MACMILLAN ENCYCLOPEDIA OF

ENERGY

EDITORIAL BOARD

Macmillan Encyclopedia of Energy

Macmillan Reference USA
1633 Broadway
New York, NY 10019

Macmillan Reference USA
27500 Drake Rd.
Farmington Hills, MI 48331-3535

Library of Congress Catalog Card Number: 00-062498

Printed in the United States of America
Printing number
1 2 3 4 5 6 7 8 9 10

Library of Congress Cataloging-in-Publication Data
Macmillan encyclopedia of energy / John Zumerchik, editor in chief.
p. cm.
Includes bibliographical references and index.
ISBN 0-02-865021-2 (set : hc). — ISBN 0-02-865018-2 (vol. 1). — ISBN 0-02-865019-0 (vol. 2). —
ISBN 0-02-865020-4 (vol. 3).
1. Power resources—Encyclopedias. I. Title: Encyclopedia of Energy. II. Zumerchik, John
TJ163.25 .M33 2000
00-062498
621.042'03 dc21

CONTENTS

Preface vii

List of Articles. ix

List of Contributors. xvii

Acronymns and Standard Abbreviations xxvii

Macmillan Encyclopedia of Energy. 1

Energy Timeline 1239

Index 1249

Editorial and Production Staff

Brian Kinsey, Charles B. Montney
Project Editors

Debra Kirby, Jeffrey Lehman, and Jane Malonis
Contributing Editors

Shawn Beall, Matthew May, Katherine Millson
Assistant Editors

Linda S. Hubbard
Editorial Director

William Drennan, Amy Lucas, Philip Jochnowitz,
Anne Janette Johnson, Jeanette Piccirelli, and Annette P. Swanson
Copyeditors

Laura Avery, Adam Goodman, Andrew Homburg, Jaime Noces, and Jennifer Smith
Proofreaders

Linda Kenny Sloan
Indexer

Julie Juengling
Permissions Assistant

Cynthia Baldwin
Art Director

Robyn V. Young
Senior Editor, Imaging and Multimedia Content

Randy Bassett
Imaging Supervisor

Ronald D. Montgomery
Manager, Data Capture Services

Beverly Jendrowski, Frances L. Monroe
Data Capture Specialists

Nikkita Bankston, Katrina Coach, Teresa Elsey, Cynthia Jones, Joseph Jorgenson, Elizabeth Pilette
Data Capture Associates

Mary Beth Trimper
Manager, Composition and Electronic Prepress

Evi Seoud
Assistant Manager, Composition Purchasing and Electronic Prepress

Dorthy Maki
Manufacturing Manager

Wendy Blurton
Senior Buyer

Datapage Technologies International
Typesetting

Macmillan Library Reference USA
Elly Dickason
Publisher

PREFACE

In the mid-1970s, many environmental organizations and energy experts predicted a bleak future of widespread economic and social disruption caused by energy scarcities and continually rising costs of fossil fuels. These predictions never came true. In fact, the opposite occurred—fossil fuel prices in the year 2000 are actually lower than the brief peaks that were reached after the OPEC oil embargo of 1973. Despite this reality, there continue to be long-term energy problems for which there are no easy answers. Foremost is availability, related to the unsustainability of ever more people demanding an ever-declining supply of fossil fuels. Second is growth in electricity demand, which requires more power plants in an era of growing political opposition to the greater use of fossil fuels and nuclear power. Third is pollution, because the more power plants and vehicles there are, the harder it is to reduce harmful emissions. And finally, there is global warming, the contentiously debated link between energy-related carbon dioxide emissions and climate change.

The three-volume *Macmillan Encyclopedia of Energy* is an interdisciplinary work with a very ambitious scope that addresses these continuing problems. Entries represent the fields of physics, chemistry, biology, geophysics, and engineering, as well as history, economics, psychology, sociology, and ethics. Traditional energy concepts such as electricity, thermodynamics, and conservation of energy are included, but so are less obvious concepts such as combustion, catalysts, propulsion, and matter and energy. The traditional fuels such as gasoline, diesel, and jet fuel are included, but so are the alternative fuels such as methanol, synthetic fuel, and hydrogen. The traditional energy sources such as coal, natural gas, petroleum, and nuclear fission are included, but so are the more forward-looking sources such as wind, solar, fuel cells, and nuclear fusion.

The *Encyclopedia* covers the major energy-using technologies and examines the advances that have been made in terms of performance and efficiency. This is an important approach because the unforeseen technology of tomorrow might be far different from what exists today, leading to an alteration in patterns of energy consumption. Only by improving how experts predict how technological innovation will alter energy consumption will we be able to avoid making erroneous energy predictions similar to those made in the 1970s. A large number of entries related to transportation are included because this sector of the economy faces the greatest challenges in the future with regard to energy use. The reliance on imported petroleum and the difficulty in switching to another energy source make the transportation sector particularly vulnerable to supply disruptions and wild price fluctuations. Subjects related to electricity are also well represented because of the rapid pace of growth in the field of electric technology.

The interdisciplinary approach of the *Encyclopedia* led to different approaches for different types of articles. Besides the historical articles covering the sources, uses, and social views of energy, many technology articles take a historical approach to discussing the invention and innovation of the technology, especially the pace of innovation and energy efficiency improvements between 1970 and 2000. Sixty biographies provide information on specific scientists and engineers who made major contributions to the understanding of energy and the related technologies. Since environmental and safety issues are important to the story of energy, these issues are addressed throughout the *Encyclopedia*.

The *Encyclopedia* includes 253 alphabetically arranged entries written by 170 authors. The text is supplemented with more than 600 photographs, illustrations, sidebars, and maps. Entries contain a set of cross-references to related entries within the set, as well as a bibliography of related books and journal articles to guide readers who want to learn more about a given topic. The front matter in Volume 1 includes a list of entry topics, and the back matter in Volume 3 contains both an extensive timeline of important dates in energy history and a comprehensive subject index.

The purpose of this publication is to provide an up-to-date reference guide for people who want full, current, trustworthy information about all aspects of energy. This includes individuals who want more information than they read in newspapers or magazines, high school and college students who want information for class discussions and papers, people looking for expert information on which to base their everyday energy-related decisions, and scholars who want a review of interdisciplinary research.

The *Encyclopedia* will help to foster a greater awareness of the historical significance of energy and the relationship between energy and technology. Decisions that are related to technology and that affect energy consumption are made on a regular basis. However, since energy is mostly invisible (one cannot see a kilowatt of electricity or a cubic foot of natural gas), these decisions are not commonly thought of as energy decisions even when the energy aspect is an essential aspect. Choices related to transportation, shelter, comfort, and recreation all entail energy decisions, but they are not themselves considered to be energy decisions. For example, driving to work instead of walking, purchasing bigger homes and vehicles instead of smaller ones, and acquiring the latest plug-in appliance all have short- and long-term energy-related consequences. Although each individual decision has an insignificant effect on overall consumption, once the behavior becomes a norm, the effect can be very significant for society since reversal is often impossible.

Production of this three-volume set has involved the efforts of many people. In particular, I want to thank my associate editors, Herm Bieber, Alan Chachich, Fred Denny, Barney Finn, John Firor, Howard Geller, David Greene, Joseph Priest, Rosalie Ruegg, Harold Wallace, and Ellen Zeman for their suggestions with regard to the organization and content of the work and for lending their expertise in evaluating and reviewing manuscripts. I would also like to thank the contributors for their cooperation in ensuring that the content and level of presentation of their entries was appropriate for our audience.

I am equally indebted to the staff at Macmillan Reference USA for their prodding and pushing that helped speed the completion of this work. In particular, I want to thank Charlie Montney and Brian Kinsey for their day-to-day management of the project and Elly Dickason, the publisher of Macmillan, for her support in producing the work you see before you.

John Zumerchik
New York City
2000

LIST OF ARTICLES

A

Acid Rain
Tom Butler

Aerodynamics
John D. Anderson

Agriculture
Hosein Shapouri
Roger Conway

Air Conditioning
Bernard A. Nagengast

Aircraft
John D. Anderson

Air Pollution
Kenneth Green

Air Quality, Indoor
Satish Kumar

Air Travel
F. Robert van der Linden

Alexanderson, Ernst Frederik Werner (1878-1975)
Albert Abramson

Alternative Fuels and Vehicles
John Zumerchik

Ampère, André-Marie (1775-1836)
Robert Sier

Animal and Human Energy
David Mulcahy

Appliances
David B. Goldstein

Archimedes (287-212 BC)
Douglas Quinney

Atmosphere
Dennis G. Baker
Anita Baker-Blocker

Auditing of Energy Use
Michael R. Muller

Automobile Performance
Charles A. Amann

Aviation Fuel
William F. Taylor

B

Batteries
John Zumerchik

Bearings
Jiang Long

Becquerel, Alexandre-Edmund (1820-1891)
Stig Steenstrup

Behavior
John Zumerchik

Bernoulli, Daniel (1700-1782)
Douglas Quinney

Bethe, Hans Albrecht (1906-   )
Silvan S. Schweber

Bicycling
John Schubert
John Zumerchik

Big Bang Theory
Don Lichtenberg

Biofuels
Deborah L. Mowery

Biological Energy Use, Cellular Processes of
David E. Harris

Biological Energy Use, Ecosystem Functioning of
Allison Brody

Black, Joseph (1728-1799)
Robert Sier

Building Design, Commercial
Jeffrey A. Johnson
William Steven Taber, Jr.

Building Design, Energy Codes and
Jeffrey A. Johnson
David Conover

Building Design, Residential
Sam Rashkin
Japhet Koteen

C

Capacitors and Ultracapacitors
Alan B. McEwen

Capital Investment Decisions
Sue McNeil

Carnot, Nicolas Leonard Sadi (1796-1832)
Joseph F. Mulligan

Carson, Rachel
Linda Sargent Wood

Catalysts
Joe W. Hightower
Burtron H. Davis

Charcoal
Herman Bieber

Chemical Energy, Historical Evolution of the Use of
Barry L. Tarmy

Clausius, Rudolf Julius Emmanuel (1822-1888)
Barbara Flume-Gorczyca

Climatic Effects
Kenneth Green

Coal, Consumption of
Anthony Fainberg

Coal, Production of
Jerry C. Tien

Coal, Transportation and Storage of
Jamie N. Heller
Stan M. Kaplan

Cogeneration
Joseph Priest

Cogeneration Technologies
Morris A. Pierce

Combustion
Kenneth Brezinsky

Communications and Energy
Gilbert Held

Conservation of Energy
Joseph Priest

Conservation Supply Curves
Alan K. Meier
Arthur H. Rosenfeld

Consumption
John Zumerchik

Control Systems, History of
Anibal T. De Almeida
Hashem Akbari

Cool Communities
Hashem Akbari
Arthur H. Rosenfeld

Culture and Energy Usage
Laura Nader

Curie, Marie Sklodowska (1867-1934)
Margaret H. Venable
T. Leon Venable

D

Demand-Side Management
Joseph Eto

Diesel, Rudolph (1858-1913)
David B. Sicilia

Diesel Cycle Engines
Robert N. Brady

Diesel Fuel
Robert N. Brady

District Heating and Cooling
Morris A. Pierce

Domestic Energy Use
Morris A. Pierce

Drivetrains
Charles A. Amann

E

Economically Efficient Energy Choices
William S. Peirce

Economic Externalities
Samuel N. Addy

Economic Growth and Energy Consumption
Scott B. Sitzer

Edison, Thomas Alva (1847-1931)
Leonard S. Taylor

Efficiency of Energy Use
Arthur H. Rosenfeld
Tina M. Kaarsberg
Joseph J. Romm

Efficiency of Energy Use, Economic Concerns and
Ronald J. Sutherland

Efficiency of Energy Use, Labeling of
James G. Mills

Einstein, Albert (1879-1955)
Abraham Pais

Elastic Energy
Joseph Priest

Electricity
Dennis Barnaal

Electricity, History of
Bernard Finn

Electric Motor Systems
John Zumerchik

Electric Power, Generation of
David E. Dismukes

Electric Power, System Protection, Control, and Monitoring of
Stanley H. Horowitz

Electric Power, System Reliability and
Roy Billinton

Electric Power Substations
John A. Palmer

Electric Power Transmission and Distribution Systems
John A. Palmer

Electric Vehicles
John Zumerchik
Fred I. Denny

Emission Control, Vehicle
Randall Guensler

Emission Control, Power Plant
Charles E. Hickman

Energy Economics
Richard L. Gordon

Energy Intensity Trends
Stephanie J. Battles

Energy Management Control Systems
Anibal T. De Almeida
Hashem Akbari

Engines
Charles A. Amann

Environmental Economics
Richard L. Stroup

Environmental Problems and Energy Use
Fred I. Denny

Ericsson, John (1803-1899)
Robert Sier

Ethical and Moral Aspects of Energy Use
Joseph R. DesJardins

Explosives and Propellants
Herman Bieber

F

Faraday, Michael (1791-1867)
Leonard S. Taylor

Fermi, Enrico (1901-1954)
Joseph F. Mulligan

Flywheels
Donald A. Bender

Fossil Fuels
Thomas S. Ahlbrandt

Fourier, Jean Baptiste Joseph (1768-1830)
David Keston

Freight Movement
John Zumerchik
Jack Lanigan, Sr.

Fuel Cells
Christopher Borroni-Bird

Fuel Cell Vehicles
Christopher Borroni-Bird

Fuller, R. Buckminster, Jr. (1895-1983)
Linda Sargent Wood

Fulton, Robert (1765-1815)
Dennis R. Diehl

Furnaces and Boilers
Esher Kweller
Roger McDonald

Futures
Frank R. Power

G

Gasoline and Additives
Sanford L. Moskowitz

Gasoline Engines
Charles A. Amann

Geography and Energy Use
Vaclav Smil

Geothermal Energy
Ernest L. McFarland

Gibbs, Josiah Willard (1839-1903)
Robert J. Deltete

Government Agencies
John Zumerchik

Government Intervention in Energy Markets
John Zumerchik

Gravitational Energy
Joseph Priest

Green Energy
Robert L. Bradley, Jr.

H

Heat and Heating
Joseph Priest

Heat Pumps
Bernard A. Nagengast

Heat Transfer
W. Dan Turner
David Claridge

Heaviside, Oliver (1850-1925)
Alan S. Heather

Helmholtz, Hermann von (1821-1894)
Joseph F. Mulligan

Hertz, Henrich Rudolph (1857-1894)
Joseph F. Mulligan

Historical Perspectives and Social Consequences
Vaclav Smil

Houdry, Eugene Jules (1892-1962)
Barry L. Tarmy

Hybrid Vehicles
Andrew F. Burke

Hydroelectric Energy
Linda Doman

Hydrogen
Joan M. Ogden

I

Import/Export Market for Energy
Douglas B. Reynolds

Industry and Business, Energy as a Factor of Production in
Douglas B. Reynolds

Industry and Business, Productivity and Energy Efficiency in
Joseph J. Romm

Insulation
David W. Yarbrough

Ipatieff, Vladimir Nikolaevitch (1867-1952)
Sanford L. Moskowitz

J

Joule, James Prescott (1818-1889)
Joseph F. Mulligan

K

Kamerlingh Onnes, Heike (1853-1926)
Paul H. E. Meijer

Kerosene
Herman Bieber

Kinetic Energy
Joseph Priest

Kinetic Energy, Historical Evolution of the Use of
Terry S. Reynolds

L

Laplace, Pierre Simon (1749-1827)
Leonard S. Taylor

Lasers
Karl J. Hejlik

Lewis, Warren K. (1882-1975)
Sanford L. Moskowitz

Lighting
Mark S. Rea

Liquefied Petroleum Gas
Roy W. Willis

Locomotive Technology
Robert E. Gallamore

Lyell, Charles (1797-1875)
Colin A. Russell

M

Magnetic Levitation
Richard D. Thornton

Magnetism and Magnets
Joseph Priest

Magnetohydrodynamics
Hugo Karl Messerle

Manufacturing
Ernst Worrell

Market Imperfections
Alan H. Sanstad

Market Transformation
Dan W. York
Mark E. Hanson

Mass Transit
Paul M. Schimek

Materials
Robert H. Doremus

Matter and Energy
Don Lichtenberg

Maxwell, James Clerk (1831-1879)
C. W. F. Everitt

Mayer, Julius Robert (1814-1878)
Joseph F. Mulligan

Mechanical Transmission of Energy
Don C. Hopkins

Meitner, Lise (1878-1968)
Joseph F. Mulligan

Methane
John Zumerchik

Methanol
John Zumerchik

Military Energy Use, Historical Aspects of
Shannon A. Brown

Molecular Energy
Ellen J. Zeman

N

Nanotechnologies
J. Storrs Hall

National Energy Laboratories
John Zumerchik

Natural Gas, Consumption of
Arlon R. Tussing

Natural Gas, Processing and Conversion of
John Zumerchik
Herman Bieber

Natural Gas, Transportation, Distribution, and Storage of
Robert V. McCormick

Nernst, Walther Hermann (1864-1941)
Tyno Abdul-Redah

Newcomen, Thomas (1663-1729)
Dennis R. Diehl

Newton, Isaac (1642-1727)
Brian S. Baigrie

Nitrogen Cycle
Deborah L. Mowery

Nuclear Energy
Ernest L. McFarland

Nuclear Energy, Historical Evolution of the Use of
David B. Sicilia

Nuclear Fission
Joseph Priest

Nuclear Fission Fuel
Brian F. Thumm

Nuclear Fusion
D. W. Ignat

Nuclear Waste
Donald Williams

O

Ocean Energy
Stuart E. Baird

Oersted, Hans Christian (1777-1851)
Leonard S. Taylor

Office Equipment
Mary Ann Piette
Bruce Nordman

Oil and Gas, Drilling for
John Zumerchik
Elena Subia Melchert

Oil and Gas, Exploration for
Raymon L. Brown

Oil and Gas, Production of
George W. Hinman
Nancy W. Hinman

Onsager, Lars (1903-1976)
P. C. Hemmer

Otto, Nikolaus August (1832-1891)
Robert Sier

P

Parsons, Charles Algernon (1854-1931)
Robert Strunz

Particle Accelerators
Don Lichtenberg

Perpetual Motion
Don C. Hopkins

Petroleum Consumption
Sanford L. Moskowitz

Piezoelectric Energy
Joseph Priest

Potential Energy
Joseph Priest

Power
Joseph Priest

Pressure
Joseph Priest

Propellers
Jeremy Kinney

Property Rights
Gary Libecap

Propulsion
Don C. Hopkins

Q

Railway Passenger Service
Louis S. Thompson

Rankine, William John Macquorn (1820-1872)
Dennis R. Diehl

Refineries
James H. Gary

Refining, History of
Sanford L. Moskowitz

Refrigerators and Freezers
J. Benjamin Horvay

Regulation and Rates for Electricity
Lawrence J. Hill

Renewable Energy
Joseph Priest

Reserves and Resources
Thomas S. Ahlbrandt

Residual Fuels
Charles W. Siegmund

Risk Assessment and Management
Ahmet E. Kocagil

Rocket Propellants
Peter A. Gorin

S

Sakharov, Andrei Dmitrievich (1921-1989)
Andrew M. Sessler

Savery, Thomas (1650-1715)
Dennis R. Diehl

Scientific and Technical Understanding
C. W. F. Everitt

Seebeck, Thomas Johann (1770-1831)
Barbara Flume-Gorczyca

Seismic Energy
Charles K. Scharnberger

Ships
A. Douglas Carmichael
Clifford A. Whitcomb

Siemens, Ernst Werner von (1816-1892)
Barbara Flume-Gorczyca

Smeaton, John (1724-1792)
Robert Sier

Solar Energy
Allan R. Hoffman

Solar Energy, Historical Evolution of the Use of
J. Bernard Moore

Spacecraft Energy Systems
Peter A. Gorin

Sperry, Elmer Ambrose (1860-1930)
Robin B. Goodale

Steam Engines
Ken Helmick

Stephenson, George (1781-1848)
Dennis R. Diehl

Stirling, Robert (1790-1878)
Robert Sier

Stirling Engines
Brent H. Van Arsdell

Storage
Joseph Priest

Storage Technology
Joseph Priest

Subsidies and Energy Costs
Richard L. Gordon

Supply and Demand and Energy Prices
Carol Dahl

Sustainable Resources
John Zumerchik

Synthetic Fuel
John Zumerchik
Herman Bieber

T

Taxation of Energy
Mark Mazur
Thomas A. Barthold

Tesla, Nikola (1856-1943)
James D. Allan

Thermal Energy
Joseph Priest

Thermodynamics
John R. Ray

Thompson, Benjamin (Count Rumford) (1753-1814)
Joseph F. Mulligan

Thomson, Joseph John (1856-1940)
Leif Gerward
Christopher Cousins

Thomson, William (Lord Kelvin) (1824-1907)
Crosbie Smith

Tires
Joseph D. Walter

Townes, Charles Hard (1915-)
Karl J. Hejlik

Traffic Flow Management
Randall Guensler

Transformers
Brian F. Thumm

Transportation, Evolution of Energy Use and
William L. Withuhn

Trevithick, Richard (1771-1833)
Karl A. Petersen

Tribology
Gary C. Barber
Barbara A. Oakley

True Energy Costs
Jonathan G. Koomey

Turbines, Gas
Ronald L. Bannister

Turbines, Steam
Ronald L. Bannister

Turbines, Wind
M. Maureen Hand

U

Units of Energy
Joseph Priest

Utility Planning
James N. Maughn

V

Volta, Alessandro (1745-1827)
Giuliano Pancaldi

W

Waste-to-Energy Technology
Michael L. Murphy

Water Heating
James D. Lutz

Watt, James (1736-1819)
Karl A. Petersen

Waves
Joseph Priest

Wheatstone, Charles (1802-1875)
Brian Bowers

Windows
Alecia Ward

Work and Energy
Joseph Priest

LIST OF CONTRIBUTORS

Tyno Abdul-Redah
Technical University, Berlin, Germany
Nernst, Walther Hermann (1864-1941)

Albert Abramson
Las Vegas, NV
Alexanderson, Ernst Frederik Werner

Samuel N. Addy
University of Alabama, Tuscaloosa
Economic Externalities

Thomas S. Ahlbrandt
U.S. Geological Survey, Denver
Fossil Fuels
Reserves and Resources

Hashem Akbari
Lawrence Berkeley National Laboratory, Berkeley, CA
Control Systems, History of (with Anibal T. De Almeida)
Cool Communities (with Arthur H. Rosenfeld)
Energy Management Control Systems (with Anibal T. De Almeida)

James D. Allan
Cypress Semiconductor Corporation, Colorado Springs
Tesla, Nikola (1856-1943)

Charles A. Amann
KAB Engineering, Bloomfield Hills, MI
Automobile Performance
Drivetrains
Engines
Gasoline Engines

John D. Anderson
National Air and Space Museum, Smithsonian Institution, Washington, DC
Aerodynamics
Aircraft

Brian S. Baigrie
University of Toronto, Ontario, Canada
Newton, Isaac (1642-1727)

Stuart E. Baird
International Council for Local Environmental Initiatives, Toronto, Ontario, Canada
Ocean Energy

Dennis G. Baker
University of Michigan, Ann Arbor
Atmosphere (with Anita Baker-Blocker)

Anita Baker-Blocker
Ann Arbor, MI
Atmosphere (with Dennis G. Baker)

Ronald L. Bannister
American Society of Engineers, Winter Springs, FL
Turbines, Gas
Turbines, Steam

Gary C. Barber
Oakland University
Tribology (with Barbara A. Oakley)

Dennis Barnaal
Luther College
Electricity

Thomas A. Barthold
Joint Committee on Taxation, U.S. Congress
Taxation of Energy (wth Mark Mazur)

Stephanie J. Battles
U.S. Energy Information Administration, Washington, DC
Energy Intensity Trends

Donald A. Bender
Trinity Flywheel Power, Livermore, CA
Flywheels

Herman Bieber
Engineering Consultant, Kenilworth, NJ
Charcoal
Explosives and Propellants
Kerosene
Natural Gas, Processing and Conversion of (with John Zumerchik)
Synthetic Fuel (with John Zumerchik)

Roy Billinton
University of Saskatchewan, Saskatoon, Canada
Electric Power, System Reliability and

Christopher Borroni-Bird
General Motors, Oakland Township, MI
Fuel Cells
Fuel Cell Vehicles

Brian Bowers
Science Museum, London, United Kingdom
Wheatstone, Charles (1802-1875)

Robert L. Bradley, Jr.
Institute for Energy Research, Houston, TX
Green Energy

Robert N. Brady
Vancouver Community College, Vancouver, Canada
Diesel Cycle Engines
Diesel Fuel

Kenneth Brezinsky
University of Illinois, Chicago
Combustion

Allison Brody
Oklahoma City, OK
Biological Energy Use, Ecosystem Functioning of

Raymon L. Brown
Oklahoma Geological Survey, Norman, OK
Oil and Gas, Exploration for

Shannon A. Brown
Washington, DC
Military Energy Use, Historical Aspects of

Andrew F. Burke
Institute of Transportation Studies, University of California, Davis
Hybrid Vehicles

Tom Butler
Institute of Ecosystem Studies, Cornell University
Acid Rain

A. Douglas Carmichael
Massachusetts Institute of Technology
Ships (with Clifford A. Whitcomb)

David Claridge
Texas A&M University
Heat Transfer (with W. Dan Turner)

David Conover
National Evaluation Services, Inc., Falls Church, VA
Building Design, Energy Codes and (with Jeffrey A. Johnson)

Roger Conway
Office of Energy Policy and New Uses, U.S. Department of Agriculture, Washington, DC
Agriculture (with Hosein Shapouri)

Christopher Cousins
University of Exeter, Exeter, United Kingdom
Thomson, Joseph John (1856-1940) (with Leif Gerward)

Carol Dahl
Colorado School of Mines
Supply and Demand and Energy Prices

Burtron H. Davis
University of Kentucky, Lexington
Catalysts (with Joe W. Hightower)

Anibal T. De Almeida
University of Coimbra, Coimbra, Portugal
Control Systems, History of (with Hashem Akbari)
Energy Management Control Systems (with Hashem Akbari)

Robert J. Deltete
Seattle University
Gibbs, Josiah Willard (1839-1903)

Fred I. Denny
Louisana State University
Electric Vehicles (with John Zumerchik)
Environmental Problems and Energy Use

Joseph R. DesJardins
St. John's University
Ethical and Moral Aspects of Energy Use

Dennis R. Diehl
Newcomen Society of the United States, Exton, PA
Fulton, Robert (1765-1815)
Newcomen, Thomas (1663-1729)

Rankine, William John Macquorn (1820-1872)
Savery, Thomas (1650-1715)
Stephenson, George (1781-1848)

David E. Dismukes
Lousiana State University
Electric Power, Generation of

Linda Doman
U.S. Energy Information Administration, Washington, DC
Hydroelectric Energy

Robert H. Doremus
Rensselaer Polytechnic Institute
Materials

Joseph Eto
Lawrence Berkeley National Laboratory, Berkeley, CA
Demand-Side Management

C. W. F. Everitt
Stanford University
Maxwell, James Clerk (1831-1879)
Scientific and Technical Understanding

Anthony Fainberg
Bethesda, MD
Coal, Consumption of

Bernard Finn
National Museum of American History, Smithsonian Institution, Washington, DC
Electricity, History of

Barbara Flume-Gorczyca
Bonn, Germany
Clausius, Rudolf Julius Emmanuel (1822-1888)
Seebeck, Thomas Johann (1770-1831)
Siemens, Ernst Werner von (1816-1892)

Robert E. Gallamore
Transportation Technology Center, Inc., Pueblo, CO
Locomotive Technology

James H. Gary
Colorado School of Mines
Refineries

Leif Gerward
Technical University of Denmark, Kongens Lyngby, Denmark
Thomson, Joseph John (1856-1940) (with Christopher Cousins)

David B. Goldstein
Natural Resources Defense Council, San Francisco, CA
Appliances

Robin B. Goodale
Houston, TX
Sperry, Elmer Ambrose (1860-1930)

Richard L. Gordon
Pennsylvania State University
Energy Economics
Subsidies and Energy Costs

Peter A. Gorin
National Air and Space Museum, Smithsonian Institution, Washington, DC
Rocket Propellants
Spacecraft Energy Systems

Kenneth Green
Reason Public Policy Institute, Austin, TX
Air Pollution
Climatic Effects

Randall Guensler
Georgia Institute of Technology
Emission Control, Vehicle
Traffic Flow Management

J. Storrs Hall
Institute for Molecular Manufacturing, Palo Alto, CA
Nanotechnologies

M. Maureen Hand
National Renewable Energy Laboratory, Golden, CO
Turbines, Wind

Mark E. Hanson
Energy Center of Wisconsin, Madison
Market Transformation (with Dan W. York)

David E. Harris
University of Southern Maine
Biological Energy Use, Cellular Processes of

Alan S. Heather
Radio Society of Great Britain, Potters Bar, United Kingdom
Heaviside, Oliver (1850-1925)

Karl J. Hejlik
Orlando, FL
Lasers
Townes, Charles Hard (1915-)

Gilbert Held
Macon, GA
Communications and Energy

Jamie N. Heller
PHB Hagler Bailly, Inc., Washington, DC
Coal, Transportation and Storage of (with Stan M. Kaplan)

Ken Helmick
General Motors, Warren, MI
Steam Engines

P. C. Hemmer
Norwegian University of Science and Technology, Trondheim, Norway
Onsager, Lars (1903-1976)

Charles E. Hickman
Tennessee Technological University (ret.)
Emission Control, Power Plant

Joe W. Hightower
Rice University
Catalysts (with Burtron H. Davis)

Lawrence J. Hill
Honolulu, HI
Regulation and Rates for Electricity

George W. Hinman
Washington State University, Pullman
Oil and Gas, Production of (with Nancy W. Hinman)

Nancy W. Hinman
University of Montana, Missoula
Oil and Gas, Production of (with George W. Hinman)

Allan R. Hoffman
U.S. Department of Energy, Washington, DC
Solar Energy

Don C. Hopkins
South Dakota School of Mines and Technology
Mechanical Transmission of Energy
Perpetual Motion
Propulsion

Stanley H. Horowitz
Consultant, Columbus, OH
Electric Power, System Protection, Control, and Monitoring of

J. Benjamin Horvay
Consulting Engineer, Isle of Palms, SC
Refrigerators and Freezers

D. W. Ignat
International Atomic Energy Agency, Vienna, Austria
Nuclear Fusion

Jeffrey A. Johnson
New Buildings Institute, White Salmon, WA
Building Design, Commercial (with William Steven Taber, Jr.)
Building Design, Energy Codes and (with David Conover)

Tina M. Kaarsberg
Office of Power Technologies, U.S. Department of Energy
Efficiency of Energy Use (with Arthur H. Rosenfeld and Joseph J. Romm)

Stan M. Kaplan
PHB Hagler Bailly, Inc., Washington, DC
Coal, Transportation and Storage of (with Jamie N. Heller)

David Keston
Siemens UK, Bracknell, United Kingdom
Fourier, Jean Baptiste Joseph (1768-1830)

Jeremy Kinney
Auburn University
Propellers

Ahmet E. Kocagil
Pensylvania State University, University Park, PA
Risk Assessment and Management

Jonathan G. Koomey
Lawrence Berkeley National Laboratory, Berkeley, CA
True Energy Costs

Japhet Koteen
U.S. Environmental Protection Agency, Washington, DC
Building Design, Residential (with Sam Rashkin)

Satish Kumar
Lawrence Berkeley National Laboratory, Berkeley, CA
Air Quality, Indoor

Esher Kweller
U.S. Department of Energy, Washington, DC
Furnaces and Boilers (with Roger McDonald)

Jack Lanigan, Sr.
Mi-Jack Products, Hazel Crest, IL
Freight Movement (with John Zumerchik)

Gary Libecap
University of Arizona, Tucson
Property Rights

Don Lichtenberg
Indiana University, Bloomington
Big Bang Theory
Matter and Energy
Particle Accelerators

Jiang Long
Global Bearings, Inc., Tulsa, OK
Bearings

James D. Lutz
Lawrence Berkeley National Laboratory, Berkeley, CA
Water Heating

James N. Maughn
Engineering Consultant, Birmingham, AL
Utility Planning

Mark Mazur
U.S. Energy Information Administration, U.S. Department of Energy, Washington, DC
Taxation of Energy (with Thomas A. Barthold)

Robert V. McCormick
Troy, MI
Natural Gas, Transportation, Distribution, and Storage of

Roger McDonald
Brookhaven National Laboratory, Upton, NY
Furnaces and Boilers (with Esher Kweller)

Alan B. McEwen
Covalent Associates, Inc., Woburn, MA
Capacitors and Ultracapacitors

Ernest L. McFarland
University of Guelph, Ontario, Canada
Geothermal Energy
Nuclear Energy

Sue McNeil
University of Illinois, Chicago
Capital Investment Decisions

Alan K. Meier
Lawrence Berkeley National Laboratory, Berkeley, CA
Conservation Supply Curves (with Arthur H. Rosenfeld)

Paul H. E. Meijer
Catholic University of America
Kamerlingh Onnes, Heike (1853-1926)

Elena Subia Melchert
U.S. Department of Energy, Washington, DC
Oil and Gas, Drilling for (with John Zumerchik)

Hugo Karl Messerle
Turramurra, Australia
Magnetohydrodynamics

James G. Mills
U.S. Federal Trade Commission, Washington, DC
Efficiency of Energy Use, Labeling of

J. Bernard Moore
Hyattsville, MD
Solar Energy, Historical Evolution of the Use of

Sanford L. Moskowitz
Chemical Heritage Foundation, Philadelphia, PA
Gasoline and Additives
Ipatieff, Vladimir Nikolaevitch (1867-1952)
Lewis, Warren K. (1882-1975)
Petroleum Consumption
Refining, History of

Deborah L. Mowery
Sandia National Laboratories, Albuquerque, NM
Biofuels
Nitrogen Cycle

David Mulcahy
Polytechnic University, Brooklyn, NY
Animal and Human Energy

Michael R. Muller
Rutgers University
Auditing of Energy Use

Joseph F. Mulligan
University of Maryland, Baltimore County
Carnot, Nicolas Leonard Sadi (1796-1832)
Fermi, Enrico (1901-1954)
Helmholtz, Hermann von (1821-1894)
Hertz, Henrich Rudolph (1857-1894)

Joule, James Prescott (1818-1889)
Mayer, Julius Robert (1814-1878)
Meitner, Lise (1878-1968)
Thompson, Benjamin (Count Rumford) (1753-1814)

Michael L. Murphy
Energy Products of Idaho, Coeur d'Alene, ID
Waste-to-Energy Technology

Laura Nader
University of California, Berkeley
Culture and Energy Usage

Bernard A. Nagengast
Consulting Engineer, Sidney, OH
Air Conditioning
Heat Pumps

Bruce Nordman
Lawrence Berkeley National Laboratory, Berkeley, CA
Office Equipment (with Mary Ann Piette)

Barbara A. Oakley
Oakland University
Tribology

Joan M. Ogden
Center for Energy and Environmental Studies, Princeton University
Hydrogen

Abraham Pais
Rockefeller University
Einstein, Albert (1879-1955)

John A. Palmer
Colorado School of Mines
Electric Power Substations
Electric Power Transmission and Distribution Systems

Giuliano Pancaldi
University of Bologna, Italy
Volta, Alessandro (1745-1827)

William S. Peirce
Case Western Reserve University
Economically Efficient Energy Choices

Karl A. Petersen
Steam Automobile Club of America, Washington, IL
Trevithick, Richard (1771-1833)
Watt, James (1736-1819)

Morris A. Pierce
University of Rochester
Cogeneration Technologies
District Heating and Cooling
Domestic Energy Use

Mary Ann Piette
Lawrence Berkeley National Laboratory, Berkeley, CA
Office Equipment (with Bruce Nordman)

Frank R. Power
National Institute of Standards and Technology, Washington, DC
Futures

Joseph Priest
Miami University, OH
Cogeneration
Conservation of Energy
Elastic Energy
Gravitational Energy
Heat and Heating
Kinetic Energy
Magnetism and Magnets
Nuclear Fission
Piezoelectric Energy
Potential Energy
Power
Pressure
Renewable Energy
Storage
Storage Technology
Thermal Energy
Units of Energy
Waves
Work and Energy

Douglas Quinney
Keele University, Keele, United Kingdom
Archimedes (287-212 BC)
Bernoulli, Daniel (1700-1782)

Sam Rashkin
U.S. Environmental Protection Agency, Washington, DC
Building Design, Residential (with Japhet Koteen)

John R. Ray
Clemson University
Thermodynamics

Mark S. Rea
Lighting Research Center, Rensselaer Polytechnic Institute
Lighting

Douglas B. Reynolds
University of Alaska, Fairbanks
Import/Export Market for Energy
Industry and Business, Energy as a Factor of Production in

Terry S. Reynolds
Michigan Technological University
Kinetic Energy, Historical Evolution of the Use of

Joseph J. Romm
Center for Energy and Climate Solutions, Annandale, VA
Efficiency of Energy Use (with Arthur H. Rosenfeld and Tina M. Kaarsberg)
Industry and Business, Productivity and Energy Efficiency in

Arthur H. Rosenfeld
Commissioner, California Energy Commission
Conservation Supply Curves (with Alan K. Meier)
Cool Communities (with Hashem Akbari)
Efficiency of Energy Use (with Tina M. Kaarsberg and Joseph J. Romm)

Colin A. Russell
The Open University, Milton Keynes, United Kingdom
Lyell, Charles (1797-1875)

Alan H. Sanstad
Lawrence Berkeley National Laboratory, Berkeley, CA
Market Imperfections

Linda Sargent Wood
University of Maryland, College Park
Carson, Rachel
Fuller, R. Buckminster, Jr. (1895-1983)

Charles K. Scharnberger
Millersville University
Seismic Energy

Paul M. Schimek
U.S. Department of Transportation, Cambridge, MA
Mass Transit

John Schubert
Adventure Cycling Association
Bicycling (with John Zumerchik)

Silvan S. Schweber
Brandeis University
Bethe, Hans Albrecht (1906-)

Andrew M. Sessler
Lawrence Berkeley National Laboratory, Berkeley, CA
Sakharov, Andrei Dmitrievich (1921-1989)

Hosein Shapouri
Office of Energy Policy and New Uses, U.S. Department of Agriculture, Washington, DC
Agriculture (with Roger Conway)

David B. Sicilia
University of Maryland, College Park
Diesel, Rudolph (1858-1913)
Nuclear Energy, Historical Evolution of the Use of

Charles W. Siegmund
Morris Plains, NJ
Residual Fuels

Robert Sier
Chelmsford, United Kingdom
Ampère, André-Marie (1775-1836)
Black, Joseph (1728-1799)
Ericsson, John (1803-1899)
Otto, Nikolaus August (1832-1891)
Smeaton, John (1724-1792)
Stirling, Robert (1790-1878)

Scott B. Sitzer
U.S. Energy Information Administration, Washington, DC
Economic Growth and Energy Consumption

Vaclav Smil
University of Manitoba, Winnipeg, Canada
Geography and Energy Use
Historical Perspectives and Social Consequences

Crosbie Smith
University of Kent, Canterbury, United Kingdom
Thomson, William (Lord Kelvin) (1824-1907)

Stig Steenstrup
Niels Bohr Institute, Copenhagen, Denmark
Becquerel, Alexandre-Edmund (1820-1891)

Richard L. Stroup
Montana State University, Bozeman
Environmental Economics

Robert Strunz
University of Limerick, Ireland
Parsons, Charles Algernon (1854-1931)

Ronald J. Sutherland
Center for the Advancement of Energy Markets, Burke, VA
Efficiency of Energy Use, Economic Concerns and

William Steven Taber, Jr.
Princeton Development Corporation, Sausalito, CA
Building Design, Commercial (with Jeffrey A. Johnson)

Barry L. Tarmy
TBD Technology, Berkeley Heights, NJ
Chemical Energy, Historical Evolution of the Use of
Houdry, Eugene Jules (1892-1962)

Leonard S. Taylor
University of Maryland, College Park
Edison, Thomas Alva (1847-1931)
Faraday, Michael (1791-1867)
Laplace, Pierre Simon (1749-1827)
Oersted, Hans Christian (1777-1851)

William F. Taylor
Consultant, Mountainsaide, NJ
Aviation Fuel

Louis S. Thompson
The World Bank, Washington, DC
Railway Passenger Service

Richard D. Thornton
Concord, MA
Magnetic Levitation

Brian F. Thumm
St. Francisville, LA
Nuclear Fission Fuel
Transformers

Jerry C. Tien
University of Missouri, Rolla
Coal, Production of

W. Dan Turner
Texas A&M University
Heat Transfer (with David Claridge)

Arlon R. Tussing
University of Alaska, Anchorage
Natural Gas, Consumption of

Brent H. Van Arsdell
American Stirling Company
Stirling Engines

F. Robert van der Linden
National Air and Space Museum, Smithsonian Institution, Washington, DC
Air Travel

Margaret H. Venable
Georgia Perimeter College
Curie, Marie Sklodowska (1867-1934) (with T. Leon Venable)

T. Leon Venable
Agnes Scott College
Curie, Marie Sklodowska (1867-1934) (with Margaret H. Venable)

Joseph D. Walter
University of Akron
Tires

Alecia Ward
Midwest Energy Efficiency Alliance, Chicago, IL
Windows

Clifford A. Whitcomb
Massachusetts Institute of Technology
Ships (with A. Douglas Carmichael)

Donald Williams
Hope College
Nuclear Waste

Roy W. Willis
Propane Education and Research Council, Washington, DC
Liquefied Petroleum Gas

William L. Withuhn
Smithsonian Institution, Washington, DC
Transportation, Evolution of Energy Use and

Ernst Worrell
Lawrence Berkeley National Laboratory, Berkeley, CA
Manufacturing

David W. Yarbrough
Tennessee Technological University
Insulation

Dan W. York
Energy Center of Wisconsin, Madison
Market Transformation (with Mark E. Hanson)

Ellen J. Zeman
Burlington, VT
Molecular Energy

John Zumerchik
New York, NY
Alternative Fuels and Vehicles
Batteries
Behavior
Bicycling (with John Schubert)
Consumption
Electric Motor Systems
Electric Vehicles (with Fred I. Denny)
Freight Movement (with Jack Lanigan, Sr.)
Government Agencies
Government Intervention in Energy Markets
Methane
Methanol
National Energy Laboratories
Natural Gas, Processing and Conversion of (with Herman Bieber)
Oil and Gas, Drilling for (with Elena Subia Melchert)
Sustainable Resources
Synthetic Fuel (with Herman Bieber)

COMMON ABBREVIATIONS AND MATHEMATICAL SYMBOLS

=	equals; double bond
≠	not equal to
≡	identically equal to; equivalent to; triple bond
~	approximately
≈	approximately equal to
≅	congruent to; approximately equal to
∝	proportional to
<	less than
>	greater than
<<	much less than
>>	much greater than
≤	less than or equal to
≥	greater than or equal to
→	approaches, tends to; yields; is replaced by
⇒	implies; is replaced by
⇐	is implied by
⇓	mutually implies
⇔	if and only if
⇄	reversible reaction
⊥	perpendicular to
∥	parallel to
$\lvert\ \rvert$	absolute value of
+	plus
–	minus
/	divided by
×	multiplied by
±	plus or minus
√	radical
∫	integral
Σ	summation
∐	product
∂	partial derivative
°	degree
°B	degrees Baumé
°C	degrees Celsius (centigrade)
°F	degrees Fahrenheit
!	factorial

"	second
'	minute
ε_0	electric constant
μ	micro-
μ_0	magnetic constant
μA	microampere
μA h	microampere hour
μC	microcoulomb
μF	microfarad
μg	microgram
μK	microkelvin
μm	micrometer
μm	micron
μm Hg	microns of mercury
μmol	micromole
μs, μsec	microsecond
μu	microunit
μΩ	microhm
σ	Stefan-Boltzmann constant
Ω	ohm
Ωcm	ohm centimeter
Ωcm/(cm/cm^3)	ohm centimeter per centimeter per cubic centimeter
A	ampere
Å	angstrom
a	atto-
A_s	atmosphere, standard
abbr.	abbreviate; abbreviation
abr.	abridged; abridgment
Ac	Actinium
ac	alternating-current
aF	attofarad
af	audio-frequency
Ag	silver
A h	ampere hour
AIP	American Institute of Physics
Al	aluminum
alt	altitude

Am	americium
AM	amplitude-modulation
A.M.	ante meridiem
amend.	amended; amendment
annot.	annotated; annotation
antilog	antilogarithm
app.	appendix
approx	approximate (in subscript)
Ar	argon
arccos	arccosine
arccot	arccotangent
arccsc	arccosecant
arc min	arc minute
arcsec	arcsecant
arcsin	arcsine
arg	argument
As	arsenic
At	astatine
At/m	ampere turns per meter
atm	atmosphere
at. ppm	atomic parts per million
at. %	atomic percent
atu	atomic time unit
AU	astronomical unit
a.u.	atomic unit
Au	gold
av	average (in subscript)
b	barn
B	boron
Ba	barium
bcc	body-centered-cubic
B.C.E.	before the common era
Be	beryllium
Bi	biot
Bi	bismuth
Bk	berkelium
bp	boiling point
Bq	becquerel
Br	bromine
Btu, BTU	British thermal unit
C	carbon
c	centi-
c.	circa, about, approximately
C	coulomb
c	speed of light
Ca	calcium
cal	calorie
calc	calculated (in subscript)
c.c.	complex conjugate
CCD	charge-coupled devices
Cd	cadmium
cd	candela
CD	compact disc
Ce	cerium
C.E.	common era
CERN	European Center for Nuclear Research
Cf	californium
cf.	confer, compare
cgs	centimeter-gram-second (system)
Ci	curie
Cl	chlorine
C.L.	confidence limits
c.m.	center of mass
cm	centimeter
Cm	curium
cm^3	cubic centimeter
Co	cobalt
Co.	Company
coeff	coefficient (in subscript)
colog	cologarithm
const	constant
Corp.	Corporation
cos	cosine
cosh	hyperbolic cosine
cot	cotangent
coth	hyperbolic cotangent
cp	candlepower
cP	centipoise
cp	chemically pure
cpd	contact potential difference
cpm	counts per minute
cps	cycles per second
Cr	chromium
cS	centistoke
Cs	cesium
csc	cosecant
csch	hyperbolic cosecant
Cu	copper
cu	cubic
cw	continuous-wave
D	Debye
d	deci-
da	deka-
dB, dBm	decibel
dc	direct-current
deg	degree
det	determinant
dev	deviation
diam	diameter
dis/min	disintegrations per minute

dis/s	disintegrations per second
div	divergence
DNA	deoxyribose nucleic acid
DOE, DoE	U.S. Department of Energy
Dy	dysprosium
dyn	dyne
E	east
e	electronic charge
E	exa-
e, exp	exponential
e/at.	electrons per atom
e b	electron barn
e/cm^3	electrons per cubic centimeter
ed. (pl., eds.)	editor
e.g.	exempli gratia, for example
el	elastic (in subscript)
emf, EMF	electromotive force
emu	electromagnetic unit
Eng.	England
EPA	U.S. Environmental Protection Agency
Eq. (pl., Eqs.)	equation
Er	erbium
EREN	Energy Efficiency and Renewable Energy Network
erf	error function
erfc	error function (complement of)
Es	einsteinium
e.s.d.	estimated standard deviation
esu	electrostatic unit
et al.	et alii, and others
etc.	et cetera, and so forth
e.u.	electron unit
eu	entropy unit
Eu	europium
eV	electron volt
expt	experimental (in subscript)
F	farad
F	Faraday constant
f	femto-
F	fermi
F	fluorine
fc	foot-candle
fcc	face-centered-cubic
Fe	iron
FERC	Federal Energy Regulatory Commission (U.S.)
fF	femtofarad
Fig. (pl., Figs.)	figure
fL	foot-lambert
fm	femtometer
Fm	fermium
FM	frequency-modulation
f. (pl., ff.)	following
fpm	fissions per minute
Fr	francium
Fr	franklin
fs	femtosecond
ft	foot
ft lb	foot-pound
ft lbf	foot-pound-force
f.u.	formula units
g	acceleration of free fall
G	gauss
G	giga-
g	gram
G	gravitational constant
Ga	gallium
Gal	gal (unit of gravitational force)
gal	gallon
g-at.	gram-atom
g.at. wt	gram-atomic-weight
Gc/s	gigacycles per second
Gd	gadolinium
Ge	germanium
GeV	giga-electron-volt
GHz	gigahertz
Gi	gilbert
grad	gradient
GV	gigavolt
Gy	gray
h	hecto-
H	henry
h	hour
H	hydrogen
h	Planck constant
H.c.	Hermitian conjugate
hcp	hexagonal-close-packed
He	helium
Hf	hafnium
hf	high-frequency
hfs	hyperfine structure
hg	hectogram
Hg	mercury
Ho	holmium
hp	horsepower
Hz	hertz
I	iodine
IAEA	International Atomic Energy Agency
ICT	International Critical Tables
i.d.	inside diameter

i.e.	id est, that is
IEA	International Energy Agency
IEEE	Institute of Electrical and Electronics Engineers
if	intermediate frequency
Im	imaginary part
in.	inch
In	indium
Inc.	Incorporated
inel	inelastic (in subscript)
ir, IR	infrared
Ir	iridium
J	joule
Jy	jansky
k, k_b	Boltzmann's constant
K	degrees Kelvin
K	kayser
k	kilo-
K	potassium
kA	kiloamperes
kbar	kilobar
kbyte	kilobyte
kcal	kilocalorie
kc/s	kilocycles per second
kdyn	kilodyne
keV	kilo-electron-volt
kG	kilogauss
kg	kilogram
kgf	kilogram force
kg m	kilogram meter
kHz	kilohertz
kJ	kilojoule
kK	kilodegrees Kelvin
km	kilometer
kMc/s	kilomegacycles per second
kn	knot
kOe	kilo-oersted
kpc	kiloparsec
Kr	krypton
ks, ksec	kilosecond
kt	kiloton
kV	kilovolt
kV A	kilovolt ampere
kW	kilowatt
kWh	kilowatt hour
kΩ	kilohm
L	lambert
L	langmuir
l, L	liter
La	lanthanum

