.[94] **a)** Cleaning every 2 days.

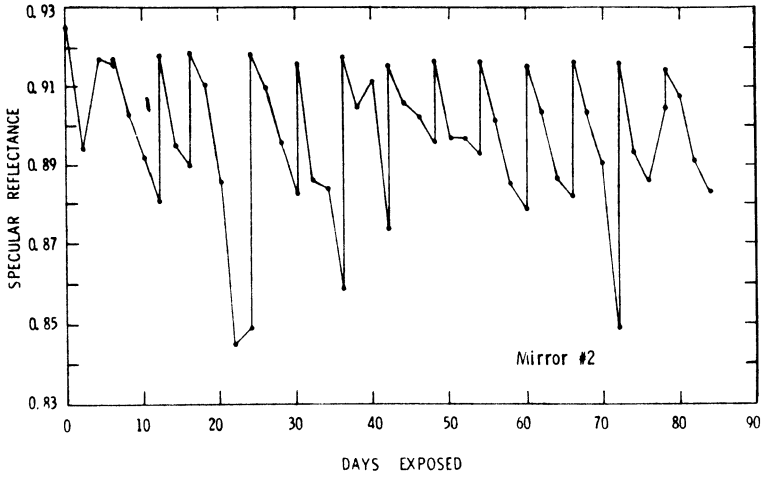


Fig. 8.42: b) Cleaning every 6 days.

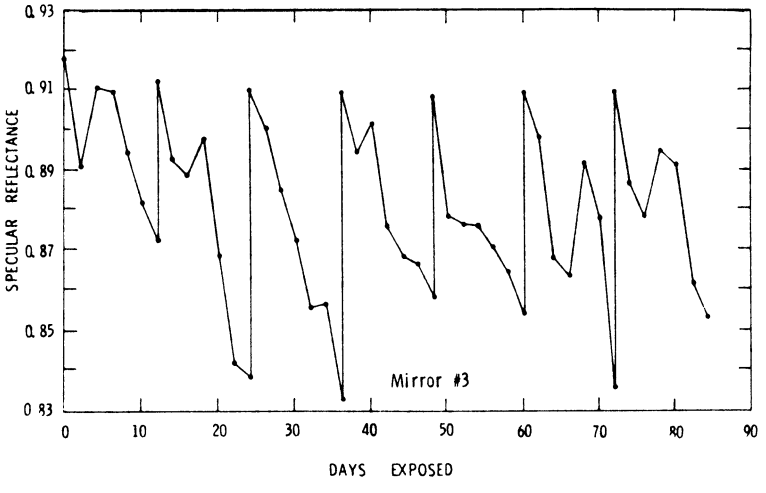


Fig. 8.42: c) Cleaning every 12 days.

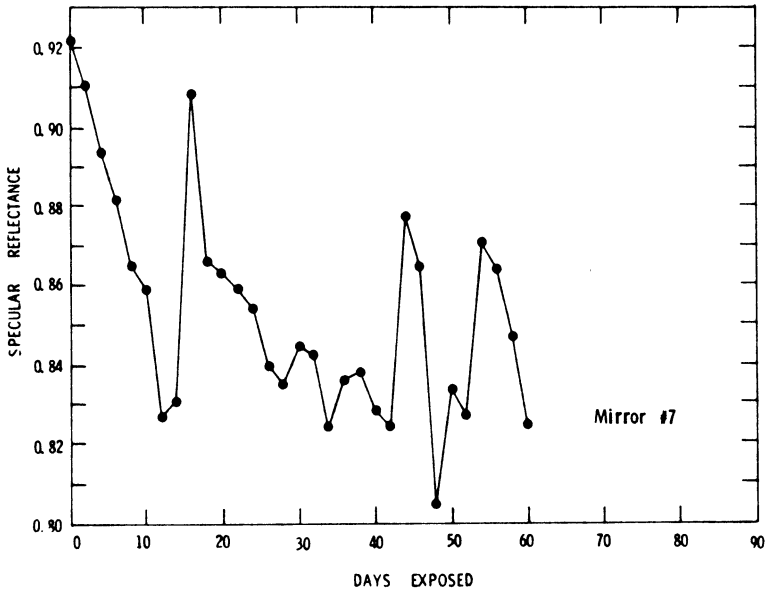


Fig. 8.42: d) No cleaning.