LA	longitudinal-acoustic
lab	laboratory (in subscript)
lat	latitude
lb	pound
lbf	pound-force
lbm	pound-mass
LED	light emitting diode
Li	lithium
lim	limit
lm	lumen
lm/W	lumens per watt
ln	natural logarithm (base e)
LO	longitudinal-optic
log	logarithm
Lr	lawrencium
LU	Lorentz unit
Lu	lutetium
lx	lux
ly, lyr	light-year
M	Mach
M	mega-
m	meter
m	milli-
m	molal (concentration)
M	molar (concentration)
m_e	electronic rest mass
m_n	neutron rest mass
m_p	proton rest mass
MA	megaamperes
mA	milliampere
ma	maximum
mb	millibarn
mCi	millicurie
Mc/s	megacycles per second
Md	mendelevium
MeV	mega-electron-volt; million electron volt
Mg	magnesium
mg	milligram
mH	millihenry
mho	reciprocal ohm
MHz	megahertz
min	minimum
min	minute
mK	millidegrees Kelvin; millikelvin
mks	meter-kilogram-second (system)
mksa	meter-kilogram-second ampere
mksc	meter-kilogram-second coulomb
ml	milliliter
mm	millimeter
mmf	magnetomotive force

mm Hg	millimeters of mercury
Mn	manganese
MO	molecular orbital
Mo	molybdenum
MOE	magneto-optic effect
mol	mole
mol %, mole %	mole percent
mp	melting point
Mpc	megaparsec
mpg	miles per gallon
mph	miles per hour
MPM	mole percent metal
Mrad	megarad
ms, msec	millisecond
mt	metric tonne
Mt	million metric tonnes
mu	milliunit
MV	megavolt; million volt
mV	millivolt
MW	megawatt
MWe	megawatts of electric power
Mx	maxwell
mμm	millimicron
MΩ	megaohm
n	nano-
N	newton
N	nitrogen
N	normal (concentration)
N	north
N, N_A	Avogadro constant
Na	sodium
NASA	National Aeronautics and Space Administration
nb	nanobarn
Nb	niobium
Nd	neodymium
N.D.	not determined
NDT	nondestructive testing
Ne	neon
n/*f*	neutrons per fission
Ni	nickel
N_L	Loschmidt's constant
nm	nanometer
No	nobelium
No.	number
NOx, NO_x	oxides of nitrogen
Np	neper
Np	neptunium
NRC	Nuclear Regulatory Commission (U.S.)
ns, nsec	nanosecond
n/s	neutrons per second
n/s cm^2	neutrons per second per square centimeter
NSF	National Science Foundation (U.S.)
ns/m	nanoseconds per meter
O	oxygen
o()	of order less than
O()	of the order of
obs	observed (in subscript)
o.d.	outside diameter
Oe	oersted
ohm^{-1}	mho
Os	osmium
oz	ounce
P	peta-
P	phosphorus
p	pico-
P	poise
Pa	pascal
Pa	protactinium
Pb	lead
pc	parsec
Pd	palladium
PD	potential difference
pe	probable error
pF	picofarad
pl.	plural
P.M.	post meridiem
Pm	promethium
Po	polonium
ppb	parts per billion
p. (pl., pp.)	page
ppm	parts per million
Pr	praseodymium
psi	pounds per square inch
psi (absolute)	pounds per square inch absolute
psi (gauge)	pounds per square inch gauge
Pt	platinum
Pu	plutonium
Q, quad	quadrillion Btus
R	roentgen
Ra	radium
rad	radian
Rb	rubidium
Re	real part
Re	rhenium
rev.	revised
rf	radio frequency
Rh	rhodium
r.l.	radiation length
rms	root-mean-square

Rn	radon
RNA	ribonucleic acid
RPA	random-phase approximation
rpm	revolutions per minute
rps, rev/s	revolutions per second
Ru	ruthenium
Ry	rydberg
s, sec	second
S	siemens
S	south
S	stoke
S	sulfur
Sb	antimony
Sc	scandium
sccm	standard cubic centimeter per minute
Se	selenium
sec	secant
sech	hyperbolic secant
sgn	signum function
Si	silicon
SI	Systäme International
sin	sine
sinh	hyperbolic sine
SLAC	Stanford Linear Accelerator Center
Sm	samarium
Sn	tin
sq	square
sr	steradian
Sr	strontium
STP	standard temperature and pressure
Suppl.	Supplement
Sv	sievert
T	tera-
T	tesla
t	tonne
Ta	tantalum
TA	transverse-acoustic
tan	tangent
tanh	hyperbolic tangent
Tb	terbium
Tc	technetium
Td	townsend
Te	tellurium
TE	transverse-electric
TEM	transverse-electromagnetic
TeV	tera-electron-volt
Th	thorium
theor	theory, theoretical (in subscript)
THz	tetrahertz
Ti	titanium
Tl	thallium
Tm	thulium
TM	transverse-magnetic
TO	transverse-optic
tot	total (in subscript)
TP	temperature-pressure
tr, Tr	trace
trans.	translator, translators; translated by; translation
u	atomic mass unit
U	uranium
uhf	ultrahigh-frequency
uv, UV	ultraviolet
V	vanadium
V	volt
Vdc	Volts direct current
vol. (pl., vols.)	volume
vol %	volume percent
vs.	versus
W	tungsten
W	watt
W	West
Wb	weber
Wb/m^2	webers per square meter
wt %	weight percent
W.u.	Weisskopf unit
Xe	xenon
Y	yttrium
Yb	ytterbium
yr	year
Zn	zinc
Zr	zirconium

A

ACID RAIN

"Acid rain" is a popular term that can include all forms of precipitation (as well as fog and cloudwater) that is more acidic than expected from natural causes. Measurement of precipitation acidity at several remote sites around the world show natural background levels of acidity to be around pH of 5.1 to 5.2 (8 to 6 ueq/l H^+ or hydrogen in concentration, respectively). This compares with present annual average values of pH 4.3 to 4.4 (50 to 40 ueq/l H^+) for most of the northeastern and midwestern United States. Note that as pH decreases, H^+ concentration or acidity increases exponentially. Individual storms, especially in the summer, can often produce pH values below 3.5 (>300 ueq/l H^+). Cloudwater and fog often show even higher concentrations of acidity and this has major implications for high-elevation ecosystems such as mountain forests and water bodies. A more appropriate term than acid rain is "acid deposition," which includes both wet and dry deposition of acidic sulfur and nitrogen compounds to the earth's surface from the atmosphere.

Acid deposition is of greatest concern wherever there are large amounts of fossil fuel combustion upwind of an area. Eastern North America, large areas of Europe, and eastern Asia all receive acidic deposition. Acidic deposition is especially a concern when poorly buffered soils, with little acid-neutralizing capacity, are impacted. In North America, large areas of eastern Canada, the Adirondack Mountains of upstate New York, and sections of New England all are considered "acid sensitive " areas, where resistant bedrocks and thin soils prevent significant neutralization of acidity.

HISTORICAL PERSPECTIVE

Acidic deposition is not a new phenomena, as E. B. Cowling (1982) has noted. In 1872, the term "acid rain" was first known to be used by Angus Smith to describe the precipitation around Manchester, England. Smith analyzed the chemistry of the rain and attributed this acid rain to combustion of coal. He also noted damage from acid rain to plants and materials. C. Crowther and H. G. Ruston (1911) demonstrated gradients in rainfall acidity decreasing from the center of Leeds, England and associated the acidity with coal combustion. E. Gorham (1957, 1958) established that acid precipitation affects the acid-neutralizing capacity of lakes and bogs. A. Dannevig (1959) of Norway recognized the relationship between acid precipitation, lake and stream acidity, and the disappearance of fish. S. Oden (1968) used trajectory analysis to demonstrate that acid precipitation in Sweden was the result of long-range transport and transformation of sulfur emissions from England and central Europe. In 1972, Likens et al. identified acid precipitation in eastern North America. G. E. Likens and F. H. Bormann (1974) demonstrated its regional distribution in the eastern United States and indicated that the transformation of nitrogen oxides (NO_x), as well as sulfur dioxide (SO_2), adds to precipitation acidity. D. W. Schindler and his colleagues (1985) performed a whole lake acidification in Canada and documented the adverse decline of the lake food web at pH levels as high as 5.8. The issue of acid rain or deposition has generated a vast amount of knowledge and understanding of atmospheric and watershed processes, and research in the field continues today.

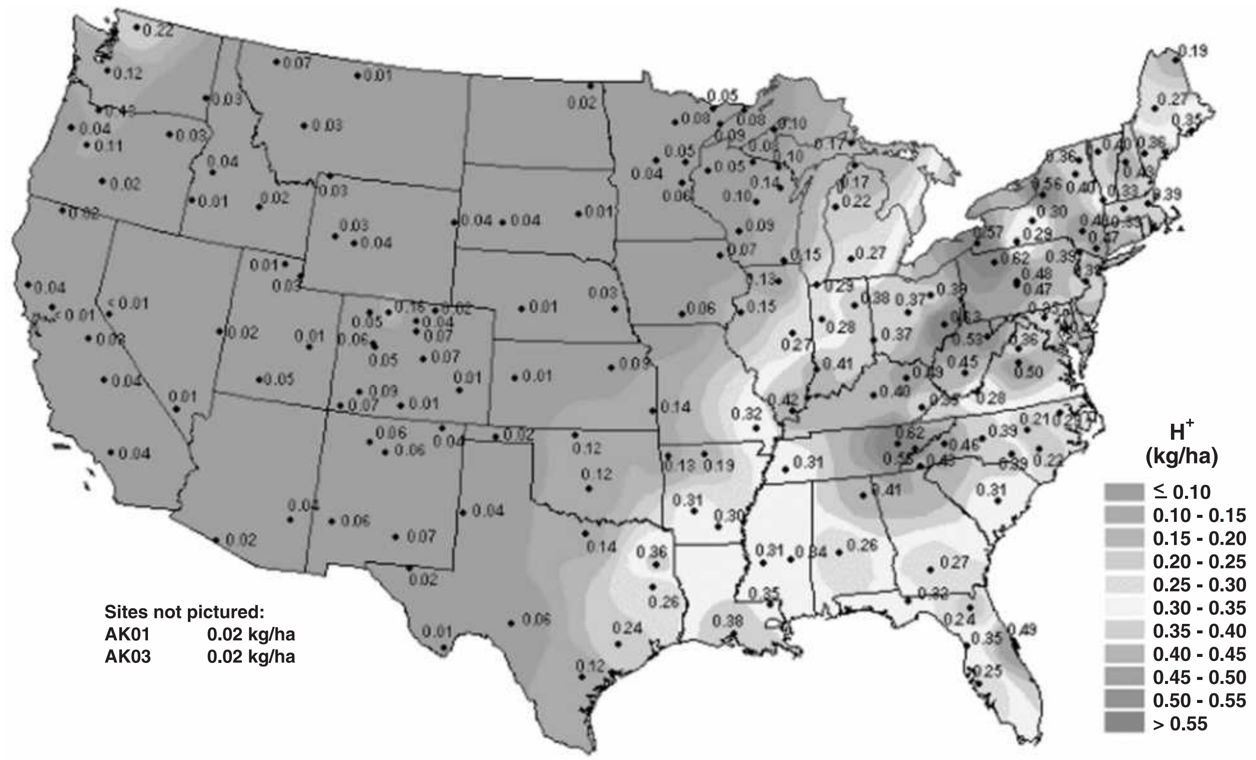


Figure 1.
Acid deposition is clearly a widespread phenomenon in the eastern third of the country.

FORMATION AND TRENDS

The formation of acidic deposition is largely from the combustion of fossil fuels and the smelting of sulfide ores. Minor natural sources exist such as the formation of hydrochloric and sulfuric acid from gaseous volcanic eruptions.

There are well over 100 gaseous and aqueous phase reactions that can lead to acid formation and more than fifty oxidizing agents and catalysts may be involved. However, in the simplest terms sulfur in fuels is oxidized to SO_2, and SO_2 in the atmosphere is further oxidized and hydrolyzed to sulfuric acid. Most nitric acid is formed by the fixation of atmospheric nitrogen gas (N_2) to NO_x (NO and NO_2) during high temperature combustion, followed by further oxidation and hydrolysis that produces nitric acid in the atmosphere. These materials can be dry-deposited onto surfaces, or be removed from the atmosphere by precipitation. The acid-generating reactions can take from hours to days depending on a wide range of atmospheric parameters such as temperature, humidity, and the presence of oxidizing agents such as hydroxl (OH) radicals, ozone (O_3) and hydrogen peroxide (H_2O_2). Depending on these conditions, and other factors such as height of release and wind speed, sulfur and nitrogen oxides can be transformed and deposited as acid deposition anywhere from a few kilometers to thousands of kilometers from their original source. Figure 1 shows the geographic distribution of acid deposition from precipitation for the United States.

The U.S. trends in emissions SO_2 and NO_x from 1900 to 1997 are shown in Figure 2. The pattern for SO_2 emissions since 1900 has shown three peaks. From 1900 to the 1920s there was a general increase

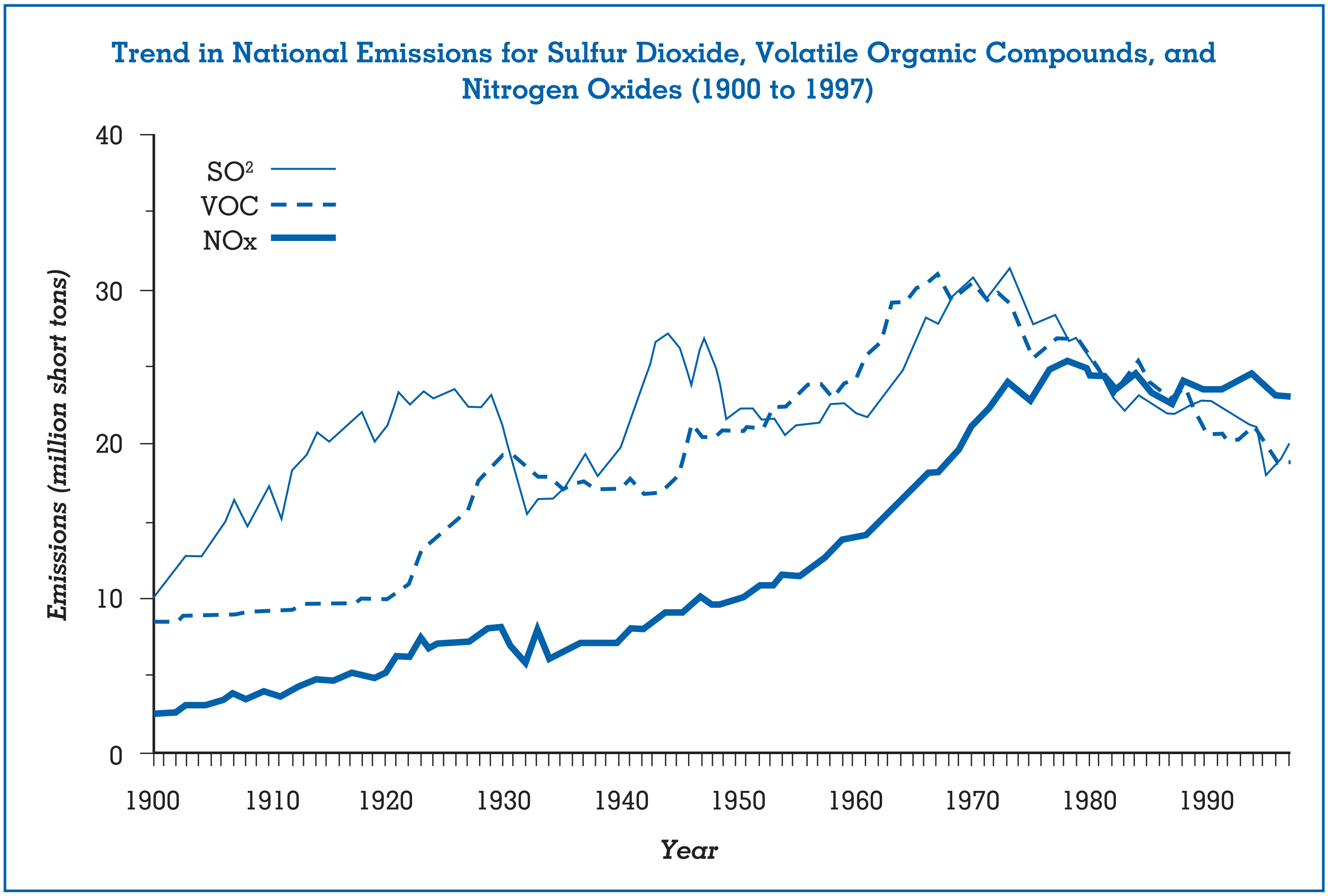


Figure 2.
Emissions of SO_2, NO_x, and volatile organic compounds in the United States, 1900–1997.

followed by a sharp decline during the Great Depression of the 1930s. World War II produced another peak, followed by a significant decline at the end of World War II. SO_2 emissions steadily rose from the early 1950s to the early 1970s. The Clean Air Act of 1970, which was directed at other air pollution concerns and not directly at acid depositon, was largely responsible for the decline in SO_2 emissions. At the time there was essentially no awareness that an "acid rain" problem existed in North America. By 1995, the implementation of the Clean Air Act Amendments (CAAA) of which specifically targeted SO_2 reductins to reduce acid deposition began to further decrease SO_2 emissions in a large part of the eastern United States, where acid deposition is most acute. NO_x emissions rose steadily until the 1970s when emissions leveled off and then showed a very slight decline.

SOURCES OF ACID DEPOSITION

Major sources for emissions of SO_2 and NO_x in the United States are presented in Figures 3 and 4 respectively. Approximately two-thirds of the SO_2 emissions are from electric utilities. Efforts to reduce SO_2 emissions both nationally and regionally have focused on electric utilities. The CAAA of 1990 have stipulated a reduction of 9.1 million metric tons (10 million short tons) of SO_2 below 1980 levels, with most of this reduction coming from coal-fired power plants. Implementation of Phase I reductions (1995–2000) has been successful and has resulted in an 18 percent decline in SO_2 emissions from electric utilities, compared with 1990 SO_2 emissions. There has been a 16 percent decline in SO_2 for this time period when all sources are considered. Phase 2 of the CAAA, which is designed to reduce SO_2 emissions from electric utilities by another 20 percent

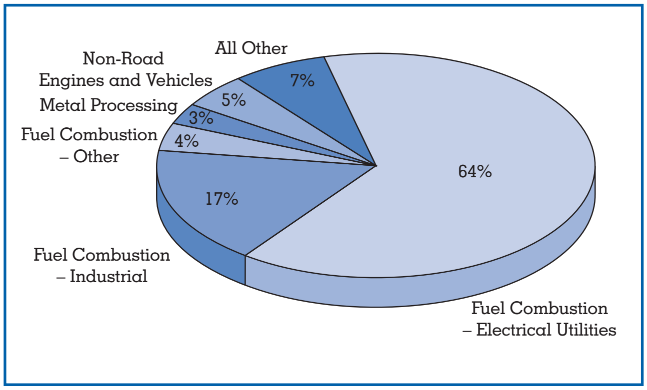


Figure 3.
National sulfur dioxide emissions by source category, 1997. Electric utilities account for almost two-thirds of SO_2 emissions, even after initial implementation of the CAAA of 1990. Total SO_2 emissions for 1997 were 18.5 million metric tons.

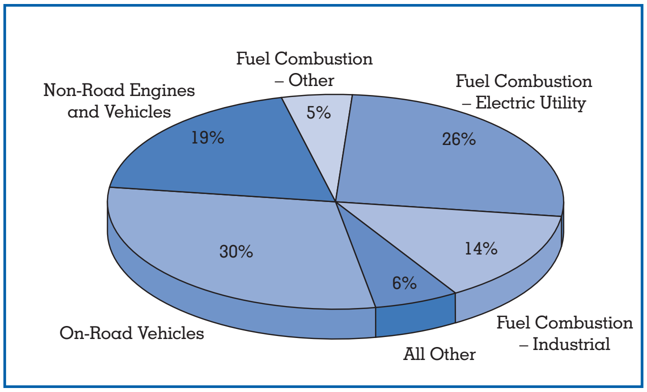


Figure 4.
National nitrogen oxide emissions by source category, 1997. Electric utilities and "on-road vehicles" account for more than half of the NO_x emissions. Total NO_x emissions for 1997 were 21.4 million metric tons.

(compared to 1990), will go into effect from 2000 to 2005.

In the United States, recent reductions in emissions of SO_2 have been achieved by a shift to burning low sulfur coal and by the introduction of SO_2 scrubbers that remove SO_2 gases from power plant stacks. Most of the reductions in SO_2, mandated by the CAAA have come from the shift to burning low sulfur coal from the western United States. Electric utilities account for about one-fourth of the NO_x emissions. However, the largest single sources of NO_x emissions are "on-road vehicles," mainly cars and trucks, which account for 30 percent of the NO_x emissions (Figure 4). Control of NO_x emissions from vehicles is technically more difficult to achieve. Utilities can meet these targets for coal-fired boilers by using low NO_x burner technology (LNBT) or by "emissions averaging." Emissions averaging for a utility requires over-control by one boiler to make up for higher emissions at another boiler.

EFFECTS OF ACID DEPOSITION

Acid deposition and the associated particulate nitrates and sulfates are implicated in the deterioration of certain sensitive ecosystems, decreased visibility, negative human health effects, and increased degradation of certain stone building materials and cultural resources, especially those made of limestone and marble. Fine particulate nitrate and sulfate particles associated with acid deposition are implicated in aggravating cardiorespiratory diseases such as asthma and chronic bronchitis, especially in urban areas.

In many cases estimating the impact of acid deposition on various ecosystems can be a difficult process because acid deposition is only one of many impacts that can effect a response. However, wet and dry acid deposition has been documented as a major factor in the following ecosystem responses.

Aquatic Effects

In both Europe and eastern North America the negative impacts of acid deposition were first documented in lakes and streams found in acid-sensitive areas. In the early 1970s the loss of fish populations and increasing acidity in rural lakes and streams were documented both in Scandinavia and North America. In the United States, studies showed increasing acidification of lakes and loss of fish populations in the Adirondack Mountains of New York. The increased dissolved inorganic aluminum leaching from watersheds due to increased acidity proved toxic to fish in this region. In addition to dissolved aluminum toxicity, increased acidification leads to a large-scale disruption of lake food webs. For example, the experimental acidification of an entire Canadian lake (pH 6.8 to pH 5.09 from 1976 to 1983) led progressively to a loss of freshwater shrimp, all minnow species, and crayfish. These were important food sources for the lake trout population. By the end of the experiment all fish reproduction had ceased. There were also large changes in the species compositon of smaller organisms (insects and crustaceans) lower in the food chain.

Another aquatic impact of acid deposition is episodic acidification. For example, one form of episodic acidification occurs during spring snow melt. When a winter snowpack first melts, acids and other soluble material are concentrated and released, causing an initial “acid pulse” of meltwater, with acidity levels that may be higher than any of the original snowfall. These highly acid episodes, which are also often associated with high dissolved aluminum concentration in runoff, can be especially damaging in streams where fish and other organisms cannot seek refuge in less acid waters. Large storms, which produce high amounts of runoff during other seasons, can also produce episodic acidification.

Historical evidence has linked acidic deposition to the acidification of surface waters in many regions of eastern North America and Europe. Thousands of lakes and streams in these areas are significantly more acid than they were a few decades ago. Large regions of eastern Canada lying on the resistant bedrock of the Precambrian Shield are sensitive to ongoing acidification. In the eastern United States, surface water acidification has occurred in the Adirondack Mountains, the Pocono/Catskill region, the mid-Appalachians, the eastern portion of the Upper Midwest, the New Jersey Pine Barrens, and to a lesser extent, the Florida panhandle. Even with reduced emissions, acidification continues today in many regions. One reason improvements have been smaller than expected is that declines in sulfur emissions have also been accompanied by declines in emissions and deposition of acid-neutralizing agents found in atmospheric dust, both in North America and Europe. The most likely causes for the declines in atmospheric dust are cleaner combustion processes, controls on particulate emissions from smokestacks, changing agricultural practices (no-till), and fewer unpaved roads.

Controlling the effects of acid deposition by the use of lime or other acid-neutralizing compounds has been tried, but mainly on an experimental basis. Adding lime to lakes usually has only a short-term effect in terms of neutralizing lake acidity. The longevity of the effect is directly related to lake’s water residence time, or how long it takes for the lake volume to be replaced with new water. Another experimental control is the liming of a lake or stream watershed. While such an approach can improve forest health, as well as reduce lake acidity for a more extended time, it is prohibitively expensive as a widespread solution to acid deposition.

The pollution rising from the smokestacks of this power plant can cause acid rain. (Photo Researchers Inc.)

Terrestial Effects

The documentation of regional level terrestrial consequences of acid deposition is complicated. For example, forested ecosystems in eastern North America can be influenced by other factors such as high atmospheric ozone concentrations, drought, insect outbreaks and disease, sometimes from non-native sources. However there is a general consensus on some impacts of acidic depositon on both soils and forests in sensitive regions.

In the eastern United States, high elevation red spruce and fir forests found in the Northeast have suffered significant injury and mortality. Significant but lesser amounts of damage have also been found in high elevation spruce-fir forests in the southern Appalachians. Damage can occur directly to trees. For example, foliar leaching of plant nutrients such as calcium, and susceptability to winter cold damage thought to be the result of exposure to highly acid cloudwater can be direct impacts. Besides direct effects, acid deposition on poorly buffered, nutrient deficient soils has caused the leaching of valuable plant nutrients such as calcium, magnesium and other base cations and the release of aluminum, which can be injurious to plants and, as mentioned earlier, toxic to aquatic life. The loss of soil base cations can have long-term deleterious effects and may delay recovery of stressed ecosystems for decades or even longer. Such long-term soil nutrient losses also occur in sensitive low elevation forested ecosystems.

The role of nitrogen in the form of nitrate (NO_3^-) from both wet and dry nitric acid deposition can have both positive and negative effects on ecosystems. NO_3^- is an important plant nutrient and in nitrogen-poor soils can lead to increased plant growth and vigor. However in many acid-sensitive soils receiving high acidic deposition, nitrogen in soils is at or near saturation and can lead to the leaching of other important plant nutrients such as base cations or the release of aluminum into solution. The relative importance of nitric acid deposition will continue to grow as substantial reductions in SO_2 emissions occur and emissions of NO_x do not appreciably decline.

CONCLUSIONS

Acid deposition is a regional problem wherever large amounts of fossil fuels are consumed. There have been significant efforts in both Europe and North America to reduce acid deposition because of its many deleterious effects. This effort has focused mainly on the reduction of SO_2 emissions. In the future acid deposition will have to be addressed in eastern Asia, where rapid industrialization and increased use of fossil fuels is likely.

In the United States the passage of the CAAAs of 1990, and their implementation starting in 1995, was an attempt to "solve" the acid rain problem mainly by reducing SO_2 emissions from electric power plants. While significant reductions in SO_2 emissions have occurred, and there already has been a reduction in deposition of sulfur and acidity, there have not been significant improvements in some sensitive regions such as the Adirondack Mountains of New York, the Green and White Mountains of New England, and the southern Appalachians. However, further reductions in SO_2 (and less so for NO_x) are expected until the year 2005. However, this too may not be enough. It is likely that further reductions in emissions of both SO_2 and NO_x, beyond those required by the CAAA, will be necessary if the goal is to protect sensitive ecosystems and public health. One reason recovery has been very limited is that acid deposition over decades has removed base cations from watersheds that are crucial to maintaining proper soil chemistry for plant growth and acid-neutralizing capacity for aquatic ecosystems. Recovery may be a slow process. Another concern is the very limited reduction in NO_x emissions over the last decade. Nitric acid is becoming a major component of acid deposition and significant reductions in NO_x emissions will probably be necessary to "solve" the acid rain problem.

Tom Butler

See also: Air Pollution; Atmosphere.

BIBLIOGRAPHY

Cowling, E. B. (1982). "Acid Precipitation in Historical Perspective." *Environmental Science and Technology* 16: 110A–123A.

Crowther, C., and Ruston, H. G. (1911). *Journal of Agricultural Science* 4:25–55.

Dannevig, A. (1959). *Jager og Fisker* 3:116–118.

Gorham, E. (1957). *Limnology and Oceanography* 2:22

Gorham, E (1958). *Phil. Trans. Royal Soc. London, Ser. B* 247:147–178.

Likens, G. E., and Bormann, F. H. (1974). "Acid Rain: A Serious Environmental Problem." *Science* 184:1176–1179

Likens, G. E.; Bormann, F. H.; and Johnson, N. M. (1972). "Acid rain." *Environment* 14:33–40.

Oden, S. (1968). "The Acidification of Air and Precipitation and Its Consequences in the Natural Environment." *Ecology Committee Bulletin No. 1*, Swedish National Science Research Council, Stockholm. Arlington, VA, Translation Consultants Ltd.

Schindler, D. W.; Mills, K. H.; Malley, D. F.; Findlay, D. L.; Shearer, J.A.; Davies, I. J.; Turner, M. A.; Linsey, G. A.; Cruikshank, D. R. (1985). "Long Term Ecosystem Stress: The Effects of Years of Experimental Acidification on a Small Lake." *Science* 228:1395–1401.

Smith, R. A. (1872). *Air and Rain: The Beginnings of a Chemical Climatology.* London: Longmans Greene.

ACOUSTICAL ENERGY

See: Waves

ADVANCED TRAFFIC MANAGEMENT SYSTEMS

See: Traffic Flow Management

AERODYNAMICS

The next time you hear an airplane flying overhead, look up, and pause for a moment. What you see is a machine that is heavier than air, but which is somehow being sustained in the air. This is due to the airflow over the airplane. This airflow exerts a lift force which counteracts the weight of the airplane and sustains it in the air—a good thing. The airflow also exerts a drag force on the airplane which retards its motion—a bad thing. The drag must be counteracted by the thrust of the engine in order to keep the airplane going. The production of thrust by the engine consumes energy. Hence, the energy efficiency of the airplane is intimately related to aerodynamic drag. This is just one of many examples where the disciplines of aerodynamics and energy interact.

DEFINITION

Aerodynamics deals with the flow of gases, particularly air, and the interaction with objects immersed in the flow. The interaction takes the form of an aerodynamic force and moment exerted on the object by the flow, as well as heat transfer to the object (aerodynamic heating) when the flow velocities exceed several times the speed of sound.

SOURCES OF AERODYNAMIC FORCE

Stop for a moment, and lift this book with your hands. You are exerting a force on the book; the book is feeling this force by virtue of your hands being in contact with it. Similarly, a body immersed in a liquid or a gas (a fluid) feels a force by virtue of the body surface being in contact with the fluid. The forces exerted by the fluid on the body surface derive from two sources. One is the pressure exerted by the fluid on every exposed point of the body surface. The net force on the object due to pressure is the integrated effect of the pressure summed over the complete body surface. In the aerodynamic flow over a body, the pressure exerted by the fluid is different at different points on the body surface (i.e., there is a distribution of variable values of pressure over the surface). At each point, the pressure acts locally perpendicular to the surface. The integrated effect of this pressure distribution over the surface is a net force—the net aerodynamic force on the body due to pressure. The second source is that due to friction between the body surface and the layer of fluid just adjacent to the surface. In an aerodynamic flow over a body, the air literally rubs over the surface, creating a frictional shear force acting on every point of the exposed surface. The shear stress is tangential to the surface at each point, in contrast to the pressure, which acts locally perpendicular to the surface. The value of the shear stress is different at different points on the body surface. The integrated effect of this shear stress distribution is a net integrated force on the body due to friction.

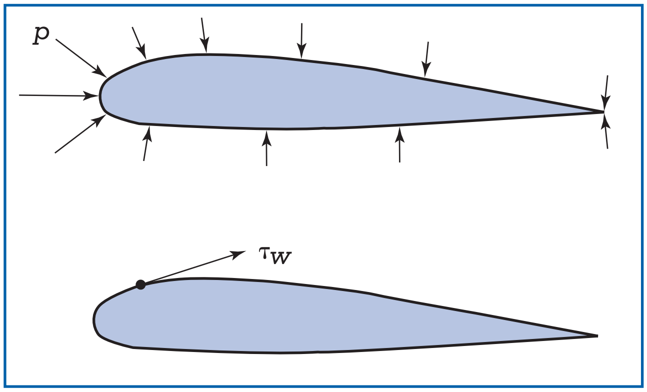


Figure 1.
Schematic of pressure and shear stress distribution over a body surface.

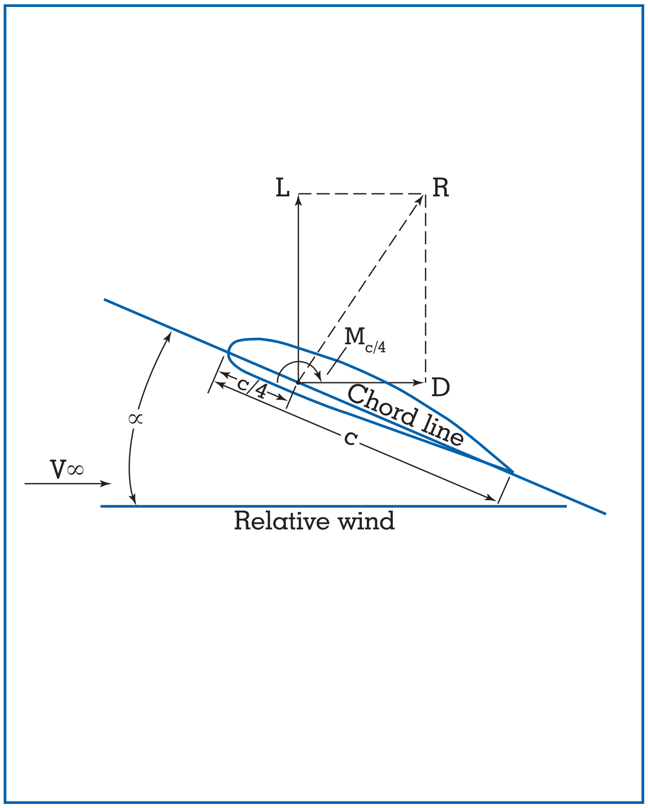


Figure 2.
Resolution of aerodynamic force into lift and drag.

The pressure (p) and shear stress (τ_ω) distributions over an airfoil-shaped body are shown schematically in Figure 1. The pressure and shear stress distributions exerted on the body surface by the moving fluids are the two hands of nature that reach out and grab the body, exerting a net force on the body—the aerodynamic force.

RESOLUTION OF THE AERODYNAMIC FORCE

The net aerodynamic force exerted on a body is illustrated in Figure 2 by the arrow labeled R. The direction and speed of the airflow ahead of the body is denoted by V_∞, called the relative wind. The body is inclined to V_∞. by the angle of attack, α. The resultant aerodynamic force R can be resolved into two components; lift, L, perpendicular to V_∞; and drag, D, parallel to V_∞. In Figure 2, R is shown acting through a point one-quarter of the body length from the nose, the quarter-chord point. Beacuse the aerodynamic force derives from a distributed load due to the pressure and shear stress distributions acting on the surface, its mechanical effect can be represented by a combination of the net force vector drawn through any point and the resulting moment about that point. Shown in Figure 2 is R located (arbitrarily) at the quarter-chord point and the moment about the quarter-chord point, $M_{c/4}$.

The aerodynamic force varies approximately as the square of the flow velocity. This fact was established in the seventeenth century—experimentally by Edme Marione in France and Christiaan Huygens in Holland, and theoretically by Issac Newton. Taking advantage of this fact, dimensionless lift and drag coefficients, C_L and C_D respectively, are defined as

$$L = \tfrac{1}{2}\rho_\infty V_\infty^2 \, S \, C_L$$
$$D = \tfrac{1}{2}\rho_\infty V_\infty^2 \, S \, C_D$$

where ρ_∞, is the ambient density in the freestream, and S is a reference area, which for airplanes is usually chosen to be the planform area of the wings (the projected wing area you see by looking at the wing directly from the top or bottom), and for projectile-like bodies is usually chosen as the maximum cross-sectional area perpendicular to the axis ofthe body (frontal area).

At flow speeds well below the speed of sound, the lift coefficient depends only on the shape and orientation (angle of attack) of the body:

$$C_L = f\,(\text{shape}, \alpha)$$

The drag coefficient also depends on shape and α, but in addition, because drag is partially due to friction, and frictional effects in a flow are governed by a powerful dimensionless quantity called Reynolds number, then C_D is also a function of the Reynolds number, Re:

$$C_D = f\,(\text{shape}, \alpha, \text{Re})$$

where $\text{Re} \cong \rho_\infty V_\infty, c/\mu_\infty$. Here, c is the reference length of the body and μ_∞ is the viscosity coefficient of the fluid. At speeds near and above the speed of sound, these coefficients also become functions of Mach number, $M_\infty \cong V_\infty/a_\infty$, where a_∞, is the speed of sound in the freestream:

$$C_L = f\,(\text{shape}, \alpha, M_\infty)$$
$$C_D = f\,(\text{shape}, \alpha, \text{Re}, M_\infty)$$

The lift and drag characteristics of a body in a flow are almost always given in terms of C_L and C_D rather than the forces themselves, because the force coefficients are a more fundamental index of the aerodynamic properties.

DRAG

One of the most important aerodynamic effects on the consumption of energy required to keep a body moving through a fluid is the aerodynamic drag. The drag must be overcome by the thrust of a propulsion mechanism, which in turn is consuming energy. Everything else being equal, the higher the drag, the more energy is consumed. Therefore, for energy efficiency, bodies moving through a fluid should be low-drag bodies. To understand how to obtain low drag, we have to first understand the nature of drag, and what really causes it.

The influence of friction on the generation of drag is paramount. In most flows over bodies, only a thin region of the flow adjacent to the surface is affected by friction. This region is called the boundary layer (Figure 3). Here, the thickness of the boundary layer is shown greatly exaggerated; in reality, for ordinary flow conditions, the boundary layer thickness, δ, on the scale of Figure 3 would be about the thickness of a sheet of paper. However, the secrets of drag production are contained in this very thin region. For example, the local shear stress at the wall, labeled in Figure 3 as τ_ω, when integrated over the entire surface creates the skin friction drag D_f on the body. The magnitude of τ_ω, hence, D_f, is determined by the nature of the velocity profile through the boundary layer, (i.e., the variation of the flow velocity as a function of distance y normal to the surface at a given station, x, along the surface). This velocity variation is quite severe, ranging from zero velocity at the surface (due to friction, the molecular layer right at the wall is at zero velocity relative to the wall) to a high velocity at the outer edge of the boundary layer. For most fluids encountered in aerodynamics, the shear stress at the surface is given by the Newtonian shear stress law:

$$\tau_\omega = \mu\,(dV/dy)_{y=0}$$

where μ is the viscosity coefficient, a property of the fluid itself, and $(dV/dy)_{y=0}$ is the velocity gradient at the wall. The more severe is the velocity variation in the boundary layer, the larger is the velocity gradient at the wall, and the greater is the shear stress at the wall.

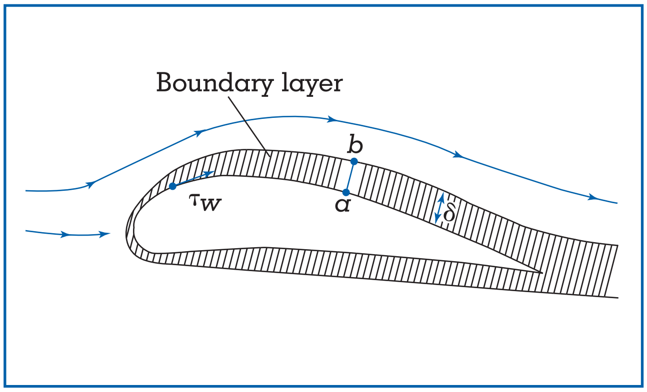


Figure 3.
The boundary layer.

The above discussion has particular relevance to drag when we note that the flow in the boundary layer can be of two general types: laminar flow, in which the streamlines are smooth and regular, and an element of the fluid moves smoothly along a streamline; and turbulent flow, in which the streamlines break up and a fluid element moves in a random, irregular, and tortuous fashion. The differences between laminar and turbulent flow are dramatic, and they have a major impact on aerodynamics. For example, consider the velocity profiles through a boundary layer, as sketched in Figure 4. The profiles are different, depending on whether the flow is laminar or turbulent. The turbulent profile is “fatter,” or fuller, than the laminar profile. For the turbulent profile, from the outer edge to a point near the surface, the velocity remains reasonably close to the freestrearn velocity; it then rapidly decreases to zero at the surface. In contrast, the laminar velocity profile gradually decreases to zero from the outer edge to the surface. Now consider the velocity gradient at the wall, $(dV/dy)_{y=0}$, which is the reciprocal of the slope of the curves shown in Figure 4 evaluated at $y = 0$. It is clear that $(dV/dy)_{y=0}$ for laminar flow is less than $(dV/dy)_{y=0}$ for turbulent flow. Recalling the Newtonian shear stress law for τ_ω leads us to the fundamental and highly important fact that laminar stress is less than turbulent shear stress:

$$\tau_{\omega\,\text{laminar}} < \tau_{\omega\,\text{turbulent}}$$

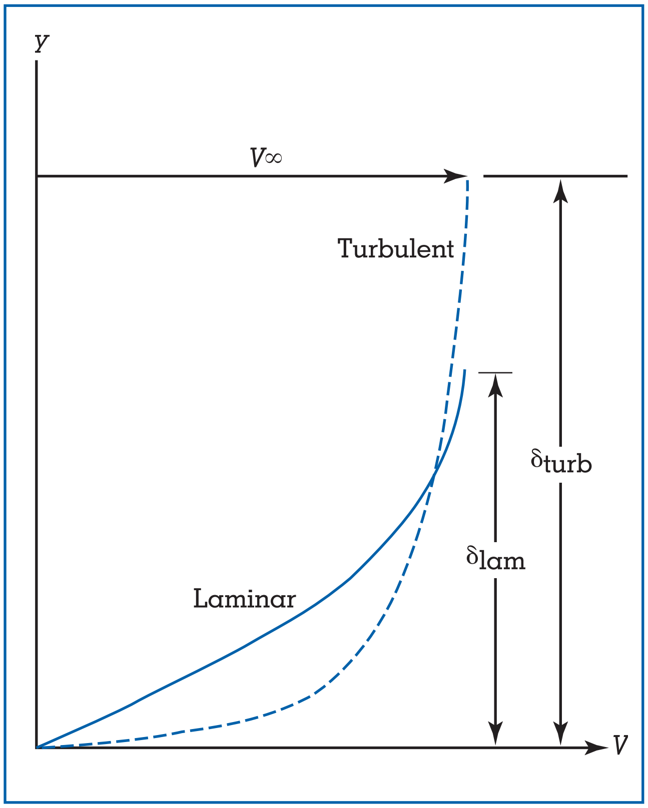


Figure 4.
Comparison of the velocity profiles for laminar and turbulent boundary layers.

This obviously implies that the skin friction exerted on an airplane wing or body will depend on whether the boundary layer on the surface is laminar or turbulent, with laminar flow yielding the smaller skin friction drag.

It appears to be almost universal in nature that systems with the maximum amount of disorder are favored. For aerodynamics, this means that the vast majority of practical viscous flows are turbulent. The boundary layers on most practical airplanes, missiles, and ship hulls, are turbulent, with the exception of small regions near the leading edge. Consequently, the skin friction on these surfaces is the higher, turbulent value. For the aerodynamicist, who is usually striving to reduce drag, this is unfortunate. Today, aerodynamicists are still struggling to find ways to preserve laminar flow over a body—the reduction in skin friction drag and the resulting savings in energy are well worth such efforts. These efforts can take the form of shaping the body in such a way to encourage laminar flow; such "laminar flow bodies" are designed to produce long distances of decreasing pressure in the flow direction on the surface (favorable pressure gradients) because an initially laminar flow tends to remain laminar in such regions. Figure 5 indicates how this can be achieved. It shows two airfoils, the standard airfoil has a maximum thickness near the leading edge, whereas the laminar flow airfoil has its maximum thickness near the middle of the airfoil. The pressure distributions on the top surface on the airfoils are sketched above the airfoils in Figure 5. Note that for the standard airfoil, the minimum pressure occurs near the leading edge, and there is a long stretch of increasing pressure from this point to the trailing edge. Turbulent boundary layers are encouraged by such increasing pressure distributions. The standard airfoil is generally bathed in long regions of turbulent flow, with the attendant high skin friction drag. Note that for the laminar flow airfoil, the minimum pressure occurs near the trailing edge, and there is a long stretch of decreasing pressure from the leading edge to the point of minimum pressure. Laminar boundary layers are encouraged by such decreasing pressure distributions. The laminar flow airfoil can be bathed in long regions of laminar flow, thus benefiting from the reduced skin friction drag.

The North American P-51 Mustang, designed at the outset of World War II, was the first production aircraft to employ a laminar flow airfoil. However, laminar flow is a sensitive phenomenon; it readily gets unstable and tries to change to turbulent flow. For example, the slightest roughness of the airfoil surface caused by such real-life effects as protruding rivets, imperfections in machining, and bug spots can cause a premature transition to turbulent flow in advance of the design condition. Therefore, most laminar flow airfoils used on production aircraft do not yield the extensive regions of laminar flow that are obtained in controlled laboratory tests using airfoil models with highly polished, smooth surfaces. From this point of view, the early laminar flow airfoils were not successful. However, they were successful from an entirely different point of view; namely, they were found to have excellent high-speed properties, postponing to a higher flight Mach number the large drag rise due to shock waves and flow separation encountered near Mach 1. As a result, the early laminar flow airfoils were extensively used on jet-propelled airplanes during the 1950s and 1960s and are still employed today on some modem high-speed aircraft.

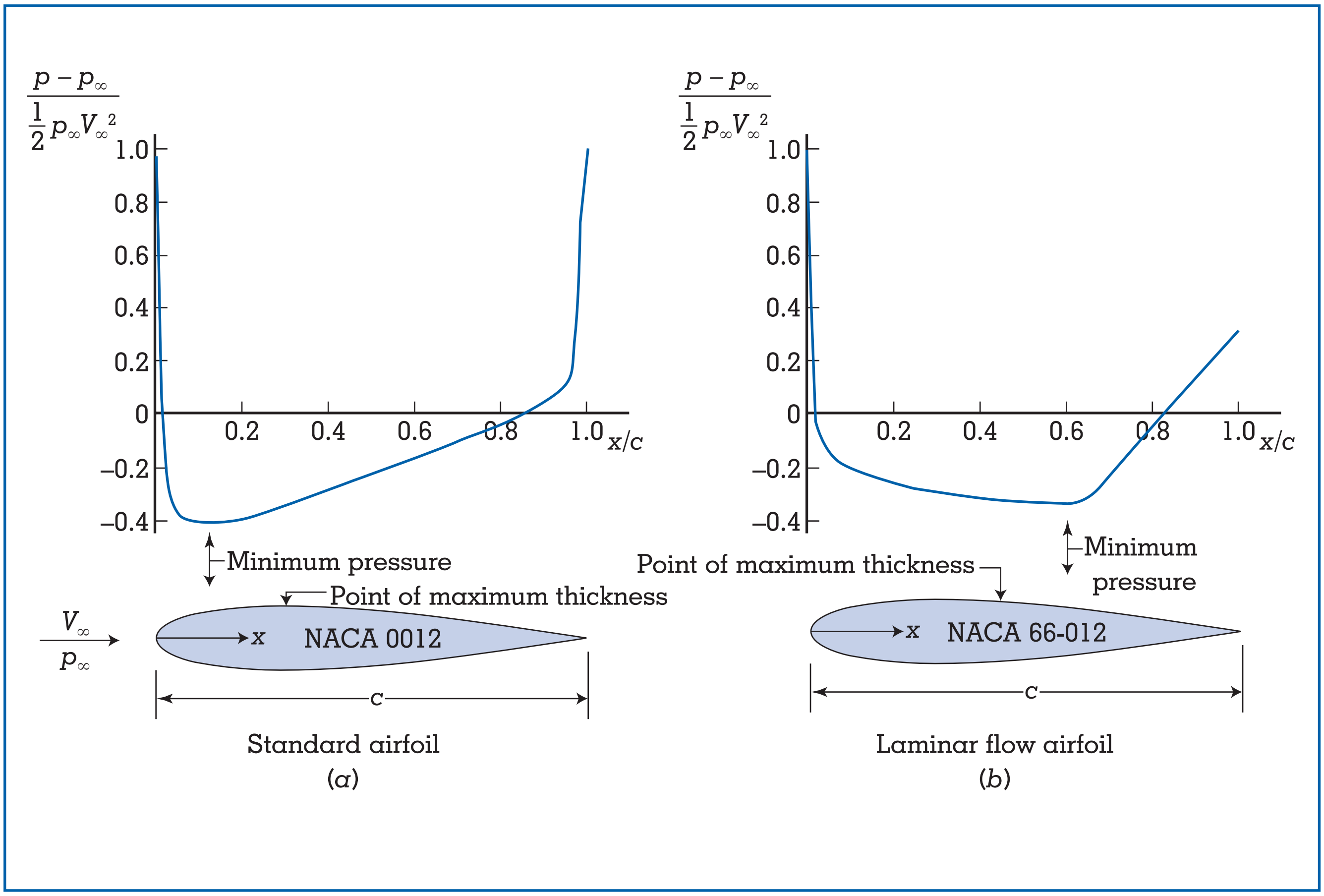


Figure 5.
Pressure distributions over (a) a standard airfoil shape and (b) a laminar low airfoil.

In reality, the boundary layer on a body always starts out from the leading edge as laminar. Then at some point downstream of the leading edge, the laminar boundary layer become unstable and small "bursts" of turbulent flow begin to grow in the flow. Finally, over a certain region called the transition region, the boundary layer becomes completely turbulent. For purposes of analysis, it is convenient to draw a picture, where transition is assumed to occur at a point located a distance x_{cr}, from the leading edge. The accurate knowledge of where transition occurs is vital to an accurate prediction of skin friction drag. Amazingly, after almost a century of research on turbulence and transition, these matters are still a source of great uncertainty in drag predictions today. Nature is still keeping some of her secrets from us.

Skin friction drag is by no means the whole story of aerodynamic drag. The pressure distribution integrated over the surface of a body has a component parallel to the flow velocity V_∞, called form drag, or more precisely pressure drag due to flow separation. In this type of drag, such as the flow over a sphere, the boundary layer does not totally close over the back surface, but rather separates from the surface at some point and then flows downstream. This creates a wake of low-energy separated flow at the back surface. The pressure on the back surface of the sphere in the separated wake is smaller than it would be if the flow were attached. This exacerbates the pressure difference between the higher pressure on the front surface and the lower pressure on the back surface, increasing the pressure drag. The bigger (fatter) the wake, the higher the form drag. Once again we see the different effects of laminar and turbulent flow. In the case where the boundary layer is laminar, the boundary layer separates near the top and bottom of the body, creating a large, fat wake, hence high pressure drag. In contrast, where the boundary layer is turbulent, it separates further around the back of the sphere, creating a thinner wake, thus lowering the pressure drag. Form drag, therefore, is larger for laminar flow than for turbulent flow. This is the exact opposite of the case for skin friction drag. To reduce form drag, you want a turbulent boundary layer.

For a blunt body, such as a sphere, almost all the drag is form drag. Skin friction drag is only a small percentage of the total drag. For blunt bodies a turbulent boundary layer is desirable. Indeed, this is the purpose of the dimples on the surface of a golf ball—to promote turbulent flow and reduce the aerodynamic drag on the ball in flight. The nose of an airplane is large compared to a golf ball. Hence, on the airplane nose, the boundary layer has already become a turbulent boundary layer, transitioning from laminar to turbulent in the first inch or two from the front of the nose. Therefore, dimples are not necessary on the nose of an airplane. In contract, the golf ball is small—the first inch or two is already too much, so dimples are placed on the golf ball to obtain turbulent flow right from the beginning.

For a body that is producing lift, there is yet another type of drag—induced drag due to lift. For example, consider an airplane wing, that produces lift by creating a higher pressure on the bottom surface and a lower pressure on the top surface. At the wing tips, this pressure difference causes the flow to try to curl around the tips from the bottom of the tip to the top of the tip. This curling motion, superimposed on the main freestream velocity, produces a vortex at each wing tip, that trails downstream of the wing. These wing tip vortices are like minitornadoes that reach out and alter the pressure distribution over the wing surface so as to increase its component in the drag direction. This increase in drag is called induced drag; it is simply another source of pressure drag on the body.

Finally, we note that if the body is moving at speeds near and above the speed of sound (transonic and supersonic speeds), shock waves will occur that increase the pressure on the front portions of the body, contributing an additional source of pressure drag called wave drag.

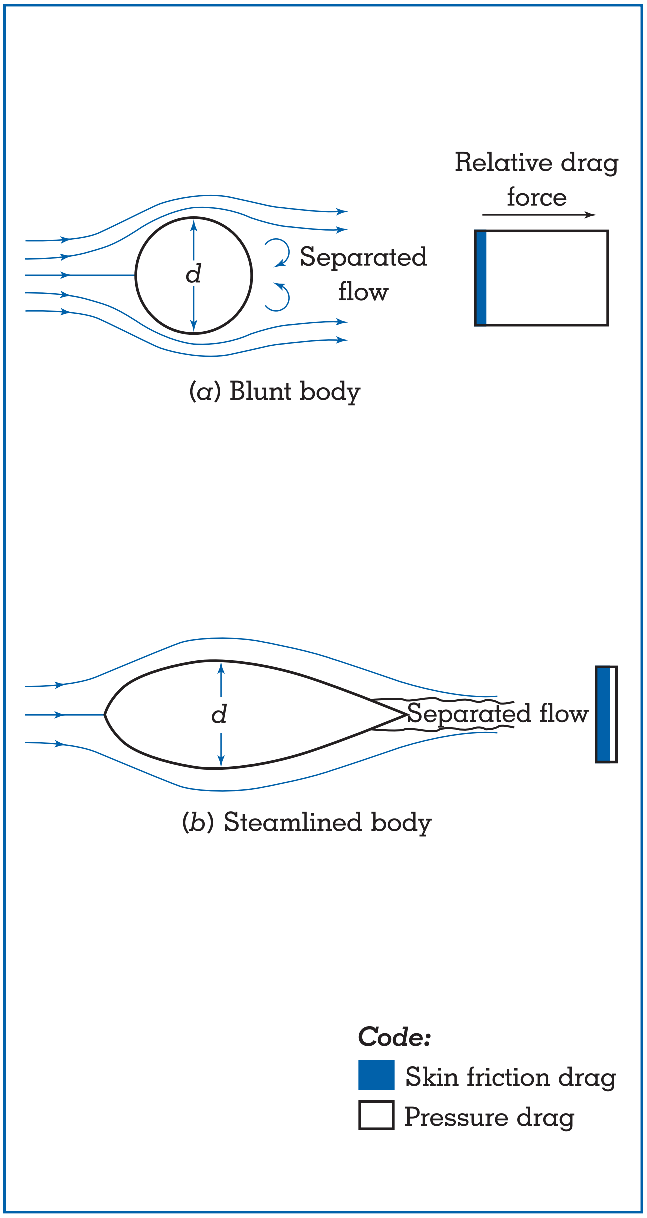


Figure 6.
Comparison of drag for a blunt body and a streamlined body.

In summary, the principal sources of drag on a body moving through a fluid are skin friction drag, form drag, induced drag, and wave drag. In terms of drag coefficients, we can write:

$$C_D = C_{D,f} + C_{D,p} + C_{D,i} + C_{D,w}$$

where C_D is the total drag coefficient, $C_{D,f}$ is the skin friction drag coefficient, $C_{D,p}$ is the form drag coefficient (pressure drag due to flow separation), $C_{D,i}$ is the induced drag coefficient, and $C_{D,w}$ is the wave drag coefficient.

STREAMLINING

The large pressure drag associated with blunt bodies such as the sphere, leads to the design concept of streamlining. Consider a body of cylindrical cross section of diameter d with the axis of the cylinder oriented perpendicular to the flow, as shown in Figure 6a. There will be separated flow on the back face of the cylinder, with a relatively fat wake and with the associated high pressure drag. The bar to the right of the cylinder denotes the total drag on the cylinder; the shaded portion of the bar represents skin friction drag, and the open portion represents the pressure drag. Note that for the case of a blunt body, the drag is relatively large, and most of this drag is due to pressure drag. However, look at what happens when we wrap a long, mildly tapered after-body on the back of the cylinder, creating a teardrop-shaped body sketched in Figure 6b. This shape is a streamlined bod, of the same thickness d as the cylinder. Flow separation on the streamlined body will be delayed until much closer to the trailing edge, with an attendant, much smaller wake. As a result, the pressure drag of the streamlined body will be much smaller than that for the cylinder. Indeed, as shown by the bar to the right of Figure 6b, the total drag of the streamlined body will be almost a factor of 10 smaller than that of the cylinder of same thickness. The friction drag of the streamlined body will be larger due to its increased surface area, but the pressure drag is so much less that it dominates this comparison.

Streamlining has a major effect on the energy efficiency of bodies moving through a fluid. For example, a bicycle with its odd shaped surfaces, has a relatively large drag coefficient. In contrast, a streamlined outer shell used for recumbent bicycles reduces the drag and has allowed the speed record to reach 67 mph. Streamlining is a cardinal principle in airplane design, where drag reduction is so important.

Streamlining has a strong influence on the lift-to-drag ratio (L/D, or C_L/C_D) of a body. Lift-to-drag ratio is a measure of aerodynamic efficiency. For example, the Boeing 747 jumbo-jet has a lift-to-drag ratio of about 20. This means it can lift 20 lb at the cost of only 1 pound of drag—quite a leverage. In airplane design, an increase in L/D is usually achieved by a *decrease* in D rather than an increase in L. Vehicles that have a high L/D are that way because they are low-drag vehicles.

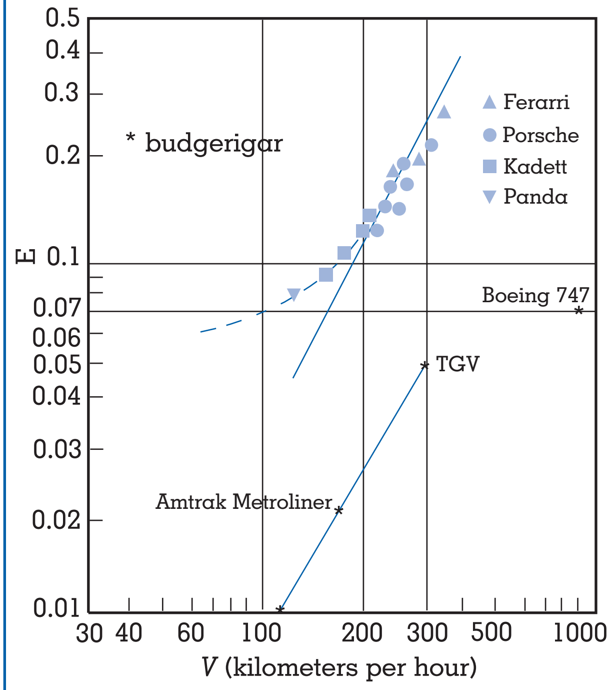


Figure 7.
Specific energy consumption versus velocity.

DRAG AND ENERGY

We now make the connection between aerodynamic drag and energy consumption, The drag of a moving vehicle must be overcome by the thrust from a propulsive mechanism in order to keep the vehicle in sustained motion. The time rate of energy consumption is defined as *power*, P. The power required to keep the vehicle moving at a given speed is the product of drag times velocity,

$$P = D V_\infty.$$

Because

$$D = ½\rho_\infty V_\infty^2 S C_D,$$

we have

$$P = ½\rho_\infty V_\infty^3 S C_D.$$

That is, the power required varies as the *cube* of the

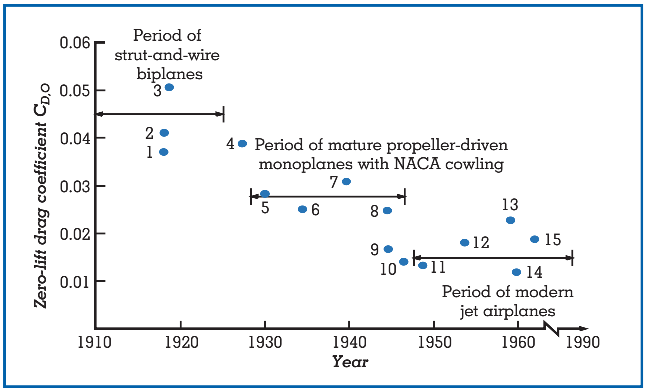


Figure 8.
Evolution of drag reduction for airplanes, drag coefficient versus years.

velocity, and directly as the drag coefficient. This clearly indicates why, as new vehicles are designed to move at higher velocities, every effort is made to reduce C_D. Otherwise, the velocity-cubed variation may dictate an amount of energy consumption that is prohibitive. Note that this is one of the realities facing civil transport airplanes designed to fly at supersonic speeds. No matter how you look at it, less drag means more energy efficiency.

The effect of aerodynamics on the energy consumption of transportation vehicles can be evaluated by using the dimensionless specific energy consumption, E, defined as E = P/WV, where P is the power required to move at velocity V and W is the weight of the vehicle, including its payload (baggage, passengers, etc.). Although power required increases as the cube of the velocity, keep in mind that the *time* required to go from point A to point B is inversely proportional to V, hence a faster vehicle operates for less time between two points. The quantity E = P/WV is the total energy expended per unit distance per unit weight; the smaller the value of E, the smaller the amount of energy required to move 1 lb a distance of 1 ft (i.e., the more energy efficient the vehicle). Representative values of E for different classes of vehicles (trains, cars, airplanes) are given in Figure 7. Using E as a figure of merit, for a given long distance trip, trains such as the Amtrak Metroliner and the French high-speed TGV are most efficient, airplanes such as the Boeing 747, are next, and automobiles are the least efficient.

DRAG OF VARIOUS VEHICLES

Let us examine the drag of various representative vehicles. First, in regard to airplanes, the evolution of streamlining and drag reduction is clearly seen in Figure 8, which gives the values of drag coefficient based on wing planform area for a number of different aircraft, plotted versus years. We can identify three different periods of airplanes, each with a distinctly lowered drag coefficient: strut-and-wire biplanes, mature propellerdriven airplanes, and modem jet airplanes. Over the past century, we have seen a 70 percent reduction in airplane drag coefficient. Over the same period, a similar aerodynamic drag reduction in automobiles has occurred. By 1999, the drag coefficients for commercialized vehicles have been reduced to values as low as 0.25. There are experimental land vehicles with drag coefficients on par with jet fighters; for example, the vehicles built for the solar challenge races, and some developmental electric vehicles.

The generic effect of streamlining on train engines is similar, with the drag coefficient again based on frontal area. The high-speed train engines of today have drag coefficients as low as 0.2.

For motorcycles and bicycles, the drag coefficient is not easy to define because the proper reference area is ambiguous. Hence, the drag is quoted in terms of the “drag area” given by D/q, where q is the dynamic pressure; $q = \frac{1}{2}\rho_\infty V_\infty^2$. A typical drag area for a motorcycle and rider can be reduced by more than fifty percent by wrapping the motorcycle in a streamlined shell.

SUMMARY

Aerodynamics is one of the applied sciences that plays a role in the overall consideration of energy. We have explained some of the more important physical aspects of aerodynamics, and illustrated how aerodynamics has an impact on energy efficiency.

John D. Anderson, Jr.

See also: Automobile Performance; Tribology.

BIBLIOGRAPHY

Allen, J. E. (1982). *Aerodynamics: The Science of Air in Motion*, 2nd ed. New York: McGraw-Hill.

Anderson, J. D., Jr., (1991). *Fundamentals of Aerodynamics,* 2nd ed. New York: McGraw-Hill.

Anderson, J. D., Jr. (1997). *A History of Aerodynamics, and Its Impact on Flying Machines.* New York: Cambridge University Press.

Anderson, J. D., Jr. (2000). *Introduction to Flight,* 4th ed. Boston: McGraw-Hill.

Tennekes, H. (1992). *The Simple Science of Flight.* Cambridge, MA: MIT Press.

von Karman, T. (1963). *Aerodynamics.* New York: McGraw-Hill. (Originally published by Cornell University Press, 1954).

AGRICULTURE

The earliest human inhabitants of the earth were hunters and gatherers. Under favorable conditions they required at least 1.5 square kilometers to provide food for one person, and in harsher environments as much as 80 to 100 square kilometers. Population pressure eventually led humans to raise plants and animals.

With the introduction of agriculture, humans began to use energy to control the growth of plants and animals, to make more efficient use of the solar energy stored in plants by photosynthesis. However, for many thousands of years, the only energy used for this purpose was human energy. The energy drawn from the biosphere was limited to the dietary energy provided by plant food and meat, and to the use of wood and grasses as fuel for heating and cooking. Later, humans learned to use animal, water, and wind energy to obtain power for transport and for simple agricultural and industrial processes. As population grew, the use of energy increased steadily, but all of it came from renewable resources.

Shifting cultivation, one of the first agricultural practices developed, is still widely used. In the early 1970s, about 36 million square kilometers of land were farmed under this system, producing food for about 250 million people. Each hectare (2.47 acres or

A pair of oxen pull a wooden plow through a muddy paddy field in Sri Lanka. (Corbis-Bettmann)

10,000 square meters) under shifting cultivation can provide adequate food for one person. Degradation of soil and vegetation usually occurs when the population density exceeds one person per four hectares. At higher population densities, shorter fallow periods and eventually annual cropping become necessary.

Annual cropping by traditional methods requires more labor, and yields are generally lower. Although animal power can help reduce human labor and provide manure for fertilizer, draught animals must be either fed by some of the crop or pastured, thereby increasing the land area required per person unless yields per unit of land increase accordingly. Average livestock ate one-fourth of the yield in the early nineteenth century.

This land per person dynamic has changed in the past one hundred years as humans moved beyond the limitations of human and animal power by drawing upon nonrenewable sources of energy in the form of fossil fuels—coal, oil, and natural gas. Rapid improvements in technology made it possible to locate and extract increasing quantities of fossil fuels with little or no increase in cost. By the 1950s and 1960s the world had come to take advantage of the large supply of fossil fuels to dramatically and economically boost agricultural production.

Energy dense fossil fuels can be converted very efficiently to heat and/or power. Use of energy for power has spread rapidly in both the industrial and agricultural sectors of the developed countries. The developing countries have followed the developed countries along the same path but at a slower pace. Although agriculture uses only a small share of the world's total energy consumption, it is generally recognized that its needs are crucial, because the existing technologies for increasing production rely so heavily on energy-intensive inputs. Farm use of energy accounts for only 4 percent of total commercial energy use in developing countries and 3 percent in the developed countries. Commercial energy includes oil, natural gas, and coal. Noncommercial sources of energy include fuel-wood, plant and animal residues, and human and animal energy. Noncommercial energy uses are very important in the developing countries, especially in rural areas.

There is a two-way relationship between agriculture and energy because agriculture is both a consumer and a producer of energy. Directly or indirectly, the whole agricultural process involves the input of energy. The ability of plants to capture and store the sun's energy through photosynthesis can be increased through the input of energy in the form of fertilizer, pesticides, and fuel to drive machinery used in production. The production of nitrogen fertilizers is very energy intensive input because it requires hydrogen-dissociation from water or fossil fuels (natural gas) to fix nitrogen in synthetic fertilizers. Production of pesticides similar to nitrogen fertilizers requires large amounts of energy. The gasoline and diesel fuel used for agricultural machinery also consumes a lot of energy, but has dropped significantly since the early 1970s because of better technology.

The role of energy in agricultural production has become crucial as population and income growth put pressure on the demand for food. By discovering new technologies (hybrid seeds, drought and disease-resistant crops, as well as genetically modified crops), dramatic increases in crop yields per hectare have been achieved in the developed countries. The pace of growth, however, has been uneven among developing countries. Asia has shown significant increases in agricultural productivity, while many countries in Africa and Latin America have had modest productivity growth. Although many developing countries have reserves of unused but potentially productive land, most have to meet rising food demands by substantially raising yields on both used and new land. The growth in yields does not come automatically. An increase in agricultural research investment is essential both to increase the yield and to expand the area under cultivation. Clearing less accessible new land for cultivation, which is the goal of many environmentalist groups, will require greater investment in inputs of commercial energy and often the provision of drainage, irrigation, and soil conservation systems. Once the land is under production, additional energy will be needed for maintenance.

CLASSIFICATION OF AGRICULTURAL PRODUCTION

World production of food and fiber could be classified into three distinct groups: traditional, transition from traditional to modern, and modern farming methods.

Traditional Agriculture

Traditional farms rely on labor, draught animals, and hand tools. The only commercial energy input in traditional farming is that required to produce hand

English farmers use a motorized tractor to drive a plow through a field around 1905. At that time most farmers still used horses for this purpose. (Corbis-Bettmann)

tools and animal implements. No commercial energy inputs are used for irrigation; the land either gets rain or it's irrigated by traditional methods.

Traditional methods continue to be the mode of operation in many parts of the world. Currently agriculture production in many of the lowest income countries in Africa, Latin America, and Asia are based on traditional methods. Agriculture is largely operated by small family farms using human and animal power and organic fertilizer with little access to or knowledge of modern inputs such as chemicals, fertilizers, hybrid seeds, or mechanical drive. Low soil fertility and inadequate or irregular rainfall sharply limit the productivity of low-input farms in developing countries.

The Transition to Modern Farming

In the transition to modern farming methods, the use of commercial energy, especially machinery and fertilizer, increases sharply. Primary tillage is usually one of the first operations to be mechanized, requiring an increase in commercial energy use not only for the production of farm machinery, but also for its operation. Improved crop varieties are often introduced during the transitional phase, requiring commercial energy for their production and distribution. To help realize their yield potential entails the use of chemical fertilizers and pesticides. Both of these inputs require commercial energy for their manufacture. In addition to the traditional irrigation methods, supplementary irrigation with mechanically powered pumps is often introduced during the transitional phase. This process substantially increases commercial energy requirements but also increases yields. The growing needs for investment during transition influence farm size. To achieve economies of scale, the number of family farms is reduced and they are replaced with larger commercial farms.

This general trend toward larger farms, greater mechanization, and greater use of commercial inputs in most countries results in greater productivity but at the cost of greater direct and indirect energy use. The combination of increased irrigation, use of high yield variety crops, and new inputs has contributed to steady increases in both absolute and percapita agricultural production.

Modern Commercial Agriculture

To feed the growing population, agricultural productivity must increase. Modern inputs are needed to increase agricultural production. Commercial fuel inputs to agriculture include mechanized land preparation, mechanized irrigation, and synthetic fertilizer. Modern commercial agriculture is greatly dependent on high yielding varieties, and modern pesticides.

The degree of utilization of the above inputs varies widely, but generally increases with economic development. These modern technologies reduce the time and labor needed for land preparation, plowing, planting, and harvesting crops. In favorable areas, it also aids double cropping management.

In the developed countries, stages in the historical development of agricultural production have been characterized by differing combinations of inputs: differences in relative proportion of land, labor and capital, and in the composition of the capital inputs. Such changes reflect primarily the changing structure of the economy and the successive advances that have been made in agricultural technology.

The well-documented case of the United States serves as an illustration. During the period 1870 to 1900 the farm population was increased through a rapid expansion of the agricultural area. The agricultural labor force increased by 60 percent, but there was a replacement of labor by nonland capital in the form of horses and mules. New and more efficient types of horse-drawn machinery including plows, cultivars, seed drills, grain harvesters, and mowers became available.

The following period, from 1900 to 1930, was a period of transition: the first half was the beginning of the end for traditional farming, based predominantly on large inputs of relatively unskilled labor, and the second half the beginning of commercial agriculture, technologically oriented and capital intensive. The crop-land use continued to increase until about 1920 but remained relatively stable thereafter, and the agricultural labor force continued to increase until about 1918. Equally significant, however, were the shifts that became evident in the composition of non-real-estate capital input: the replacement of horses and mules by tractors, automobiles and trucks; of manpower by motor-driven machinery and equipment; and the purchase of production inputs such as fertilizer, lime, seed, and feed required for the application of improved production techniques.

Between 1935 and 1960 outputs per man-hour of labor increased about 4.5 times, and crop production per hectare of crop-land almost doubled. Inputs of labor were decreased by 50 percent, inputs of land remained relatively stable, but inputs of non-real-estate capital inputs were nearly tripled. Among these capital inputs, those of seed, feed, and livestock purchased increased by about four times, and those of mechanical power and machinery by more than 2.5 times.

It has been estimated that of the total U.S. increase in farm output between 1940 and 1955, 43 percent is attributable to increased crop yields per hectare, 27 percent to increases in value added by livestock production, 23 percent to reduction in farm-produced power, and 7 percent to changes in the amount of capital used. While it is not possible to isolate the effect of a single input, it is estimated that increased use of fertilizer accounted for more than half of the increase in crop production per hectare. Other important causes of increased crop yields include the use of hybrid maize and other improved plant varieties, irrigation, better soil tillage practices, more timely planning, cultivation and harvesting operations, and better weed, insect, and disease control.

ENERGY USE IN TRADITIONAL AND MODERN AGRICULTURE

Commercial energy plays a major role in modern methods of production. Energy use even with modern methods varies by crop. Crops requiring irrigation use much more energy than rain-fed crops. About 42 percent of total energy is used for irrigation, 20 percent for the manufacture and operation of farm machinery, 18 percent for fertilizers, and 7 percent for drying. For modern rice production in the United States, the commercial energy input is more than 500 times that in traditional production and more than 10 times in transitional production in the Philippines. For modern corn production in the United States, total commercial energy use is only about half that in modern rice production, mainly because little or no irrigation is required. Regardless of the crop produced, energy use is much higher with modern production method. For example, commercial energy input in the United States is more than 174 times that in traditional production in the Mexico.

With modern production methods 1500 kilograms of petroleum per hectare are needed for rice and 700 kilograms for maize. However, with this commercial energy use, yields of 5.8 metric tons per hectare have been obtained for rice and 5 metric tons per hectare for maize—about five times those obtained with traditional methods, Thus, 20 to 25 people can be fed on an all-grain diet from a single hectare compared with 4 to 6 people by traditional methods.

Total energy used to produce a kilogram of rice on Philippine traditional farms requires three grams of petroleum and in transitional farms 55 grams. In the U.S. modern rice farming system the requirement increases to 258 grams of petroleum. Similarly, the ratio for corn production in Mexico for traditional farming is equal to four grams of petroleum, and for modern corn farming in the United States is equal to 137 grams of petroleum.

Despite higher dependence on energy in modern methods of production, agriculture is responsible for only a small part of total energy use. Total energy use as well as energy used in agricultural production in selected developing and developed countries and for the largest food producers in the world; the United States and China are compared in Table 1. In 1996, agricultural shares of energy use for all developed countries were 2 percent, whereas in the developing countries the proportion was slightly higher. Agricultural shares of total commercial energy use in the developing countries ranged from 0.67 percent in Egypt to 5.08 percent in Brazil. For the developed countries agricultural share of energy ranged from 1.09 percent in the United States to 3.39 percent for the European Community.

COMMERCIAL ENERGY USE IN AGRICULTURE

Agriculture uses a variety of commercial energy forms, directly for operating machinery and equipment on the farm and indirectly through the use of fertilizer and pesticides produced off the farm. In addition, commercial energy is used in manufacturing of farm machinery and farm equipment. During 1972 to 1973 farm machinery manufacture and operation, the largest user of commercial energy in agriculture, accounted for 51 percent of the world total and ranged from 8 percent in the Far East to 73 percent in Oceania. Chemical fertilizer was second with 45 percent of the world total and ranged from 26 percent in Oceania to 84 percent in the Far East. However, in the developing countries chemical fertilizer was first.

About 1.5 billion hectares of crop land are planted for annual crops and permanent crops. About one third of the arable and permenant crop-land is located in the United States, India, and China. Out of 1.5 billion hectares, 263 million hectares are irrigated. Irrigated lands in China and India account for the largest share—about 110 million hectares.

Reports by the International Energy Agency indicate that total renewable energy use in developing countries and in agriculture is high, but no data is available. The available data on selected countries shows that the use of renewable energy in developing countries varies from 5 percent in Egypt to 86 percent in Nigeria. Energy share of the renewable energy in the developed countries ranged from 1 percent in Japan to 6 percent in Australia.

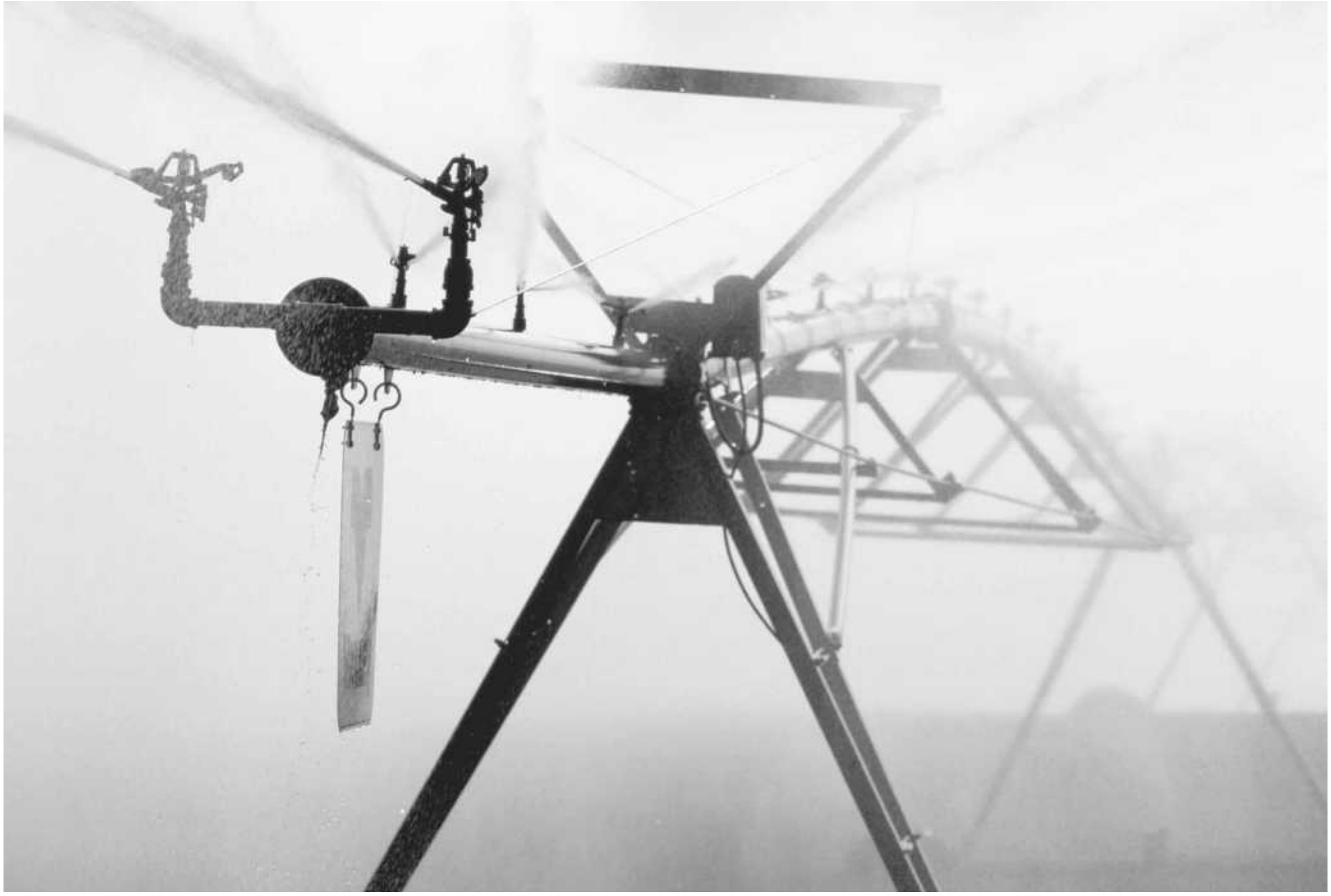
To measure the efficiency of energy use in agri-

	1996 Total energy use				
	Fossil	Renewable	Energy used in agriculture	Energy share of Renewable	Energy share of Agriculture
	Million tons of oil equivalent			percent	percent
Developing Countries:					
Bangladesh	22.32	15.76	0.48	71	
Brazil	138.23	33.57	7.02	24	5.08
China	865.86	206.08	26.52	24	3.06
Egypt	25.2	1.19	0.17	5	0.67
India	350.26	188.65	8.1	54	2.31
Indonesia	99.76	43.83	1.67	44	1.67
Nigeria	73.12	63		86	
Pakistan	46.82	21.09	0.84	45	1.79
Total developing	**3270.3**	**824.73**		**25**	
Developed Countries:					
Australia	66.11	3.99	1.5	6	2.27
Canada	181.92	8.58	4.13	5	2.27
EC(15) 1/	1010.54	26.37	23.15	3	2.29
Japan	337.08	3.42	11.44	1	3.39
USA	1443	30.09	15.8	2	1.09
Total developed	**3479.11**	**96.32**	**72.61**	**3**	**2.09**
World	**6749.41**	**921.05**		**14**	

Table 1.
Total fossil and renewable energy use, energy use in agriculture, and agriculture energy share for selected developed and developing countries, 1996.
Note: European Community (15) includes Austria, Belgium-Luxembourg, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal, Spain, Sweden, and United Kingdom.
SOURCE: *Energy Statistics and Balances of OECD and Non-OECD countries, 1995–96,* International Energy Agency. Paris, 1998.

culture, energy use in agriculture is compared with agriculture's share of gross domestic products (GDP). For many developing countries, the share of agricultural GDP is much larger than the share of energy used in agriculture, which suggests that much of the agricultural sector is still engaged in subsistence, low input, labor-intensive farming. For some middle income countries such as Brazil, Egypt, and Indonesia the two ratios are in better balance. Developed countries have a very close balance between agriculture's share of GDP and share of energy use in agriculture. Energy use share is lower than agricultural GDP share except for Japan and the European Union which both rely heavily on intensive production methods.

Energy use per hectare of arable and permanent crop-land can be used as an indicator of the efficiency of energy use in agriculture, although definite conclusions from this indicator are hard to draw unless countries with the same scale of operations are compared. Low energy use per hectare can reflect low-input, low-yield agriculture, which is the case for some developing countries such as Bangladesh, India, Indonesia, Nigeria, and Pakistan. On the other hand, low energy use (dry-land farming with less fertilizer and pesticide use) per hectare can also indicate the extensive farming practices of countries such as Australia and Canada where low yield relative to high-input use countries have not prevented these countries from being competitive agricultural exporters. Japan and the European Union stand out as being very high energy users relative to their land base. The United States has a rela-



Irrigation with mechanically powered pumps substantially increases commercial energy requirements but also increases yields. (Corbis-Bettmann)

tively low energy use per hectare, especially when compared to the European countries.

INPUT USE

Land Use and Irrigated Area

Using data from the United Nations Food and Agriculture Organization (FAO) for arable and permanent crop-land for 1966 to 1996, developing countries have experienced the most significant growth in land area. The highest growth was realized in Brazil, China, and India. In contrast, most developed countries have experienced a decline in arable and permanent crop-land, with Australia and Canada being the exceptions. For some developing countries, such as Egypt and India, increases in area as well as yields have been important in increasing output. For developed countries, increases in output have been a function of higher yields rather than increases in area. The U.S. agricultural area increased more than 12 million hectares during the 1970s and 1980s and declined in the 1990s, with an overall decline of one-half million hectares.

Irrigated areas increased in the last three decades in all countries except Japan. Growth has been rapid in Bangladesh and Brazil. Growth in irrigated areas in the United States has been among the slowest of the selected developed countries. The share of agricultural area under irrigation varies considerably, reflecting resource endowments and crop composition. In 1996, Egypt had 100 percent of its area under irrigation, followed by Pakistan with 81 percent and Japan with 63 percent. More than a third of land is irrigated in China, India, and Bangladesh. Canada, Australia, and the United States, with large-scale, extensive farming practices, have lower shares of irrigated areas.

Fertilizer Use

Fertilizer use (measured in tons of active ingredients) has increased most rapidly in developing countries, where yields generally lagged behind those in developed. Among developed countries, fertilizer use per hectare is very high in Japan and the European countries. Fertilizer use is also high in some of the developing countries, including Egypt, China, Bangladesh, and Pakistan. Fertilizer use in the United States is low compared to the European countries, as is true for other large crop producers such as Canada, Australia, and Brazil.

Pesticide Use

Data on pesticide use, including insecticides, herbicides, fungicides, and other chemicals is scarce. Similar to fertilizer use, pesticide use is the highest in the United States followed by the European countries, and Australia. Pesticide use per hectare is much higher than in other developed countries because farmers are trying to maximize the yield per hectare. Among developing countries pesticide use is also high in Egypt, Brazil, and Pakistan. The trend in developing countries is for more use of chemicals in agriculture to increase yield per hectare.

Machinery Use

Machinery use has followed a similar pattern to fertilizer use, with the largest increases occurring in developing countries and smaller increases or declines in developed countries. The data do not reflect changes in quality or complexity of agricultural machinery. The relative intensity of use is highest in Japan followed by the European countries. Machinery use in the United States is much lower than in Europe and has remained steady because the average farm is much larger in the United States, which translates into better utilization of farm equipment. Machinery use increased in all developing countries. Growth has been rapid in Pakistan, Indonesia, and India, followed by Japan. In the developing countries, machinery use per hectare is lowest in Bangladesh and highest in Egypt.

Roger Conway
Hosein Shapouri

See also: Biological Energy Use, Ecosystem Functioning of; Chemical Energy, Historical Evolution of the Use of; Nitrogen Cycle.

BIBLIOGRAPHY

Dovring, F. (1988). *Farming for Energy: The Political Economy of Energy Sources in the United States.* New York: Praeger Publishers.

Food and Agricultural Organization of the United Nations. (1977). *The State of Food and Agriculture, 1976*, pp. 79–110. Rome: United Nations.

Food and Agricultural Organization Homepage. (1999). *Data on Production, Trade, and Agricultural Inputs.* Rome: United Nations.

International Energy Agency. (1998). *Energy Statistics and Balances of OECD and Non-OECD Counties, 1995-96.* Paris: United Nations

Stanhill, G. (1984). *Energy and Agriculture.* Germany: Springer-Verlag.

Stout, B. A. (1979). *Energy for World Agriculture.* Rome: United Nations.

Stout, B. A. (1984). *Energy Use and Management in Agriculture.* North Scituate, MA: Breton Publishers.

U.S. Congress. (1991). *Energy in Developing Countries. Office of Technology Assessment.* Washington, DC: U.S. Government Printing Office.

U.S. Department of Agriculture. Economic Research Service. (1997). *Agricultural Resources and Environmental Indicators, 1996–97.* Washington, DC: U.S. Government Printing Office.

World Bank. (1998). *World Development Indicators, 1998.* Washington, DC: Author.

AIR CONDITIONING

Air conditioning is the treatment of air to control simultaneously its temperature, humidity, cleanliness, and distribution to meet the requirements of a conditioned space. Common use of the term "air conditioning" applies it to the cooling of air; however, true air conditioning treats all aspects of indoor atmospheric comfort.

An air conditioning system uses an assembly of equipment to treat air. Normally the assembly includes a heating system for modifying winter indoor temperature and humidity; a refrigeration system for modifying summer temperature and humidity, a means to maintain indoor air quality (i.e., air filters and fresh air intake); a method of distribution of conditioned air; and a control system, such as a thermostat, to maintain desired comfort conditions.

Air conditioning systems fall into two broad classes. Comfort air conditioning, which accounts for most applications, is used to modify and maintain the indoor environment for habitation. Process air conditioning is the modification of the indoor environment to enhance an industrial or a scientific process.

Most air conditioning systems utilize a vapor-compression refrigeration system (Figure 1) to transfer the indoor heat to a suitable heat sink such as the outdoors. Vapor-compression refrigeration systems employ a cycle in which a volatile liquid, the refrigerant, is vaporized, compressed, liquefied, and expanded continuously in an enclosed system. A compressor serves as a pump, pressurizing the refrigerant and circulating it through the system. Pressurized refrigerant is liquefied in a condenser, liberating heat. Liquid refrigerant passes through an expansion device into an evaporator where it boils and expands into a vapor, absorbing heat in the process.

Some air conditioning systems use an absorption refrigeration system (Figure 2). Absorption refrigeration systems work by evaporating refrigerant in an evaporator, with the refrigerant vapor then absorbed by an absorbent medium, from which it is subsequently expelled by heating in a generator and changed back into liquid in a condenser. Absorption systems may use a pump to help circulate the refrigerant. The most common absorption systems used for air conditioning use water as an absorbent and ammonia as a refrigerant or lithium bromide salt as an absorber and water as a refrigerant.

For heating purposes, most air conditioning systems use fossil-fueled furnaces to heat air, or boilers to heat water or produce steam. Forced air systems use a blower fan and ductwork to distribute conditioned air to points of use. Air quality is enhanced in forced-air systems through the use of filters. The filters are normally placed in the return air, just before the heating and cooling components. Provision may be made for fresh outdoor air to be added to recirculated room air. A hydronic heating system uses hot water to convey heat from a boiler to radiators, convectors, or wall and floor panel coils. Most steam heating systems use boiler-produced steam to heat buildings via radiators, convectors, or heating coils placed in ductwork. Heating systems may employ humidifiers in winter to counter the drying effect of heated air, particularly in forced-air systems.

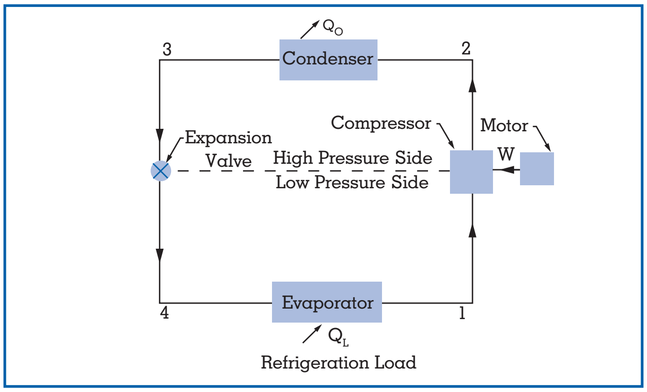


Figure 1.
Equipment diagram for refrigeration using a basic vapor compression cycle.

Summer comfort cooling can increase electrical loads as much as 50 percent over average consumption. The most uncomfortable summer days, those with both high temperature and high relative humidity, increase both sensible and latent system load. The sensible load increases because the difference between indoor and outdoor temperature is greater, requiring the cooling system to move heat to a higher level. The latent load increases as humidity rises, since the cooling system must extract more moisture from the air.

Many air conditioning systems use simple thermostats to cycle equipment; however, more sophisticated control systems employing electronics and microprocessors can reduce energy consumption.

Automotive air conditioning systems provide simple heating and cooling/dehumidification functions. There is no provision for filtration of the air. All current automotive air conditioning utilizes vapor-compression refrigeration systems coupled to the automobile's engine.

Air conditioning systems are categorized by the method used to control cooling or heating in a conditioned space. They are further described based on their terminal cooling or heating media. The most common type for cooling is the refrigerant based all-air system which uses a refrigeration cycle to transfer heat from indoor air to outdoor air with heat exchangers commonly called *coils*. Most heating systems in residences and small buildings are all-air-using fossil-fueled furnaces. Hydronic (hot water) or steam heating systems also are common, particularly in large buildings. Many systems employ unitary equipment—that is, they consist of

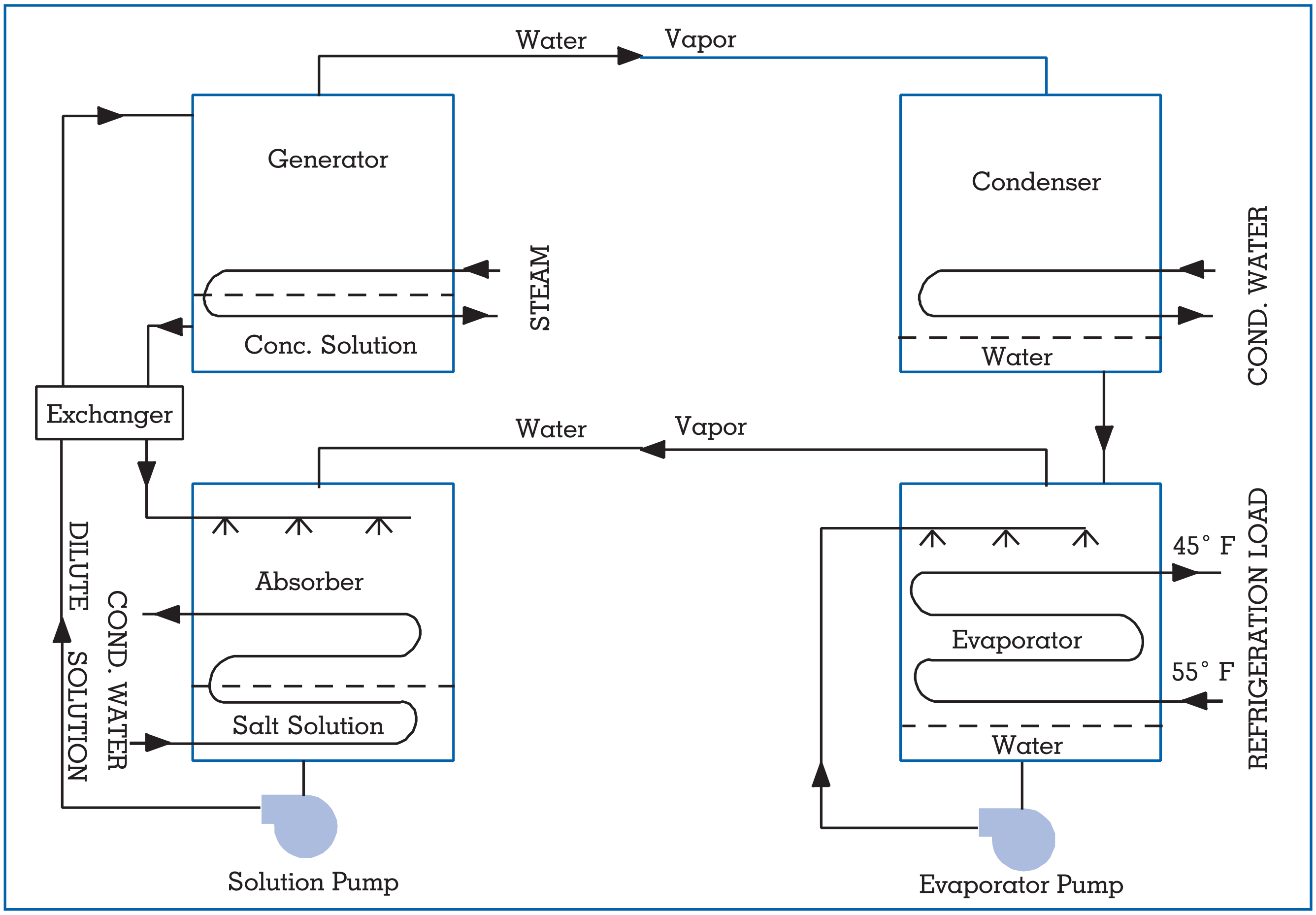


Figure 2.
Absorption type refrigeration using water as a refrigerant and lithium-bromide salts as an absorber.

one or more factory-built modules. Larger buildings may require built-up systems made of various system components that require engineering design tailored for the specific building. Sometimes multiple unitary units are used in large buildings for ease of zone control where the building is divided into smaller areas that have their own thermostats controlling that space.