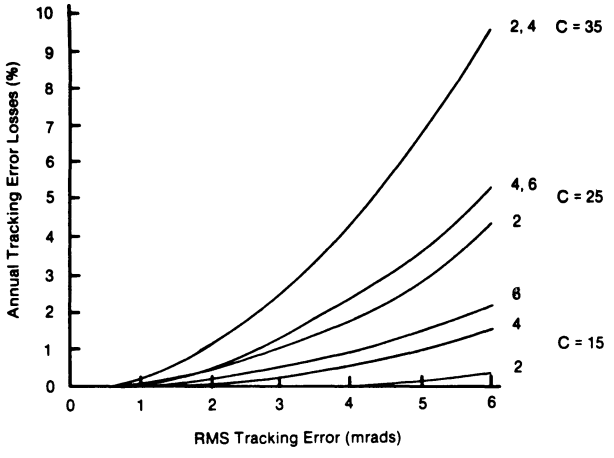


Fig. 8.43: Effect of tracking errors on annual collectible energy.[80]

tracking accuracies in the range of 1 to 2 mrad have been demonstrated at the Central Receiver Test Facility in Albuquerque, New Mexico.[97] Tracker improvements have been achieved by more sophisticated electronics and by change in sensor design, thereby reducing the impact of diffuse insolation. Also, some new trackers have been developed that measure directly the flux at the receiver, thus eliminating many of the alignment problems experienced with trackers mounted on the aperture. While tracking accuracy of the new trackers is best for high levels of beam irradiance, it remains good down to beam irradiance levels of 200W/m^2, and only the beam-normal irradiance seems to matter, not the angle of incidence or the diffuse irradiance.[96]

To provide a measure of the importance of tracking accuracy, Fig. 8.43 shows the decrease in annual collected energy as a function of tracking error for parabolic troughs. Curves are shown for several values of concentration ratio, C, and mirror contour error, $\sigma_{contour}$. The higher the concentration, the greater the sensitivity to tracking errors. Tracking errors of 5 mrad are quite serious; they cause loss of about 1% for C equal to 15 and 7% for C equal to 35. But if the tracking error is less than 1 mrad, then the loss is negligible (a fraction of 1%), even for C as high as 35.

8.7 SUMMARY

The main features of the principal concentrator types are summarized in Tables 8.7 to 8.9. Flat plates have been included for the sake of comparison. For each collector, the practical range of operating temperatures is indicated. Costs are also mentioned since no discussion of solar technologies is complete without reference to economics.

Table 8.7

Nontracking collectors

Collector Type	Approximate Maximum Operating Temperature (°C)	Cost in 1980$ ($/m^2) Now	Goal	Comments
Shallow solar pond	40 to 60	160[a] (complete system, including storage for one day)		Plastic covers may need to be replaced every 5 years or so. Needs sunny climate for good performance.
Deep solar pond (salt gradient)	40 to 90	30 to 60[b] (includes storage)		Collector and long-term storage in one unit. For seasonal storage, depth should be about 3 m. Low cost, but low efficiency (10% to 20%).
Flat plate	40 to 80	100 to 150[c]		Best known and most developed of all collector types.
Nonevacuated CPC fixed-tilt, or summer-to-winter tilt adjustment	80 to 120	150[c]		
Evacuated tubes (with reflector enhancement including CPC)	100 to 200	200[c] to 300	100	Many opportunities for cost reduction by mass production and for performance improvements through R&D.

[a] Personal communication, Solar Energy Group, Lawrence Livermore Laboratory, Livermore, CA 94550, 1980.

[b] A 3-m-deep salt gradient solar pond was built in Miamisburg, Ohio for $35/m^2 in 1978. Solar ponds without long-term storage or in locations with free salt should cost considerably less.

[c] *Solar Age Magazine*. 1979. *Solar Products Specification Guide*. Church Hill, Harrisville, NH: Solar Vision, Inc.

Table 8.8

One-axis tracking collectors

Collector Type	Approximate Maximum Operating Temperature (°C)	Cost in 1980$ ($/m^2) Now	Goal	Comments
Inflated cylindrical reflector	140	50 to 70[a]		Does not need continuous tracking, but does require weekly tilt adjustments; plastic cover may need to be replaced every 5 years or so.
Parabolic trough	300	150 to 250[b]	110[c]	Continuous accurate tracking; sensitive to dirt.
Line-focus Fresnel reflector	250			May combine advantages of parabolic trough and of central receiver for temperatures below 250°C.
Fixed line-focus reflector with tracking receiver	250			Problems with dirt accumulation on reflectors.