Selection of air conditioning equipment depends on balancing various criteria, including performance, capacity, special requirements, initial and continuing operating costs; reliability; flexibility and maintainability.

Air conditioning systems have been traditionally compared and rated by cooling and/or heating capacity and, more recently, energy efficiency. Capacity is expressed in British thermal units per hour (Btuh) or in watts. Energy efficiency is expressed as the operating efficiency using the term "energy efficiency ratio" (EER). In the U.S. EER is expressed as British thermal units per hour per watt (Btuh/w). For window air conditioners, an energy usage label is required by the 1975 Energy Policy and Conservation Act. Unitary air conditioners are rated at operating conditions standardized by the Air Conditioning and Refrigeration Institute (ARI). Standard ratings provide the consumer comparisons between competing models. This information is published annually in ARI's *Unitary Directory*.

Higher EERs in air conditioning equipment are not the sole answer to reducing energy consumption. Proper sizing, reduction of building air leakage, increasing insulation, reducing unnecessary internal energy usage, proper conditioned air distribution, use of night temperature setback or increase, and proper maintenance of equipment can be greater contribute factors for reducing energy cost. For example, a properly sized low-efficiency system operating in a well-sealed and insulated building can consume less energy than a high-efficiency system in a poorly insulated building with lots of outside air infiltration from poorly fitting windows and doors.

HISTORY

The invention of air conditioning is actually a progression of the applied ideas of many individuals starting in the early nineteenth century and dramatically accelerating in the twentieth century. Most air conditioning developments occurred in the twentieth century.

Because of the limitations of energy sources and the technology to top that energy, mechanical air conditioning was not a practical possibility until the dawn of the Scientific Age and the Industrial Revolution. In 1813, Sir John Leslie of England made one of the earliest proposals to use artificial cooling for human comfort. Jean Frederic Marquis de Chabannes followed in 1815 with a British patent for use of a centrifugal fan to force heated or cooled air through ducts to rooms. The major breakthrough came in 1834 when Jacob Perkins invented the vapor compression refrigeration system, making it possible to cool air mechanically. David Boswell Reid designed the first building air conditioning system, using water sprays, for the British Houses of Parliament in about 1836.

In the United States, Dr. John Gorrie, unaware of the work of Perkins, proposed that mechanical refrigeration be used for comfort cooling, and he constructed mechanical systems for cooling his patients at his home in Florida in about 1842.

Although limited experiments, such as Gorrie's were being conducted, there was little understanding of the science involved in cooling and dehumidification. The first engineering textbook for heating and cooling, *Guide to Calculating and Design of Ventilating and Heating Installations* by Hermann Rietschel, was published in Germany in 1894. Rietschel's chapter on room cooling was the earliest comprehensive example of a real scientific approach to comfort cooling.

Rietschel's engineered approach was introduced in the United States by consulting engineer Alfred Wolff, who designed the first modern energy-saving air conditioning system for the New York Stock Exchange in 1901. Wolff's huge system used waste steam from the building power plant to power an absorption-type refrigeration system to provide comfort cooling. The fact that the system used steam that would have been thrown away meant that the energy needed for the cooling plant was free! The system operated successfully for twenty years. Wolff designed several other comfort cooling systems for large buildings before his death in 1909.

Textile engineer Stuart Cramer first published the term "air conditioning" in 1906, and G. B. Wilson defined air conditioning as the control of temperature, humidity, and air quality in 1908.

Process or industrial air conditioning was proposed, and a few examples installed, by the late nineteenth century. Willis Haviland Carrier devoted his engineering career to air conditioning, catering to industrial needs beginning in 1902. Carrier took an engineering, scientific, and business approach to air conditioning, becoming its greatest early-twentieth-century practitioner. Carrier patented dewpoint control, used for precise humidity control, in 1907. At about the same time, Carrier devised a psychrometric chart for calculating air conditions, which became an essential engineering tool in use to this day. He founded the first design and manufacturing company exclusively devoted to air conditioning. In the 1920s Carrier expanded his interest to comfort cooling.

Before the twentieth century, few homes or public venues experienced mechanical comfort cooling. A curious public was exposed enmasse to the pleasures of summer cooling at the Louisiana Purchase Exposition in St. Louis in 1904, where the Missouri State Building had an air conditioned amphitheater. A hospital in Boston had air conditioned wards in 1906. Some hotels installed cooling systems for lobbies, meeting halls, and restaurants after 1907. Motion picture theaters began to install mechanical comfort cooling systems after 1915.

Air conditioned theaters produced a two-pronged demand for comfort cooling. Consumers liked it, asked for it, and patronized those theaters offering it. Increased attendance at cooled theaters showed that the installation and operating costs were worthwhile expenditures, causing more theater owners to decide to purchase comfort cooling systems.

The aforementioned applications of air conditioning were possible because a commercial advantage was present. Limited refrigeration technology contributed to high installation costs. These costs could

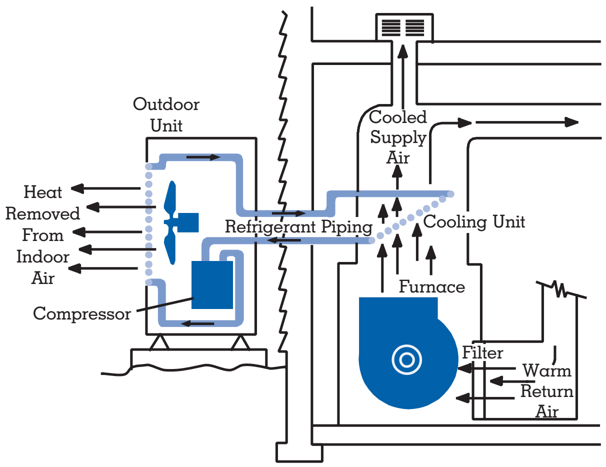


Figure 3.
Schematic drawing of a central air conditioning system.

be reduced by using ice instead of mechanical refrigeration, however, ice-type systems did not dehumidify as well, and could present higher operation costs, dependent upon the cost of ice. Still, it often paid to incur the costs of air conditioning where a commercial advantage was present.

A commercial advantage was present in many industrial processes. Uncontrolled heat and humidity impacted some products such as chocolate, pasta, textiles, and tobacco. Air conditioning allowed for uniform and continuous production despite weather conditions, reduced spoilage and waste, and thus saved money—enough in many cases to easily justify the installed and operating costs.

Thus, most of the uses of air conditioning before the 1930s concentrated on applications that had a viable financial payback. There was no obvious financial advantage in air conditioning homes, and this branch of air conditioning developed later. Home air conditioning passed from luxury to necessity only when the installed cost decreased and when air conditioning systems became worry-free.

The perfection of the electric household mechanical refrigerator by the 1930s provided a technology applicable to home air conditioning. This technology allowed mass production of lower cost, reliable package air conditioners for homes. It was no accident that the first room cooler was introduced by the major refrigerator manufacturer, Frigidaire, in 1929, followed swiftly by others such as General Electric, Kelvinator, and Westinghouse. These early package air conditioners were not central systems and thus were applicable to homes and small commercial establishments without any modification of an existing heating system. These simple cooling devices, forerunners of the window air conditioners introduced in the late 1930s, were the beginnings of affordable air conditioning.

Central air conditioning systems for homes were available in the early 1930s from several manufacturers (Figure 3). These were combined with automatic fossil-fueled heating systems, a new innovation of the time. Reliable, thermostatically controlled automatic oil, gas, or coal-firing systems had only become available after the late 1920s. Before that, most homes and buildings had used coal that was hand-fired in all but the largest installations.

The safety of refrigeration and air cooling systems had always been an issue due to the toxicity or flammability of most refrigerants. In fact, increasing prevalence of refrigeration caused accidents and deaths, and the bad publicity and restrictive legislation were becoming serious threats to the growth of refrigeration and air conditioning. Fortunately, a solution was found when Thomas Midgley, Albert Henne, and Robert McNary invented chlorofluorocarbon refrigerants (CFCs) for Frigidaire. They were introduced in 1930 and, with the realization of their overall importance to health, safety, and the future of refrigeration and air conditioning, CFCs were made available to the entire industry. The CFCs made it possible to engineer air conditioning systems for any application without fear of safety issues. All other refrigerants used for small refrigeration and most air conditioning systems were soon completely displaced. The CFC refrigerants were applied to air conditioning systems in the early 1930s. One early use was in air conditioning for passenger trains. By 1936 all long-distance dining and sleeping cars on U.S. railroads were air conditioned.

Most homes were not air conditioned by the 1940s, and the costs involved in installing air conditioning in existing homes were prohibitive until the window air conditioner was introduced. By the late 1940s the cost and reliability of a window air condi-

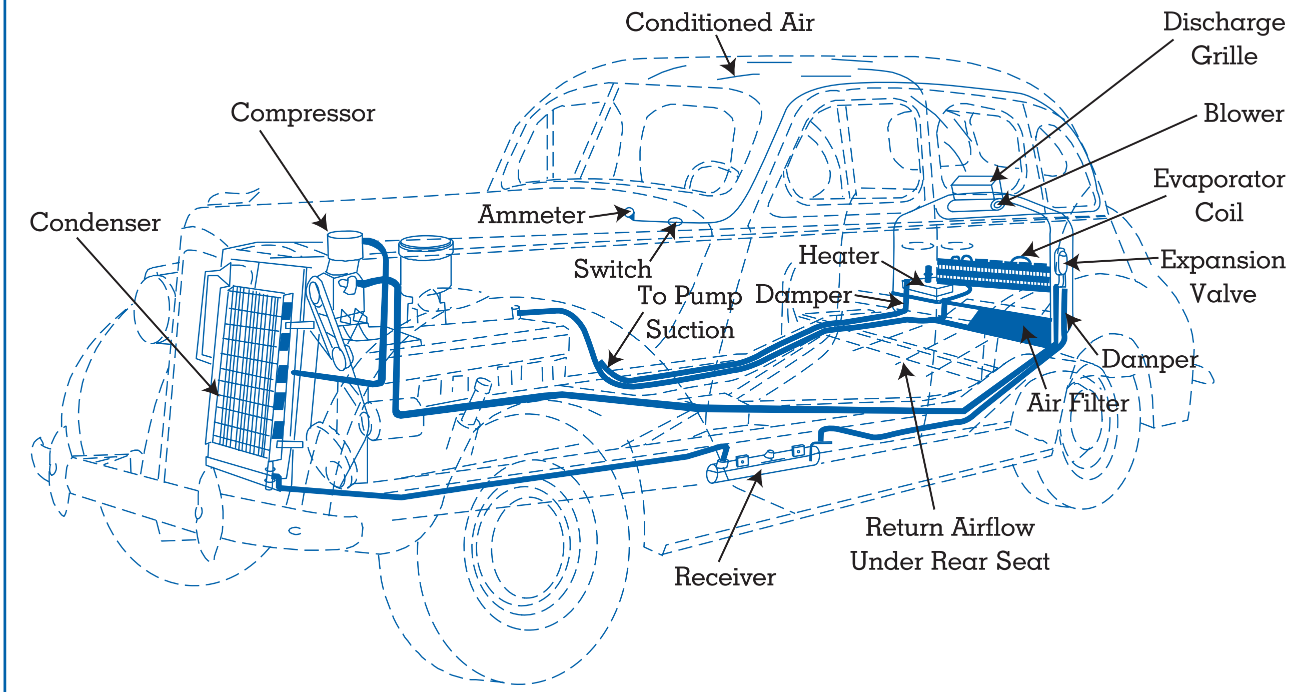


Figure 4.
The first automotive air conditioning system (1938).

tioner were such that middle-class homeowners could afford them. Sales rocketed from about 50,000 in 1946 to more than 1 million in 1953. The window air conditioner has become so relatively inexpensive that it ceased to be a luxury. Sixteen million window air conditioners were manufactured in 1992. The experience of window air conditioned homes no doubt played a part in later demand for centrally air conditioned homes as homeowners moved on to newer homes. The window air conditioner, combined with automatic central heating, now made it possible to live, work, and play indoors oblivious to the environment outside. In fact, the ability to control the indoor environment so effectively is credited with helping to reverse population migration out of the southern United States after 1960.

Automotive air conditioning systems using vapor compression refrigeration systems and CFC refrigerants were proposed in 1932 and debuted in the late 1930s, but system problems retarded its popularity until the 1950s. Auto air conditioning soared in the 1960s; for example, American car installations tripled between 1961 and 1964.

By the 1960s, central residential air conditioning was becoming increasingly popular. Equipment had developed to the point that a number of manufacturers were producing and marketing unitary air conditioning equipment. Most new homes were designed to include central air conditioning. This was true of commercial buildings also. In fact, the use of air conditioning changed architectural design. By removing all environmental restraints, air conditioning made it possible to design almost any type of building. Hollow ventilation cores were no longer necessary, and the “windowless” building made its debut. Residential construction changed too. Deep porches for shading were no longer necessary, and a home site was no longer dependent on prevailing summer breezes.

During the 1960s, air conditioning system control began to shift from simple electrical or pneumatic control to electronic and rudimentary computer control. This new technology was applied to commercial, institutional and commercial buildings where the high cost and complicated nature justified its use.

The oil embargo of the early 1970s gave impetus to increasing energy efficiency. Manufacturers and designers responded by developing and using higher- efficiency compressors for cooling. Condenser size was increased to lower system pressures, resulting in energy savings. Heating boilers and furnaces were redesigned for higher efficiencies. Building construction methods were revised to include more insulation and to reduce outside air infiltration. Existing buildings were scrutinized for energy inefficiencies. Some building owners went to the extreme of closing up outside air intakes; the unintended effect being stale building air and occupant complaints.

Overall efficiency of air conditioning equipment steadily rose starting in the mid-1970s, attributed to consumer demand, government mandate, and incentive programs. For example, the average efficiency, as expressed in seasonal energy efficiency ratio, of new central air conditioners increased about 35 percent between 1976 and 1991. After national standards took effect in 1992, efficiency has increased as much as another 15 percent.

Increased system efficiency has resulted in higher equipment costs but lower operating costs. A higher initial cost for equipment can often be justified by the monetary savings from lower energy consumption over the life of the equipment.

CFC PHASEOUT

The air conditioning industry was challenged to reinvent one of its vital system components, refrigerants, when CFCs were targeted as a prime cause of high-level atmospheric ozone depletion. Ozone, an active form of oxygen, is present in the upper atmosphere. One of its functions is to filter out solar ultraviolet radiation, preventing dangerous levels from reaching the ground.

A hypothesis that halocarbons, including CFCs, diffused into the upper atmosphere, could break down, and an ozone-destroying catalytic reaction could result, was published in 1974. Computer modeling of the hypothesis showed that such destruction could happen. The resulting increase in UV radiation would have adverse health and biological system consequences. This scenario so alarmed nations that various measures to control CFC emissions were undertaken. Although there were scientific uncertainties concerning the depletion hypothesis, the United States became the first nation to ban nonessential CFC use in aerosol sprays, in 1978.

Other nations followed with various control measures, but CFCs were also widely used as refrigerants, fire-extinguishing agents, insulation components, and solvents. The UN Environment Program began working on a worldwide CFC control scenario in 1981, culminating in the Montreal Protocol of 1987 which called for a phaseout of CFC production over time. In the United States, an excise tax on CFCs was passed, steadily increasing year by year so as to make CFCs increasingly more costly to use. As a result of various measures, CFC consumption by 1999 had decreased more than 50 percent.

Phaseout of the CFC refrigerants posed a challenge for the refrigeration and air conditioning industries. If the primary refrigerants in use were to be replaced, what was to be used instead? Some previously used refrigerants, such as hydrocarbons and ammonia, were proposed; however, safety and litigation fears eliminated them from serious consideration. The producers of CFCs, pursuing their vested interest, conducted extensive research, resulting in a number of alternative refrigerants that are widely accepted today. These alternatives do not contain chlorine, the element responsible for the ozone-depleting reaction.

Replacement of halocarbon refrigerants has increased the cost of refrigeration and air conditioning systems since the new refrigerants are more costly, system components need to be redesigned for the new refrigerants, and service and installation are more complicated. For example, most refrigeration and air conditioning systems used one of four refrigerants before 1987, but now there are more than a dozen alternatives.

RECENT TRENDS IN AIR CONDITIONING

Energy efficiency had always been a goal of building owners simply because they were trying to reduce operating costs. Sometimes there was a trade-off in personal comfort. Today's emphasis on energy conservation also must consider a balance of comfort in all forms: temperature and humidity control, and indoor air quality. Air conditioning engineers now have awareness that energy must be saved, but at the same time reasonable comfort, and therefore productivity, must be maintained. Modern air conditioning system and equipment design coupled with responsible building architecture have resulted in indoor environments that minimize the trade-offs. The percentage of new homes built with central air conditioning increased from 45 percent in 1975 to 80 percent in 1995. As of 1997, 41 percent of all U.S. households used central air conditioning, and an additional 30 percent used room air conditioners. By 1990, 94 percent of new cars sold in the United States had air conditioning systems.

The trend has been toward progressively higher energy efficiencies. Per-capita end-use energy consumption began decreasing in the early 1970s and has trended downward ever since. National minimum efficiency requirements for room and central air conditioners were enacted through the National Appliance Energy Conservation Act (NEACA) in 1987. The requirements took effect in 1990 for room units and in 1992 for central air conditioners.

For air conditioning systems used in commercial buildings, minimal efficiency targets are published in the form of Standards by the American Society of Heating, Refrigerating and Air Conditioning Engineers (ASHRAE). These standards influenced building energy codes and legislated minimum standards included in the U.S. Energy Policy Act of 1992. ASHRAE, in addition to developing standards for design of equipment, publishes the most comprehensive air conditioning design and application information in their *Handbook*, published annually.

Several innovations have contributed to more efficient and comfortable air conditioning. In large buildings, indoor environmental conditions can be determined and even corrected from a central, computer-controlled console. Programmable, microprocessor control systems are seeing increasing use in large buildings, and their cost is decreasing, improving the probability that they will be used in residences. Microprocessors permit design of air conditioning systems that can respond to occupant and building needs at a scale only dreamed of before. Direct digital control of conditioned air outlets permits buildings to be divided into mini zones. The result is customized energy use to match the zone's needs.

Motion detectors and daylight detectors reduce artificial lighting use, reducing summer cooling loads.

Variable-speed fan drives permit conditioned air distribution to be matched more closely to a building's needs. High-efficiency electric motors are used to drive the fans, saving as much as half the energy once used. Both variable-speed and high- efficiency motors are being applied even in residential air conditioning systems.

New compressor technology employing rotary scroll-type compressors is replacing previously used reciprocating technology. Scroll compressors operate at higher efficiencies over wider operating conditions. Electrical and electronic technology are making variable-capacity compressors cost-effective for the twenty-first century. Compressor performance can be optimized for a wider operation range, and matching compressor capacity to the actual demand for cooling increases the system efficiency, saving energy.

Thermal storage is being used to reduce energy needs. Cooling systems are being designed to "store" cooling at night, when energy cost and consumption may be lower, and release the stored cooling during the day. Building design itself has begun to change so the energy storage capability of the building mass itself can be used. Shading and "low e" glass is being used to reduce heat gains and losses through windows.

Energy recovery ventilators, which move energy between outgoing stale building air and incoming fresh air, are being used to improve indoor air quality without large increases in energy consumption.

Cooling systems that use refrigeration systems to cool and partially dehumidify air are being combined with desiccant dehumidification to reduce energy costs. Dessicant systems use a regenerable moisture absorbant to extract humidity from the air to be conditioned. Saturated absorbant is exposed to a higher air temperature, releasing the absorbed moisture. The absorbant is then recycled. Some systems use rotating wheels to continuously recycle the absorbant. Absorption-type refrigeration systems are used in these hybrid systems where the waste heat can be used to continuously regenerate the desiccant.

Research continues in both industry and government to find ways to innovate and improve air conditioning systems. For example, ASHRAE maintains an ongoing research program covering energy conservation, indoor air quality, refrigerants and environmentally safe materials. The society's journal maintains an HVAC&R search engine on the Internet covering more than 700 related web sites. In addition, the U.S. Department of Energy funds research and development on advanced air conditioning technologies such as innovative thermally activated heat pumps.

THE FUTURE OF AIR CONDITIONING

Precise and sophisticated control of air conditioning systems will be the trend of the near future. The explosion of innovation in computer and electronic technology will continue to impact the design and operation of air conditioning systems. Buildings will become "intelligent," their internal systems responding to changing environmental and occupancy conditions. "Cybernetic" building systems will communicate information and control functions simultaneously at multiple levels for various systems, including heating, cooling, and ventilation energy management, fire detection, security, transportation, and information systems. "Interoperability" in control systems will allow different controls to "talk to each other" in the same language. Wireless sensors will be used, allowing easy retrofit of older buildings.

The recent trend of integrating building design with the environment will result in energy savings as old concepts of natural ventilation, shading, and so on are reapplied. Technological innovation will permit increased use of solar technology as costs decrease. Hybrid cooling systems using both electricity and gas will be used in greater numbers.

The current debate over controversial global warming theories will continue. The impact of carbon dioxide levels in the atmosphere, whether they are increasing or not over time, and the effect on climate and economics will continue to be discussed. A solution, if it is needed, may evolve—or not.

Science, technology, and public need will continue to interact in ways that cannot be accurately predicted, each providing a catalyst for change at various times. However, the history of the development of air conditioning has shown that the trend has been beneficial. No doubt it will continue to be.

Bernard A. Nagengast

See also: Air Quality, Indoor; Building Design, Commercial; Building Design, Energy Codes and; Building Design, Residential; Energy Management Control Systems; Heat and Heating; Heat Pumps; Insulation; Refrigerators and Freezers; Water Heating.

BIBLIOGRAPHY

Air Conditioning and Refrigeration Institute. (1999). *ARI Unitary Directory.* <http://www.ari.org>.

American Society of Heating, Refrigerating and Air Conditioning Engineers. (1978). *ASHRAE Composite Index of Technical Articles, 1959–1976.* Atlanta: Author.

American Society of Heating, Refrigerating, and Air Conditioning Engineers. (1989). *CFCs: Time of Transition.* Atlanta: Author.

American Society of Heating, Refrigerating and Air Conditioning Engineers. (1989) *Standard 62-1989, Ventilation for Acceptable Indoor Air Quality.* Atlanta: Author.

American Society of Heating, Refrigerating, and Air Conditioning Engineers. (1989). *Standard 90.1-1989, Energy-Efficient Design of New Buildings Except Low-Rise Residential Buildings.* Atlanta: Author.

American Society of Heating, Refrigerating, and Air Conditioning Engineers. (1993) *Standard 90.2-1993, Energy-Efficient Design of New Low-Rise Residential Buildings.* Atlanta: Author.

American Society of Heating, Refrigerating, and Air Conditioning Engineers. (1995) *Standard 100-1995, Energy Conservation in Existing Buildings.* Atlanta: Author.

American Society of Heating, Refrigerating, and Air Conditioning Engineers. (1996). *ASHRAE Handbook: Heating, Ventilating, and Air Conditioning Systems and Equipment.* Atlanta: Author.

American Society of Heating, Refrigerating, and Air Conditioning Engineers. (1997). *Refrigerants for the 21st Century.* Atlanta: Author.

American Society of Heating, Refrigerating, and Air Conditioning Engineers. (1999). *ASHRAE Handbook: Heating, Ventilating, and Air Conditioning Applications.* Atlanta: Author.

Bhatti, M. S. (1999). "Riding in Comfort: Part II: Evolution of Automotive Air Conditioning." *ASHRAE Journal* 41 (Sept.):44–52.

Chartered Institution of Building Services Engineers. (1994). *The International Dictionary of Heating, Ventilating, and Air Conditioning.* London: Author.

Cooper, G. (1998). *Air Conditioning America: Engineers and the Controlled Environment, 1900–1960.* Baltimore: Johns Hopkins University Press.

Donaldson, B., and Nagengast, B. (1994). *Heat & Cold: Mastering the Great Indoors: A Selective History of Heating, Ventilation, Refrigeration, and Air Conditioning.* Atlanta: American Society of Heating, Refrigerating, and Air Conditioning Engineers.

Elliott, C. D. (1992). *Technics and Architecture*. Cambridge, MA: MIT Press.

Howell, R.; Sauer, H.; and Coad, W. (1997). *Principles of Heating, Ventilating, and Air Conditioning*. Atlanta: American Society of Heating, Refrigerating, and Air Conditioning Engineers.

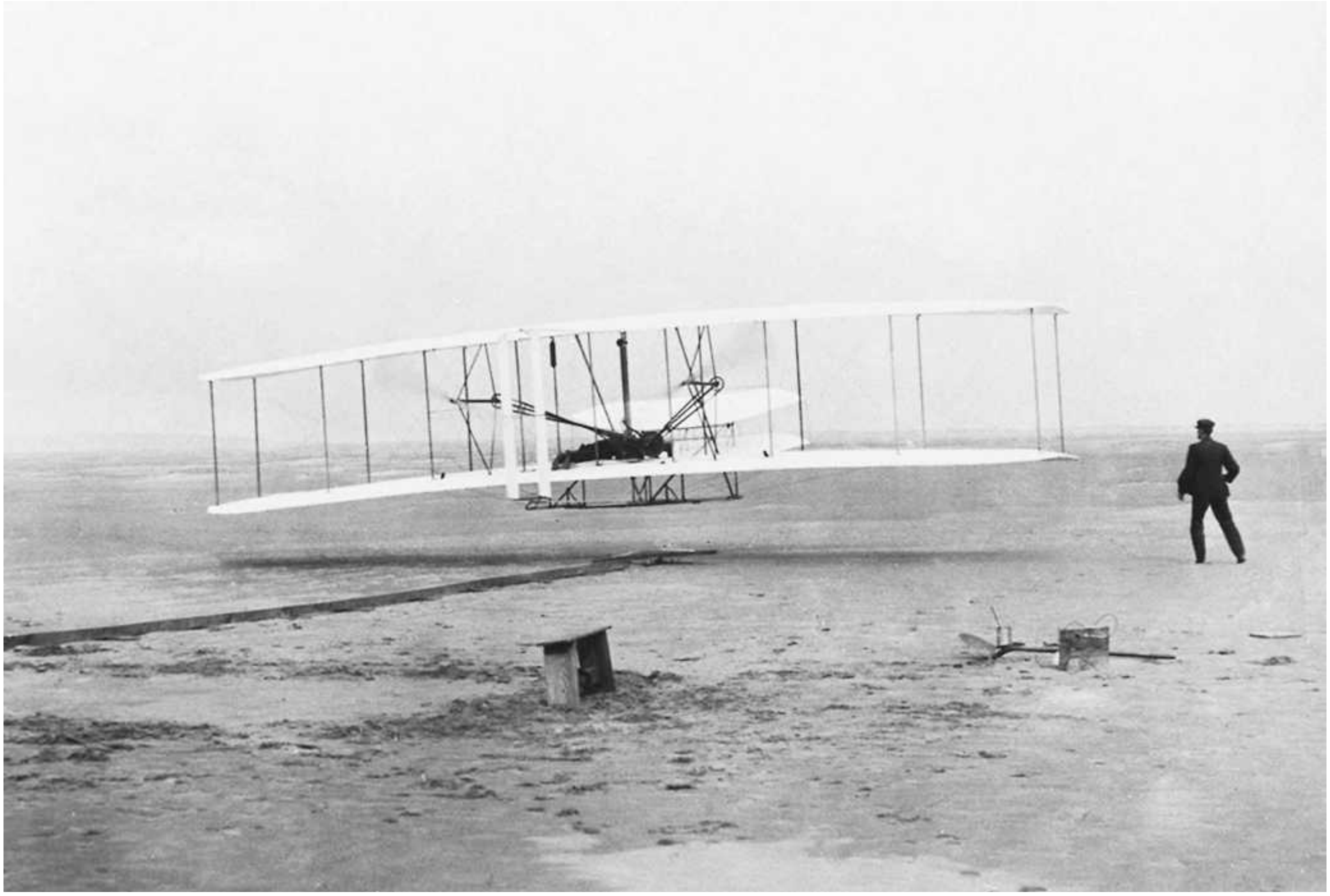
Hunn, B. D. (1996). *Fundamentals of Building Energy Dynamics*. Cambridge, MA: MIT Press.

Lorsch, H. G. (1993). *Air Conditioning Design*. Atlanta: American Society of Heating, Refrigerating, and Air Conditioning Engineers.

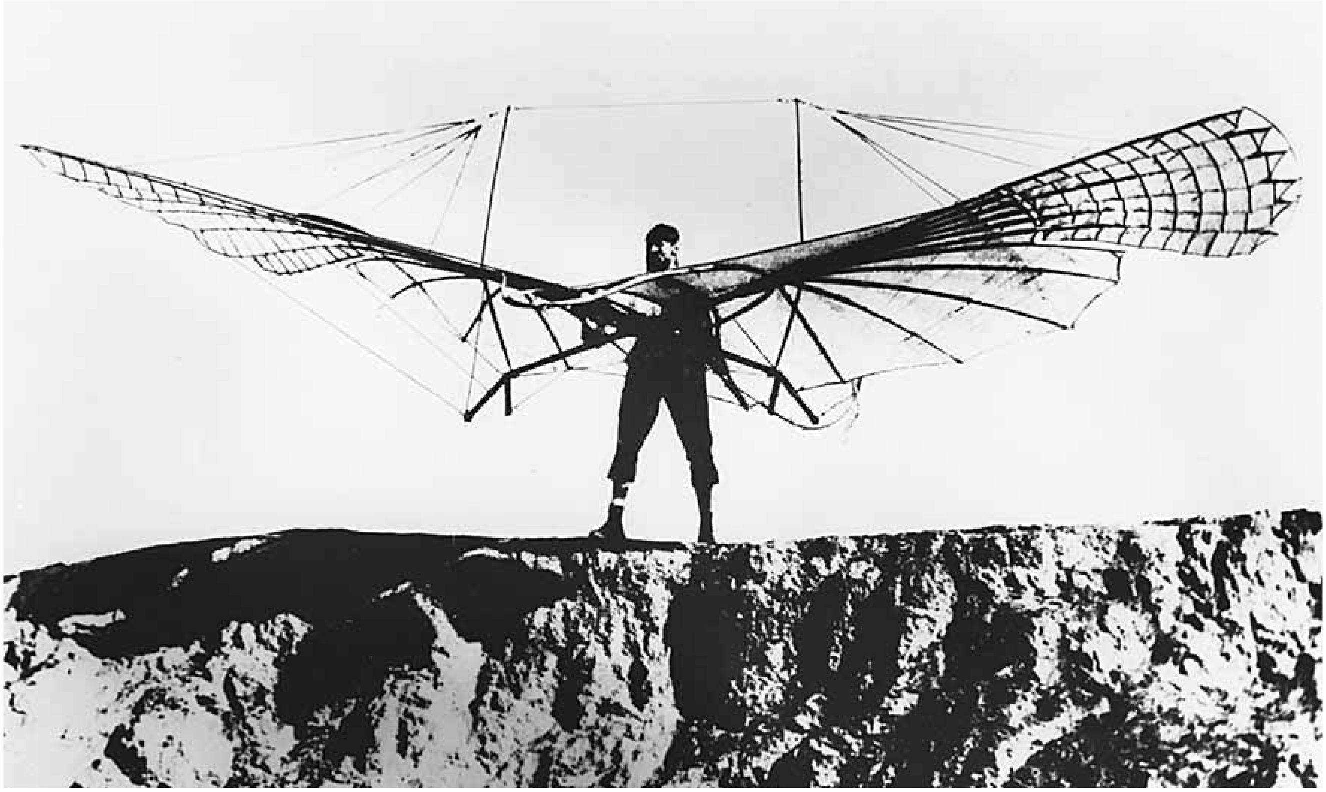
AIRCRAFT

Today the airplane is part of everyday life, whether we see one gracefully winging overhead, fly in one, or receive someone or something (package, letter, etc.) that was delivered by one. The invention and development of the airplane is arguably one of the three most important technical developments of the twentieth century—the other two being the electronics revolution and the unleashing of the power of the atom.

The first *practical* airplane was invented by Orville and Wilbur Wright, two bicycle shop proprietors from Dayton, Ohio. On December 17, 1903, the Wright Flyer lifted from the sand of Kill Devil Hill near Kitty Hawk, North Carolina, and with Orville at the controls, flew a distance of 120 ft above the ground, staying in the air for 12 sec. It was the first successful, sustained flight of a heavier-than-air piloted airplane. The photograph of the Flyer as it is lifting off the ground, with Wilbur running alongside to keep the right wing tip from digging into the sand, is the most famous photograph in the annals of the history of aeronautics. There were three more flights that morning, the last one covering a distance of 852 ft above the ground, and remaining in the air for 59 sec. At that moment, the Wright brothers knew they



First flight of the Wright Flyer, December 17, 1903. (Library of Congress)



Otto Lilienthal, with his collapsible glider, near Rathenow, East Germany. (Corbis-Bettmann)

had accomplished something important—a feat aspired to by many before them, but heretofore never achieved. But they had no way of knowing the tremendous extent to which their invention was to dominate the course of events in the twentieth century—technically, socially, and politically.

The history of the technical development of the airplane can be divided into four eras: pre-Wright; the strut-and-wire biplane; the mature propeller-driven airplane, and the jet-propelled airplane. We will organize our discussion in this article around these four eras.

THE PRE-WRIGHT ERA

Before the Wright brothers' first successful flight, there were plenty of attempts by others. Indeed, the Wrights did *not* invent the first airplane. They inherited a bulk of aeronautical data and experience achieved by numerous would-be inventors of the airplanes over the previous centuries. In many respects, when the Wright brothers began to work on the invention of the practical airplane, they were standing on the shoulders of giants before them.

From where and whom did the idea of the modern configuration airplane come? The modern configuration that we take for granted today is a flying machine with fixed wings, a fuselage, and a tail, with a separate mechanism for propulsion. This concept was first pioneered by Sir George Cayley in England in 1799. Cayley is responsible for conceiving and advancing the basic idea that the mechanisms for lift and thrust should be separated, with fixed wings moving at an angle of attack through the air to generate lift and a separate propulsive device to generate thrust. He recognized that the function of thrust was to overcome aerodynamic drag. In his own words, he stated that the basic aspect of a flying machine is "to make a surface support a given weight by the application of power to the resistance of air."

To key on Cayley's seminal ideas, the nineteenth century was full of abortive attempts to actually build and fly fixed-wing, powered, human-carrying flying machines. Cayley himself built several full-size aircraft over the span of his long life (he died in 1857 at the age of eighty-three), but was unsuccessful in achieving sustained flight. Some of the most important would-be inventors of the airplane were William Samuel Henson and John Stringfellow in England, Felix Du Temple in France, and Alexander Mozhaiski in Russian. They were all unsuccessful in achieving sustained flight. In regard to the nature of airplane performance and design, we note that these enthusiastic but unsuccessful inventors were obsessed with horsepower (or thrust). They were mainly concerned with equipping their aircraft with engines powerful enough to accelerate the machine to a velocity high enough that the aerodynamic lift of the wings would become large enough to raise the machine off the ground and into the air. Unfortunately, they all suffered from the same circular argument—the more powerful the engine, the more it weighs; the heavier the machine is, the faster it must move to produce enough lift to get off the ground; the faster the machine must move, the more powerful (and hence heavier) the engine must be—which is where we entered this circular argument. A way out of this quandary is to develop engines with more power without an increase in engine weight, or more precisely, to design engines with large horsepower-to-weight ratios. The thrust-to-weight ratio, T/W, for the entire aircraft, is a critical parameter in airplane performance and design. In the nineteenth century, inventors of flying machines functioned mainly on the basis of intuition, with little quantitative analysis to guide them. They knew that, to accelerate the aircraft, thrust had to be greater than the drag; that is, T - D had to be a positive number. And the larger the thrust and the smaller the drag, the better things were. In essence, most of the nineteenth-century flying machine inventors were obsessed with brute force—given enough thrust (or horsepower) from the engine, the airplane could be wrestled into the air. The aviation historians call such people "chauffeurs." They were so busy trying to get the flying machine off the ground that they paid little attention to how the machine would be controlled once it got into the air; their idea was that somehow the machine could be chauffeured in the air much as a carriage driven on the ground. This philosophy led to failure in all such cases.

The antithesis of the chauffeur's philosophy was the "airman's" approach. In order to design a successful flying machine, it was necessary to first get up in the air and experience flight with a vehicle unencumbered by a power plant; that is, you should learn to fly before putting an engine on the aircraft. The person who introduced and pioneered the airman's philosophy was Otto Lilienthal, a German mechanical engineer, who designed and flew the first successful gliders in history. Lilienthal first carried out a long series of carefully organized aerodynamic experiments, covering a period of about twenty years, from which he clearly demonstrated the aerodynamic superiority of cambered (curved) airfoils in comparison to flat, straight surfaces. His experiments were extensive and meticulously carried out. They were published in 1890 in a book entitled "Der Vogelflug als Grundlage der Fliegekunst" ("Bird Flight as the Basis of Aviation"); this book was far and away the most important and definitive contribution to the budding science of aerodynamics to appear in the nineteenth century. It greatly influenced aeronautical design for the next fifteen years, and was the bible for the early work of the Wright brothers. Lilienthal's aerodynamic research led to a quantum jump in aerodynamics at the end of the nineteenth century.

The last, and perhaps the most dramatic, failure of the pre-Wright era was the attempt by Samuel P. Langley to build a flying machine for the U.S. government. Intensely interested in the physics and technology of powered flight, Langley began a series of aerodynamic experiments in 1887, using a whirling arm apparatus. At the time, he was the director of the Allegheny Observatory in Pittsburgh. Within a year he seized the opportunity to become the third Secretary of the Smithsonian Institution in Washington, D.C. Langley continued with his aeronautical experiments, including the building and flying of a number of elastic-powered models. The results of his whirling arm experiments were published in 1890 in his book *Experiments in Aerodynamic*. In 1896, Langley was successful in flying several small-scale, unmanned, powered aircraft, which he called aerodromes. These 14-ft-wingspan, steam-powered aerodromes were launched from the top of a small houseboat on the Potomac River, and they flew for about a minute, covering close to 1 mi over the river. These were the first steam-powered, heavier-than-air machines to successfully fly—a historic event in the history of aeronautics that is not always appreciated today.

This was to be the zenith of Langley's success. Spurred by the exigency of the Spanish-American War, Langley was given a $50,000 grant from the War

Department to construct and fly a full-scale, person-carrying aerodrome. He hired an assistant, Charles Manly, who had just graduated from the Sibley School of Mechanical Engineering at Cornell University. Together, they set out to build the required flying machine. The advent of the gasoline-powered internal-combustion engine in Europe convinced them that the aerodrome should be powered by a gasoline-fueled reciprocating engine turning a propeller.

By 1901 Manly had assembled a radically designed five-cylinder radial engine. It weighed 200 lb, produced a phenomenal 52.4 hp, and was the best airplane power plant designed until the beginning of World War I. The full-scale aerodrome, equipped with his engine, was ready in 1903. Manly attempted two flights; both resulted in the aerodrome's falling into the water moments after its launch by a catapult mounted on top of a new houseboat on the Potomac River.

Langley's aerodrome and the fate that befell it are an excellent study in the basic aspects of airplane design. Despite excellent propulsion and adequate aerodynamics, it was the poor structural design that resulted in failure of the whole system.

ERA OF STRUT-AND-WIRE BIPLANES

The 1903 Wright Flyer ushered in the era of successful strut-and-wire biplanes, an era that covers the period from 1903 to 1930. There is no doubt in this author's mind that Orville and Wilbur Wright were the first true aeronautical engineers in history. With the 1903 Wright Flyer, they had gotten it all right—the propulsion, aerodynamic, structural, and control aspects were carefully calculated and accounted for during its design. The Wright brothers were the first to fully understand the airplane as a whole and complete system, in which the individual components had to work in a complementary fashion so that the integrated system would perform as desired.

Let us dwell for a moment on the Wright Flyer as an airplane design. The Wright Flyer possessed all the elements of a successful flying machine. Propulsion was achieved by a four-cylinder in-line engine designed and built by Orville Wright with the help of their newly hired mechanic in the bicycle shop, Charlie Taylor. It produced close to 12 hp and weighed 140 lb, barely on the margin of what the Wrights had calculated as the minimum necessary to get the flyer into the air. This engine drove two propellers via a bicycle-like chain loop. The propellers themselves were a masterpiece of aerodynamic design. Wilbur Wright was the first person in history to recognize the fundamental principle that a propeller is nothing more than a twisted wing oriented in a direction such that the aerodynamic force produced by the propeller was predominately in the thrust direction. Wilbur conceived the first viable propeller theory in the history of aeronautical engineering; vestiges of Wilbur's analyses carry through today in the standard "blade element" propeller theory. The Wrights had built a wind tunnel, and during the fall and winter of 1901 to 1902, they carried out tests on hundreds of different airfoil and wing shapes. Wilbur incorporated these experimental data in his propeller analyses; the result was a propeller with an efficiency that was close to 70 percent (propeller efficiency is the power output from the propeller compared to the power input to the propeller from the engine shaft). This represented a dramatic improvement of propeller performance over contemporary practice. For example, Langley reported a propeller efficiency of only 52 percent for his aerodromes. Today, a modern, variable-pitch propeller can achieve efficiencies as high as 85 to 90 percent. In 1903, the Wrights' propeller efficiency of 70 percent was simply phenomenal. It was one of the lesser-known but most compelling reasons for the success of the Wright Flyer. With their marginal engine linked to their highly efficient propellers, the Wrights had the propulsion aspect of airplane design well in hand.

The aerodynamic features of the Wright Flyer were predominately a result of their wind tunnel tests of numerous wing and airfoil shapes. The Wrights were well aware that the major measure of aerodynamic efficiency is the lift-to-drag ratio L/D. They knew that the lift of an aircraft must equal its weight in order to sustain the machine in the air, and that almost any configuration could produce enough lift if the angle of attack was sufficiently large. But the secret of "good aerodynamics" is to produce this lift with as small a drag as possible, that is, to design an aircraft with as large an L/D value as possible. To accomplish this, the Wrights did three things:

1. They chose an airfoil shape that, based on the collective data from their wind tunnel tests, would give a high L/D. The airfoil used on the

Wright Flyer was a thin, cambered shape, with a camber ratio (ratio of maximum camber to chord length) of 1/20, with the maximum camber near the quarter-chord location. (In contrast, Lilienthal favored airfoils that were circular arcs, i.e., with maximum camber at midchord.) It is interesting that the precise airfoil shape used for the Wright Flyer was never tested by the Wright brothers in their wind tunnel. By 1903, they had so much confidence in their understanding of airfoil and wing properties that, in spite of their characteristic conservative philosophy, they felt it unnecessary to test that specific shape.

2. They chose an aspect ratio of 6 for the wings. (Aspect ratio is defined as the square of the wing span divided by the wing area; for a rectangular wing, the aspect ratio is simply the ratio of the span to the chord length.) They had experimented with gliders at Kitty Hawk in the summers of 1900 and 1901, and they were quite disappointed in their aerodynamic performance. The wing aspect ratio of these early gliders was 3. However, their wind tunnel tests clearly indicated that higher-aspect-ratio wings produced higher values of L/D. (This was not a new discovery; the advantage of high-aspect-ratio wings had been first theorized by Francis Wenham in 1866. Langley's whirling arm data, published in 1890, proved conclusively that better performance was obtained with higher-aspect-ratio wings. Based on their own wind tunnel results, the Wrights immediately adopted an aspect ratio of 6 for their 1902 glider, and the following year for the 1903 flyer. At the time, the Wrights had no way of knowing about the existence of induced drag; this aerodynamic phenomenon was not understood until the work of Ludwig Prandtl in Germany fifteen years later. The Wrights did not know that, by increasing the aspect ratio from 3 to 6, they reduced the induced drag by a factor of 2. They only knew from their empirical results that the L/D ratio of the 6-aspect-ratio wing was much improved over their previous wing designs.
3. The Wrights were very conscious of the importance of parasite drag, which in their day was called head resistance. They used empirical formulas obtained from Octave Chanute to estimate the head resistance for their machines. (Octave Chanute was a well-known civil and railroad engineer who had become very interested in aeronautics. In 1893 he published an important survey of past aeronautical work from around the world in a book entitled "Progress in Flying Machines." It has become a classic; you can still buy reprinted copies today. From 1900, Octave Chanute was a close friend and confidant of the Wright brothers, giving them much encouragement during their intensive inventive work in 1900 to 1903.) The Wrights choice of lying prone while flying their machines, rather than sitting up, or even dangling underneath as Lilienthal had done, was a matter of decreasing head resistance. In early 1903, they even tested a series of wooden struts in an airstream in order to find the cross-sectional shape that gave minimum drag. Unfortunately, they did not appreciate the inordinately high drag produced by the supporting wires between the two wings.

The Wrights never quoted a value of L/D for their 1903 Wright Flyer. Modern wind tunnel tests of models of the Wright Flyer carried out in 1982 and 1983 as reported by Culick and Jex at Cal Tech indicate a maximum L/D of 6. This value is totally consistent with values of $(L/D)_{max}$ measured by Gustave Eiffel in 1910 in his large wind tunnel in Paris for models of a variety of aircraft of that time. It has been estimated that the Fokker E-111, an early World War I aircraft had an $(L/D)_{max}$ of 6.4. In 1903 the Wrights had achieved a value of $(L/D)_{max}$ with their flyer that was as high as that for aircraft designed 10 years later.

The control features of the Wright Flyer are also one of the basic reasons for its success. The Wright brothers were the first to recognize the importance of flight control around all three axes of the aircraft. Pitch control, obtained by a deflection of all or part of the horizontal tail (or the forward canard such as the Wright Flyer), and yaw control, obtained by deflection of the vertical rudder, were features recognized by investigators before the Wrights; for example, Langley's aerodrome had pitch and yaw controls. However, no one except the Wrights appreciated the value of roll control. Their novel idea of differentially warping the wing tips to control the rolling motion of the airplane, and to jointly control roll and yaw for coordinated turns, was one of their most important contributions to aeronautical engineering. Indeed, when Wilbur Wright finally carried out the first public demonstrations of their flying machines in LeMans, France, in August 1908, the two technical features of the Wright machines most appreciated and immediately copied by European aviators were their roll control and their efficient propeller design.

Finally, the structural features of the Wright Flyer were patterned partly after the work of Octave Chanute and partly after their own experience in designing bicycles. Chanute, inspired by the gliding flights of Lilienthal, carried out tests of gliders of his own design beginning in 1896. The most important technical feature of Chanute's gliders was the sturdy and lightweight Pratt-truss method of rigging a biplane structure. The Wright brothers adopted the Pratt-truss system for the Wright Flyer directly from Chanute's work. Other construction details of the Wright Flyer took advantage of the Wrights' experience in designing and building sturdy but lightweight bicycles. When it was finished, engine included, the empty weight of the Wright Flyer was 605 lb. With a 150-lb person on board, the empty weight-gross weight ratio was 0.8. By comparison, the empty weight of the Fokker E-111 designed 10 years later was 878 lb, and the empty weight-gross weight ratio was 0.65, not greatly different from that of the Wright Flyer. Considering that 10 years of progress in aircraft structural design had been made between the 1903 flyer and the Fokker E-111, the structural design of the 1903 Wright Flyer certainly seems technically advanced for its time. And the fact that the flyer was structurally sound was certainly well demonstrated on December 17, 1903.

In summary, the Wright brothers had gotten it right. All the components of their system worked properly and harmoniously—propulsion, aerodynamics, control, and structures. There were no fatal weak links. The reason for this was the natural inventiveness and engineering abilities of Orville and Wilbur Wright. The design of the Wright Flyer is a classic first study in good aeronautical engineering. There can be no doubt that the Wright brothers were the first true aeronautical engineers.

The Wright Flyer ushered in the era of strut-and-wire biplanes, and it basically set the pattern for subsequent airplane design during this era. The famous World War I fighter airplanes—such as the French Nieuport 17 and the SPAD XIII, the German Fokker D. VII, and the British Sopwith Camel—were in many respects "souped-up" Wright flyers.

First, the wing warping method of roll control used by the Wrights was quickly supplanted by ailerons in most other aircraft. (The idea of flaplike surfaces at the trailing edges of airplane wings can be traced to two Englishmen: M. P. W. Boulton, who patented a concept for lateral control by ailerons in 1868; and Richard Harte, who also filed for a similar patent in 1870). Ailerons in the form of triangular "winglets" that projected beyond the usual wingtips were used in 1908 by Glenn Curtiss on his June Bug airplane; flying the June Bug, Curtiss won the Scientific American Prize on July 4, 1908, for the first public flight of 1,000 m or longer. By 1909, Curtiss had designed an improved airplane, the Gold Bug, with rectangular ailerons located midway between the upper and lower wings. Finally, in 1909 the Frenchman Henri Farman designed a biplane named the Henri Farman III, which included a flaplike aileron at the trailing edge of all four wingtips; this was the true ancestor of the conventional modern-day aileron. Farman's design was soon adopted by most designers, and wing warping quickly became passé. Only the Wright brothers clung to their old concept; a Wright airplane did not incorporate ailerons until 1915, six years after Farman's development.

Second, the open framework of the fuselage, such as seen in the Wright Flyer, was in later designs enclosed by fabric. The first airplane to have this feature was a Nieuport monoplane built in 1910. This was an attempt at "streamlining" the airplane, although at that time the concept of streamlining was only an intuitive process rather than the result of real technical knowledge and understanding about drag reduction.

Third, the demands for improved airplane performance during World War I gave a rebirth to the idea of "brute force" in airplane design. In relation to the thrust minus drag expression T – D, designers of World War I fighter airplanes, in their quest for faster speeds and higher rates of climb, increased the thrust rather than decreasing the drag. The focus was on more powerful engines. The SPAD XIII, one of the best and most famous aircraft from World War I, had a Hispano-Suiza engine that produced 220 hp—the most powerful engine used on a fighter aircraft at that time. Because of this raw power, the SPAD XIII had a maximum velocity of 134 mph which made it one of the fastest airplanes during the war. The SPAD XIII typifies the strut-and-wire biplane: the struts and wires produced large amounts of drag, although this was not fully understood by most airplane designers at that time. In fact, in the March 1924 issue of the Journal of the Royal Aeronautical Society, the noted British aeronautical engineer Sir Leonard Bairstow was prompted to say, "Our war experience showed that, whilst we went forward as regard to horsepower, we went backwards with regard to aerodynamic efficiency." Aircraft design during World War I was an intuitive "seat-of-the-pants" process. Some designs were almost literally marked off in chalk on the concrete floor of a factory, and the completed machines rolled out the front door two weeks later.

ERA OF THE MATURE, PROPELLER-DRIVEN AIRPLANE

The period from 1930 to 1950 can be classified as the era of the mature, propeller-driven airplane. During this time, airplane design matured, new technical features were incorporated, and the speed, altitude, efficiency, and safety of aircraft increased markedly. The 1930s are considered by many aviation historians as the "golden age of aviation" (indeed, there is currently a gallery at the National Air and Space Museum with this title). Similarly, the 1930s might be considered as a golden age for aeronautical engineering—a period when many improved design features, some gestating since the early 1920s, finally became accepted and incorporated on "standard" aircraft of the age.

The maturity of the propeller-driven airplane is due to nine major technical advances, all of which came to fruition during the 1930s.

Cantilevered-Wing Monoplane

Hugo Junkers in Germany during World War I and Anthony Fokker in Holland during the 1920s pioneered the use single-wing aircraft (monoplanes). This was made possible by the introduction of thick airfoils, which among other advantages allowed room for a large cantilevered spar that structurally supported the wing internally. This eliminated the need for the biplane box structure with its external supporting struts and wires. Consequently, the drag of monoplanes was less than that of comparable strut-andwire biplanes.

The All-Metal Airplane

The vast majority of airplanes before 1930 were constructed from wood and fabric, with some having a steel tube frame mechanism for the fuselage, over which fabric was stretched. Although Hugo Junkers designed and built the first all-metal airplane in 1915, this design feature was not adopted by others for many years. The case for the all-metal airplane was strengthened when the famous Notre Dame football coach Knute Rockne was killed on March 31, 1931, in the crash of a Fokker tri-motor transport. This shook the public's faith in the tri-motor design, and essentially led to its demise in the United States. Such concern was misdirected. Later investigation showed that the wooden wing spar (the entire wing of the Fokker tri-motor was made from wood) had rotted, and the crash was due to this structural failure. What better case could be made for all-metal construction?

Air-Cooled Engines and the NACA Cowling

Propeller-driven airplanes have two types of reciprocating engines—liquid cooled, or air-cooled engines. Since the early days of flight, liquid-cooled engines had the advantage of being longer and thinner, allowing them to be enclosed in relatively streamlined housings with less frontal drag. However, such engines were more vulnerable to damage during combat—a bullet through any part of the liquid cooling system would usually spell failure of the engine. Also, liquid-cooled engines were heavy due to all the machinery and cooling jackets that were associated with the liquid cooling mechanism. In contrast, air-cooled engines, where the cylinder heads are directly exposed to, and cooled by, the airstream over the airplane, are lighter. They require fewer moving parts, and therefore tend to be more reliable. The development of the powerful and reliable Pratt and Whitney Wasp series and the Curtiss-Wright Cycline series of air-cooled radial engines during the late 1920s and the 1930s resulted in the widespread adoption of these engines. But with the cylinders exposed directly to the airstream, the drag created by these was inordinately large.

This set the stage for a major technical development during this era, namely the National Advisory Committee for Aeronautics (NACA) cowling for radial piston engines. Such engines have their pistons arranged in a circular fashion about the crankshaft, and the cylinders themselves are cooled by airflow over the outer finned surfaces. Until 1927, these cylinders were usually directly exposed to the main airstream of the airplane, causing inordinately high

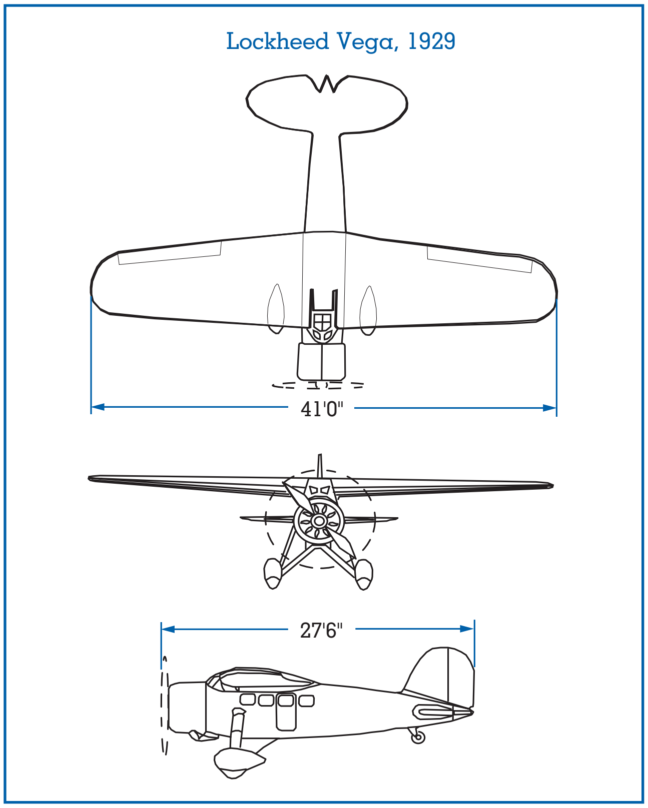


Figure 1.

drag. Engineers recognized this problem, but early efforts to enclose the engines inside an aerodynamically streamlined shroud (a cowling) interfered with the cooling airflow, and the engines overheated. One of the earliest aeronautical engineers to deal with this problem was Colonel Virginius E. Clark (for whom the famous Clark-Y airfoil is named). Clark designed a primitive cowling in 1922 for the Dayton-Wright XPSI airplane; it was marginal at best, and Clark had no proper aerodynamic explanation as to why a cowling worked. The first notable progress was made by H. L. Townend at the National Physical Laboratory in England. In 1927, Townend designed a ring of relative short length that wrapped around the outside of the cylinders. This resulted in a noticeable decrease in drag, and at least it did not interfere with engine cooling. Engine designers who were concerned with the adverse effect of a full cowling on engine cooling were more ready to accept a ring.

The greatest breakthrough in engine cowlings was due to the National Advisory Committee for Aeronautics in the United States. Beginning in 1927, at the insistence of a group of U.S. aircraft manufacturers, the NACA Langley Memorial Laboratory at Hampton, Virginia, undertook a systematic series of wind tunnel tests with the objective of understanding the aerodynamics of engine cowlings and designing an effective shape for such cowlings. Under the direction of Fred E. Weick at Langley Laboratory, this work quickly resulted in success. Drag reduction larger than that with a Townend ring was obtained by the NACA cowling. In 1928, Weick published a report comparing the drag on a fuselage-engine combination with and without a cowling. Compared with the uncowled fuselage, a full cowling reduced the drag by a stunning 60 percent. By proper aerodynamic design of the cowling, the airflow between the engine and the inside of the cowling resulted in enhanced cooling of the engine. Hence, the NACA cowling was achieving the best of both worlds. One of the first airplanes to use the NACA cowling was the Lockheed Vega, shown in Figure 1. Early versions of the Vega without a cowling had a top speed of 135 mph; after the NACA cowling was added to later versions, the top speed increased to 155 mph. The Lockheed Vega went on to become one of the most successful airplanes of the 1930s. The Vega 5, equipped with the NACA cowling and a more powerful engine, had a top speed of 185 mph. It was used extensively in passenger and corporate service. In addition, Amelia Earhart and Wiley Post became two of the most famous aviators of the 1930s—both flying Lockheed Vegas. Not only is the Vega a classic example of the new era of mature propeller-driven airplanes, but its aesthetic beauty supported the popular adage "If an airplane looks beautiful, it will also fly beautifully."

Variable-Pitch and Constant-Speed Propellers

Before the 1930s, a weak link in all propeller-driven aircraft was the propeller itself. For a propeller of fixed orientation, the twist of the propeller is designed so that each airfoil section is at its optimum angle of attack to the relative airflow, usually that angle of attack that corresponds to the maximum lift-to-drag ratio of the airfoil. The relative airflow seen by each airfoil section is the vector sum of the forward motion of the airplane and the rotational motion of the propeller. When the forward velocity of the airplane is changed, the angle of attack of each airfoil section changes relative to the local flow direction. Thus a fixed-pitch propeller is operating at maximum efficiency only at its design speed; for all other speeds of the airplane, the propeller efficiency decreases.

The solution to this problem was to vary the pitch of the propeller during the flight so as to operate at near-optimum conditions over the flight range of the airplane—a mechanical task easier said than done. The aerodynamic advantage of varying the propeller pitch during flight was appreciated as long ago as World War I, and H. Hele-Shaw and T. E. Beacham patented such a device in England in 1924. The first practical and reliable mechanical device for varying propeller pitch was designed by Frank Caldwell of Hamilton Standard in the United States. The first production order for Caldwell's design was placed by Boeing in 1933 for use on the Boeing 247 transport. The 247 was originally designed in 1932 with fixed-pitch propellers. When it started flying in early 1933, Boeing found that the airplane had inadequate takeoff performance from some of the airports high in the Rocky Mountains. By equipping the 247 with variable-pitch propellers, this problem was solved. The new propellers increased its rate of climb by 22 percent and its cruising velocity by over 5 percent. Later in the 1930s, the variable-pitch propeller, which was controlled by the pilot, developed into the constant-speed propeller, where the pitch was automatically controlled so as to maintain constant rpm over the flight range of the airplane. Because the power output of the reciprocating engine varies with rotational speed, by having a propeller in which the pitch is continuously and automatically varied to constant engine speed, the net power output of the engine-propeller combination can be maintained at an optimum value.

High-Octane Aviation Fuel

Another important advance in the area of propulsion was the development of high-octane aviation fuel, although it was eclipsed by the more visibly obvious breakthroughs in the 1930s such as the NACA cowling, retractable landing gear, and the variable-pitch propeller. Engine "pinging," an audible local detonation in the engine cylinder caused by premature ignition, had been observed as long ago as 1911. An additive to the gasoline, tetraethyl lead, was found by C. F. Kettering of General Motors Delco to reduce this engine knocking. In turn, General Motors and Standard Oil formed a new company, Ethyl Gasoline Corporation, to produce "ethyl" gasoline with a lead additive. Later, the hydrocarbon compound of octane was also found to be effective in preventing engine knocking. In 1930, the Army Air Corps adopted 87-octane gasoline as its standard fuel; in 1935, this standard was increased to 100 octane. The introduction of 100-octane fuel allowed much higher compression ratios inside the cylinder, and hence more power for the engine. For example, the introduction of 100-octane fuel, as well as other technological improvements, allowed Curtiss-Wright Aeronautical Corporation to increase the power of its R-1820 Cycline engine from 500 to 1,200 hp in the 1930s.

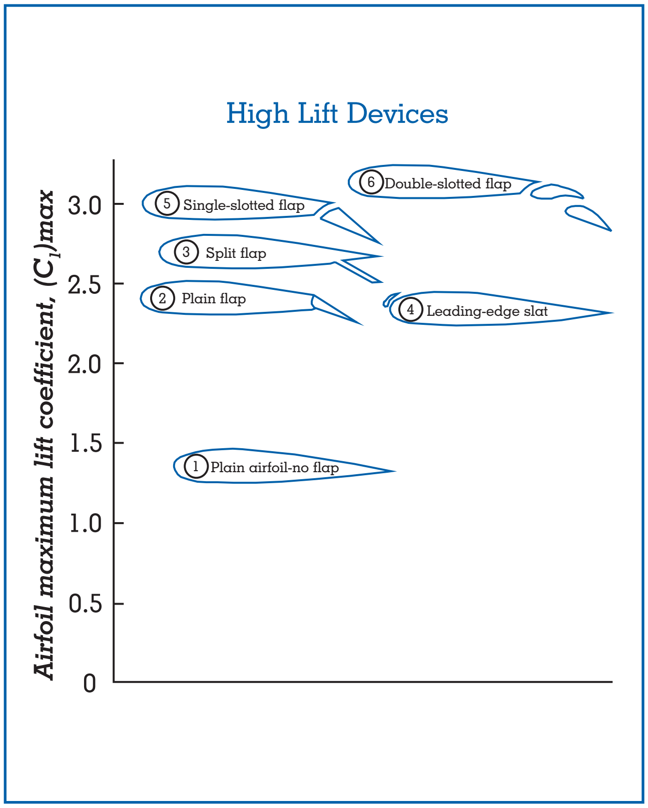


Figure 2.

High-Lift Devices

When a new airplane is designed, the choice of wing area is usually dictated by speed at takeoff or landing (or alternatively by the desired takeoff or landing distances along a runway). The wing area must be large enough to provide sufficient lift at takeoff or landing; this criterion dictates the ratio of airplane weight to wing area, that is, the wing loading W/S—one of the most important parameters in airplane performance and design. After the airplane has taken off and accelerated to a much higher cruising speed, the higher-velocity airflow over the wing creates a larger pressure difference between the upper and lower wing surfaces, and therefore the lift required to sustain the weight of the airplane can be created with a smaller wing area. From this point of view, the extra wing required for takeoff and landing is extra baggage at cruising conditions, resulting in higher structural weight and increased skin friction drag. The design of airplanes in the era of strut-and-wire biplanes constantly suffered from this compromise. A partial solution surfaced in the late 1920s and 1930s, namely, the development of high-lift devices such as flaps, slats, and slots. Figure 2 illustrates some of the standard high-lift devices employed on aircraft since the 1920s, along with a scale of lift coefficient indicating the relative increase in lift provided by each device. By employing such high-lift devices, sufficient lift can be obtained at takeoff and landing with wings of smaller area, allowing airplane designers the advantage of high wing loadings at cruise. High-lift devices were one of the important technical developments during the era of the mature propeller-driven airplane.

Pressurized Aircraft

Another technical development of the late 1930s is the advent of the pressurized airplane. Along with the decrease in atmospheric pressure with increasing altitude, there is the concurrent decrease in the volume of oxygen necessary for human breathing. The useful cruising altitude for airplanes was limited to about 18,000 ft or lower. Above this altitude for any reasonable length of time, a human being would soon lose consciousness due to lack of oxygen. The initial solution to the problem of sustained high-altitude flight was the pressure suit and the auxiliary oxygen supply breathed through an oxygen mask. The first pilot to use a pressure suit was Wiley Post. Looking like a deep-sea diver, Post set an altitude record of 55,000 ft in his Lockheed Vega in December 1934. This was not a practical solution for the average passenger on board an airliner. The answer was to pressurize the entire passenger cabin of the airplane, so as to provide a shirtsleeve environment for the flight crew and passengers. The first airplane to incorporate this feature was a specially modified and structurally strengthened Lockheed IOE Electra for the Army Air Corps in 1937. Designated the XC-35, this airplane had a service ceiling of 32,000 ft. It was the forerunner of all the modern pressurized airliners of today.

Superchargers for Engines

Along with pressurization for the occupants, high-altitude aircraft needed "pressurization" for the engine. Engine power is nearly proportional to the atmospheric density; without assistance, engine power dropped too low at high altitudes, and this was the major mechanical obstacle to high-altitude flight. Assistance came in the form of the supercharger, a mechanical pump that compressed the incoming air before it went into the engine manifold. Supercharger development was a high priority during the 1930s and 1940s; it was a major development program within NACA. All high-performance military aircraft during World War II were equipped with superchargers as a matter of necessity.

Streamlining

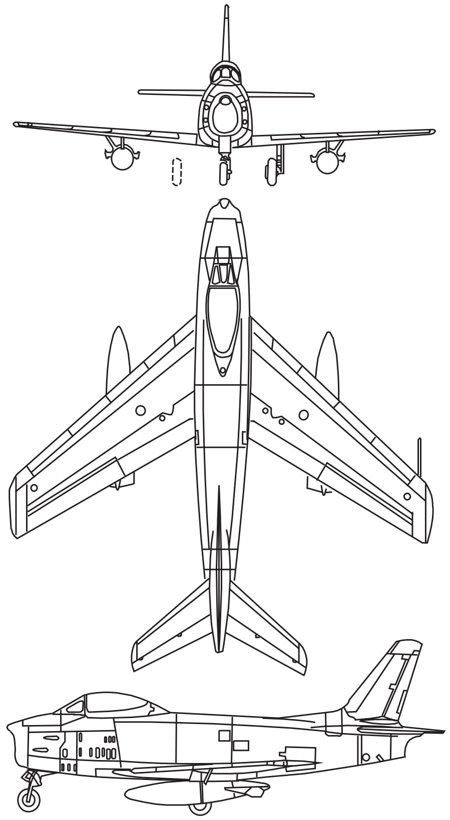
One of the most important developments in the era of the mature propeller-driven airplane was the appreciation of the need for streamlining the airframe. The rather box-like shape of the World War I vintage SPAD was characteristic of airplanes of that day. There was little if any attempt to shape the airplane into a streamlined configuration. The Douglas DC-3, however, was designed and began airline service in the mid-1930s. Here is streamlining personified. By comparison, the zero-lift-drag coefficient for the SPAD is 0.04, whereas that for the DC-3 is about 0.025, a considerable improvement. Part of the concept of streamlining was to retract the landing gear flush with the external airframe.

Summary

The Douglas DC-3 epitomizes the mature, propeller-driven aircraft of the 1930s. Here you see a cantilever wing monoplane powered by radial engines enclosed in NACA cowlings, and equipped with variable-pitch propellers. It is an all-metal airplane with retractable landing gear, and it uses flaps for high lift during takeoff and landing. For these reasons, the 1930s can indeed be called the golden age of aeronautical engineering.

ERA OF THE JET-PROPELLED AIRPLANE

The jet engine was invented independently by two people: Frank Whittle in England and Dr. Hans von Ohain in Germany. In 1928, as a student at the Royal Air Force technical college at Cranwell, Frank Whittle wrote a senior thesis entitled "Future



North American F-86H

Figure 3.

Developments in Aircraft Design" in which he expounded on the virtues of jet propulsion. It aroused little interest. Although Whittle patented his design for a gas-turbine aircraft engine in 1930, it was not until five years later that he formed, with the help of friends, a small company to work on jet engine development. Named Power Jets Ltd., this company was able to successfully bench-test a jet engine on April 12, 1937—the first jet engine in the world to successfully operate in a practical fashion. It was not the first to fly. Quite independently, and completely without the knowledge of Whittle's work, Dr. Hans von Ohain in Germany developed a similar gas-turbine engine. Working under the private support of the famous airplane designer Ernst Heinkel, von Ohain started his work in 1936. On August 27, 1939, a specially designed Heinkel airplane, the He 178, powered by von Ohain's jet engine, successfully flew; it was the first gas turbine-powered, jet-propelled airplane in history to fly. It was strictly an experimental airplane, but von Ohain's engine of 838 lb of thrust pushed the He 178 to a maximum speed of 360 mph. It was not until almost two years later that a British jet flew. On May 15, 1941, the specially designed Gloster E.28/39 airplane took off from Cranwell, powered by a Whittle jet engine. It was the first to fly with a Whittle engine. With these first flights in Germany and Britain, the jet age had begun.

The era of jet-propelled aircraft is characterized by a number of design features unique to airplanes intended to fly near, at, or beyond the speed of sound. One of the most pivotal of these design features was the advent of the swept wing. For a subsonic airplane, sweeping the wing increases the airplane's critical Mach number, allowing it to fly closer to the speed of sound before encountering the large drag rise caused by the generation of shock waves somewhere on the surface of the wing. For a

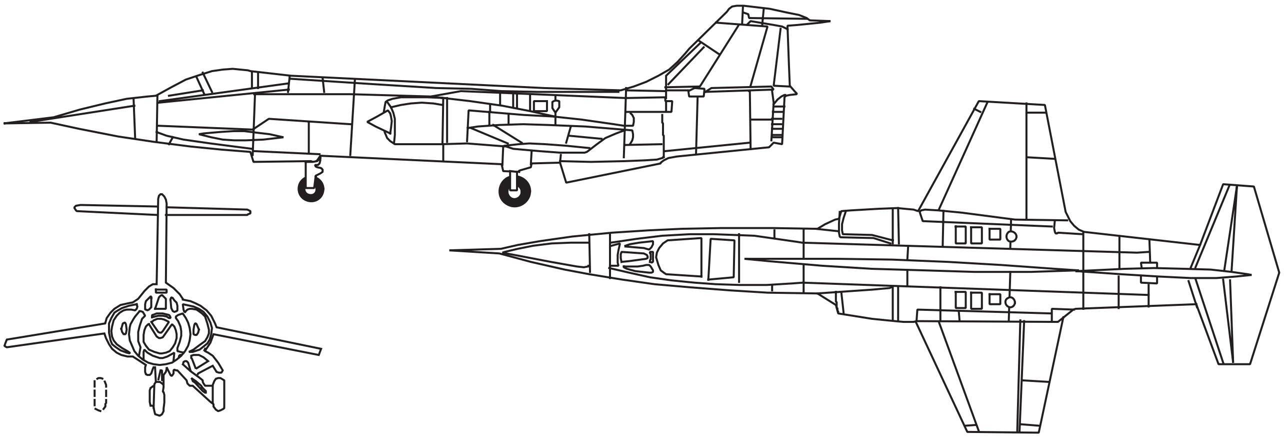
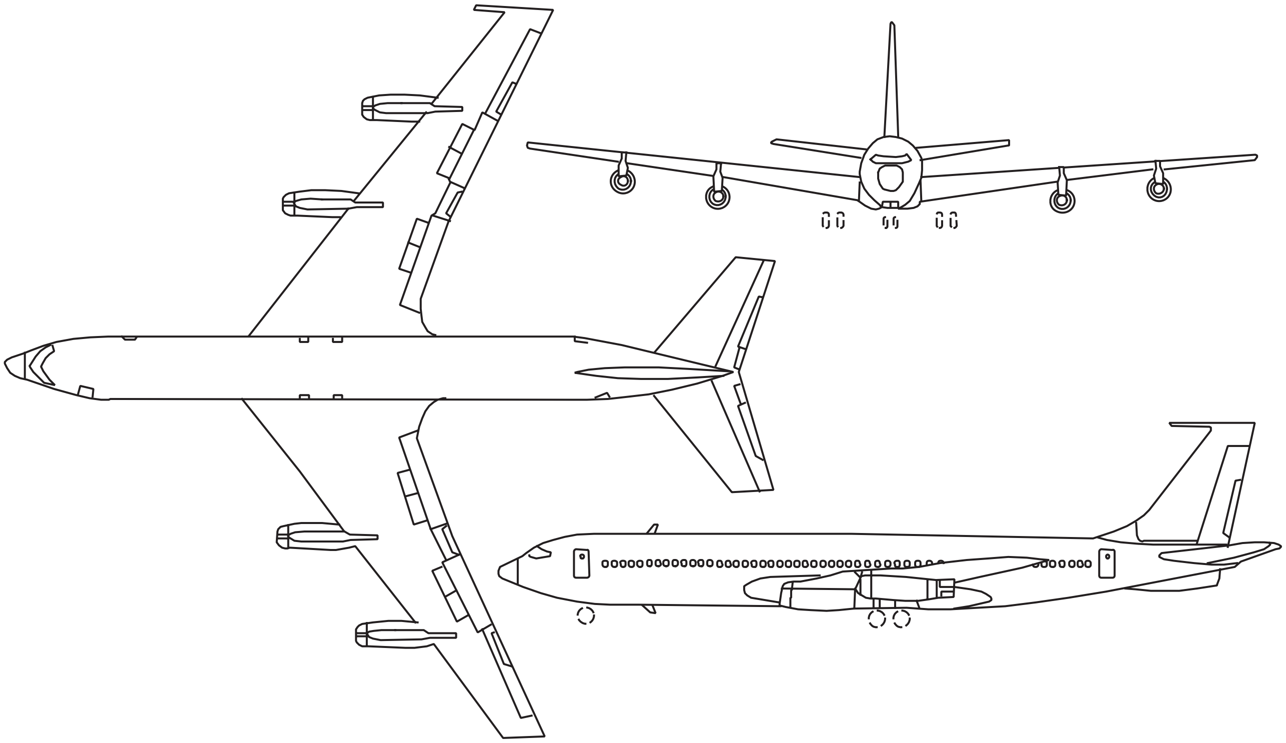

Lockheed F-104

Figure 4.

supersonic airplane, the wing sweep is designed such that the wing leading edge is inside the Mach cone from the nose of the fuselage; if this is the case, the component of airflow velocity perpendicular to the leading edge is subsonic (called a subsonic leading edge), and the resulting wave drag is not as severe as it would be if the wing were to lie outside the Mach cone. In the latter case, called the supersonic leading edge, the component of flow velocity perpendicular to the leading edge is supersonic, with an attendant shock wave generated at the leading edge. In either case, high subsonic or supersonic, an airplane with a swept wing will be able to fly faster than one with a straight wing, everything else being equal.

The concept of the swept wing for high-speed aircraft was first introduced in a public forum in 1935. At the fifth Volta Conference, convened on September 30, 1935, in Rome, Italy, the German aerodynamicist Adolf Busernann gave a paper in which he discussed the technical reasons why swept wings would have less drag at high speeds than conventional straight wings. Although several Americans were present, such as Eastmann Jacobs from NACA and Theodore von Karman from Cal Tech, Busernann's idea went virtually unnoticed; it was not carried back to the United States with any sense of importance. Not so in Germany. One year after Busemann's presentation at the Volta Conference, the swept-wing concept was classified by the German Luftwaffe as a military secret. The Germans went on to produce a large bulk of swept-wing research, including extensive wind tunnel testing. They even designed a few prototype swept-wing jet aircraft. Many of these data were confiscated by the United States after World War II, and made available to U.S. aircraft companies and government laboratories. Meanwhile, quite independently of this German research, Robert T. Jones, a NACA aerodynamicist,



Boeing 707

Figure 5.

had worked out the elements of swept-wing theory toward the end of the war. Although not reinforced by definitive wind tunnel tests in the United States at that time, Jones's work served as a second source of information concerning the viability of swept wings.

In 1945, aeronautical engineers at North American Aircraft began the design of the XP-86 jet fighter; it had a straight wing. The XP-86 design was quickly changed to a swept-wing configuration when the German data, as well as some of the German engineers, became available after the war. The prototype XP-86 flew on October 1, 1947, and the first production P-86A flew with a 35° swept wing on May 18, 1948. Later designated the F-86, the swept-wing fighter had a top speed of 679 mph, essentially Mach 0.9—a stunning speed for that day. Shown in Figure 3, the North American F-86 Sabre was the world's first successful operational swept-wing aircraft.

By the time the F-86 was in operation, the sound barrier had already been broken. On October 14, 1947, Charles (Chuck) Yeager became the first human being to fly faster than the speed of sound in the Bell X-1 rocket-powered airplane. In February 1954, the first fighter airplane capable of sustained flight at Mach 2, the Lockheed F-104 Starfighter, made its first appearance. The F-104, Figure 4, exhibited the best qualities of good supersonic aerodynamics—a sharp, pointed nose, slender fuselage, and extremely thin and sharp wings. The airfoil section on the F-104 is less than 4 percent thick (maximum thickness compared to the chord length). The wing leading edge is so sharp that protective measures must be taken by maintenance people working around the aircraft. The purpose of these features is to reduce the strength of shock waves at the nose and trailing edges, thus reducing supersonic wave drag. The F-104 also had a straight wing with a very low aspect ratio rather than a swept wing. This exhibits an

Aircraft	*Payload Capacity*	*Top Speed*	*Cruising Speed*	*Fuel Consumption* (1000 mile flight)	*4D Ratio*
737	128 passengers	M=0.89	M=0.83	1600 gal.	17
737 Stretch	188 passengers	0.89	0.83	1713	17
747-400	413 passengers	0.87	0.82	6584	21
SST Concorde	126 passengers	2.2	2.0	6400	8
Turboprop DHC-8	37 passengers	325 MPH	305 MPH	985	15
Fighter Plane F-15		M=2.5		750	6
Military Cargo Plane C-17	120,000 lbs.	0.80	0.76	5310	14

Table 1.
Comparison of Some Aircraft Figures
Payload capacity for civil transports is given in number of passengers; for the C-17 it is in pounds. Payloads and cruising speeds were unavailable for the fighter planes.

alternative to supersonic airplane designers; the wave drag on straight wings of low aspect ratio is comparable to that on swept wings with high aspect ratios. Of course, this low-aspect-ratio wing gives poor aerodynamic performance at subsonic speeds, but the F-104 was point-designed for maximum performance at Mach 2. With the F-104, supersonic flight became an almost everyday affair, not just the domain of research aircraft.

The delta wing concept was another innovation to come out of Germany during the 1930s and 1940s. In 1930, Alexander Lippisch designed a glider with a delta configuration; the leading edges were swept back by 20°. The idea had nothing to do with high-speed flight at that time; the delta configuration had some stability and control advantages associated with its favorable center-of-gravity location. When Busemann introduced his swept-wing ideas in 1935, Lippisch and his colleagues knew they had a potential high-speed wing in their delta configuration. Lippisch continued his research on delta wings during the war, using small models in German supersonic wind tunnels. By the end of the war, he was starting to design a delta wing ramjet-powered fighter. Along with the German swept-wing data, this delta wing technology was transferred to the United States after the war; it served as the basis for an extended wind tunnel test program on delta wings at NACA Langley Memorial Laboratory.

The first practical delta wing aircraft was the Convair F-102. The design of this aircraft is an interesting story in its own right—a story of the interplay between design and research, and between industry and NACA. The F-102 was designed as a supersonic airplane. Much to the embarrassment and frustration of the Convair engineers, the prototype F-102 being tested at Edwards Air Force Base during October 1953 and then again in January 1954 exhibited poor performance and was unable to go supersonic. At the same time, Richard Whitcomb at NACA Langley was conducting wind tunnel tests on his "area rule" concept, which called for the cross-sectional area of the fuselage to be reduced in the vicinity of the wing. By so doing, the transonic drag was substantially reduced. The Convair engineers quickly adopted this concept on a new prototype of the F-102, and it went supersonic on its second flight. Convair went on to produce 975 F-102s; the practical delta wing airplane was a finally a reality.

The area rule was one of the most important technical developments during the era of jet-propelled airplanes. Today, almost all transonic and supersonic aircraft incorporate some degree of area rule. For his work on the area rule, Whitcomb received the Collier Trophy, the highest award given in the field of aeronautics.

One of the most tragic stories in the annals of airplane design occurred in the early 1950s. Keying on England's early lead in jet propulsion, de Havilland Aircraft Company designed and flew the first commercial jet transport, the de Havilland Comet. Powered by four de Havilland Ghost jet engines, the

Comet carried 36 passengers for 2,000 mi at a speed of 460 mph, cruising at relatively high altitudes near or above 30,000 ft. The passenger cabin was pressurized; indeed, the Comet was the first pressurized airplane to fly for extended periods at such high altitudes. Inasmuch as good airplane design is an evolutionary process based on preceding aircraft, the de Havilland designers had little precedent on which to base the structural design of the pressurized fuselage. The Comet entered commercial service with BOAC (a forerunner of British Airways) in 1952. In 1954, three Comets disintegrated in flight, and the airplane was quickly withdrawn from service. The problem was later found to be structural failure of the fuselage while pressurized. De Havilland used countersunk rivets in the construction of the Comet; reaming the holes for the rivets produced sharp edges. After a number of pressurization cycles, cracks in the fuselage began to propagate from these sharp edges, leading eventually to catastrophic failure. At the time, de Havilland had a massive lead over all other aircraft companies in the design of commercial jet aircraft. While it was in service, the Comet was very popular with the flying public, and it was a moneymaker for BOAC. Had these failures not occurred, de Havilland and England might have become the world's supplier of commercial jet aircraft rather than Boeing and the United States.

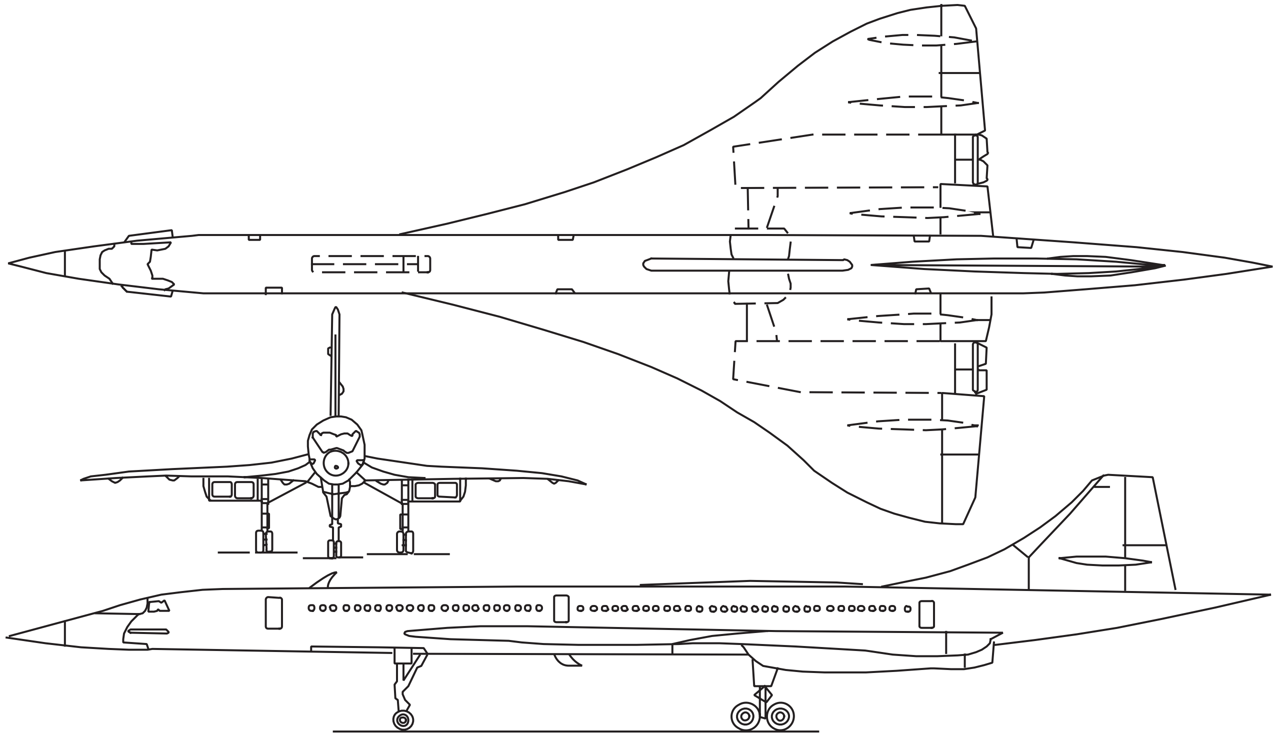
In 1952, the same year as the ill-fated de Havilland Comet went into service, the directors of Boeing Company made a bold and risky decision to privately finance and build a commercial jet prototype. Designated the model 367-80, or simply called the Dash 80 by the Boeing people, the prototype first flew on July 15, 1954. It was a bold design that carried over to the commercial field Boeing's experience in building swept-wing jet bombers for the Air Force (the B-47 and later the B-52). Later renamed the Boeing 707, the first production series of aircraft were bought by Pan American Airlines and went into service in 1958. The Boeing 707 (Figure 5), with its swept wings and podded engines mounted on pylons below the wings, set the standard design pattern for all future large commercial jets. The design of the 707 was evolutionary because it stemmed from the earlier experience at Boeing with jet bombers. But it was almost revolutionary in the commercial field, because no other airliner had ever (not even the Comet) looked like that. Boeing's risky gamble paid off, and it transformed a predominately military aircraft company into the world's leader in the design and manufacture of commercial jet transports.

Boeing made another bold move on April 15, 1966, when the decision was made to "go for the big one." Boeing had lost the Air Force's C-5 competition to Lockheed; the C-5 at the time was the largest transport airplane in the world. Taking their losing design a few steps further, Boeing engineers conceived of the 747, the first wide-body commercial jet transport. Bill Allen, president of Boeing at that time, and Juan Trippe, president of Pan American Airlines, shared the belief that the large, wide-body airplane offered economic advantages for the future airline passenger market, and they both jointly made the decision to pursue the project. It was an even bolder decision than that concerning the 707.

The gamble paid off. The Boeing 747 first flew in February 1969, and it entered service for the first time in January 1970 on Pan American's New York—London route. Boeing is still producing 747s.

What about commercial transportation at supersonic speeds? In the 1960s this question was addressed in Russia, the United States, England, and France. The Tupolev Design Bureau in Russia rushed a supersonic transport design into production and service. The Tu-144 supersonic transport first flew on December 31, 1968. More than a dozen of these aircraft were built, but none entered extended service, presumably due to unspecified problems. One Tu-144 was destroyed in a dramatic accident at the 1973 Paris Air Show. In the United States, the government orchestrated a design competition for a supersonic transport; the Boeing 2707 was the winner in December 1966. The design turned into a nightmare for Boeing. For two years, a variable-sweep wing supersonic transport (SST) configuration was pursued, and then the design was junked. Starting all over again in 1969, the design was caught up in an upward spiral of increased weight and development costs. When the predictions for final development costs hit about $5 billion, Congress stepped in and refused to appropriate any more funds. In May 1971, the SST development program in the United States was terminated. Only in England and France was the SST concept carried to fruition.

The first, and so far only, supersonic commercial transport to see long-term regular service was the Anglo-French Concorde (Figure 6). In 1960 both



Anglo-French Aerospatial/BAC Concorde supersonic transport

Figure 6.

the British and French independently initiated design studies for a supersonic transport. It quickly became apparent that the technical complexities and financial costs were beyond the abilities of either country to shoulder alone. On November 29, 1962, England and France signed a formal agreement aimed at the design and construction of a supersonic transport. The product of this agreement was the Aerospatiale-British Aerospace Corporation's Concorde. Designed to cruise at Mach 2.2 and carry 125 passengers, the Concorde first flew on March 2, 1969. It first exceeded Mach 1 on October 1, 1969, and Mach 2 on November 4, 1970. Originally, orders for 74 Concordes were anticipated. When the airlines were expected to place orders in 1973, the world was deep in the energy crisis. The skyrocketing costs of aviation jet fuel wiped out any hope of an economic return from flying the Concorde, and no orders were placed. Only the national airlines of France and Britain, Air France and British Airways, went ahead, each signing up for seven aircraft after considerable pressure from their respective governments. After a long development program, the Concorde went into service on January 21, 1976. In the final analysis, the Concorde was a technical, if not financial, success. It was in regular service from 1976 until the entire fleet was grounded in the wake of the first concorde crash in August 2000. It represents an almost revolutionary airplane design in that no such aircraft existed before it. The Concorde designers were not operating in a vacuum. Examining Figure 6, we see a supersonic configuration which incorporates good supersonic aerodynamics—a sharp-nosed slender fuselage and a cranked delta wing with a thin airfoil. The Concorde designers had at least fifteen years of military airplane design experience with such features to draw upon. Today, we know that any future second-generation SST will have to be economical in service and environmentally acceptable. The design of such a vehicle is one of the great challenges in aeronautics.

In summary, the types of aircraft in use today cut across the flight spectrum from low-speed, propeller-driven airplanes with reciprocating engines, moderate speed turboprop airplanes (propeller driven by gas turbine engines), and high-speed jet-propelled airplanes. For low-speed flight, below about 250 mph, the reciprocating engine/propeller combination has by far the best propulsive efficiency. For moderate speeds (250–400 mph) the turboprop is superior. This is why most high-performance commuter aircraft are powered by turboprops. For high speeds (at least above 500 mph) the jet engine is the only logical powerplant choice; a propeller rapidly looses efficiency at higher flight speeds. In short, a reciprocating engine/propeller combination is a high efficiency, but comparably low thrust powerplant, and a jet engine is a lower efficiency but higher thrust powerplant. The turboprop is a middle-ground compromise between thrust and efficiency. The wide variety of airplanes in use today draw on the technology developed in both the era of the mature, propeller-driven airplane and the era of the jet-propelled airplane. In the future, airplane design will continue to be influenced by the desire to fly faster and higher, but moderated by the need for environmental effectiveness, economic viability, and energy efficiency.

John D. Anderson

See also: Aerodynamics; Air Travel; Efficiency of Energy Use; Propellers.

BIBLIOGRAPHY

Anderson, J. D., Jr. (1997). *A History of Aerodynamics, and Its Impact on Flying Machines,* New York: Cambridge Universtiy Press.

Anderson, J. D., Jr. (1999). *Aircraft Performance and Design.* Boston: McGraw-Hill.

Anderson, J. D., Jr. (2000). *Introduction to Flight,* 4th ed. Boston: McGraw-Hill.

Loftin, L. K. (1985) *Quest for Performance: The Evolution of Modern* Aircraft, Washington, DC: NASA.

Greenwood, J. T. ed. (1989). *Milestones of Aviation.* Washington, DC: National Air and Space Museum, Smithsonian Institution.

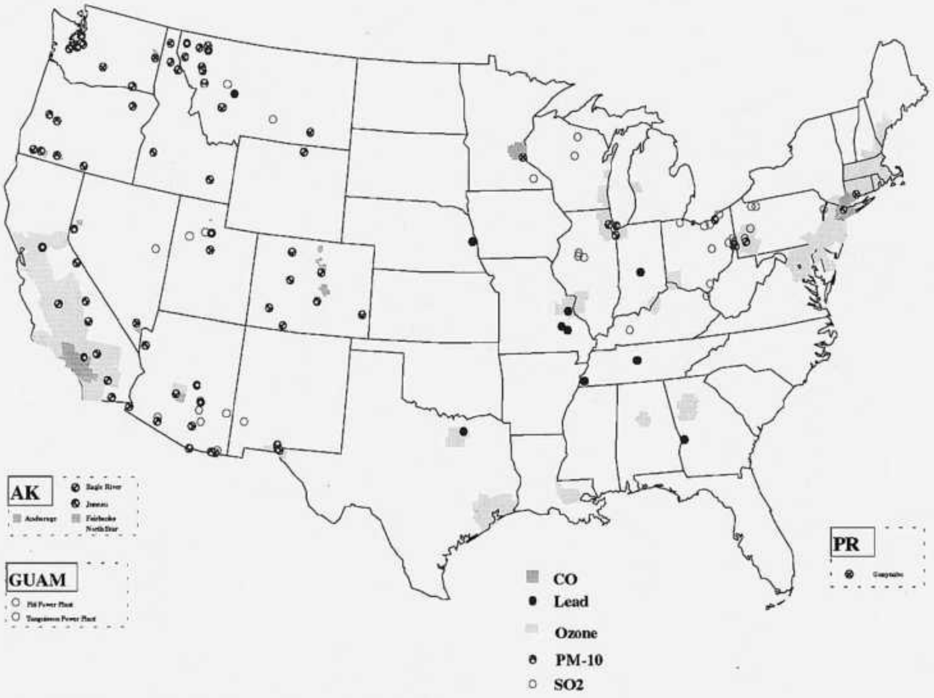
AIR POLLUTION

The pernicious effects of air pollution were first documented long ago. As early as 61 C.E., Seneca, a Roman philosopher and noted essayist, wrote, "As soon as I had gotten out of the heavy air of Rome, and from the stink of the chimneys thereof, which being stirred, poured forth whatever pestilential vapors and soot they had enclosed in them, I felt an alteration to my disposition" (Miller and Miller, 1993). With technology being as simple as it was in Rome, however, there was not much that could be done about the problem.

Aerial view of Los Angeles obscured by smog. (Phototake)

Thirteen hundred years later, controls on the use of coal in London were passed, marking the recorded start of air pollution control. But such controls were not enough to prevent the buildup of pollutants as by-products of industrialization; air pollution was common to all industrialized nations by 1925. Air pollution is still a significant problem in urban centers worldwide. In the United States, pollutant emissions and air pollution concentrations have been falling, for the most part, since the 1970s.

The major constituents of unpolluted air (not including water) at ground level are nitrogen (78.08%) and oxygen (20.95%). The next most abundant constituents are argon (at 0.934%) and carbon dioxide (about 0.034%), followed by the other noble gases:



Location of nonattainment areas for criteria pollutants, September 1998.

Note: Incomplete data, not classified, and Section 185(a) areas are not shown.
* Ozone nonattainment areas on map are based on the pre-existing ozone standard.
Nonattainment designations based on the revised 8-hour ozone standard will not be designated until 2000.
** PM-10 nonattainment areas on map are based on the pre-existing PM-10 standards.
Nonattainment designations based on the revised PM-10 standards have not yet been made.
Source: U.S. EPA, *National Air Quality and Emissions Trends Report, 1997.*

Source: U.S. Environmental Protection Agency.

neon (0.002%), helium (0.0005%), krypton (0.0001%), and xenon (0.000009%). A variety of chemicals known as trace gases (some quite toxic) also are found in unpolluted air, but at very low concentrations.