[a] Personal communication, Solar Energy Group, Lawrence Livermore Laboratory, Livermore, CA 94550, 1980.

[b] Based on *Solar Age Magazine*. 1980. *Solar Products Specifications Guide*. Church Hill, Harrisville, NH: Solar Vision.

[c] Cost goals for 1990 (in 1980$) from *Solar Thermal Program Multiyear Plan*, Golden, CO: Solar Energy Research Institute. Consistent cost goals have not been developed for all collector types.

Table 8.9

Two-axis tracking collectors

Collector Type	Approximate Maximum Operating Temperature (°C)	Cost in 1980$ ($/m^2) Now	Goal	Comments
Parabolic dish or point-focus Fresnel lens	1500 (possibly more)	1000[a]	130[b]	Good if energy can be used directly in focal zone (e.g., photovoltaics or solar thermal power); otherwise, transporting heat to point of use is problematic.
Central receivers	1000 (possibly more)	500[c] (200[d] + tower[e]	100[b] + tower[e]	Optical transport of energy.
Fixed-hemispherical reflector, tracking receiver	400			Problems with heat transport to point of use, and with dirt accumulation on reflector.

[a] *Solar Age Magazine.* 1979. *Solar Products Specifications Guide.* Church Hill, Harrisville, NH: Solar Vision, Inc.

[b] Cost goals for 1990 (in 1980$) from *Solar Thermal Program Multiyear Plan.* Golden, CO: Solar Energy Research Institute. Consistent cost goals have not been developed for all collector types.

[c] Average cost of heliostats for Barstow solar power plant.

[d] Incremental cost of heliostats for Barstow after tooling costs, etc., have been paid. (P. Eicker, Sandia Livermore Laboratories, personal communication, 1980.)

[e] Cost of tower is estimated to be approximately 10% of heliostat cost.

8.8 REFERENCES

1. R. Winston, "Light collection within the framework of geometrical optics," *J. Opt. Soc. Am.* **60**, 245 (1970).
2. A. Rabl, "Comparison of solar concentrators," *Solar Energy* **18**, 93 (1976).
3. W. T. Welford and R. Winston, *The Optics of Nonimaging Concentrators* (Academic Press, New York, 1978).
4. H. Hinterberger and R. Winston, *Rev. Sci. Instr.* **37**, 1094 (1966).
5. M. Ploke, "Lichtführungseinrichtungen mit starker Konzentrationswirkung," *Optik* **25**, 31 (1967).
6. V. K. Baranov and G. K. Melnikov, *Soviet Journal of Optical Technology* **33**, 408 (1966).
7. R. Winston and H. Hinterberger, "Principles of cylindrical concentrators for solar energy," *Solar Energy* **17**, 255 (1975).
8. A. Rabl and R. Winston, "Ideal concentrators for finite sources and restricted exit angles," *Applied Optics* **15**, 2880 (1976).
9. R. Winston, "Dielectric compound parabolic concentrators," *Applied Optics* **15**, 291 (1976).
10. W. R. McIntire, "New reflector design which avoids losses through gaps between tubular absorbers and reflectors," *Solar Energy* **25**, 215 (1981).
11. A. Rabl, J. O'Gallagher, and R. Winston, "Design and test of non-evacuated solar collectors with compound parabolic concentrators," *Solar Energy* **25**, 335 (1980).
12. A. Rabl, "Radiation transfer through specular passages," *Int. J. Heat Mass Transfer* **20**, 323 (1977).
13. A. Rabl, "Optical and thermal properties of compound parabolic concentrators," *Solar Energy* **18**, 497 (1976).
14. W. R. McIntire, "Truncation of nonimaging cusp concentrators," *Solar Energy* **23**, 35 (1979).
15. A. Rabl, *Active Solar Collectors and Their Applications* (Oxford University Press, New York, 1985).
16. R. Winston and W. T. Welford, "Geometrical vector flux and some new nonimaging concentrators," *J. Optical Soc. America* **69**, 532 (1979).
17. D. R. Mills and J. E. Giutronich, "Ideal prism solar concentrators," *Solar Energy* **21**, 423 (1978).
18. D. R. Mills, "The place of extreme asymmetrical non-focusing concentrators in solar energy utilization," *Solar Energy* **21**, 431 (1978).
19. J. J. O'Gallagher et al., "Absorption enhancement in solar collectors by multiple reflections," *Solar Energy* **24**, 323 (1980).
20. A. Rabl, N. B. Goodman, and R. Winston, "Practical design considerations for CPC solar collectors," *Solar Energy* **22**, 373 (1979).