When certain substances in the air rise to the level at which they can harm plant or animal life, damage property, or simply degrade visibility in an area more than it would be in the absence of human action, those substances are considered to be pollutants. Such pollutants enter the atmosphere via both natural processes and human actions.

The pollutants most strongly damaging to human, animal, and sometimes plant health include ozone, fine particulate matter, lead, nitrogen oxides (NO_x), sulfur oxides (SO_x), and carbon monoxide. Many other chemicals found in polluted air can cause lesser health impacts (such as eye irritation). VOC compounds comprise the bulk of such chemicals. Formaldehyde is one commonly mentioned pollutant of this sort, as is PAN (peroxyacyl nitrate). Such chemicals are common components of photochemical smog, a term that refers to the complex mixture of chemicals that forms when certain airborne chemicals given off by plant and human activity react with sunlight to produce a brownish mixture of thousands of different chemical species.

Finally, there are also pollutants that do not cause direct health impacts but that may have the potential to cause harm indirectly, through their actions on the overall ecology, or as they function as precursor chemicals that lead to the production of other harmful chemicals. The major indirect-action pollutants include volatile organic carbon (VOC) compounds that act as precursors to more harmful species; chemicals called halocarbons; and chemicals called greenhouse gases.

ENERGY USE AND POLLUTION

Energy use is the predominant source of global air pollution, though other human actions produce significant amounts of pollution as well. The burning of biomass for agriculture, such as the burning of crop stubble is one such nonenergy source, as is the use of controlled-burn management of forest fires. Natural events such as forest fires as well as volcanic outgassing and eruption also contribute to air pollution. Finally, even living things produce considerable quantities of emissions considered "pollutants." Animals and insects produce a large share of the world's methane emissions, while plants emit significant quantities of volatile organic carbon compounds—enough, in some areas to produce elevated ozone levels with no other pollutant emissions at all.

Ozone

Ozone, or O_3, a chemical consisting of three atoms of oxygen, is a colorless, odorless gas produced by a variety of chemical reactions involving hydrocarbons and nitrogen oxides in the presence of sunlight. Often there is confusion about ozone's effects because it is found at two different levels of the atmosphere, where it has two very different effects. At ground level, ozone is known to be a respiratory irritant, implicated in causing decreased lung function, respiratory problems, acute lung inflammation, and impairment of the lungs' defense mechanisms. Outdoor workers, elderly people with pre-existing lung diseases, and active children who spend significant amounts of time in areas with elevated ozone levels are thought to be particularly at risk.

Low-altitude ozone forms when sunlight reacts with "precursor" chemicals of both human and nonhuman origin. These precursors include volatile organic carbon compounds, carbon monoxide, and nitrogen oxides. Volatile organic carbon compounds are created both naturally (by plants) and through human activity such as fuel use, biomass burning, and other industrial activities such as painting and coating. Nitrogen oxides are generated by stationary sources of energy use such as power plants and factories, and mobile sources such as cars, trucks, motorcycles, bulldozers, and snowmobiles. Because of the diversity of ozone precursors, sources vary by region.

At high altitudes, ozone forms naturally from the interaction of high-energy solar radiation and normal diatomic oxygen, or O_2. High-altitude ozone is not considered a pollutant, but is actually beneficial to life on Earth, as it screens out solar radiation that can damage plants, or cause skin cancer and cataracts in animals. High-altitude ozone can be destroyed through the action of chemicals called halocarbons, which are commonly used in refrigeration and air conditioning. Ozone levels across the United States fell, on average, by four percent between 1989 and 1998.

VOC Source Breakdown	*Tg/yr*
Human emissions sources	98
Biomass burning	51
Continental biogenic sources	500
Oceans	30-300
Total	**750**

Table 1.
VOC Amounts by Source
Note: One Tg (teragram) is equivalent to 1 million metric tons. Continental biogenic sources includes animal, microbial, and foliage emissions
SOURCE: Finlayson-Pitts & Pitts (2000)

Volatile Organic Carbon Compounds

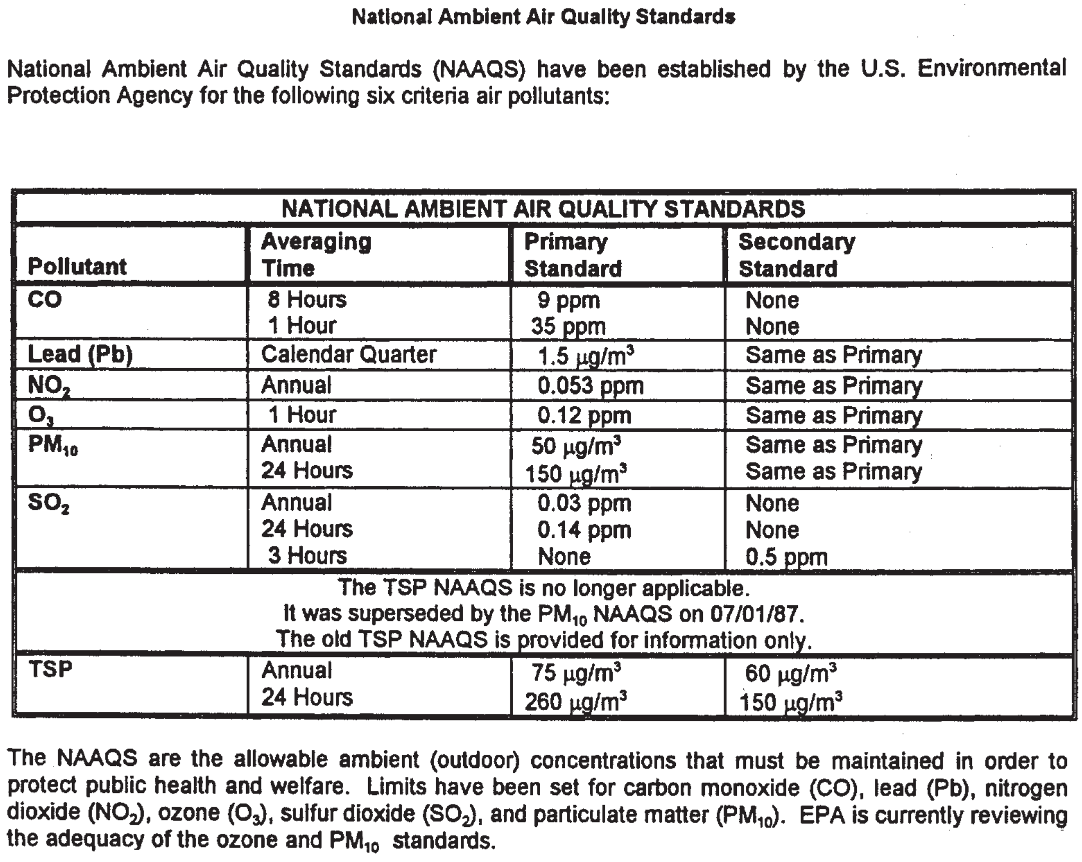
There are many different chemical species in the VOC compounds group, including such commonly known chemicals as formaldehyde and acetone. The common feature that VOCs share is that they are ring-shaped (organic) molecules consisting principally of carbon, hydrogen, oxygen, and nitrogen. VOCs are released into the environment as a result of human activity, but also because of natural biological processes in plants and animals. Table 1 shows the percentage of global contribution to VOC concentrations from all sources.

VOCs react in the presence of sunlight to produce photochemical smog, a mixture of organic chemicals that can irritate the eyes and other mucous membranes. VOCs also constitute a major precursor chemical leading to ozone production. VOC levels across the United States fell, on average, by 20.4 percent between 1989 and 1998.

Particulate Matter

Particulate matter, generated through a range of natural and manmade processes including combustion and physical abrasion, has been implicated in increased mortality for the elderly, as well as those members of the population with damaged respiratory systems. Studies also have linked particulate matter with aggravation of preexisting respiratory and cardiovascular disease, resulting in more frequent and/or serious attacks of asthma in the elderly or in children.

The particulate matter of most concern consists of



National Ambient Air Quality Standards

National Ambient Air Quality Standards (NAAQS) have been established by the U.S. Environmental Protection Agency for the following six criteria air pollutants:

NATIONAL AMBIENT AIR QUALITY STANDARDS

Pollutant	Averaging Time	Primary Standard	Secondary Standard
CO	8 Hours 1 Hour	9 ppm 35 ppm	None None
Lead (Pb)	Calendar Quarter	1.5 µg/m^3	Same as Primary
NO$_2$	Annual	0.053 ppm	Same as Primary
O$_3$	1 Hour	0.12 ppm	Same as Primary
PM$_{10}$	Annual 24 Hours	50 µg/m^3 150 µg/m^3	Same as Primary Same as Primary
SO$_2$	Annual 24 Hours 3 Hours	0.03 ppm 0.14 ppm None	None None 0.5 ppm
The TSP NAAQS is no longer applicable. It was superseded by the PM$_{10}$ NAAQS on 07/01/87. The old TSP NAAQS is provided for information only.			
TSP	Annual 24 Hours	75 µg/m^3 260 µg/m^3	60 µg/m^3 150 µg/m^3

The NAAQS are the allowable ambient (outdoor) concentrations that must be maintained in order to protect public health and welfare. Limits have been set for carbon monoxide (CO), lead (Pb), nitrogen dioxide (NO$_2$), ozone (O$_3$), sulfur dioxide (SO$_2$), and particulate matter (PM$_{10}$). EPA is currently reviewing the adequacy of the ozone and PM$_{10}$ standards.

Table 2.
National Ambient Air Quality Standards. (NAAQS).

particles that are most likely to be trapped in tiny air sacs of the lung (alveoli) after inhalation. Studies suggest that such particles are in the microscopic range, from less than 1 micrometer in diameter to about 2.5 microns in average diameter.

Although it has been shown that exposure to certain air pollutants can aggravate preexisting lung ailments such as asthma, no causal link has been identified between exposure to low-level air pollution and asthma. In fact, while asthma levels have been rising, air pollution levels have been declining, suggesting that there is probably a different cause for the increase in asthma rates. While various indoor air pollutants have been suggested as the cause, no definitive cause for the increase in asthma rates has yet been found.

Particulate matter is generated by stationary sources such as power plants and factories and by mobile sources such as cars, trucks, motorcycles, bulldozers, and snowmobiles. Particulate matter levels across the United States fell, on average, by 25 percent between 1989 and 1998.

Carbon Monoxide

Carbon monoxide, or CO, is a highly toxic chemical that chemically binds to hemoglobin, rendering it incapable of carrying oxygen to the tissues of the body. CO is produced by the incomplete combustion of fossil fuels. Carbon monoxide levels across the United States fell, on average, by 39 percent between 1989 and 1998.

Lead

Lead is an element used in many industrial processes and also has been used in fuels and coatings. Tetraethyl lead was added to gasoline to improve performance as a motor fuel, and elemental lead was extensively used in paints and coatings to improve coverage and durability until the 1970s, when phase-out efforts began to reduce lead emissions to the environment.

Long-term exposure to airborne lead was shown to lead to a variety of health problems, primarily neurological. Atmospheric lead concentrations fell dramatically through the 1970s and continue to fall: Atmospheric lead concentrations fell, on average, by 56 percent between 1989 and 1998. The amount of lead found in the bloodstream of children growing up in urban environments also fell dramatically.

Sulfur Oxides

Sulfur oxide emissions enter the atmosphere from a variety of sources, some of human origin, others of natural origin. The main sulfur oxide is sulfur dioxide, or SO_2.

High concentrations of SO_2 can produce temporary breathing difficulties in asthmatic children and in adults who are active outdoors. Sulfur dioxide also can directly damage plants and has been shown to decrease crop yields. In addition, sulfur oxides can be converted to sulfuric acid and lead to acid rain. Acid rain can harm ecosystems by increasing the acidity of soils as well as surface waters such as rivers, lakes, and streams. Sulfur dioxide levels fell, on average, by 39 percent between 1989 and 1998.

Nitrogen Dioxide

Nitrogen dioxide (NO_2) is a reddish-brown gas that is formed through the oxidation of nitrogen oxide (NO). The term "nitrogen oxides," or NO_x is used to encompass NO_2 as well as NO and the other oxides of nitrogen that lead to NO_2 production.

Nitrogen oxides are generated by both human and nonhuman action, but the major sources of NO_x are high-temperature combustion processes such as those occurring in power plants and automobile engines. Natural sources of NO_x include lightning, chemical processes that occur in soil, and the metabolic activities of plants.

Short-term exposure to elevated levels of NO_x have been shown to cause changes in the function of human lungs, while chronic exposures have been linked to increased susceptibility to lung infections and to lasting changes in lung structure and function. Nitrogen oxides also are of concern because of their roles as ozone precursors, and through their contribution to acid rain in the form of nitric acid.

Nitrogen dioxide concentrations have changed little since 1989, and alterations in measuring techniques make it difficult to accurately assess the trend. Data suggest that NO_2 concentrations may have increased by 2 percent between 1989 and 1998.

Human industry is one of the leading causes of air pollution. (Corbis Corporation)

AIR POLLUTION REGULATION IN THE UNITED STATES

Air pollution in the United States is regulated at federal, state, and local levels. Allowable concentrations of the major air pollutants are set by the U.S. Environmental Protection Agency (EPA) under the auspices of the Clean Air Act. States and localities implement pollution control plans in accordance with the provisions of the Clean Air Act in regions where air pollutant concentrations exceed the federal standards. Some states and localities have air pollution standards of their own, and in the past, such standards have occasionally been more stringent than those of the EPA.

The EPA sets two kinds of national ambient air quality standards. The primary standard is set at a level intended to protect human health with an adequate margin of safety. The secondary standard, usually less stringent, is set based on protecting the public welfare, which can include factors other than health impacts, such as reduced visibility, and damage to crops.

Pollutant	*Primary Standard (Health Related)*		*Secondary Standard (Welfare Related)*
	Type of Average	*Allowable Concentration*	
CO	8-hour	9 parts per million	No secondary standard
	1-hour	35 parts per million	No secondary standard
Pb	Maximum quarterly average	1.5 micrograms/cubic-meter of air	Same as primary standard
NO_2	Annual arithmetic mean	0.053 parts per million	Same as primary standard
O_3	Maximum daily 1-hr average	0.12 parts per million	Same as primary standard
	4th Maximum daily 8-hr average	0.08 parts per million	Same as primary standard
PM_{10}	Annual arithmetic mean	50 micrograms/cubic-meter of air	Same as primary standard
	24-hour	150 micrograms/cubic-meter of air	Same as primary standard
$PM_{2.5}$	Annual arithmetic mean	15 micrograms/cubic-meter of air	Same as primary standard
	24-hour	65 micrograms/cubic-meter of air	Same as primary standard
SO_2	Annual arithmetic mean	0.03 parts per million	3-hour/0.50 parts per million
	24-hour	0.14 parts per million	

Table 3.
National Ambient Air Quality Standards in Effect as of December 1999
SOURCE: U.S. EPA National Air Quality and Emissions Trends Report, 1998, Table 2-1, p 9.

Table 3 shows the current health-related national ambient air quality standards set by the EPA as of 1999. Regions that violate these standards may be classified as "nonattainment" areas by the EPA, and can face sanctions if they do not promulgate pollution control plans that are acceptable to the agency.

Environmental regulations to curtail air pollution have had a major impact on energy producers, manufacturers of energy-using products, and energy consumers. Before the Clean Air Act of 1970, coal-fired steam turbines were the least expensive means available for utilities to generate electricity. Coal still remains the least expensive fossil fuel, yet the high capital costs of installing the technology to comply with ever more stringent environmental regulations have resulted in many utilities reconsidering their options for new electric-power-generating facilties. The combined cycle natural gas turbine is being favored not only because of the much easier compliance with current air quality regulations but also because of the likelihood of stricter future regulations. A typical coal power plant burns more than 70 lb of coal each second, and considering that these plants number in the hundreds, coal-burning emissions are likely to be a major future target of legislators and regulators eager to curtail emissions further.

Transportation is another sector that is a major contributor to air pollution. To improve air quality, particularly in urban areas, regulations require the use of reformulated gasoline and alternative fuels to reduce emissions of nitrogen oxides and carbon monoxide. Reflecting the higher cost of refining, these regulations have added at least 5 to 10 cents to the price of gasoline at the pump.

Besides cleaner fuels, vehicle makers have developed many emission-reducing technologies—both in "cleaner combustion" and in catalytic converter technologies—to comply with ever stricter tailpipe emission standards. The U.S. EPA stringent standards proposed in 1999 for model year 2004 vehicles will result in new vehicles emitting less than 1 percent of the VOC and NO_x emissions of their 1960s counterparts.

Kenneth Green

See also: Acid Rain; Air Quality, Indoor; Atmosphere; Automobile Performance; Climatic Effects; Emission Control, Vehicle; Emission Control, Power Plant; Environmental Economics; Environmental Problems and Energy Use; Gasoline and Additives; Transportation, Evolution of Energy Use and; Turbines, Gas.

BIBLIOGRAPHY

Finlayson-Pitts, B. J., and Pitts, J. N., Jr. (2000). *Chemistry of the Upper and Lower Atmosphere: Theory, Experiments, and Applications.* San Diego: Academic Press.

Klaassen, C. D. (1996) *Casarett & Doull's Toxicology: The Basic Science of Poisons*, 5th ed. New York: McGraw-Hill.

Miller, F. W., and Miller, R. M. (1993). *Environmental Hazards: Air Pollution—A Reference Handbook.* Santa Barbara, CA: ABC-CLIO.

U. S. Environmental Protection Agency. (1998). "National Air Quality and Emissions Trends Report, 1998." <http://www.epa.org/>.

AIR QUALITY, INDOOR

While most of us are aware of the dangers posed by outdoor air pollution, awareness of airborne chemical and biological pollutants present indoors and its implications for human health is more limited. Some indoor air gases and pollutants such as radon, asbestos, carbon monoxide, biological contaminants, and volatile organic compounds (VOCs) pose a serious threat to our health and well-being. Over the past several decades, our exposure to indoor air pollutants is believed to have increased due to a variety of factors, including the construction of more tightly sealed buildings; reduced ventilation rates to save energy; use of synthetic building materials and furnishings; and use of chemically formulated personal care products, pesticides, and household cleaners. Since an average person spends increasing amount of time indoors, it is important to understand the health risks posed by prolonged exposure to indoor pollutants and the energy and comfort implications of different methods to control and mitigate these pollutants in order to ensure acceptable indoor air quality.

Acceptable indoor air quality (IAQ) is defined as "air in which there are no known contaminants at harmful concentrations as determined by cognizant authorities and with which a substantial majority (80%) of the people exposed do not express dissatisfaction" (ASHRAE, 1989). Some of these indoor air contaminants are particulates, vapors, and gases that may be generated by occupants and their activities, building materials, furniture, equipment and appliances present in indoor space, operations and maintenance activities, or brought in from outside. Examples of indoor pollutants are certain gases (radon, carbon monoxide, and carbon dioxide), volatile organic compounds or VOCs (environmental tobacco smoke, formaldehyde, solvents, and fragrances, etc.), bioaerosols (mold spores, pollen, bacteria, animal dander, etc.), and particles from buildings, furnishings and occupants (fiberglass, paper dust, lint from clothing, carpet fibers, etc.).

As the science of indoor air quality has matured, indoor air professionals have realized that many indoor air contaminants and the associated health effects are linked to specific types of buildings and their characteristics. For example, radon is primarily an indoor air concern in homes because of the ease with which it can be transported inside residential construction from the soil beneath. On the other hand, Sick Building Syndrome (SBS) primarily afflicts office building occupants who experience acute health and comfort effects that appear to be linked to time spent in a specific building.

It has been estimated that hundreds of billions of dollars per year is lost due to decreased workplace productivity and increased health costs that can be saved by maintaining good indoor air quality in commercial buildings. The financial benefits of improving IAQ can accrue from reducing costs for health care and sick leave, as well as the costs of performance decrements at work caused by illness or adverse health symptoms and of responding to occupant complaints and costs of IAQ investigations.

Indoor air quality problems have grown with the increased use of heating, ventilating, and air-conditioning (HVAC) systems in commercial and residential buildings. Greater use of HVAC systems have also resulted in the closer examination of the energy impacts of maintaining good indoor air quality. The trade-off between reducing heating and cooling loads of the HVAC system by recirculating as much indoor air as possible and providing an optimum amount of fresh outdoor air forms the underpinnings of many IAQ standards and guidelines in climate-controlled commercial buildings.

HVAC SYSTEM

A wide variety of HVAC systems are used in residential and commercial buildings to thermally condition and ventilate the occupied spaces. While HVAC systems can provide enhanced levels of thermal comfort in very hot or very cold weather conditions, thermodynamic processes required to condition outdoor air and deliver conditioned air to occupied spaces to maintain indoor comfort conditions are fairly energy intensive.

A typical HVAC system uses a combination of heating, cooling, humidification (adding moisture) and dehumidification (removing moisture) processes to thermally condition air. This conditioned air, which is a mixture of outdoor air and recirculated indoor air, is known as supply air. The supply airstream typically passes through filters, heat exchangers that add or remove heat from the supply airstream, a supply fan, air ducts, dampers that are used to regulate the rate of airflow, and finally diffusers located either in the ceiling or floor to the occupied space. The return air is drawn from the occupied spaces and flows back to the mechanical rooms either through return air ducts or through the plenum between suspended ceiling and the floor of the next-higher story. A portion of the return air is exhausted to the outdoors, and the remainder is mixed with the fresh outdoor air and resupplied to the space after filtering and thermal conditioning. In general, the supply air contains more recirculated air than fresh outdoor air to keep the energy cost of air conditioning down.

In an all-air system, the indoor temperature can be controlled either by a constant air volume (CAV) system, which varies the temperature of the air but keeps the volume constant, or by a variable air volume (VAV) system, which maintains a constant temperature and varies the volume of the air supplied to internal spaces.

To save energy, many HVAC systems employ a mechanism for regulating the flow of outdoor air called an economizer cycle. An economizer cycle takes advantage of milder outdoor conditions to increase the outside air intake and in the process reduces the cooling load on the system. Controlling the rate of flow of outdoor air appears simple, in theory, but often works poorly in practice. The small pressure drop required to control the flow rate of outdoor air is rarely controlled and monitored. Quite often, the damper system used to regulate the airflow is nonfunctional, disconnected from the damper actuators, or casually adjusted by building operators (Institute of Medicine, 2000).

Outdoor airflow concerns are some of the many reasons that make maintaining good indoor air quality in a controlled indoor environment a particularly challenging task. Maintaining a safe, comfortable indoor environment for workers and residents is challenging under the best of circumstances because apart from temperature control, HVAC systems are also responsible for moisture control and ventilation of buildings. Hot, humid climates, in particular, present some of the biggest challenges, and solutions that are tested and proven in temperate climates may actually worsen IAQ problems in hot, humid climates (Odom and DuBose, 1991).

INDOOR AIR POLLUTANTS: GENERATION, MITIGATION, AND EXPOSURE

To maintain acceptable indoor air quality, the concentration of pollutants known to degrade indoor air quality and affect human health must be controlled. If the origin of the contaminant is known, it is more effective to exercise source control over any mitigation strategy. If the origin of the contaminants is not known, building ventilation and air cleaning and filtration are the two most commonly used processes to dilute or remove all types of contaminants from the indoor air and maintain acceptable indoor environmental conditions.

Source Control

Source control is one of the most important methods to achieve healthy indoor air. Methods vary depending on the pollutants and can range from simple solutions (not using pressed wood furniture that use formaldehyde) to complex and costly solutions (identifying moisture infiltration through the building envelope). Source control may also require behavioral changes on the part of the affected population, a remedy that may be achieved either through environmental regulation or by raising public awareness. A case in point is the changing perception toward environmental tobacco smoke (ETS). Banning smoking in public spaces or discouraging smoking in homes and in front of children can reduce the risks of second hand smoke greatly. Other examples are selecting furniture that doesn't contain formaldehyde and using paints and carpets that don't emit chemicals that are known to degrade indoor air. If the air contaminants are being transported inside from outdoor sources, precautions should be taken to either plug the pathway to stop such transport (e.g., sealing the construction joints and cracks in the basement to reduce radon infiltration) or take steps to minimize the infiltration of contaminants (e.g., positioning the fresh air intake away from loading docks and parking spaces).

For radon, moisture, ETS, asbestos, lead-based paint, building materials and products that emit VOCs, pesticides, and household products, one can develop effective source control strategy as well. For example, assuring building envelope integrity at the time of construction of the building would help control and restrict the airflow into the building. The envelope (walls, roof, floor or slab system) should also control moisture infiltration by installing a continuous vapor barrier on the warm side of the insulation system. Operating the building at a slight positive pressure can also help in keeping the moisture outdoors. All these precautions can help control the growth of mold and mildew that thrive under moist conditions and can adversely affect the health of building occupants.

Ventilation

In addition to minimizing the emissions of pollutants from indoor sources, ventilation with outside air must be provided at an adequate rate to maintain acceptable IAQ. The ventilation rate—the rate of outside air supply—is usually defined per unit of floor area (liters per second per sq. meter), number of occupants (liters per person), or indoor air volume (air changes per hour). For indoor-generated particles, the effects of ventilation rate is highly dependent on particle size because the depositional losses of particles increases dramatically with particle size. The predicted change in pollutant concentrations with ventilation rate is greatest for an "ideal" gaseous pollutant that is not removed by deposition or sorption on surfaces (Institute of Medicine, 2000). Ventilation rate may have a very significant indirect impact on indoor concentrations of some pollutants because they affect indoor humidities, which in turn modify indoor pollutant sources.

Buildings are ventilated mechanically with the HVAC systems where it is a controlled process, as well as via air infiltration and through the openable windows and doors where it is largely an uncontrolled process. However, as discussed earlier, mechanical ventilation is one of the most energy-intensive methods of reducing indoor pollutant concentrations primarily because of the need to thermally condition air before it can be circulated inside the occupied spaces. It is estimated that the ventilation needs are responsible for consuming 10 percent of all the energy consumed in buildings in developed countries.

On average, buildings with air conditioning that have inadequate supply of fresh air are far more likely to suffer from poor indoor air quality than naturally ventilated buildings. On the other hand, one can find serious IAQ problems in homes and apartment buildings that are naturally ventilated as well.

There are two commonly used techniques to control odors and contaminants. Both depend on ventilation to achieve their goals. One of them relies on the concept of "ventilation effectiveness," which is defined as the ability of the ventilation system to distribute supply air and dilute internally generated pollutants by ensuring a consistent and appropriate flow of supply air that mixes effectively with room air. The second technique isolates odors and contaminants by maintaining proper pressure relationship between outdoors and indoors and between different indoor spaces. This is accomplished by adjusting the air quantities that are supplied to and removed from each room. In many large commercial buildings, particularly in warm humid climates, the design intent is to pressurize the building slightly with the mechanical ventilation system in order to prevent undesirable infiltration of unconditioned air, moisture, and outdoor air pollutants. On the other hand, smoking rooms, bathrooms, and laboratories are often depressurized so that pollutants generated within these rooms do not leak into the surrounding rooms.

Often, local dedicated exhaust ventilation is used in rooms with high pollutant or odor sources as it is more efficient in controlling indoor pollutant concentrations than general ventilation of the entire space (U.S. Environmental Protection Agency, 1991; U.S. Department of Energy, 1998). In practice, however, indoor-outdoor pressure differences are often poorly controlled, and many buildings are not pressurized (Persily and Norford, 1987). There is considerable uncertainty in predicting the rate of dilution of indoor contaminants in actual complex indoor environment, with rates of pollutant loss by deposition on indoor surfaces being one of the largest sources of uncertainty.

Air Cleaning and Filtration

Particle air cleaning is any process that is used intentionally to remove particles from the indoor air. Filtration and electronic air cleaning are the two most

Pollutants	*Major Sources*	*Health Effects*
Biological Contaminants		
1. **Infectious communicable bioaerosols** contain bacteria or virus within small droplet nuclei produced from the drying of larger liquid droplets and can transmit disease.	Human activity such as coughing and sneezing; wet or moist walls, ceilings, carpets, and furniture; poorly maintained humidifiers, dehumidifiers, and air conditioners; bedding; household pets.	Eye, nose, and throat irritation; dizziness; lethargy; fever. May act as asthma trigger; may transmit humidifier fever; influenza, common cold, tuberculosis and other infectious diseases.
2. **Infectious non-communicable bioaerosols** are airborne bacteria or fungi that can infect humans but that have a non-human source.	Cooling towers and other sources of standing water (e.g., humidifiers) are thought to be typical sources of Legionella in buildings.	The best known example is Legionella, a bacterium that causes Legionnaires Disease and Pontiac Fever.
3. **Non-infectious bioaerosols** include pollens, molds, bacteria, dust mite allergens, insect fragments, and animal dander.	The sources are outdoor air, indoor mold and bacteria growth, insects, and pets.	The health effects of non-infectious bioaerosols include allergy symptoms, asthma symptoms, and hypersensitivity pneumonitis.
Carbon Monoxide (CO) is a colorless and odorless gas that can prove fatal at high concentrations. High carbon monoxide concentration is more likely to occur in homes.	Unvented kerosene and gas space heaters; leaking chimneys and furnaces; back-drafting from furnaces, gas water heaters, woodstoves, and fireplaces; automobile exhaust from attached garages; environmental tobacco smoke.	At low concentrations, fatigue in healthy people and chest pain in people with heart disease. At higher concentrations, impaired vision and coordination; headaches; dizziness; nausea. Fatal at very high concentrations.
Carbon dioxide (CO_2) is one of the gaseous human bioeffluents in exhaled air. Indoor concentrations are usually in the range of 500 ppm to a few thousand ppm.	Humans are normally the main indoor source of carbon dioxide. Unvented or imperfectly vented combustion appliances can also increase indoor CO_2 concentrations.	At typical indoor concentrations, CO_2 is not thought to be a direct cause of adverse health effects; however, CO_2 is an easily-measured surrogate for other occupant-generated pollutants.
Environmental tobacco smoke (ETS) is the diluted mixture of pollutants caused by smoking of tobacco and emitted into the indoor air by a smoker. Constituents of ETS include submicron-size particles composed of a large number of chemicals, plus a large number of gaseous pollutants.	Cigarette, pipe, and cigar smoking.	Eye, nose, and throat irritation; headaches; lung cancer; may contribute to heart disease; build-up of fluid in the middle ear; increased severity and frequency of asthma episodes; decreased lung function. ETS is also a source of odor and irritation complaints.
Fibers in indoor air include those of asbestos, and man-made mineral fibers such as fiberglass, and glass wool.	Deteriorating, damaged, or disturbed insulation, fireproofing, acoustical materials, and floor tiles	No immediate symptoms, but long-term risk of chest and abdominal cancers and lung diseases.

Table 1.
Reference Guide to Major Indoor Air Pollutants

Pollutants	*Major Sources*	*Health Effects*
Moisture is not a pollutant but it has a strong influence on indoor air quality. In some situations, high relative humidity may contribute to growth of fungi and bacteria that can adversely affect health.	Water vapor is generated indoors due to human metabolism, cooking and taking showers, unvented combustion activities and by humidifiers; water and moisture leaks through roof or building envelope; improperly maintained HVAC equipment	Condensation of water on cool indoor surfaces (e.g., windows) may damage materials and promote the growth of microorganisms. The presence of humidifiers in commercial building HVAC systems has been associated with an increase in various respiratory health symptoms.
Particles are present in outdoor air and are also generated indoors from a large number of sources including tobacco smoking and other combustion processes. Particle size, generally expressed in microns (10-6 m) is important because it influences the location where particles deposit in the respiratory system (U.S. Environmental Protection Agency 1995), the efficiency of particle removal by air filters, and the rate of particle removal from indoor air by deposition on surfaces.	Some particles and fibers may be generated by indoor equipment (e.g. copy machines and printers). Mechanical abrasion and air motion may cause particle release from indoor materials. Particles are also produced by people, e.g., skin flakes are shed and droplet nuclei are generated from sneezing and coughing. Some particles may contain toxic chemicals.	Increased morbidity and mortality is associated with increases in outdoor particle concentrations (U.S. Environmental Protection Agency 1995). Of particular concern are the particles smaller than 2.5 micrometers in diameter, which are more likely to deposit deep inside the lungs (U.S. Environmental Protection Agency 1995). Some particles, biological in origin, may cause allergic or inflammatory reactions or be a source of infectious disease.
Volatile Organic Compounds (VOCs): VOCs are a class of gaseous pollutants containing carbon. The indoor air typically contains dozens of VOCs at concentrations that are measurable.	VOCs are emitted indoors by building materials (e.g., paints, pressed wood products, adhesives, etc.), equipment (photocopying machines, printers, etc.), cleaning products, stored fuels and automotive products, hobby supplies, and combustion activities (cooking, unvented space heating, tobacco smoking, indoor vehicle use).	Eye, nose, and throat irritation; headaches, nausea. Some VOCs are suspected or known carcinogens or causes of adverse reproductive effects. Some VOCs also have unpleasant odors or are irritants. VOCs are thought to be a cause of non-specific health symptoms.
Radon (Rn) is a naturally occurring radioactive gas. Radon enters buildings from underlying soil and rocks as soil gas is drawn into buildings.	The primary source of radon in most buildings is the surrounding soil and rock, well water, earth-based building materials.	No immediate symptoms but estimated to contribute to between 7,000 and 30,000 lung cancer deaths each year. Smokers are at higher risk of developing radon-induced lung cancer.

Table 1 (cont.).
Reference Guide to Major Indoor Air Pollutants
SOURCE: Excerpted from U.S. Environmental Protection Agency (1995) and Indoor Environmental Quality Appendix to International Performance Measurement & Verification Protocol (U.S. Department of Energy, 1998)

common examples. Typically, portable air cleaning devices are used in rooms, while in typical commercial HVAC systems the filter is placed upstream of many of the HVAC components in the path of supply airstream to filter particles. Two facts that are applicable to both air cleaners and filters are

- Efficiency of any air cleaner or filter is a function of the particle size present in the indoor air and the velocity and volume of air flowing through the device.
- Pressure drop is a concern wherever filters and, to a lesser extent, air cleaners are employed in the path of normal forced ventilation system.

The technologies for removing particles include mechanical filters; electrostatic precipitators, which charge particles and then collect them onto a surface with the opposite charge; and ion generators, which charge particles and thereby facilitate their deposition. Among mechanical filters, high efficiency particulate air (HEPA) filters are highly efficient in removing particles of a wide range of sizes. However, there is little evidence of either direct health benefits or reduced concentration of pollutants resulting from air cleaning or filtration applications (American Thoracic Society, 1997). New stricter standards that also allow filter selection based on offending contaminants and their particle sizes found in buildings are more likely to show direct health benefits.

Table 1 describes some of the common indoor air pollutants found in buildings, their sources, and their adverse health effects on human beings.

INDOOR AIR QUALITY AND ENERGY

Energy conservation measures instituted following the energy crisis of 1974 resulted in the elimination of openable windows and in the recycling of as much air as possible to avoid heating or cooling outside air. The amount of outdoor air considered adequate for proper ventilation has varied substantially over time. The current guideline, widely followed in the United States, was issued by ASHRAE in 1989. To achieve good IAQ in all-air systems, large but finite amount of fresh air needs to be brought in, heated or cooled depending on the climate and season, and distributed to various parts of the building. The energy implications are obviously huge because the temperature and humidity of the supply air stream must be maintained within a very narrow range to satisfy the thermal comfort requirements of the building's occupants. Furthermore, temperature and humidity are among the many factors that affect indoor contaminant levels. Quality design, installation, and testing and balancing with pressure relationship checks are critically important for the proper operation of all types of HVAC systems and for maintaining good IAQ (U.S. Environmental Protection Agency, 1991; ASHRAE, 1989).

Energy professionals and equipment manufacturers are more cognizant of the indoor air problems and are coming out with new products and strategies that reduce energy use without degrading indoor air quality. According to the U.S. Department of Energy (1998), some of these energy efficient technologies can be used to improve the existing IAQ inside buildings:

Using outdoor air economizer for free cooling—An air "economizer" brings in outside air for air conditioning a commercial building. This strategy is particularly effective in mild weather where the temperature and humidity content of outside air is suitable for increasing the rate of outside air supply above the minimum setpoint. Generally, IAQ will improve due to the increase in average ventilation rate. Care must be exercised in using the economizer cycle in regions where outdoor air is of suspect quality. Also, in very humid regions, one must employ enthalpy-based control systems to take advantages of free cooling with the economizer cycle without encountering mold and mildew problems.

Heat recovery from exhaust air—If a heat recovery system allows an increase in the rate of outside air supply, IAQ will usually be improved. Proper precautions must be taken to ensure that moisture and contaminants from the exhaust air stream are not transferred to the incoming air stream. An innovative way of recovering heat and reducing the dehumidification cost is to use the waste heat to recharge the desiccant wheels that are then used to remove moisture from the supply air. In this method, the energy savings have to be substantial to offset the high cost of the desiccant wheels.

Nighttime pre-cooling using outdoor air—Nighttime ventilation may result in decreased indoor concentrations of indoor-generated pollutants when occupants arrive at work. Once again, proper precautions must be taken to ensure that outdoor air with the right level of moisture content is used for this purpose, otherwise condensation on heating, ventilation, and air conditioning equipment or building components may result, increasing the risk of growth of microorganisms.

Using radiant heating/cooling systems—Because of the higher thermal capacity of water compared to air, water is a better heat transfer medium. In hydronic climate conditioning systems, heat exchangers transmit heat from water to indoor environment, or vice-versa. These heat exchangers can either be convectors or radiators depending on the primary heat transfer process. The decoupling of ventilation from heating and cooling can save energy and improve IAQ. However, one must take appropriate measures to avoid condensation problems.

CONCLUSIONS

Research done by experts in the field as well as in laboratories has helped them understand the relationship between IAQ, ventilation, and energy. More research is needed to link specific health symptoms with exposure to specific or a group of pollutants. The policy challenge will be to raise awareness of indoor air quality so that healthy, comfortable environments can be provided by energy efficient technology.

Satish Kumar

See also: Air Conditioning; Building Design, Commercial; Building Design, Residential; Furnances and Boilers.

BIBLIOGRAPHY

American Society of Heating, Refrigerating, and Air Conditioning Engineers (ASHRAE). (1989). *Ventilation for Acceptable Indoor Air Quality*. Standard 62. Atlanta: Author.

American Society of Heating, Refrigerating, and Air Conditioning Engineers (ASHRAE). (1996). *1996 ASHRAE Handbook: HVAC Systems and Equipment*. Atlanta: Author.

American Society of Heating, Refrigerating, and Air Conditioning Engineers (ASHRAE). (1999). *Method of Testing General Ventilation Air Cleaning Devices for Removal Efficiency by Particle Size*. Standard 52.2. Atlanta: Author.

American Thoracic Society. (1997). "Achieving Healthy Indoor Air." *American Journal of Respiratory and Critical Care Medicine* 156(Suppl. 3):534–564.

Blank, D. M. (1998). "Earning It; What's in the Office Air? Workers Smell Trouble." *New York Times*, February 22, Sunday Section: Money and Business/Financial Desk.

Barbosa, D. (2000). "3M Says It Will Stop Making Scotchgard." *New York Times*, May 17.

Dorgan, C. E., and Dorgan, C. B. (1999). "Developing Standards of Performance (SOP) for IAQ in Building." In *Proceedings of the 8th International Conference on Indoor Air Quality and Climate*. Edinburgh, Scotland.

Fisk, W. J., and Rosenfeld, A. H. (1998). "Potential Nationwide Improvements in Productivity and Health from Better Indoor Environments." In *Proceedings of the ACEEE 1998 Summer Study of Energy Efficiency in Buildings*. Washington, DC: American Council for an Energy-Efficient Economy.

Institute of Medicine. (2000). "Clearing the Air: Asthma and Indoor Air Exposures." Washington, DC: National Academy Press.

"Is Your Office Killing You?" (2000). *Business Week*, no. 3684, p. 114.

Liddament, M. W. (1999). "A review of Ventilation and the Quality of Ventilation Air." In *Proceedings of the 8th International Conference on Indoor Air Quality and Climate*. Edinburgh, Scotland.

Mendell, M. J. (1993). "Non-specific Health Symptoms in Office Workers: a Review and Summary of the Epidemiologic Literature." *Indoor Air* 3(4):227–236.

National Academy of Science. (1999). *Health Effects of Exposure to Radon: BEIR VI, Committee on Health Risks of Exposure to Radon*. Washington, DC: National Academy Press.

Odom, D., and DuBose, G. H. (1991). "Preventing Indoor Air Quality Problems in Hot, Humid Climates: Problem Avoidance Guidelines." Denver, CO: CH2M Hill, in cooperation with the Disney Development Company.

Persily, A., and Norford, L. (1987). "Simultaneous Measurements of Infiltration and Intake in an Office Building." *ASHRAE Transactions* 93(2):42–56.

Roulet, C.; Rossy, J.; and Roulet, Y. (1999). "Using Large Radiant Panels for Indoor Climate Conditioning." *Energy and Buildings* 30:121–126.

Seppanen, O. (1999). "Estimated Cost of Indoor Climate in Finnish Buildings." In *Proceedings of the 8th International Conference on Indoor Air Quality and Climate*. Edinburgh, Scotland.

Smith, K. R. (1999). "The National Burden if Disease from Indoor Air Pollution in India." In *Proceedings of the 8th International Conference on Indoor Air Quality and Climate*. Edinburgh, Scotland.

U.S. Department of Energy. (1998). *Indoor Air Quality Appendix to International Performance Measurement & Verification Protocol*. Washington, DC: Author.

U.S. Environmental Protection Agency. (1991). *Building Air Quality—A Guide for Building Owners and Facility Managers*. Washington, DC: Author.

U.S. Environmental Protection Agency. (1992). *A Citizen's*

Guide to Radon, The Guide to Protecting Yourself and Your Family From Radon, 2nd ed. Washington, DC: Author.

U.S. Environmental Protection Agency. (1995). *The Inside Story: A Guide to Indoor Air Quality*. Washington, DC: Author.

U.S. Environmental Protection Agency. (2000). "EPA and EM." Headquarters Press Release, Washington, DC, May 16.

AIR RESISTANCE

See: Aerodynamics

AIR TRAVEL

Since its inception in the 1920s, commercial air travel has steadily grown. At first available only for the discretionary exclusive use of the wealthy, air travel is now an indispensable tool of business and the world economy. Today, with the competitive marketplace dictating routes and fares, it is now possible for almost anyone in the developed world to fly almost anywhere. It is critical to remember that the airline industry is just that—a business—and the key to business success is efficiency. No matter how romantic the image of air travel may be, airlines exist only as long as they are able to provide a reliable service for a reasonably profitable price. Starting in the United States in 1978 and gradually spreading around the globe, the economic deregulation of the airline industry has profoundly changed the face of this critical enterprise by greatly expanding the market for air travel while exposing it to the vicissitudes of the free market.

In 1999 an estimated 1.5 billion people from around the world chose to fly on commercial carriers. Within the United States alone, 530 million passengers chose to fly. During that year airliners took off and landed 7.3 million times on domestic flights alone. These aircraft consumed over 13 billion gallons of jet fuel. According to the Air Transport Association the U.S. airline industry spent more than $10 billion for fuel for its members' domestic and international flights.

ORIGINS

With more than $100 billion in assets in U.S. carriers alone, the airlines have constantly sought ways to improve their ability to make money. They have searched continuously for better, more efficient aircraft in the endless search for greater yield and productivity. In Europe the airline industry was nationalized for most of its history. Consequently, the emphasis was more on service and less on profits. In the United States the experience was different. Although the early commercial airlines were completely dependent on government subsidies administered as air mail contracts, federal policy encouraged the airlines to develop passenger and freight service to offset the subsidies. In the early 1930s the airlines were given bonuses for carrying passengers and for developing faster and safer aircraft. The government's successful plan was to wean the airlines off the subsidies gradually by encouraging the development of larger, faster, and more efficient aircraft. The immediate result of this far-sighted policy was the creation of the Boeing 247, the world's first modern airliner, and the subsequent Douglas series of commercial aircraft. In fact, according to former American Airlines president C. R. Smith, the 21-passenger Douglas DC-3 was the first aircraft capable of making a profit just by carrying passengers. By 1952 the air mail subsidy was withdrawn, as it was no longer needed.

THE JET AGE

During the 1940s and 1950s, the airline industry developed a new generation of large four-engine aircraft that expanded air travel around the globe. Douglas DC-4s, 6s, and 7s, along the Boeing Stratoliners and the graceful Lockheed Constellations, became the industry standard for speed, comfort, and profitability. While these piston-engine aircraft dominated the world's air routes, a new generation of much faster aircraft powered by jet engines was poised to revolutionize air travel. First developed by de Havilland in Great Britain, the sleek Comet entered service in 1952 but experienced serious technical problems that were not overcome until a new series of aircraft built in the United States seized the market. The Boeing 707 and the Douglas DC-8 entered service in the late 1950s and found immediate success throughout the world. These aircraft offered high subsonic speeds significantly faster than their piston-engine counterparts and much larger passenger-carrying capability. The combination of the greater reliability of the jet and the use of much cheaper kerosene-based fuel instead of expensive high-octane gasoline greatly increased efficiency and productivity.

By the end of the 1960s jets had virtually replaced all piston-engine airliners on most of the world's air routes. Newer, more powerful and more efficient turbofan engines replaced the first generation of thirsty and polluting turbojet engines. In 1970 the arrival of the Boeing 747 dramatically increased aircraft productivity by greatly decreasing seat-mile costs. Equipped with four new-generation high-bypass turbofans that were much quieter, more powerful, and more efficient, the 747 and the wave of wide-bodied aircraft that followed carried twice the number of passengers previously possible, with lower operating costs.

THE DEREGULATION REVOLUTION

By the late 1970s, economists and members of the U.S. Congress realized that despite the great efficiencies now possible because of technological advancements, air fares were not decreasing. In a bipartisan effort, fifty years of regulation dating back to the days of the Post Office contract were ended with the deregulation of the U.S. airline industry in October 1978. While the federal government continued to regulate safety matters, the airlines were now free to charge whatever fares the market would allow and were now free to enter or exit any market they wished. The immediate result was a flood of discount fares as millions of people, many of whom had never flown before, took to the sky. New airlines emerged almost overnight, while numerous local and regional carriers sought to compete directly with the major airlines.

These were heady days for the consumer but chaos for the airlines. Passengers could now compare prices as the airlines drastically cut fares to attract business. Despite a national economy in recession, ridership increased from 254 million to 293 million between 1978 and 1979. But problems loomed. The airline industry is capital-intensive, with high barriers to entry, high operating costs, and thin profit margins. Ruthless competition did not produce long-term stability as most of the new and many of the old airlines fell by the wayside, unable to pay for expensive aircraft and unable to fight against dominant carriers with experience and large cash reserves and access to credit. No longer protected by the government, airlines were free to succeed or fail against the invisible hand of market forces. Many failed. Many more were consumed during a merger mania in the 1980s that sought to cut overhead costs by consolidating companies. Aviation enthusiasts lamented this logical business step as many grand old names such as Pan American, Eastern, Braniff, Piedmont, Frontier, and others disappeared. An oligopoly of larger, more efficient companies survived, ironically little different from the oligopoly of companies that dominated before 1978.

Efficiency became the watchword after deregulation. With competition opened up to all comers, the airlines fought hard to make money by cutting costs and increasing productivity. Coupled with an unexpected crisis in the supply of fuel in 1979, there was an industrywide effort to acquire the latest, most efficient equipment using the latest engine technology.

Following the success of the large high-bypass turbofan engines built for the massive wide-bodied airliners, engine manufacturers turned their attention to building a new generation of smaller engines using the same advanced technology. The result was a new series of engines that produced 20,000 to 40,000 pounds of thrust while using much less fuel than earlier engines of comparable power.

These new engines, from Pratt & Whitney, Rolls-Royce, and the international consortium of General Electric and Snecma, found wide acceptance on new and old narrow-bodied airframes. Boeing produced its 757 in 1982, while McDonnell Douglas updated its proven DC-9 series with the improved MD-80 line. First flown in 1965, the Boeing 737 had been a slow seller. With deregulation and the replacement of its low-bypass engines with two GE/Snecma CFM56 power plants, the 737 quickly became a sales leader because of its high serviceability, low operating costs, and high productivity. It became the aircraft of choice for a host of new entrant and expanding airlines, such as Southwest. The 737 in its many forms is today the most successful jet airliner in history, with well over 4,500 sold.

In Europe, the international consortium of Airbus Industrie took notice of the postderegulation environment in the United States and the changing environment in Europe as well and produced its own efficient narrow-bodied airliner, the A320, in 1988. Incorporating the latest technology, the A320 is the first commercial airliner equipped with a fully computerized "glass cockpit" and digital fly-by-wire control system that cuts weight, enhances safety, and maximizes efficiency. So successful is this revolutionary aircraft that Airbus has sold more than 2,000 A320s in just over ten years.

AIR TRAVEL TODAY

The sudden rush of growing traffic because of the resulting lower fares taxed the aviation infrastructure to its limits. Aircraft utilization soared from six hours per day to twelve hours a day as load factors increased. This led immediately to much congestion at overcrowded airports and severe air traffic problems. Airlines sought to maximize productivity by funneling passengers though centralized airports, the "hub and spokes," where traffic could be efficiently concentrated and dispersed throughout the airlines' route networks. While maximizing productivity for the airlines, the hub and spokes system strained the facilities of these airports during peak hours of very high traffic often causing serious delays and inconveniencing passengers. The sheer volume of aircraft all trying to converge on the same airport at the same time often overwhelmed the air traffic control system. The installation of flow control whereby the movement of aircraft is controlled throughout the system has alleviated some of the worst problems. By holding aircraft at the gate of the departing airport until the route and the gate of the receiving airport will be clear has cut fuel consumption by preventing the massive stacking of aircraft that used to occur at overwhelmed airports. It has also improved the flow of traffic around bad weather and helped to rationalize the system. Problems with aging equipment, the difficult introduction and integration of new technologies, and the issues of an overworked and understaffed labor force still hamper the effectiveness of the nation's air traffic control.

By the early 1990s some semblance of order had returned to the industry. After the trying period of consolidation in the late 1980s, a solid core of well-managed, stable airlines emerged despite the cyclical nature of the business caused by variations in the economy and unpredictable fluctuations in the price of fuel. A solid oligopoly of large airlines now dominates the market and, while problems remain, much improvement has been made. Airports are gradually catching up to the demand for their services by spending billions of dollars to upgrade their facilities, while gradual improvements in air traffic control and more rational scheduling are reducing the worst congestion problems. In the late 1990s the airlines spent heavily in acquiring new, efficient aircraft of all types. Boeing and Airbus, the two largest manufacturers of airliners, are building hundreds of aircraft each year to satisfy the continuous demand for ever-better transports.

FUTURE DEVELOPMENTS

As the new century unfolds, Airbus and Boeing are heavily engaged in market studies to determine the scale and scope of the next generation of airliners. Air travel is steadily growing domestically and internationally, though the rate of growth is slowing. Through 2020, the U.S. Department of Energy projects that the growth in jet fuel consumption should outpace that of all other liquid fuels. The *Annual Energy Outlook* for 1999 foresees that by 2020 the consumption of gasoline will drop from 65 to 61 percent, diesel from 18 to 14 percent, while jet fuel use will rise from 13 to 17 percent.

It is estimated that more than 80 percent of the American population has flown at least once; if that is true, it is effectively a fully mature market. Outside the United States, the demand for air travel is expanding as well, although it has been tempered by economic downturns, particularly in the Asian market. These factors are leading Airbus and Boeing to the conclusion that there will exist a need for an aircraft larger than any now flying. They differ, however, on the question of when that aircraft will be needed and what form it should take.

Airbus Industries is staking its plans on the assumption that the market now exists for 1,300 aircraft that can each seat 550 to 650 passengers. Their marketing experts point to the heavily traveled North Atlantic route, particularly that between New York and London, as well as several high-density routes in the Asia-Pacific region as evidence for this new market and are assuming a steady 5 percent growth in traffic. This aircraft will be needed to handle the increased traffic expected between major population centers and will be necessary to help overburdened airports and air traffic control systems to handle more people, but with a manageable number of aircraft. The large airliner, bigger than the current 747-400 series that carries 416 and 524 passengers, will also enjoy the lowest seat-mile costs of any aircraft, thereby addressing the airlines' ever-present need to improve revenue yield.

Similarly, Boeing predicts a 4.7 percent growth in between 2000 and 2020 but anticipates market fragmentation instead of consolidation as airlines choose to expand service away from the traditional population centers and begin serving many more city pairs not previously served. The competitive marketplace will dictate an increased number of flights from many more cities; therefore the demand will be for more frequency of smaller aircraft. Boeing argues that this is already happening over the North Atlantic, where in 1980 the large four-engine 747 dominated travel. Today smaller, twin-aisle aircraft such as the 767, 777, and A330, dominate as airlines seek to bypass congested hubs. Boeing believes that governments will keep pace with the rise in traffic by building new airports, improving existing facilities, and investing in improved air traffic control.

Boeing does concede that the market for a very large aircraft will exist, particularly along the Pacific Rim, but that that will not occur until the second decade of the twenty-first century, ten years after Airbus's forecast. Boeing believes that there will be a market for 1,030 large aircraft between 2000 and 2020. Of this number, only 425 will be for aircraft with more than 500 seats, one-third that of Airbus's prediction. Boeing argues that in the future, market fragmentation will occur in the United States as well because of increased competition and the availability of existing long-range, large-capacity airliners such as the 777, which can seat up to 394 passengers.

Only time and the marketplace will determine which manufacturer is correct. One thing that both manufacturers agree on is that the airlines will always demand more efficient aircraft in their constant struggle to maintain profitability. The competition between the two manufacturers will help ensure that the best products will be available to the customer and the passengers.

F. Robert van der Linden

See also: Aircraft; Aviation Fuel; Efficiency of Energy Use, Economic Concerns and; Engines; Kerosene; Subsidies and Energy Costs; Supply and Demand and Energy Prices; Transportation, Evolution of Energy Use and.

BIBLIOGRAPHY

Air Transport Association Home Page. <http://www.airtransport.org/default.htm>.

Airbus Industrie Home Page. <http://www.airbus.com>.

Davies, R.E.G. (1982). *Airlines of the United States Since 1914*. Washington, DC: Smithsonian Institution Press.

Dienel, H.-L., and Lyth, P., eds. (1998). *Flying the Flag: European Commercial Air Transport Since 1945*. New York: St. Martin's Press.

Energy Information Administration Home Page. Department of Energy. <www.eia.doe.gov>.

FAA Statistical Handbook of Aviation. Department of Transportation, Federal Aviation Administration. <http://www.bts.gov/NTL/data/tables>.

Heppenheimer, T. A. (1995). *Turbulent Skies: The History of Commercial Aviation*. New York: John Wiley and Sons.

Lynn, M. (1997). *Birds of Prey: Boeing vs Airbus—A Battle for the Skies*. New York: Four Walls Eight Windows.

Morrison, S., and Winston, C. (1995). *The Evolution of the Airline Industry*. Washington, DC: The Brookings Institution.

Ott, J., and Neidl, R. E. (1995). *Airline Odyssey: The Airline Industry's Turbulent Flight into the Future*. New York: McGraw-Hill.

ALCOHOL

See: Biofuels

ALEXANDERSON, ERNST FREDERIK WERNER (1878–1975)

Ernst F. W. Alexanderson was a Swedish American engineer and inventor who is best remembered for his pioneering work on the high frequency alternator that made long-distance radio communication possible. He was born on January 25, 1878, in Uppsala, Sweden. His father, Aron M. Alexanderson, taught classical languages at the University of Uppsala and was later chair of classical languages at the University of Lund. Alexanderson's mother was the former Amelie von Heidenstam. The young Alexanderson was educated at Lund High School and then at the University of Lund (between 1896 and 1897). He continued his studies at the Royal Institute of Technology in Stockholm, and later, in Berlin, he studied under the instruction of Adolf K. H. Slaby, the inventor of a primitive form of radio communication.

Alexanderson was anxious to put his knowledge to practical use. America, which seemed at that time to be the fountainhead of many important technological advancements, beckoned. Arriving in New York in 1901, he immediately went to work as a draftsman for the C. & C. Electrical Company in New Jersey. Alexanderson sought out and was quickly befriended by the esteemed inventors Thomas Edison and Charles Steinmetz. In 1904, after passing General Electric's engineering exams, he became a member of that company's engineering staff.

Alexanderson's big break came when he was commissioned by Reginald Fessenden, a pioneering wireless operator, to build a generator that could produce alternating, high frequency currents. These currents would be used to generate a continuous, dependable wave for radio transmission and thus enable a broadcast of more complexity. On Christmas Eve of 1906, Alexanderson's invention was used to broadcast the first radio show that featured singing and conversation.

Guglielmo Marconi, the Italian engineer and inventor, visited Alexanderson in 1915 and bought one of his 50-foot alternators for the transatlantic Marconi Center in New Jersey. Within a few years, Alexanderson's alternators were to be found in numerous countries. Using a 200-foot Alexanderson alternator, Marconi broadcast radio transmissions during World War I that were heard all over Europe.

In 1916, Alexanderson made another important contribution to radio broadcasting when he unveiled his tuned radio receiver, which allowed for selective tuning. It quickly became an integral part of radio broadcasting.

Alexanderson's alternator played an important part in history when President Woodrow Wilson used it to broadcast his 1918 ultimatum to Germany, ending the war. Afterwards, the Marconi company sought to buy exclusive world rights to the alternator, but was rebuffed by the U.S. government. Wishing to keep control of the invention within American hands, the government set up the Radio Corporation of America (RCA) in 1919, with Alexanderson as its chief head engineer. Concurrently, Alexanderson continued to work for General Electric, an association that lasted forty-six years.

In 1919, he made history with yet another of his inventions when his multiple-tuned antenna, antistatic receiver, and magnetic amplifier were used to transmit the first two-way radio conversation. This great event took place 900 miles out to sea, between the Trans-Atlantic Marconi Company station at New Brunswick and the steamship *George Washington,* with President Woodrow Wilson on board as a witness.

The magnetic amplifier was outmoded by another Alexanderson invention, the electronic modulator, which used vacuum tubes to help generate high frequency transmitters of great power. Alexanderson also helped to create the amplidyne, a direct current generator. By the use of compensating coils and a short circuit across two of its brushes, the amplidyne uses a small power input to precisely control a large power output. Its system of amplification and control originally was designed for use in steel mills, but later hundreds of other applications, including an adaption to fire antiaircraft guns during World War II. Alexanderson also held patents for his inventions of telephone relays, radiant energy guided systems for aircraft, electric ship propulsion, automatic steering, motors and power transmission systems, railway electrification systems, as well as inventions in the fields of radio and television.

In 1924 Alexanderson began his television research. By 1927 his group was able to broadcast mechanical television into the home and in 1930 General Electric gave the first large screen demonstration of television in a theatre in Schenectady. His team then transferred to RCA where they helped develop our modern system of television.

Alexanderson retired from General Electric in January 1948, although he remained a consultant engineer to the company. In 1952 Alexanderson was consulted for RCA. In all, Alexanderson held 322 patents for his inventions.

Ernst Alexanderson's enormous contribution to technology was acknowledged more than once. Some of the honors and awards he received during his long life are: The Gold Medal of the Institute of Radio Engineers in 1919, the Order of the Polonia Restituta in 1924, the John Ericcson Medal in 1928, the Edison Medal of the American Institute of Electrical Engineers in 1944, the Cedergren Medal of the Royal Institute of Technology of Sweden in 1945, and the Valdemar Poulsen Gold Medal and the Royal Danish Medal, both in 1946. He received honorary degrees from Union College, Schenectady, New York in 1926, and the University of Uppsala in 1938. He was a member, fellow, and later president of the Institute of Radio Engineers, a member and president of the Institute of Radio Engineers, and a member of the Swedish Royal Academy and Sigma Xi. In his spare time and retirement, Alexanderson enjoyed sailing, and was elected the first Commodore of the Lake George Yacht Club in New York. Alexanderson died at the age of ninety-seven on May 14, 1975.

Albert Abramson

BIBLIOGRAPHY

Brittain, J. E. (1992). "Alexanderson: Pioneer in American Electrical Engineering." Baltimore: Johns Hopkins Press.

IEEE. (1975). "Ernst Alexanderson: Radio and TV Pioneer at 97." *Spectrum* 12(August):109.

Krebs, A. (1975). "Dr. Ernst Alexanderson, Radio Pioneer, Dies at 97." *New York Times Biography Series* 6:547–548.

"Last of the Pioneers Is Gone—Alexanderson Dies at Age 97." (1975). *Radio-Electronics,* July, p. 6.

"Our Respects to Ernest Frederik Werner Alexanderson." (1945). *Broadcasting-Telecasting,* December, pp. 48–50.

ALTERNATING CURRENT

See: Electricity

ALTERNATING CURRENT MOTOR

See: Electric Motor Systems

ALTERNATIVE FUELS AND VEHICLES

The term alternative fuels first appeared in the energy literature in the late 1970s as a way to refer to nonconventional fuels—fuels that are not gasoline, diesel or aviation fuel. Alternative fuels excludes all fuels refined from petroleum that are normally liquid at ambient conditions, such as gasoline through heavy fuel oil. It does include the highly volatile fractions: liquefied petroleum gas (propane), liquefied natural gas, and compressed natural gas. The category also comprises all fuels made from other fossil fuels, such as coal and oil shale, biofuels originating from plant material, and chemically derived fuels such as methanol and hydrogen. The nonfossil plant-derived fuels such as ethanol and bio-diesel (from vegetable oils), and hydrogen made from water via solar powered electrolysis, are the only renewable energy alternative fuels. Electric vehicles are considered alternative-fuel vehicles since only about 3 percent of electricity comes from burning petroleum.

Alternative fuels should not be confused with alternative energy, which is another term whose origins date back to the 1970s. Alternative energy and alternative fuel both exclude petroleum energy, yet alternative energy goes further by excluding all fossil fuel sources and nuclear. However, sometimes energy sources such as hydroelectric, which accounts for about 10 percent of U.S. electricity production, may be considered alternative even though it has been a major energy source for centuries.

Of all the fossil fuels, petroleum is considered the least sustainable fuel option (most finite and fastest rate of depletion), which is the main reason for the development of alternative fuels. Resources of natural gas and coal are far greater, and depletion slower, which makes alternative fuels developed from these sources more sustainable. The nonfossil plant-derived fuels of ethanol and bio-diesel (from vegetable oils) are the only renewable energy alternative fuels and, in theory, are considered the most sustainable. However, there is not enough agricultural land for biofuels alone to ever become a total replacement for petroleum at the year 2000 world petroleum consumption rate of about 30 million barrels a day

In the United States, the leading use of alternative fuels is not as standalone fuels, but as additives to petroleum-based gasoline and diesel fuel. For example, gasoline sold in much of the United States is 10 percent ethanol or 10 percent methyl tertiary butyl ether (MTBE).

RECENT HISTORY

The first major government investment in an alternative fuel was for the purpose of energy security and oil import independence. Beginning in the late 1970s, billions of dolars were spent on synthetic fuels (converting coal and oil shale into gasoline and diesel). When oil prices began to fall in the early 1980s, and it became apparent that the costs of producing synthetic fuels would remain well above that of petroleum fuels, the program was abandoned.

Interest in alternative fuels grew again in the late 1980s in response to urban air quality problems. Early in the 1990s, environmental regulations calling for oxygenated fuels (nonpetroleum fuels containing oxygen blended with gasoline) to cut carbon monoxide emissions went into effect that significantly increased the sales of MTBE and ethanol. In 1995, alternative fuels comprised about 3 percent of all fuels consumed in the United States (4.4 billion gallons versus 142 billion for gasoline and diesel). MTBE and ethanol consumption is greatest since these fuels are blended with gasoline as required by environmental regulations for carbon monoxide reduction. However, when the 1999 National Research Council study found that there was no statistically significant reduction in ozone and smog based on the data available, the continued requirement of using MTBE was questioned, especially since MTBE leaking from storage tanks can quickly contaminate groundwater.

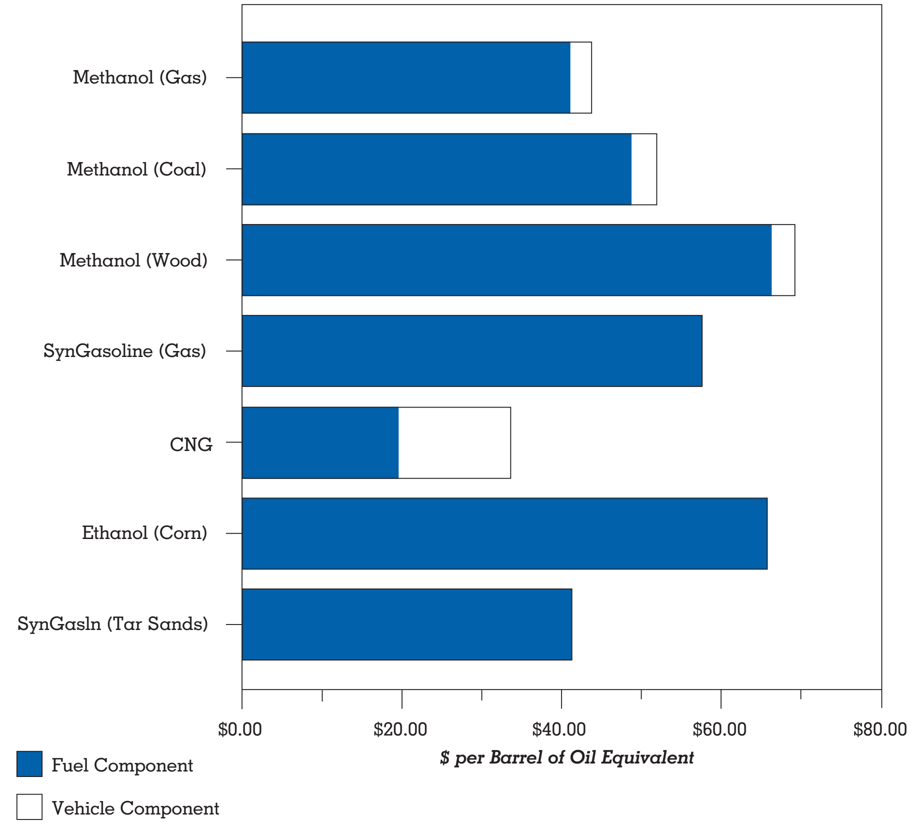


Figure 1.
NRC estimated costs of alternative fuels, including incremental vehicle costs, compared to oil at $20 per barrel.
SOURCE: National Research Council (1990), Table D-4.

MARKET BARRIERS

Alternative fuels are advocated as a way to improve the environment, enhance energy security, and replace dwindling petroleum reserves. Thus, the federal government continues to generously fund research and development for alternative fuels either as a replacement for, or for blending with, conventional fuel. Among the federal subsidies and regulations to promote alternative fuel use are the Energy Policy Act of 1992 (requiring alternative fuel vehicles in fleets and providing tax breaks for people who buy these vehicles), the Intermodal Surface Transportation Efficiency Act of 1991 (providing grants for purchasing alternative-fuel vehicles and for building refueling stations), and the Alternative Motor Fuels Act of 1988 (allowing automakers to sell more large, higher-profit conventional cars with poorer fuel economy if they also sell alternative fuel vehicles). However, even with subsidies and favorable regulations, alternative fuels face significant hurdles before becoming practical replacements for conventional fuels. Foremost is cost, followed by safety, practicality and reliability, and finally, the development of infrastructure (production, distribution and retailing availability).

Cost-Competitive

Natural-gas-derived fuels are the most cost-competitive because natural gas does not need to be refined like gasoline and diesel fuel from petroleum (Figure 1). Ethanol, a heavily subsidized alternative fuel, is not as cost-competitive as natural-gas-derived fuels. If not for the subsidies and environmental regulations requiring oxygenates, ethanol would not be used at all. The chances of ethanol ever becoming cost-competitive in the free market are slim since extensive land is needed to raise high-energy-yield plants for fuel, and the energy that must be expended to raise, harvest and dry the plants for the fermentation alcohol results in a low net energy yield.

When gasoline-powered automobiles are modified to burn a fuel such as ethanol alone, they are known as dedicated ethanol vehicles—risky investments for buyers who have concerns about future availability. For example, Brazil's Proalcool program promoted and heavily subsidized ethanol, and thus dedicated ethanol vehicles, from 1975 to 1988. Once the subsidies were curtailed and then eliminated (estimates of the costs of the subsidy to the government range from $7 to $10 billion), shortages resulted. Many of the owners of ethanol-dedicated vehicles either had to junk or retrofit the vehicles to run on gasoline, and the sales of ethanol-dedicated vehicles went from 50 percent of the market in 1988 to 4 percent by mid-1990.

Aside from the difference in fuel costs, the cost of redesigning and equipping vehicles engines and fuel tanks to run on alternative fuels has to be considered. Responding to the desire to switch fuels for cost reasons, or refueling security when the alternative fuel is not readily available, several auto makers offer flexible fuel vehicles that run primarily on compressed natural gas, but also gasoline when compressed natural gas is not available. These vehicles are sold at a premium and have shown little success in attracting buyers since low fuel prices ensure their return on investment will be poor in comparison to standard gasoline vehicles.

Practicality and Reliability

When the energy density of the alternative fuels is considerably less than gasoline and diesel fuel, it greatly impacts the practicality of the fuel for transportation. Most of the alternative fuels have a much lower energy density (Table 1). For vehicles, much more storage space is required to accommodate much larger fuel tanks to achieve comparable range or, for gaseous fuels, storage tanks that can withstand greater compression. Moreover, it will always take longer to refuel a vehicle using a lower-energy-density liquid fuel or a gaseous fuel.

The lower energy density of alternative fuels is even more problematic for aircraft. Methanol has been suggested as a jet fuel replacement. But using methanol would seriously curtail range and payload since the plane's weight is a principle determinaent of how much fuel is needed. A typical four-engine commercial jet will carry 775,000 pounds of aviation fuel to maximize range; to achieve the same range with methanol would require one million more pounds of fuel.

Since most of the alternative fuel vehicles burn cleaner, experience has found that this reliability is equal or better than that of comparable gasoline or diesel fuel vehicles.

Energy Density	*lb./ gal. @ 60 degrees F.*
No. 2 Diesel Fuel	6.7 - 7.4
Gasoline	6.0 - 6.5
Methanol	6.63
Ethanol	6.61
MTBE	6.19
Propane (LPG)	4.22
Methane	1.07

Table 1.
Energy Densities of Fuels

Infrastructure Needs

A vast petroleum production, refining, distribution and retailing operation exists to deliver gasoline and diesel fuel. The major oil companies have invented billion of dollars in the setup and delivery of liquid fuels that can be stored in underground tanks; thus, any alternative fuel that requires massive new investments in infrastructure will face considerable market resistance. Moreover, there are personal investments in over 200 million vehicles on the road that are designed to consume either gasoline or diesel fuel, and a dauntingly immense and specialized infrastructure of industry building these vehicles and small businesses maintaining them. Since so much of the economy has a vested interest in the internal combustion engine burning gasoline or diesel fuel, a market transition to alternative fuels and vehicles is likely to be gradual.

In the free market, as long as petroleum supplies are plentiful, there is little incentive for oil companies to transition to any of the alternative fuels, which is a major reason that the U.S. Department of Energy projects petroleum consumption will rise from 18.6 million barrels per day in 1997 to 22.5-26.8 million barrels by 2020. As the crude oil reserves dwindle, the marketplace will either transition to the electrification the transportation system (electric and fuel cell vehicles and electric railways), or see the development of alternative fuels. Any short-term transition to an alternative fuel is likely to meet environmental air quality regulations. Beyond 2020, the transition is likely to occur due to the depletion of oil reserves resulting in steeply rising gasoline and diesel prices, or from advances in technologies that make alternative fuels and alternative transportation more attractive.

John Zumerchik

See also: Biofuels; Capital Investment Decisions; Hydrogen; Kinetic Energy, Historical Evolution of the Use of; Methanol; Natural Gas, Processing and Conversion of.

BIBLIOGRAPHY

Greene, D. L. (1996). *Transportation and Energy*. Landsdowne, VA: Eno Transportation Foundation, Inc.

Hadaller, O. J., and Momenthy, A. M. (1993). "*Characteristics of Future Aviation Fuels.*" In *Transportation and Global Climate Change*, edited by D. L. Greene and D. J. Santini. Washington DC: American Council for an Energy Efficient Economy.

Howes, R., and Fainberg, A. (1991*). The Energy Sourcebook: A Guide to Technology, Resources and Policy*. New York: American Institute of Physics.

Lorenzetti, M. S. (1995). *Alternative Motor Fuels: A Nontechnical Guide*. Tulsa, OK: Pennwell Press, Inc.

National Research Council, Energy Engineering Board. (1990). *Fuels to Drive Our Future*. Washington DC: National Academy Press.

ALTERNATOR

See: Electric Motor Systems

AMPERE

See: Units of Energy

AMPÈRE, ANDRÉ-MARIE (1775–1836)

André-Marie Ampère was born in Lyons, France, the son of a wealthy merchant. Ampère's education was determined by his father, Jean-Jacques, who followed Jean Jacques Rousseau's theories of education. Ampère was left to educate himself, as his inclinations dictated, among the books of his father's extensive library. At an early age Ampère discovered a talent for mathematics, working out the early books of Euclid by himself. On finding that some of the books he wished to consult in the library in Lyons were in Latin, he taught himself the language. Ampère's mother was a devout Catholic, who ensured he was thoroughly instructed in the faith.

Ampère's domestic life was beset with tragedy. In 1787 the French Revolution began; Jean-Jacques assumed the post of *Juge de Paix,* a role with considerable police powers. When Lyons fell to the troops of the Republic in 1793, Jean-Jacques was tried and guillotined. In 1799 Ampère married, supporting his wife by teaching mathematics in Lyons, where their son was born the next year. Weakened by childbirth, his wife died in 1803. Ampère moved to Paris and took a post in mathematics at the École Polytechnique. In 1806 he remarried, but this union was ill-advised. After the birth of a daughter, his wife and mother-in-law made life for Ampère so unbearable that he was forced to seek a divorce. Ampère persuaded his mother and sister to come to Paris and take charge of his household. Ampère's expectations of his children were never realized, and in domestic life he faced constant money problems. In 1808 Ampère was named Inspector General of the newly formed university system, a post he held until his death. In 1824 Ampère was elected to the Chair of Experimental Physics at the Collège de France.

As a deeply religious person whose personal life was beset by a series of calamities, Ampère searched in the field of science for certainty. He constructed a philosophy that enabled him to retain a belief in the existence of both God and an objective natural world. Ampère's philosophy contained two levels of knowl-

André-Marie Ampère. (Public domain)

edge of the external world. There are phenomena witnessed through the senses, and the objective causes of phenomena—*noumena*—that can only be apprehended through intellectual intuition. Although Ampère's philosophical system was the one continuing intellectual passion of his life, he also devoted himself to other fields of scientific research. From 1800 to around 1814, mathematics was Ampère's primary interest, with his spare time spent in chemical investigations. From 1820 to 1827 he carried out the scientific work for which he is best known, pioneering the science of electrodynamics.

In 1820 Hans Christian Oersted discovered electromagnetism. A report of Oersted's work was delivered before a sceptical meeting of the Académie des Sciences held on September 4, 1820. Oersted's work was contrary to established ideas, based on Coulomb's work of the 1780s, that there could not be any interaction between electricity and magnetism. Ampère however, immediately accepted Oersted's discovery, and set to work, reading his first paper on the subject to the Académie on September 18, 1820. Oersted's discovery suggested to Ampère, that two wires, each conducting current, might effect one another. Deducing the pattern of magnetic force around a current carrying wire to be circular, Ampère went on to visualize the resultant force if the wire were coiled into a helix. One week later Ampère announced to the Académie, his discovery of the mutual attraction and repulsion of two helices. In doing so Ampère presented a new theory of magnetism as electricity in motion.

Ampère's researches followed his own philosophy on the nature of science and scientific explanation. The phenomenon of electromagnetism had been discovered by Oersted, and the relationship between two current-carrying wires by Ampère; what remained was the discovery of the noumenal causes of the phenomenon. In his first memoir on electodyamics Ampère investigated the phenomenon and provided factual evidence to show that magnetism was electricity in motion. He concluded that two electric currents attract one another when moving parallel to one another in the same direction; they repel each other when they are parallel but in opposite directions. Ampère felt that electrical phenomena could be explained in terms of two fluids, a positive one flowing in one direction and a negative fluid going in the other.

In 1820 Ampère described the magnetism of a needle placed within a helical coil. With the assistance of Augustin Fresnel, Ampère unsuccessfully attempted to reverse the procedure by wrapping a coil around a permanent magnet. They did not investigate the effect of moving the magnet within the coil. If magnetism is only electricity in motion, then, Ampère argued, there must be currents of electricity flowing through ordinary bar magnets. It was Fresnel who pointed out the flaw in Ampère's noumenal explanation. Since iron was not a good conductor of electricity, there should be some heat generated, but magnets are not noticeably hot. In a letter to Ampère, Fresnel wrote, since nothing was known about the physics of molecules, why not assume that currents of electricity move around each molecule. Ampère immediately accepted the suggestion, assuming each molecule to be traversed by a closed electric current free to move around its center. The coercive force, which is low in soft iron, but considerable in steel, opposes this motion and holds them in any position in which they happen to be. Magnetism consists in giving these molecular currents a parallel direction;

the more parallel the direction, the stronger the magnet. Ampère did not say why molecules should act this way, but it was sufficient that this electrodynamic model provided a noumenal foundation for electrodynamic phenomena.

In 1821 Michael Faraday sent Ampère details of his memoir on rotary effects, provoking Ampère to consider why linear conductors tended to follow circular paths. Ampère built a device where a conductor rotated around a permanent magnet, and in 1822 used electric currents to make a bar magnet spin. Ampère spent the years from 1821 to 1825 investigating the relationship between the phenomena and devising a mathematical model, publishing his results in 1827. Ampère described the laws of action of electric currents and presented a mathematical formula for the force between two currents. However, not everyone accepted the electrodynamic molecule theory for the electrodynamic molecule. Faraday felt there was no evidence for Ampère's assumptions and even in France the electrodynamic molecule was viewed with skepticism. It was accepted, however, by Wilhelm Weber and became the basis of his theory of electromagnetism.

After 1827 Ampère's scientific work declined sharply; with failing health and family concerns, he turned to completing his essay on the philosophy of science. In 1836 Ampère died alone in Marseilles during a tour of inspection.

Robert Sier

See also: Faraday, Michael; Oersted, Hans Christian.

BIBLIOGRAPHY

Hofmann, J. R. (1995). *André-Marie Ampère*. Oxford: Blackwell.

Williams, L. P. (1970). "Andre-Marie Ampere." In *Dictionary of Scientific Biography,* edited by C. C. Gillispie. New York: Scribners.

ANIMAL AND HUMAN ENERGY

Culture is the primary mechanism of human behavior and adaptation. Cultures are passed on socially from generation to generation. Tools and the ways they are made and used both shape and are shaped by the social organizations of which they are a part. In his classic study, *The Science of Culture*, Leslie A. White suggested that cultures grow in relation to the degree of efficiency of their tools in liberating energy from their natural environments. A working hypothesis in present day anthropology suggests that, in general, as tools became more efficient and numerous, populations increased in size and density, and new forms of social organization had to develop to cope with these population changes. It would be useful to consider three very general evolutionary stages in the development of our topic: hunting-gathering societies, early agricultural societies, and state agricultural societies.

HUNTING-GATHERING SOCIETIES

There is archaeological evidence that the earliest stone tools were used in Africa more than two million years ago. These "pebble tools" were often made from flat, ovoid river stones that fit in the palm of one's hand. Another stone was used to chip off a few adjacent flakes to form a crude edge. However crude the edge, it could cut through the thick hide of a hunted or scavenged mammal when fingernails and teeth could not, and thus provide the group with several dozen to several hundred pounds of meat—a caloric and protein windfall. It is a common misconception that stone tools are crude and dull. Human ancestors learned to chip tools from silicon-based stones such as flint and obsidian that have a glass-like fracture. Tools made from such stones are harder and sharper than finely-honed knives of tool steel. As early as 500,000 years ago, well-made hand axes were being used in Africa, Europe and Asia. There is evidence that people were using fire as early as one million years ago. With the advent of the use of fire, it is believed that the amount of energy used by early man doubled from 2,000 kilocalories per person per day (energy from food) to about 4,000 kilocalories per person per day (Figure 1). Thus, fire and eventually the apparatus employed to make it were important energy liberating tools. Cooking food allowed humans to expand their diet. Controlled fire could also be used to scare game into the open to be hunted. Present-day hunters and gatherers burn berry patches in a controlled way, to encourage new plants and more berries. Fire also subsidized body heat, allowing people to colonize colder regions.

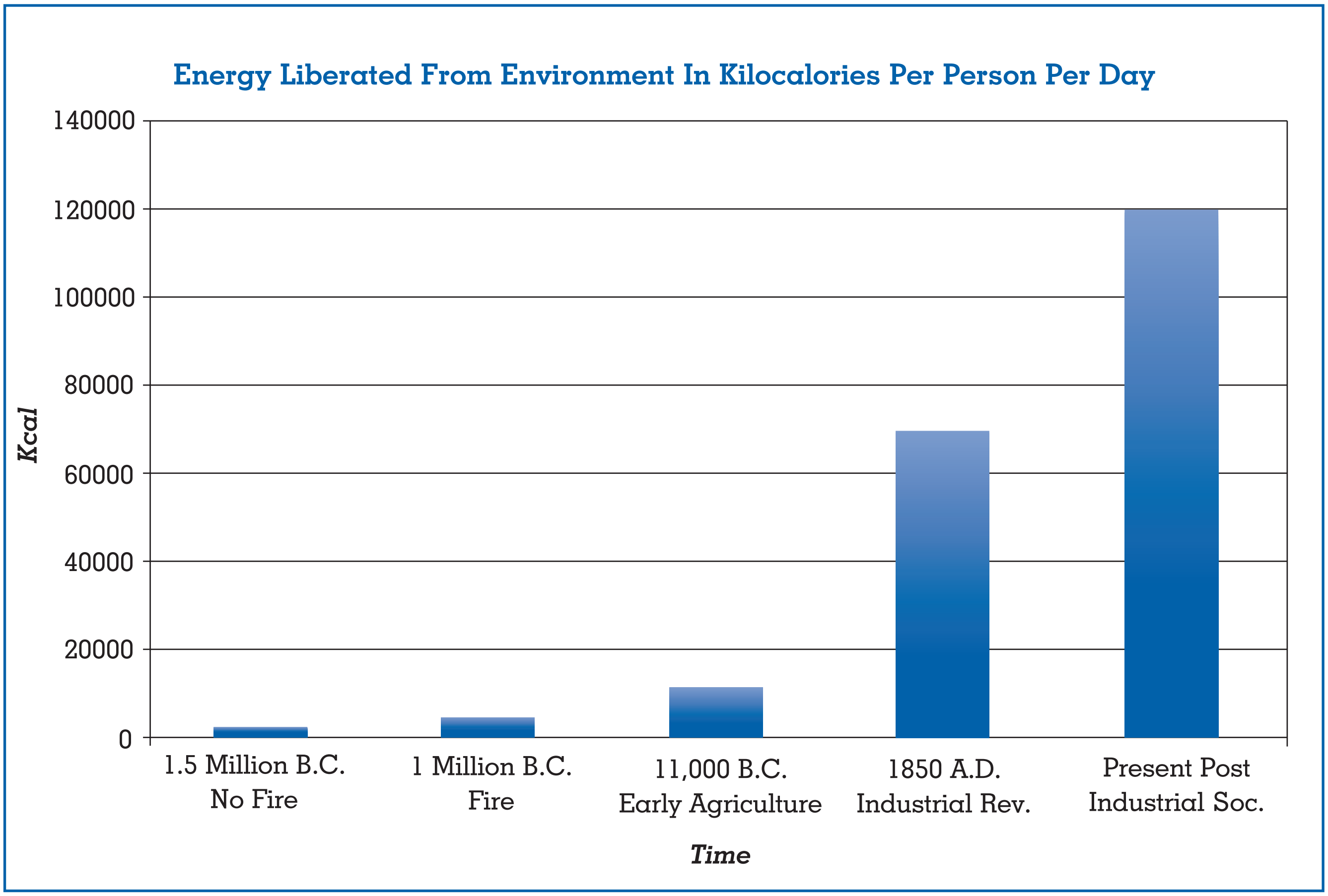


Figure 1.

SOURCE: Based on data from Earl Cook's *Man, Energy and Society* (1976, pp. 166–167).

The Neanderthal people who lived from about 130,000 to 35,000 years ago invented gracile and efficient "blade tools" that allowed them to make lighter and sharper spearheads. The Upper Paleolithic hunters who succeeded the Neanderthals made many smaller, lighter, and super-efficient blade tools called *burins*, each of which could be used for a special purpose. These burins could be transported easily from campsite, thus saving energy. Upper Paleolithic hunters also made many small and efficient tools from bone, ivory, shell and wood. The Upper Paleolithic invention of the *atlatl* or "throwing stick" allowed hunters to throw spears and darts at animals over longer distances and with greater force and accuracy than comparable missiles thrown by hand.

Just at the end of the Upper Paleolithic a Mesolithic cultural period emerged and with it several new innovations made hunting even more energy efficient. The bow could launch very lightweight arrows at game, and poison for arrows decreased the time spent stalking wounded prey. Finally, the domestication of the dog, itself a master hunter, lightened the hunter's task.

Anthropological research with modern hunter-gatherers suggests an ideal type or model for this kind of society. They were nomadic and exhibited low population size and density—on the order of thirty people per thousand square miles. Paramount in maintaining this low size and density was an imperative common to all hunter-gatherer women. A nomad woman had to move herself, all that her family owned (which was very little), and her children at a moments notice. Modern hunters and gatherers often have to walk twenty miles a day, so mothers cannot carry more than one small child. Faced with this restriction, women are careful to space their children so that the two or rarely three children they have will not burden them when they must move. There is every reason to believe that this pattern also existed in remote times.

There was little social differentiation. No one was particularly rich or poor. Leadership was informal and transitory and based on cooperation and sharing, especially with regard to meat. Although homicide was known, intragroup conflict was usually resolved by breaking the group into two or more subgroups that avoided each other until the conflict was resolved or forgotten.

Warfare was not a major cultural focus of hunter-gatherer communities and most anthropologists are of the opinion that when it did occur it was due to population pressure exerted by agricultural and/or state societies. Although hunter-gatherer social structure was informal and free-form, it was efficient with regard to hunting big animals. Studies of modern hunter-gatherer societies indicate that very little work is done in many such groups. With some exceptions, hunter-gatherers worked approximately four hours per day per adult.

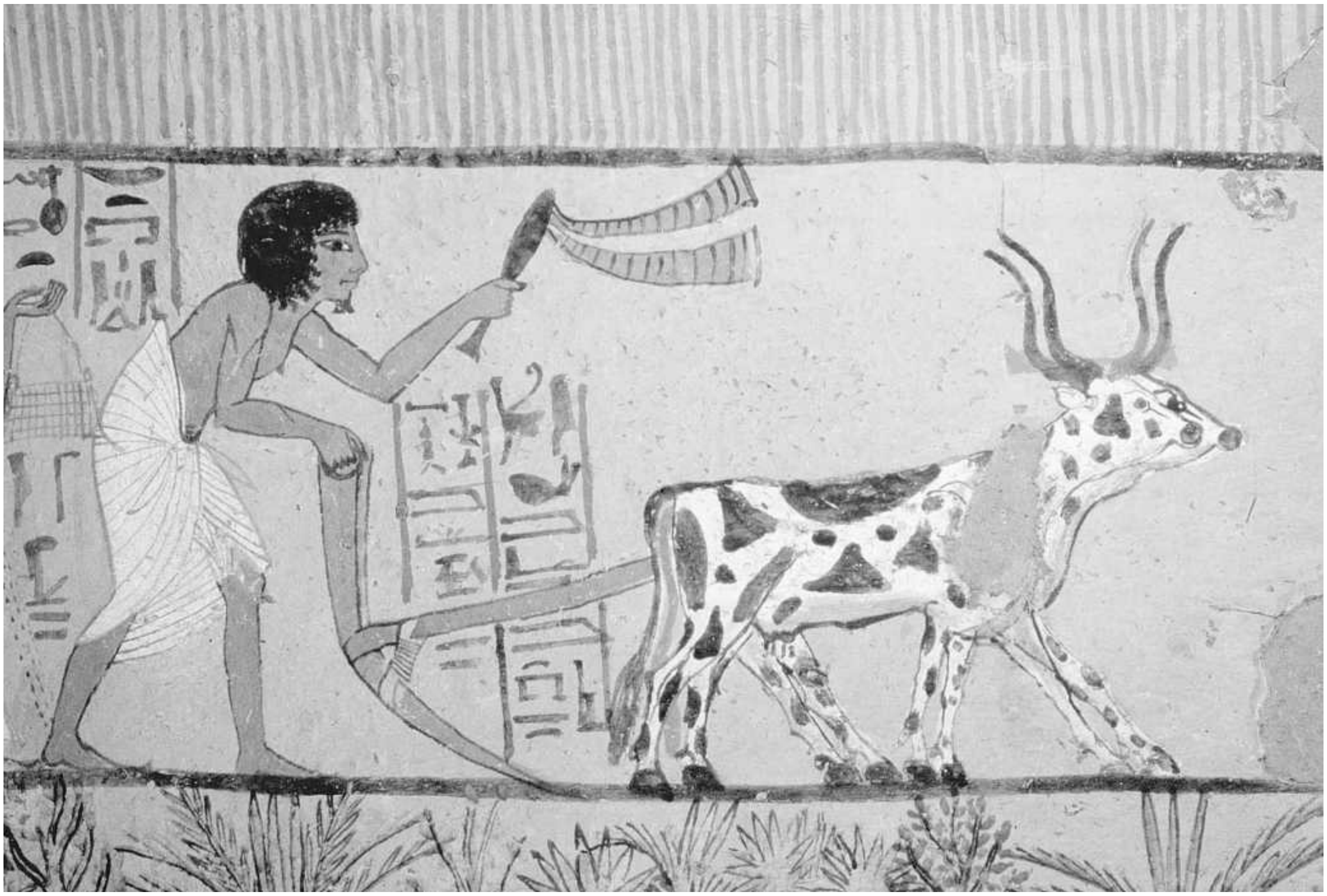
EARLY AGRICULTURAL SOCIETIES

Beginning about eleven thousand years ago in the Middle East and somewhat later in other parts of the world, people began to experiment with the more sedentary subsistence patterns of growing crops and animal husbandry, and thus the Neolithic revolution began. Scholars cite two reasons for this change. First, the last glacial period began to wane as did many of the big mammals associated with it. Second, hunter-gatherer tools and organization became so efficient that human population numbers became relatively large and depleted the supply of game and wild food plants. Population expansion resulted in competition for space and directed humans toward the invention of more intensive modes of food production.

Early Neolithic peoples domesticated the more productive local plants, cared for them in densely planted plots, protected them from animals and other plants (weeds) and harvested the results. Likewise they tamed, bred and cared for local animals and ate them as they deemed fit. In the cases of cattle, horses, sheep and goats, milk and its products became staple foods. In some places larger domestic animals became beasts of burden. For very sound ecological reasons, agriculture allowed even early farmers to liberate much more caloric energy from plants than could be liberated from hunting and gathering and, thus, many more humans could be supported per square mile. In ecosystems, the most numerous organisms (measured in mass) are plants. Animals that eat plants (herbivores) are less abundant, while animals that hunt and eat these herbivores are least numerous. The reason for this decrease in "biomass" as one proceeds up the food chain lies in the fact that there is a considerable loss of energy due to inefficiency as animals search out, eat, digest and change the plants they consume into heat, growth and kinetic energy. The same can be said with regard to the killing and utilization of herbivores by hunting carnivores. Thus when people began cultivating plants, in effect, they stopped being carnivores and became herbivores and their population size and density increased. Since agricultural peoples were sedentary, they lived in more or less permanent settlements and their women could and did have more children. In early agricultural societies the amount of energy liberated per person per day rose to about 11,000 kilocalories.

In woodland areas early farmers developed *swidden* agriculture. They cut down forest plots using new and more efficient axes that were ground and polished from hard and dense stone. The plots were left to dry and then were burned, allowing nutrients to return to the soil. One to three plantings could be grown on these plots every year. Other plots were then prepared and cropped until the original plot had been overgrown and the process was repeated. Swidden cycles lasted from about thirty to one hundred years.

In general there is a correlation between human population size and complexity of social organization. In the 1960s, Robert Carneiro showed that there is a rough positive correlation between the number of organizational traits (N) in single community societies and their population (P), expressed by the equation . Neolithic populations increased and there was an organizational reaction. Society became stratified based on access to the extra calories of agricultural surplus. It was now possible for some people to "own" more land and animals than other people. With sedentary lifeways it was possible to accrue and own duplicate things and heavy things. Coercive institutions like war evolved to protect what one had and as the means to get more. Although more work was done in early agricultural societies, the people in these societies



This ancient Egyptian fresco evinces the long-standing relationship between humans and other animals for agricultural purposes. (Corbis Corporation)

probably had, on the whole, more leisure time than people today. The cultural florescence of the Iroquois Indians of New York State is a case in point. Around 1000 C.E. a Proto-Iroquoian culture became evident in the archaeological record. This *Owasco* culture was characterized primarily by swidden agriculture of maize, beans and squash supplemented by hunting and gathering. Over the next six hundred years village size increased as did size of dwellings. Villages became palisaded and cannibalism became evident, indicating that warfare was an institution of growing importance. When Europeans began to establish contact with the Iroquois in the 1600s, they found a people with sufficient leisure to engage in elaborate warfare patterns that were connected with a rich and complex ritual life. At first contact the Iroquois were building advanced political and governmental institutions and establishing larger orbits of political influence based on the collection of tribute.

STATE AGRICULTURAL SOCIETIES

Late Neolithic times saw the evolution of a technical innovation that fostered the growth of societies that were monumental both in population size and organization. This innovation was irrigation. Hence, historians such as Karl Wittfogel speak of the "hydraulic" theory of the state. In Old World societies like Egypt, Mesopotamia and Northern China, farmers began to grow crops on the flood plains of great river systems, taking advantage of the water and nutrients that these rivers deposited on an annual basis. Irrigation works and the subsequent population increase they stimulated required more irrigation and an evolving bureaucratic organization to manage workers, works and increasing surpluses. The animal-drawn plow became very important at this juncture and thus there was a dramatic rise in the calories that farmers could wrest from their environments. A man pushing a lever could generate about 1/20 horsepower, while an ox pulling a load (or plow) could generate 1/2 horsepower. This new technology could support still larger populations than could simple agriculture. The large-scale civilizations which evolved in all three areas supported hundreds of people per square mile. Similar state societies based on irrigation evolved independently in the New World even though Conquest States like the Aztecs and the Incas of Peru lacked both the plow and the wheel and had no beasts of burden except the dog and the llama.

Modern State Societies depend on irrigation farming as well as dry field agriculture based on mechanized energy subsidies. At first steam power accounted for these subsidies. By 1800 Watt's steam engine could generate 40 horsepower. Later the internal combustion engine replaced steam. At the height of the Industrial Revolution (c.1850) each person used about 70,000 kilocalories per day. Today we have entered into a new phase of social ecology subsidized by an ever-increasing use of fossil fuels. As a result people in developed countries now use about 140,000 kilocalories per person per day.

David Mulcahy

BIBLIOGRAPHY

Carneiro, R. L. (1967). "On the Relationship between Size of Population and Complexity of Social Organization." *Southwestern Journal of Anthropology* 23:234-243.

Cohen, M. (1977). *The Food Crisis in Prehistory.* New Haven: Yale University Press.

Cook, E. (1976). *Man, Energy and Society.* San Francisco: W. H. Freeman and Company.

Funk, R. E. (1983). "The Northeastern United States." In *Ancient North Americans,* Chapter 8, ed. Jesse D. Jennings. San Francisco: W. H. Freeman and Company.

Kottak, C. P. (1997). *Anthropology: The Exploration of Human Diversity*, 7th ed. New York: McGraw-Hill.