21. W. R. McIntire, "Elimination of the optical losses due to gaps between absorbers and their reflectors," *Proc. of the 1980 Annual Meeting of the American Section of the International Solar Energy Society*, 2-6 June 1980, Phoenix, Ariz., p. 600.

22. R. Winston, "Cavity enhancement by controlled directional scattering," *Applied Optics* **19**, 195 (1980).

23. W. R. McIntire, "Stationary concentrators for tubular evacuated receivers: Optimization and comparison of reflector designs," *Proc. of the 1980 Annual Meeting of AS of ISES*, 2-6 June 1980, Phoenix, Ariz., p. 505.

24. B. Window and I. M. Bassett, "Optical collection efficiencies of tubular solar collectors with specular reflectors," *Solar Energy* **26**, 341 (1981).

25. H. Tabor, "Stationary mirror systems for solar collectors," *Solar Energy* **2**, 17 (1958).

26. K. G. T. Hollands, "A concentrator for thin-film solar cells," *Solar Energy* **13**, 149 (1971).

27. M. K. Selcuk, "Analysis, development and testing of a fixed tilt solar collector employing reversible vee-trough reflectors and vacuum tube receivers," *Solar Energy* **22**, 413 (1979).

28. S. L. Grassie and N. R. Sheridan, "The use of planar reflectors for increasing the energy yield of flat plate collectors," *Solar Energy* **19**, 663 (1977).

29. D. K. McDaniels et al., "Enhanced solar energy collection using reflector solar-thermal collector combinations," *Solar Energy* **17**, 277 (1975).

30. S. C. Seitel, "Collector performance enhancement with flat reflectors," *Solar Energy* **17**, 291 (1975).

31. D. C. Larson, "Optimization of flat plate collector–flat mirror system," *Solar Energy* **24**, 203 (1980).

32. H. F. Chiam, "Planar concentrators for flat plate solar collectors," *Solar Energy* **26**, 503 (1981).

33. A. Rabl and P. Bendt, "Effect of circumsolar radiation on performance of focusing collectors," *ASME Journal of Solar Energy Engineering* **104**, 237 (1982).

34. Grether and Hunt, "Description of the LBL reduced data base," informal report, Lawrence Berkeley Laboratory, Berkeley, CA 94720.

35. F. Biggs and C. N. Vittitoe, *The Helios Model for the Optical Behavior of Reflecting Solar Concentrators*, Albuquerque, N.M.: Sandia Laboratories, report SAND 76-0347, 1979.

36. R. B. Pettit and B. L. Butler, "Semiannual review ERDA thermal power systems, dispersed power systems, distributed collectors, and research and development: Mirror materials and selective coatings," Albuquerque, N.M.: Sandia Laboratories, report SAND 77-0111.

37. H. Cramer, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, N.J., 1947).

38. W. J. Adams, *The Life and Times of the Central Limit Theorum* (Kaedman, New York, 1947).

39. C. N. Vittitoe and F. Biggs, "Six-Guassian representation of the angular brightness distribution for solar radiation," *Solar Energy* **27**, 469 (1981).

40. R. B. Pettit, "Characterization of the reflected beam profile of solar mirror materials," *Solar Energy* **19**, (1977).

41. P. Bendt et al., "Optical analysis and optimization of line focus solar collectors," Golden, Colo.,: Solar Energy Research Institute, SERI/TR-36-092, 1979.

42. M. Riaz, "A theory of concentrators of solar energy on a central receiver for electric power generation," *J. Engineering for Power* **98**, 375)1976). This reference derives closed analytical formulas for shading and blocking effects.

43. L. L. Vant-Hull, "An educated ray trace approach to solar tower optics," *Proc. of the Society of Photo-Optical Instrumentation Engineers* **85**, 111 (1976).

44. P. Leary and J. Hankins, "A user's guide for MIRVAL computer code for comparing designs of heliostat receiver optics for central receiver solar power plants," Livermore, Calif.: Sandia Laboratories, report SAND 77-8280, 1977.

45. T. A. Dellin and M. J. Fish, "A user's manual for DELSOL—A computer code fo calculating the optical performance, field layout, and optimal system design for solar central receiver plants," Albuquerque, N.M.: Sandia Laboratories, report SAND 79-8215, 1979.

46. K. D. Bergeron and C. J. Chiang, "SCRAM: A fast computational model for the optical performance of point focus solar central receiver systems," Albuquerque, N.M.: Sandia Laboratories, report SAND 80-0433, 1980.

47. M. D. Walzel, F. W. Lipps, and L. L. Vant-Hull, "A solar flux density calculation for a solar tower concentrator using a two-dimensional hermite function expansion," *Solar Energy* **19**, 239 (1977).

48. F. W. Lipps and L. L. Vant-Hull, "A cell-wise method for the optimization of large central receiver systems," *Solar Energy* **20**, 505 (1978).

49. P. J. Eicker, Sandia Livermore Laboratories, letter of 18 April 1979 to J. Thornton, Solar Energy Research Institute.

50. C. L. Pitman and L. L. Vant-Hull, "Atmospheric transmission model for a solar beam propagating between a heliostat and a receiver," *Progress in Solar Energy*, (American Solar Energy Society, Boulder, Colo., 1982), p. 1267.

51. E. M. Kritchman, A. A. Friesem, and G. Yekutieli, "Highly concentrating fresnel lenses," *Applied Optics* **18**, 2688 (1979).

52. M. Collares-Pereira, "High temperature solar collector with optimal concentration non-focussing Fresnel lens with secondary concentrator," *Solar Energy* **23**, 409 (1979).

53. M. J. O'Neill, "A unique new Fresnel lens solar concentrator," International Solar Energy Congress, May 1979, Atlanta, Ga. Also U. S. patent No. 4,069,812, "Solar Concentrator and Energy Collection System" (1978).

54. E. M. Kritchman, A. A. Friesem, and G. Yekutieli, "Efficient lens for solar concentration," *Solar Energy* **22**, 119 (1979).

55. A. M. Clausing, "The performance of a stationary' reflector/tracking absorber solar concentrator," ISES Solar Energy Conference, Vol. 2, pp. 304, Winnipeg, Canada, August 1976.

56. J. F. Kreider, "Thermal performance analysis of the stationary reflector/tracking absorber (SRTA) solar concentrator," *J. Heat Transfer* **97**, 451 (1975).

57. H. Tabor and H. Zeimer, "Low cost focusing collector for solar power units," *Solar Energy* **6**, 55 (1962).

58. J. W. Gerich, "A nontracking inflated cylindrical solar concentrator," ISES Congress, Atlanta, Ga., May 28–June 1, 1979.

59. I. R. Edmonds et al., "The design and performance of ideal solar concentrators based on the prism-assisted cylindrical reflector," *Solar Energy* **30**, 537 (1983).

60. J. L. Russell, "Central station solar power," *J. of Power Engng.* (Nov. 1974).

61. P. J. Eggers et al., General Atomic report GA-A14209 (Rev.) UC-62.

62. L. W. Brantley, Jr., "A pressure stabilized solar collector," *Proceedings of the 1977 Annual Meeting of the American Section of ISES*, June 1977.

63. J. W. Gerich, "An inflated cylindrical solar concentrator," Proceedings of the 1978 Annual Meeting of the American Section of ISES, Denver, Colo., Aug. 28–31, 1978, p. 889.

64. P. G. McCormick, "Optical evaluation of cylindrical elastic concentrators," *Solar Energy* **26**, 519 (1981).

65. J. D. Walton, "Development of the spiral Fresnel concentrator," presented at the International Symposium on Solar Thermal Power and Energy Systems, 15–20 June 1980, Marseille, France.

66. U. H. Kurzweg, "Maximized solar flux concentration achievable with Axicon collectors," *Solar Energy* **25**, 221 (1980).

67. E. Kritchman, "Optimized second stage concentrator," *Applied Optics***21**, 751 (1982).

68. E. Kritchman, "A brief comparison of second stage elements," *Applied Optics***20**, 3824 (1982).

69. E. Kritchman, "CEC second stage concentrator," *Applied Optics***21**, 751 (1982).

70. P. I. Cooper and R. V. Dunkle, "A nonlinear flat-plate collector model," *Solar Energy* **26**, 133 (1981).

71. F. Kreith and J. F. Kreider, *Principles of Solar Engineering*, 1st ed. (McGraw-Hill, New York, 1978).

72. J. J. O'Gallagher et al., "A new evacuated CPC collector tube," *Solar Energy* **29**, 575 (1982).

73. SANDIA 1980. "Line focus solar thermal energy technology development. FY 1979 annual report." Albuquerque, N. M.: Sandia National Laboratories, report SAND 80-0865. NM 87185.

74. "A pipeline to the sun from Owens-Illinois." SUNPAK No. 13.11 OW, 1979. Owens-Illinois, Inc. Development Center, 1020 North Westwood, Toledo, Ohio 43607.

75. H. W. Gaul and A. Rabl, "Incidence angle modifier and average optical efficiency of parabolic trough collectors," *Transactions of the ASME J. of Solar Energy Engineering* **102**, 16 (1980).

76. *ASHRAE Standard 93-77*, "Collector test procedure of the American Society of Heating, Refrigeration and Air Conditioning Engineers," 343 W. 43rd St., New York, N.Y.

77. W. A. Beckman, "Duct and pipe losses in solar energy systems," *Solar Energy* **21**, 53 (1978).

78. F. de Winter, "Heat exchanger penalties in double loop solar water heating systems," *Solar Energy* **17**, 335 (1975).

79. A. Rabl, "Yearly average performance of the principal solar collector types," *Solar Energy* **27**, 215 (1981).

80. V. E. Dudley and R. M. Workhoven, *Summary Report: Concentrating Solar Collector Test Results Collector Module Test Facility*, Albuquerque, N. M.: Sandia Laboratories, report SAND 78-0977, 1979.

81. T. Sullivan, "Solar heat transfer fluids," *Solar Age*, Dec. 1980, p. 33.

82. J. Vresk et al., "Final reliability and materials design guidelines for solar domestic hot water systems," Argonne National Laboratory, Sept. 1981, report ANL/SDP-11.

83. B. O. Seraphin and G. E. Carver, Annual Report, May 1, 1978 to April 30, 1979. "Chemical vapor deposition of refractory metal reflectors for spectrally selective solar absorbers," Optical Sciences Center, University of Arizona, Tucson, Ariz. (1979).

84. B. L. Butler and R. S. Claassen, "Survey of solar materials," *Transactions of the ASME J. of Solar Energy Engineering* **102**, 175 (1980).

85. G. Jorgensen, "Long-term glazing performance, SERI report SERI/TP-31-193.

86. J. Vitko, Jr., "Optical studies of second surface mirrors proposed for use in solar heliostats," Sandia Laboratories, Livermore, Calif., report SAND 78-8228.

87. B. Atkinson and R. Caesar, "The Volkspanel Model T," *Solar Age* April 1983, 33.

88. R. E. Peterson and J. W. Ramsey, "Thin film coatings in solar-thermal power systems," *J. Vacuum Sci. Technol.* **12**, 174 (1975).

89. E. K. Beauchamp, "Low reflectance films for solar collector cover plates," Albuquerque, N. M.: Sandia Laboratories, report SAND 75-0035, 1975.

90. A. Rabl, "Prisms with total internal reflection as solar reflectors," *Solar Energy* **19**, 555 (1977).

91. G. Smestad and P. Hamill, "Concentration of solar radiation by white painted transparent plates. *Applied Optics* **21**, 1298 (1982).

92. P. Call, "National program plan for absorber surfaces R&D," Golden. Colo.: Solar Energy Research Institute, report SERI/TR-31-103, 1979.

93. R. S. Berg, "Heliostat dust buildup and cleaning studies," Albuquerque, N. M.: Sandia Laboratories, report SAND 78-0510, 1978.

94. J. M. Freese, "Effects of outdoor exposure on the solar reflectance properties of silvered glass mirrors," Albuquerque, N. M.: Sandia Laboratories, report SAND 78-1649, 1978.

95. R. L. Champion, "Cleaning and maintenance," *Proceedings of the Solar Thermal Concentrating Collector Technology Symposium*, June 14–15, 1978, edited by B. F. Gupta and F. Kreith, SERI report SERI/TP-34-048.

96. R. C. Gee, "An experimental performance evaluation of line-focus sun trackers," Golden, Colo.: Solar Energy Research Institute, report SERI/TR-632-646, 1982.

97. J. T. Holmes, "Heliostat operation at the central receiver test facility," *Proc. STTF Testing for Long-Term Systems Performance Workshop*, Albuquerque, N. M.: Jan. 7–9, 1981, p. 179.

INDEX

A

B

C

D

E

F

G

H

I

K

L

M

N

O

P

R

S

T

U

V

W

Y