Jones, S., et al., eds. (1992). *The Cambridge Encyclopedia of Human Evolution.* Cambridge, Eng.: Cambridge University Press.

Robbins, R. H. (1997). *Cultural Anthropology: A Problem-Based Approach*, 2nd ed. Itasca, IL: F. E. Peacock Publishers, Inc.

Sahlins, M. (1972). *Stone Age Economics.* Chicago: Aldine Atherton Press.

Stein, P. L., and Rowe, B. M. (1996). *Physical Anthropology*, 6th ed. New York: McGraw-Hill.

White, L. A. (1949). *The Science of Culture: A Study of Man and Civilization.* New York: Grove Press.

Wittfogel, K. A. (1970). *Oriental Despotism, A Comparative Study of Total Power.* New Haven: Yale University Press.

APPLIANCES

INTRODUCTION

Appliances refers to a vast array of devices that account for about 35 percent of overall energy consumption in the United States. As was pointed out in an early California Energy Commission hearing determining the scope of the law that delegated to that Commission authority to regulate the efficiency of appliances, *Webster's Unabridged Dictionary* (2nd ed., 1979) defines *appliance* as "something applied to a particular use; a device or machine…." This broad definition allows products that use energy directly or indirectly for virtually any purpose to be considered as appliances.

This review adopts the broad perspective, recognizing that the primary policy mechanisms applied to improve energy efficiency—minimum efficiency standards, incentive programs, normative and informative labeling programs, and technology-driven market forces—can address a very wide variety of products.

Examples of products considered appliances under U.S. federal law, most recently modified by the Energy Policy Act of 1992, include residential products such as refrigerators, freezers, clothes washers, dishwashers, water heaters, heating and cooling equipment, televisions, computers and their power supplies, showerheads, toilets, plumbing fittings, and cooking products. Also considered appliances under U.S. federal law commercial and industrial appliances such as air conditioning and heating systems and their components, water heaters and storage tanks, boilers, lighting system components such as fluorescent lamps, ballasts and fixtures, in addition, incandescent lamps, and motors. These products are involved in virtually all energy use in the residential and commercial sectors, which exceeds 35 percent of the U.S. total, including the upstream effect of electricity consumption. Excluding motors and lighting, appliance energy use is about one–fifth of the U.S. total.

Each of these products is different; each uses a different technology, is subject to different market forces, different production characteristics, distribution practices, and pricing systems. Yet, the dominant trends in energy efficiency, and in the policies used to affect it, show remarkable similarity across end uses.

For most appliances, substantial efficiency improvements have taken place over the last thirty years. Most of these improvements have been responsive to policy initiatives at the state or federal level, although in a few cases technology improvements that were aimed at some other goal also included substantial energy efficiency improvements as a by-product. The range of efficiency improvement has varied dramatically, from reductions of over 75 percent in energy use of refrigerators that comply with the 2001 Department of Energy standard compared with similiar (but somewhat smaller) products in 1972, to water heaters, where current products appear to be no higher in efficiency than they were in the 1940s.

Energy policies and the progress of technology are deeply intertwined for appliances. Appliances are sold to provide a specific energy service, and that service is the focus of their marketing and promotion as well as of consumer acceptance. Energy efficiency is not widely perceived by the marketplace to be important. Even when the payback period of energy efficiency investment by the consumer is short, surveys have shown that energy efficiency is never the first, or even second or third most important feature determining consumer choice.

This trend is self-reinforcing in the broader market and creates a vicious circle. Manufacturers recognize that the consumer will not accept a product that costs more to achieve greater energy efficiency, even if the payback period is as short as three years. Because of this observation, manufacturers do not produce high– efficiency options for the consumer marketplace in sufficient quantities to make their prices competitive. The consumer, therefore, does not perceive efficiency as a product differentiation feature. This reinforces an indifference to energy efficiency. In recognition of this indifference, policy interventions—in the form of mandatory efficiency standards, incentive programs such as rebates for products meeting specified efficiency levels or competitive acquisition of products that achieve the highest cost/efficiency results, bulk purchases of efficient products, and normative labeling—have led to the introduction of vastly improved technologies.

Indeed, before widespread government policy interventions into appliance efficiency, trends in efficiency were as likely to represent decreases in efficiency as increases. The same market barriers that recently have impeded the introduction of new technologies for efficiency have, in the past, caused products to be redesigned with new technologies for lower energy efficiency in order to cut first costs.

Refrigerators declined in efficiency following World War II due, apparently, to the substitution of lower-cost, lower-efficiency motors with improved thermal insulation to protect them from waste heat generation, and there is some evidence that water heater efficiency declined between the end of World War II and 1975, and that industrial dry-type transformers declined in efficiency between 1970 and 1993.

Energy consumption also can increase through the introduction of new product features, a process that should be distinguished from reductions in efficiency. The move from black and white to color television, and the introduction of automatic defrost on refrigerators, are examples of additions of features that compromise energy conservation. These must be balanced against other features or technologies, such as electric components for television and micro-channel heat exchangers for air conditioners, that improve product performance while also improving efficiency.

The effects of trends toward lower efficiency and higher energy-consuming features and the countervailing force of policy interventions is illustrated in Figure1, which displays the history of refrigerator energy consumption.

During the period before 1972, when energy policy ignored efficiency, energy consumption grew at an annual rate of about 6 percent. Some of this trend was due to the introduction of new features such as automatic defrosting and through-the-door ice service, some of it was due to increases in product size and in freezer compartment size, but two percentage points of the annual growth rate in energy was due to actual decreases in efficiency.

Following multiple energy policy initiatives of the post-1973 era, this trend was arrested and energy consumption began to decline, even as size and features continued to increase. The effect of the 1979 California efficiency standard is clearly apparent in the figure. Further inflection points in the curve appear when California implemented its 1987 standards, when the 1990 federal standards went into effect, and particularly when the 1993 amendments went into effect. In addition, a gradual slope toward decreasing energy consumption coincided with utility-sponsored incentive programs whose largest effects were felt in the mid-1980s and early 1990s.

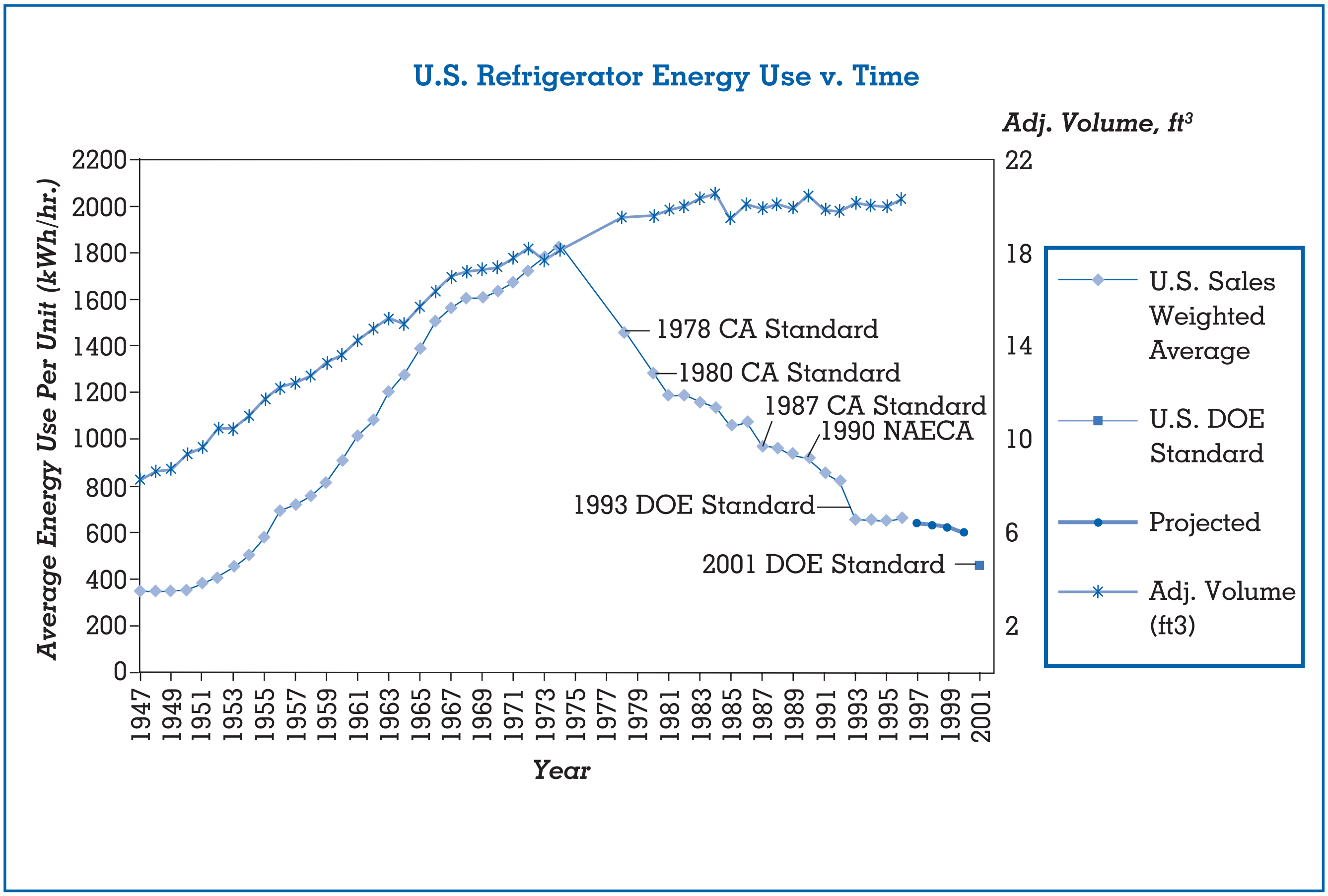


Figure 1.

SOURCE: S. M. Berman, et al. (1976). "Electrical Energy Consumption in California: Data Collection and Analysis." Lawrence Berkeley Laboratory, UCID 3847 (for 1947–1975 data). Association of Home Appliance Manufacturers (for 1972 and 1978–1995 data).

In sum, for appliances, energy efficiency technology improvements are not always adopted even after they become well understood and technology changes do not necessarily lead to improved energy efficiency.

HISTORY OF APPLIANCE EFFICIENCY IN THE UNITED STATES

During the era from the end of World War II to 1973, energy consumption for appliances increased dramatically. Overall, residential electricity consumption was increasing at an annual rate of 9 percent in the United States, while residential energy consumption was increasing by about 3 percent. These increases were driven by a number of factors that all compounded the growth: increasing population, increasing number of households due to declining household size, increasing saturation of products, increasing levels of energy service being provided by many of the products, and, in several cases, declining energy efficiency. Some, new products were introduced that had not existed previously, but if we define end uses broadly, this has not been a major effect.

Increasing saturation was a particularly important trend. While the vast majority of families owned refrigerators after World War II, virtually everyone owned a refrigerator in 1973, and over 10 percent of households had more than one. While televisions were relatively unknown in 1947, radios provided a similar service and consumed noticeable amounts of energy. Some other trends in saturation are noted in Table 1.

Appliances began to attract serious analytic interest toward the end of this era. The Northeast blackout of

Product	*1970*	*1982*	*1990*	*1996*
Washers	71.1%	73.6%	92.6%	95.3%
Dryers	41.6%	65.3%	80.6%	82.4%
Dishwashers	18.9%	44.5%	53.9%	56.9%
Microwave Ovens	small	25.6%	82.7%	90.5%

Table 1.
Trends in Appliance Saturation

1965 inspired engineers to take a second look at the consequences of continued exponential growth in electricity demand. On the West Coast, environmental concerns over the siting of power plants led to early analysis of the possibilities of policy interventions such as appliance efficiency standards.

The energy crisis of 1973 greatly accelerated efforts at analyzing the impact of appliances on regional and national energy consumption, and stimulating policy responses. On the West Coast, a major analysis of power plant siting options and efficiency and renewable energy alternatives led to the passage of the Warren-Alquist Act of 1974. This act established a California Energy Commission with the authority to forecast the needs for power plants under different scenarios of energy efficiency, analyze efficiency options, and propose efficiency standards. California adopted its first appliance standard in 1976.

In the east, a New York Public Service Commissioner testified before Congress in mid-1973 in support of appliance efficiency standards, even before the energy crisis, and New York State began to adopt its own appliance efficiency standards in 1976.

At the national level, the Ford Administration in response to the embargo began to take active steps to develop energy policy for presentation in the 1975 State of the Union Address. The Federal Energy Administration staff worked with state officials to provide a framework for a broad national energy policy. This policy, announced by President Gerald Ford, led to an executive order and ultimately, the Energy Policy and Conservation Act of 1975.

This act selected some dozen key residential appliances and proposed industry-wide voluntary targets for energy efficiency improvement. If industry could not meet these voluntary targets, which averaged a 20 percent reduction in energy use compared to then-current figures, mandatory standards would be established. The legislation also required the establishment of nationally uniform test procedures for appliances, and mandatory labeling of the energy performance results obtained from the tests.

But before the effectiveness of the voluntary program could be determined other important events intervened. First, states began to adopt efficiency regulations of their own. The state proceedings generated considerable controversy, with manufacturers uniformly opposing the energy standards.

The prospect of a patchwork of state standards became a cause of great concern to manufacturers. When President Jimmy Carter was elected, he proposed that mandatory standards be set by the Department of Energy (DOE) to replace the voluntary efficiency targets. While opposing mandatory standards at the federal level, manufacturers acquiesced to the National Energy Conservation and Policy Act (NECPA) of 1978, which required DOE to set appliance efficiency standards for residential products. In return, manufacturers were able to obtain a requirement for DOE to evaluate the impacts on manufacturers that standards would impose and to consider them in setting the standards.

In addition, manufacturers obtained language allowing federal standards to preempt state efforts in most cases. Pursuant to this legislation, the Carter Administration proposed appliance standards in mid-1980, but was unable to issue a final rule before the Reagan Administration took over.

By the end of the 1970s, important new trends began to be manifest. First, one appliance manufacturer, Carrier, changed its initial advocacy position from opposition to appliance standards to support of those standards, because it found that the higher efficiency and higher profitability products that it would have preferred to sell were not doing well in states without standards, but were selling well in states with standards.

In addition, public interest organizations began to be active participants in the efficiency standards debates. Organizations such as the Natural Resources Defense Council and the American Council for an Energy-Efficient Economy were participating in regulatory and legislative proceedings on appliance efficiency standards and policies, supporting stronger standards than those proposed by state or federal officials.

Another important trend that had become established by 1980 was governmental support, primarily at the federal level but also in several states, of research and development efforts to improve end use energy efficiency. These efforts, funded primarily by the Department of Energy at the federal level, and the New York State Energy Research and Development Authority at the state level, but also by other organizations, focused both on expanding the technological opportunities for efficiency and on analyzing the markets for energy–efficient products and services, and market barriers that were impeding progress. Already by 1980, the intellectual basis of technologies that would prove very important commercially over the next twenty years had been initiated. Products such as low-emissivity coatings for windows, compact fluorescent lamps, electronic ballasts for fluorescent lamps, and high-efficiency compressors for refrigerators, are some of the noteworthy examples.

While initial policy efforts had focused at direct users of energy in the residential sector, increasing analysis identified other opportunities for efficiency, which began to be the target of standards efforts and later utility programs. State standards began to be established for shower heads and plumbing fittings, as well as for ballasts powering fluorescent lamps.

By the early 1980's the federal labeling requirements had taken effect and all residential appliances were labeled with yellow "Energy Guide" stickers. Initial studies, however, suggested that theses stickers were "not particularly effective in specific purchase decisions." More recent analysis has found that despite high levels of awareness of the label, significant comprehension problems exist with it.

The movement toward federal appliance efficiency standards stalled in the 1980s as the Reagan Administration, which opposed standards from an ideological perspective, began. That administration's approach was made evident by its refusal to finalize the DOE's 1980 standards proposal, and in 1983, by the issuance of a federal rule that determined that no standards were necessary. Both the delay and the "no standard" determination were challenged by NRDC, with the support of several large states, through the courts.

While these disputes were being settled, other activities were taking place. Utilities began to offer rebates to encourage the purchase of more efficient products, focusing first on refrigerators and air conditioners. State energy offices began considering adopting their own appliance efficiency standards. In California, the Energy Commission in 1983 adopted stringent standards for refrigerators, freezers, central air conditioners, and heat pumps. Following California's lead, several other states became interested in adopting appliance efficiency standards, since there was now a state model on which they could draw. By the end of 1986, six states had adopted new standards for one or more products.

The proliferation of state standards—and the 1985 court decision overturning DOE's "no-standard" stance—led manufacturers to accept federal standards. In 1986, the appliance industry offered to negotiate with NRDC, seeking a compromise that would adopt national appliance efficiency standards, but provide enhanced federal preemption of state efforts. NRDC, working directly with state energy offices, utilities, and other environmental and consumer organizations, reached an agreement with manufacturers over legislation that would adopt specific efficiency regulations explicitly in the law, and provide a schedule according to which DOE would consider increasingly stringent regulation in the future. The legislation passed Congress rapidly and overwhelmingly. It is known as the National Appliances Energy Conservation Act (NAECA) of 1987.

States also began to look at new products. Massachusetts promulgated legislation requiring its state energy office to set standards for fluorescent and incandescent lamps, and introduced legislation requiring standards for electric motors. Transformers were later added to the Massachusetts list.

Another forum for advancing the efficiency of appliances used in the commercial sector was the American Society of Heating, Refrigerating, and Air Conditioning Engineers (ASHRAE), which had first issued voluntary standards for energy efficiency of commercial buildings, including the efficiency of equipment installed in those buildings, in 1975, responding (albeit slowly) to the Northeast blackout. The ASHRAE/EIS standards, issued in 1989, became the basis for negotiations on national standards for commercial-sized heating, cooling, and water heating equipment, which were eventually incorporated into national law in the Energy Policy Act of 1992 (EPAct).

EPAct also nationalized state efforts that had gotten underway to regulate water consumption of toilets, plumbing fittings, and showerheads. The faucet and showerhead standards reduce hot water use, thereby saving energy for consumers. The toilet efficiency standards do not save energy in the home, but they do reduce municipal expenditures for water supply, disposal, and clean-up.

The late-1980s also saw the initiation of more widespread programs operated by utilities to promote energy–efficient appliances in the marketplace. Many of these focused on reducing peak loads by offering rebates for more efficient air conditioners and chillers in both residential and commercial applications.

As the 1990s began, a new concept in voluntary programs, market transformations, was developed and implemented for an increasing number of product. This approach invokes a strategic intervention in the marketplace, intended to produce significant improvements in energy efficiency features, generally through the introduction of new technologies, structured in such a way as to promote long-lasting effects.

Some of the first market transformation programs in the United States included the Energy Star program for computers and related equipment, established by the Environmental Protection Agency, which set a voluntary specification for efficiency based on the technologies used in laptop computers. In a laptop, the hard drive and display screens power down after brief periods of inactivity to low stand-by levels in order to conserve battery life. These same features were to be incorporated in desktop computers to qualify for an Energy Star label. The cost of the improvements was very low, and the Energy Star specification gained very high market penetration relatively quickly.

At the same time, a consortium of utilities working with public interest organizations, state energy offices, and the federal government, organized the Super Efficiency Refrigerator Program, a competitive challenge to refrigerator manufacturers to produce a product that saved at least 25 percent in energy use while eliminating chlorofluorocarbons. The winning manufacturer—the one that bid the lowest life cycle cost product—could receive up to $30 million obtained from utility subscriptions.

The progress of these programs, and encouraging results from similar programs in Sweden, led to an increasing number of market transformation programs during the 1990s.

Market transformation involves coordination between large numbers of market players. It recognizes that because of returns to scale in manufacturing, uniform specifications are necessary to create the climate necessary to make investment in energy efficiency improvements. This coordination was developed at the regional level through informal collaboration in California; the Northwest Energy Efficiency Alliance in Oregon, Washington, Montana, and Idaho; the Northeast Energy Efficiency Partnerships in the New England area; by a joint agreement between EPA and DOE to promote the Energy Star program at the national governmental level; and the nationwide Consortium for Energy Efficiency (CEE), an organization consisting of utilities, public interest organizations, and state energy offices.

Currently, CEE offers programs for air conditioners, heat pumps, clothes washers, dishwashers, refrigerators, industrial motors and drives, gas furnaces and boilers, lamps, light fixtures, and transformers.

In the early part of the 1990s deliberations on appliance efficiency standards appeared to be heading toward greater consensus. Manufacturers, efficiency advocates, and states joined together to discuss a negotiated joint proposal for the second DOE revision under NAECA of refrigerator standards, which was to be issued in 1995. All major parties submitted a joint proposal to DOE in late 1994.

But an ideological shift in Congress disrupted this process. In the 104th Congress, industrial opponents of appliance efficiency standards found sympathetic support, and passed a one-year moratorium on appliance efficiency standards in 1995. The moratorium held back DOE efforts on appliance standards for nearly two years. The refrigerator standard that was to be issued early in 1995 was delayed until 1997, and the effectiveness date set back three years until 2001. Progress toward new standards on ballasts, water heaters, air conditioners, clothes washers, and other products was delayed.

As of mid-2000, DOE was actively pursuing revised standards on clothes washers, water heaters, and residential central air conditioners, in addition to fluorescent lighting ballasts. These choices were the result of a formal prioritization proceeding, in which DOE, with stakeholder input, decided to concentrate its budgetary resources on the products that had the potential for the largest energy savings and economic value. Standards on fluorescent lamp ballasts and clothes washers had become a foregone conclusion as energy advocates and industry representatives reached negotiated agreements on joint support of new standards that would be promulgated by DOE in 2000.

The benefits from appliance efficiency improvements have been substantial to date. As of mid-2000,

	1970-1975 Energy Use		*2000-2001 Energy Use*	
Appliance	*Average*	*Best*	*Average*	*Best*
Refrigerator	1,725 kWh/yr	1,325 kWh/yr	Standard = 475 kWh/yr	
Clothes Washer	3.81 kWh/cycle		~2 kWh/cycle	0.7 kWh/cycle
Central Air Conditioners	7 EER	9.5 EER	9.8 EER	13EER
Dishwashers	4.2 kWh/cycle		~2 kWh/cycle	~1 kWh/cycle

Table 2.
Examples of Efficiency Changes

total projected energy savings from current standards by the year 2015, when most of the stock of appliances will have turned over, is 3.5 quads (a quad is a unit of thermal energy equal to one quadrillion Btu's) of primary energy per year, almost 4 percent of total U.S. energy used for all purposes. The net economic savings from these standards exceeds $175 billion.

The overwhelming bulk of these savings is due to appliance and equipment efficiency standards, although the synergistic relationship between standards and voluntary programs makes any assignment of credit somewhat arbitrary. Savings could be considered significantly larger if the base case against which it is being compared allows efficiencies to declines. Although this phenomenon has occurred several times, it is virtually never present in the economic models used to evaluate standards.

CURRENT STATUS

Despite considerable gains in the energy efficiency of most appliances over the past thirty years, substantial additional savings remain feasible and cost effective. A study by ACEEE and NRDC estimated that an aggressive set of new appliance efficiency standards could save one and one-half quads of energy, or thirty metric tons carbon equivalent (MTCE)of greenhouse pollution by 2010; appliance turnover would raise the savings from the same standards to over three quads and almost sixty MTCE by the year 2020. An unpredictable fraction of this potential can or will be achieved by other efficiency programs.

Additional savings from next generation standards or from additional products are likely. For virtually all products that have had efficiency regulations, new technologies became available following the implementation of the standards that were not available before, and in some cases were not even foreseeable. Some of this progress is illustrated in Table 2.

These savings can be complemented by additional efficiency improvements brought forth by market transformation programs. For products such as room air conditioners, dishwashers, and residential lighting systems, the federal government, through the Energy Star program, and the Consortium for Energy Efficiency have issued specifications for higher levels of efficiency than required by standards. These specifications can be promoted in the market through utility-paid incentives, tax credits (such as those currently available in the state of Oregon), and the provision of information and marketing support.

Such programs are underway at the regional, national level, or international level, developed by government agencies, by utilities, or by non-governmental organization.

For distribution transformers, a voluntary standard assembled by the trade association NEMA (National Electrical Manufacturers Association) achieves substantial savings, particularly in standby energy, with a payback period typically of three years. The loss rate from the transformers is small, but the throughput of electricity accounts for a large fraction of total energy use in buildings, particularly when utility-owned transformers are considered.

A number of new technology promotion options are being explored in the lighting area. Work is underway on market transformation programs based on bulk procurement for improved efficacy incandescent light bulbs and for compact fluorescent lamps and fixtures. For incandescent lamps, adaptations of the infrared reflective lamp coating that is already in use on reflector bulbs are encouraged by the EPAct requirement.

The Department of Energy is issuing an Energy Star specification for compact fluorescent screw-in lamps, and EPA is revising its specification for energy efficiency (effectively compact fluorescent). These Energy Star specifications can serve as the base for marketing efforts and utility incentive programs. EPA recently issues an Energy Star specification for electronic equipment calling for greatly reduced standby losses in the small AC/DC transformers; Lawrence Berkeley National Laboratory has worked on developing a one-watt standby loss specification for such transformers worldwide.

Given the rapid growth of low voltage portable electronic equipment, including cellular phones, tape recorders and CD players, VCRs, televisions and associated equipment, and other products using rechargeable batteries. Considering the ever growing number of products that operate on standby, the potential savings from such a standard and other products using rechargeable batteries, the potential savings from such a standard could be quite large.

While tremendous progress has been made towards improving the efficiency of home refrigerators, a large number of commercial refrigeration units, both in the form of refrigerated beverage vending machines and in the form of supermarket-style refrigerators, use technologies essentially unchanged in decades. These products consume several times the energy use per unit volume of refrigerated storage compared to residential units. Work is underway for an Energy Star specification for these products as well as buyer group-based programs that will encourage the highest currently feasible efficiencies.

An additional potential for significant and growing energy savings is in the consumer electronics area. The explosive growth of laptop computers has demonstrated the feasibility of products that use about an order of magnitude less energy than their desktop counterparts, even when they are fully on. The rapid proliferation of video display screens in both home and office environments makes this a particularly attractive area for technology development and for policy encouragement of technological development aimed at energy efficiency.

In summary, there is no evidence to suggest that we have reached the point of diminishing returns in improving appliance efficiency in the United States. Instead, each step forward for a given product tends to reveal additional measures not previously analyzed that could reduce energy use even further, often while improving the quality of the energy service provided. The primary constraint on progress appears to be the intellectual effort of identifying the opportunities and developing programs directed at pursuing them.

David Goldstein

BIBLIOGRAPHY

Blumstein, C.; Kreig, B.; Schipper, L.; and York, C. (1980). "Overcoming Social and Institutional Barriers to Energy Efficiency." *Energy* 5(4):355–72.

Blumstein, C., and Harris, J. (1993). "The Cost of Energy Efficiency." Letter to the Editor. *Science*. 261.

Geller, H. S., and Goldstein, D. B. (1999). "Equipment Efficiency Standards: Mitigating Global Climate Change at a Profit." *Physics & Society*. 28(2).

Greening, L. A.; Sanstad, A. H.; and McMahon, J. E. (1997). "Effects of Appliance Standards on Product Price and Attributes: An Hedonic Pricing Model." *Journal of Regulatory Economics*, 11(2):181–194.

Greening, L. A., et al. (1996). "Retrospective Analysis of National Energy-Efficiency Standards for Refrigerators." *Proceedings of the 1996 ACEEE Summer Study on Energy Efficiency in Buildings*. Washington, DC: American Council for an Energy-Efficient Economy.

Goldstein, D. B. (1983). "Refrigerator Reform: Guidelines for Energy Gluttons." *Technology Review*. February/March.

Golove, W. H., and Eto, J. H. (1996). *Market Barriers to Energy Efficiency: A Critical Reappraisal of the Rationale for Public Policies to Promote Energy Efficiency*. Lawrence Berkeley National Laboratory. LBNL-38059.

Howarth, R., and Andersson B. (1993). "Market Barriers to Energy Efficiency." *Energy Economics* 15(4).

Hinnels, M., and McMahon, J. E. (1997) "Stakeholders and Market Transformation: An Integrated Analysis of Costs and Benefits." *Proceeding of the 1997 ACEEE Summer Study on Energy Efficiency in Buildings*. Washington, DC: American Council for an Energy-Efficient Economy.

Jaffe, A., and Stavins, R. (1994). "The Energy Efficiency Gap: What Does it Mean?" *Energy Policy* 22(10):804–810.

Koomey, J. G.; Mahler, S. A.; Webber C. A.; and McMahon, J. E. (1997). *Projected Regional Impacts of Appliance Efficiency Standards for the U.S. Residential Sector*." Lawrence Berkeley National Laboratory. LBNL-39511.

Levine, M. D.; Hirst, E.; Koomey, J. G.; McMahon, J. E.; and Sanstad, A. H. (1994) *Energy Efficiency, Market Failures, and Government Policy*. Lawrence Berkeley National Laboratory. LBL-35376.

Lovins, A. (1976). "Energy Strategy: The Road Not Taken?" *Foreign Affairs*. 55(1):65–96

Lovins, A. (1992). *Energy-Efficient Buildings: Institutional Barriers and Opportunities*. Boulder, CO: E-Source, Inc.

McMahon, J. E.; Pickle, S.; Turiel, I.; Chan, P.; Chan, T.; and Webber, C. (1996) "Assessing Federal Appliance and Lighting Performance Standards." *Proceedings of the 1996 ACEEE Summer Study on Energy Efficiency in Buildings*, 9:159–165. Washington, DC: American Council for an Energy-Efficient Economy.

Meier, A., ed. (1997). *Energy and Buildings: A Special Issue devoted to Energy Efficiency Standards for Appliances* 26.

Nadel, S., and Goldstein, D. (1996). "Appliance and Equipment Efficiency Standards: History, Impacts, Current Status and Future Directions." *Proceedings of the 1996 ACEEE Summer Study on Energy Efficiency in Buildings,* 2.163–2.172. Washington, DC: American Council for an Energy-Efficient Economy.

Sanstad, A., and Howarth, R. (1994). "'Normal Markets, Market Imperfections, and Energy Efficiency." *Energy Policy* 22(10):811–818.

Sutherland, R. (1991). "Market Barriers to Energy-Efficiency Investments." *Energy Journal* 12(3):15–34.

U.S. Department of Energy. (1995). *Technical Support Document: Energy Efficiency Standards for Consumer Products: Refrigerators, Refrigerator-Freezers, & Freezers.* DOE/EE-0064. Washington DC: Department of Energy.

ARC LAMP

See: Lighting

ARCHIMEDES (287 B.C.E.–212 B.C.E.)

Archimedes was a native of Syracuse, Sicily, the son of the astronomer Pheidias. The many achievements accredited to him include: showing that the value of π lies between the values 3 10/71 and 3 1/7 (this he obtained by circumscribing and inscribing a circle with regular polygons having 96 sides); showing that the problem of squaring the circle and rectifying its circumference were equivalent; developing a number system based on powers of myriad (10,000) to deal with large numbers; and establishing methods for finding the area under a parabola, a result that needed the integral calculus of Gottfried von Leibnitz and Isaac Newton by 2,000 years. His name is also attached to many fundamental ideas in hydrostatics and the use of levers.

Little is known of his early life other than that he studied in Alexandria and became friends with Conon, with whom he corresponded for many years. This correspondence is the source of much that is known of Archimedes mathematics. A good deal of his work survived only in Arabic translations of the Greek originals, and was not translated into Latin until 1543. Perhaps due to the high regard contemporaries had for his geometrical work, much of it survived. It was standard reading for scholars into the late seventeenth century, and would have been read by Leibnitz and Newton.

It is thought that Archimedes had a lower regard for his mechanical work; however, this is difficult to validate because few writings about his mechanical devices remain. Archimedes used mechanics as a tool to think about abstract problems, rather than as a field of study itself. Contemporaries such as Plato frowned upon such a link between geometry and mechanics; they considered it as a corruption of the purity of geometry.

Despite his preference for pure geometry, Archimedes was not adverse to dramatic demonstrations of his discoveries of force-enhancing devices such as levers. Reports tell that he was able to manipulate a fully laden ship single-handed, using a series of levers and pulleys, after which he is said to have exclaimed, "Give me a place to stand on and I will move the earth." Applications of these ideas were exploited by Hieron II in the Punic wars when Marcellus, a Roman General, attacked Syracuse in 214 B.C.E. Marcellus's pride and joy was a primitive siege engine mounted on eight galleys lashed together, but Archimedes built a variety of far more advanced machines to defeat him. These included catapults that could launch massive stones to crash down on the fleet and sink Marcellus's galleys, and other devices, using systems of levers and counterweights, capable of lifting an entire galley until it was upright on its stern, and then plunging it to the bottom.

Archimedes' mechanical skill, together with his theoretical knowledge, enabled him to construct many ingenious machines. During his time in Egypt, he invented a hand-cranked manual pump, known as Archimedes' screw, that is still used in many parts of the world. Its open structure is capable of lifting fluids even if they contain large amounts of debris.

Archimedes' fascination with geometry was beautifully described by Plutarch:

> Oftimes Archimedes' servants got him against his will to the baths, to wash and anoint him; and yet

> being there, he would ever be drawing out of the geometrical figures, even in the very embers of the chimney. And while they were anointing of him with oils and sweet savours, with his fingers he drew lines upon his naked body; so far was he taken from himself, and brought into ecstasy or trance, with the delight he had in the study of geometry. Archimedes discovered fundamental theorems concerning the center of gravity of plane figures and solids. His most famous theorem, called Archimedes' Principle, gives the weight of a body immersed in a liquid.

The reference to Archimedes' Principle is in connection with another problem posed by Hieron II. The story tells of how Hieron, suspecting that a disreputable jeweler had adulterated a gold crown with silver, asked Archimedes to determine whether the crown was pure gold or not. Legend has it that Archimedes discovered a solution while in his bath, yelled "Eureka-I've found it" and ran off to the palace, neglecting to dress first! It is not know whether the goldsmith was guilty, but for the sake of the story it is usually assumed that he was.

The result that Archimedes discovered was the first law of hydrostatics, better known as Archimedes' Principle. Archimedes studied fluids at rest, *hydrostatics*, and it was nearly 2,000 years before Daniel Bernoulli took the next step when he combined Archimedes' idea of pressure with Newton's laws of motion to develop the subject of fluid dynamics.

As enigmatics Archimedes was in life, he is perhaps better remembered for his death. An account is given by Livy (59 B.C.E.–17 C.E.) *History of Rome from its Foundation, Book XXV*. It tells how Archimedes, while intent on figures that he had traced in the dust, and regardless of the hideous uproar of an army let loose to ravage and despoil a captured city, was killed by a soldier who did not know who he was. Another version, by Plutarch, recounts that Archimedes was intent on working out some problem by a diagram, and having fixed both his mind and eyes upon the subject of his speculation, noticed neither the entry of the Romans nor that the city was taken. A soldier unexpectedly came up to him and commanded that Archimedes accompany him. When he declined to do this before he had finished his problem, the enraged soldier drew his sword and ran him through. Yet a third account by John Tzetzes in the twelfth century *Book of Histories (Chiliades), Book II*, tells a similar story with a slight twist. It says that when Archimedes refused to stand clear of one of his diagrams when disturbed by a Roman soldier Archimedes cries out "Somebody give me one of my engines." The Roman, knowing that Archimedes' engines had defeated Marcellus's fleet, became frightened and slew him.

Douglas Quinney

See also: Bernoulli, Daniel.

BIBLIOGRAPHY

Bell, E. T. (1965). *Men of Mathematics*. London: Penguin

Dijksterhuis, E. J. (1987). *Archimedes*. Princeton, NJ.; Princeton University Press.

Fauvel, J., and Gray, J. (1987). *The History of Mathematics*. London: Macmillan.

Hollingdale, S. (1983). *Makers of Mathematics*. London: Pelican.

ARCHITECTURE

See: Building Design, Commercial; Building Design, Residential

ARGONNE NATIONAL LABORATORY

See: National Energy Laboratories

ATMOSPHERE

Man is able to directly utilize only a small portion of the energy of the Earth's atmosphere. Indeed, excessive concentrated energy in the atmosphere—hurricanes and tornadoes—represents risks to mankind. Most human demands created by atmospheric conditions involve consumption of energy to maintain comfort. The ambient surface air temperature, for instance, determines how much energy is needed for heating or cooling demands and the level of outside ambient illumination determines the need for artificial lighting. Electricity derived from fossil fuels powers the industrialized world Petroleum products directly power most forms of transportation. Pollution emitted by massive fossil fuel consumption affects man's well-being and quality of life on both immediate and long-term time scales, and mitigation this anthropogenic (manmade) pollution using emission-control devices requires even greater energy consumption.

ATMOSPHERIC COMPOSITION

Unpolluted air contains about 78 percent molecular nitrogen, 21 percent molecular oxygen, 1 percent argon, up to 3 percent water vapor, and a host of trace gases, including carbon dioxide, carbon monoxide, methane, nitrous oxide, helium, krypton, and radon. Oxygen is constantly released to the atmosphere by green plants during photosynthesis. Plants and animals excrete carbon dioxide during respiration. Water evaporates from the surface of the Earth and travels as a vapor through the atmosphere, eventually condensing and falling as precipitation. The atmosphere-ocean-geosphere and biosphere have maintained a natural chemical balance over many millennia, although steadily increasing anthropogenic trace gas emissions may have the potential to change this natural balance in the future.

Air pollution is produced by various natural and anthropogenic sources. Natural sources inject large amounts of particles into the atmosphere, including inorganic minerals, pollen, small seeds, bacteria, fungi, and effluvia from animals, especially insect parts. These natural particles usually have diameters greater than 10^{-5} cm. Many anthropogenic particles are continuously injected into the atmosphere, including latex and soot. Particles produced from combustion generally have diameters smaller than 10^{-5} cm. Tiny hygroscopic particulates from both natural and anthropogenic sources play an important role in the atmosphere, serving as condensation nuclei for water droplet and ice crystal formation. The period of time that particles remain in the atmosphere is influenced by their height and weight. Small particles in the stratosphere remain aloft much longer than small particles in the lower troposphere.

Polluted air often contains carbon monoxide and volatile organic carbon (VOC) gases, including ketones and aldehydes, as well as oxides of sulfur and oxides of nitrogen. Anthropogenic emissions of these gases arise from incomplete combustion and subsequent photochemical alterations in the atmosphere. Anthropogenic emissions also inject a number of relatively inert gases into the troposphere, including chlorofluorocarbons, sulfur hexafluoride, and carbon tetrachloride. Trees have been found to be a major natural source for VOCs.

Pollutants have various atmospheric residence times, with reactive gases and large aerosols being rapidly removed from air. In the London air pollution episode of December 1952, the residence time for sulfur dioxide was estimated to be five hours; daily emissions of an estimated 2,000 tons of sulfur dioxide were balanced by scavenging by fog droplets, which were rapidly deposited. Most relatively inert gases remain in the atmosphere for extended periods. Sulfur hexafluoride, used extensively in the electric power industry as an insulator in power breakers because of its inertness, has an estimated atmospheric lifetime of 3,200 years.

Emissions from fossil fuel combustion have caused increasing air pollution problems. Four major types of problems have been recognized: acid deposition and acid rain, air pollution episodes involving sulfur-rich smog from coal burning, photochemical smogs from gasoline-powered vehicles, and the threat of global warming as a result of increasing levels of carbon dioxide (a "greenhouse gas") in the atmosphere.

North American and Western European countries responded to acid rain, acid deposition, and acidic sulfurous smog episodes (which caused excess mortality and morbidity as well as greatly decreased visibility) by passing emission control laws. Sulfurous smogs are now rare in North America and Western Europe. Despite emission controls, acid rain and acid deposition are believed by many scientists to be the cause of forest decline, known as "neuartige Waldschäden" in Europe. This forest decline has been detected throughout Central Europe at all elevations and on all soil types. Evidence suggests that nitrates in acid rain play an important role in Waldschäden. Asian countries continue to burn fuel with few emission controls. A few tragic consequences include erosion of the Taj Mahal by acidic air pollutants, and occasional snoot-laden smog blanketing the Indian Ocean north of the equator.

Photochemical smogs arise worldwide because of the action of sunlight on emissions from gasoline-powered vehicles. Decreased visibility, increased morbidity, and crop damage as a result of photochemical smogs led to introduction of the catalytic converter on automobiles in the United States. This has had only a small impact on the occurrence of photochemical smogs in the United States.

Global warming has attracted growing worldwide concern, leading to the Kyoto Accords of 1997, which agreed that rich industrial nations would reduce greenhouse gas emissions. Legally binding reductions for each major greenhouse gas were set, with emphasis on reducing carbon dioxide emissions by 2008–2012. The Kyoto Accords were signed by President Clinton in 1998, with a carbon dioxide emissions reduction objective of 7 percent for the United States, although the U.S. Senate failed to ratify them. China, Brazil, India, and Mexico were among nations exempted from the Kyoto Accords. Canada (with an emissions reduction objective of 6%) and the European Union (with an emissions reduction objective of 8%) have developed (and implemented) some strategies to reduce carbon dioxide emissions. Norway has begun a program to sequester carbon in the ocean.

The concentration of chlorofluorocarbons in the atmosphere has been steadily increasing since they began being manufactured. It has been discovered that chlorofluorocarbons are slowly destroyed by chemical reactions in the stratosphere, especially heterogeneous reactions in polar stratospheric clouds above Antarctica. The chlorine released during these reactions in turn destroys stratospheric ozone, the most prominent result being the creation of the infamous “ozone hole,” a zone with greatly diminished stratospheric ozone centered over Antarctica during winter. This ozone depletion occurs in the winter and early spring—when the sun’s radiation strikes the Antarctic stratosphere. Ozone levels recover

Recognition of the threat of stratospheric ozone depletion posed by chlorofluorocarbons and chlorofluorohydrocarbons led 131 countries to sign the Montreal Protocol in 1987. Production of chlorofluorocarbons was banned as of January 1, 1996, because of their potential to further deplete stratospheric ozone. Chlorofluorohydrocarbons will be phased out of production by 2030; HCFC-22 will be phased out by 2020. However, large amounts of chlorofluorocarbon refrigerants produced over many decades remain in use worldwide, awaiting future release.

EARTH’S RADIATION BALANCE

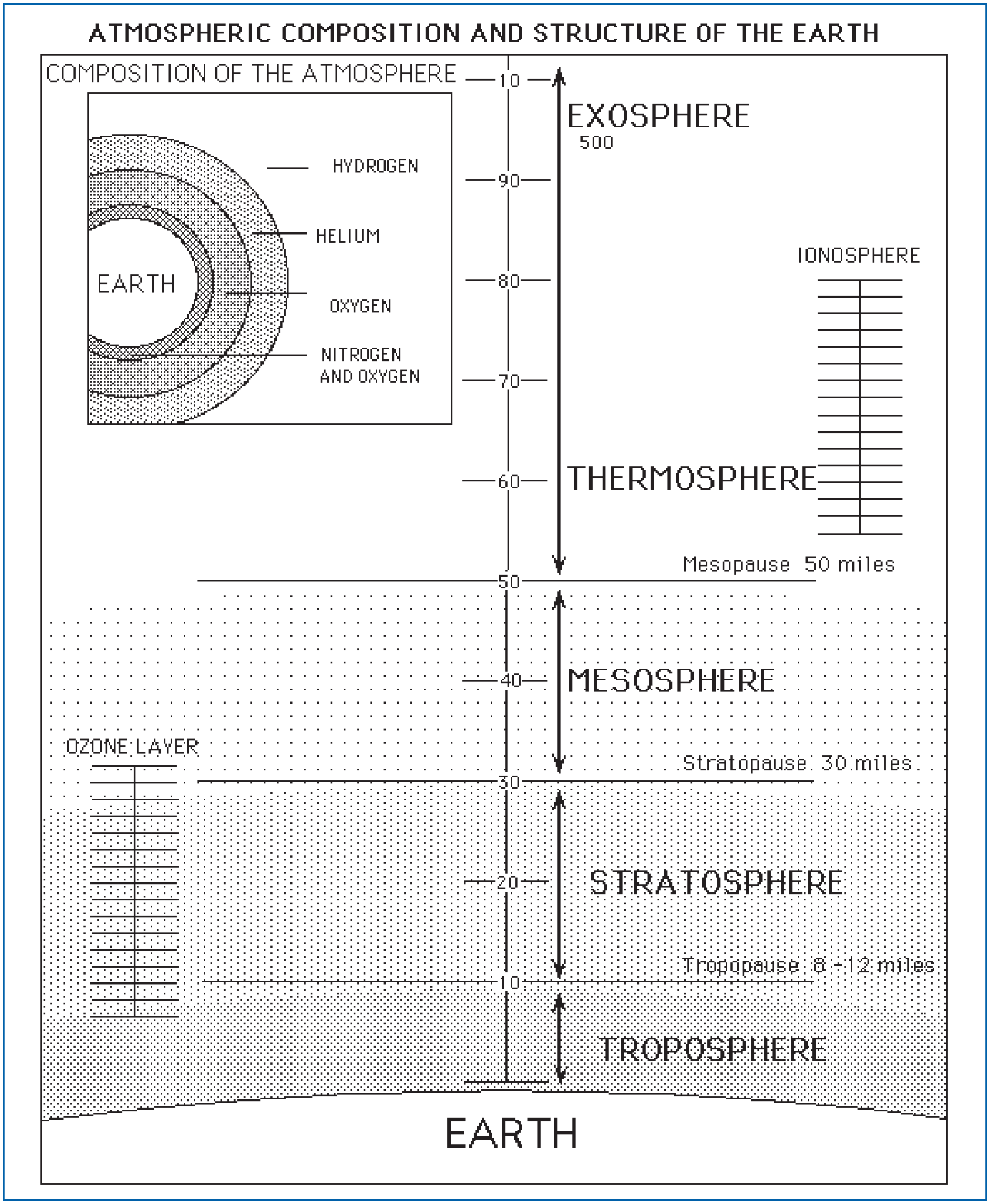
Solar radiation continually reaches the earth, warming the atmosphere, ocean, and land surfaces on the sunlit portions of the planet. Although the sun emits a continuous spectrum of electromagnetic energy, its peak emissions are in the visible wavelengths, with a maximum at 500 nm wavelength. The average amount of solar energy received globally at the top of the atmosphere is relatively constant, about 1,353 W/m^2. Clouds and particles reflect some incident solar radiation back into space. Some large volcanic eruptions inject copious numbers of particles, which attenuate solar radiation reaching Earth’s surface. When a volcanic eruption injects large amounts of sulfur into the stratosphere, sulfuric acid aerosols slowly form in the stratosphere, where they remain for months; these aerosols also reflect incident solar radiation.

About 51 percent of solar energy incident at the top of the atmosphere reaches Earth’s surface. Energetic solar ultraviolet radiation affects the chemistry of the atmosphere, especially the stratosphere where, through a series of photochemical reactions, it is responsible for the creation of ozone (O_3). Ozone in the stratosphere absorbs most of the short-wave solar ultraviolet (UV) radiation, and some long-wave infrared radiation. Water vapor and carbon dioxide in the troposphere also absorb infrared radiation.

Considerable energy is radiated back from Earth’s surface into space as long-wave infrared radiation. The atmosphere absorbs some of this infrared radiation, preventing its loss to space. This trapping is sometimes referred to as “the Greenhouse Effect.”

THE HYDROLOGICAL CYCLE

Water is constantly evaporated from rivers, lakes, and oceans, and released from vegetation through evapotranspiration. Water vapor travels through the atmosphere, eventually forming small droplets or ice crystals in clouds. Some particles grow sufficiently



Composition and structure of the atmosphere. (Gale Group)

Earth's atmosphere as seen from outer space. (National Aeronautics and Space Administration)

large, and fall as rain or snow. Most precipitation occurs over the world's oceans. Much of the rain and snow falling over continental areas rapidly runs off into major river channels, returning water to the oceans. Some snow is deposited in glaciated areas, including high mountain peaks on the continents and on the Greenland and Antarctic ice sheets, where it may remain for millennia. About 75 percent of the Earth's fresh water is currently stored in glaciers ice sheets. Calving of icebergs from the ice sheets and periodic glacial retreat in major mountain ranges return some of this long-frozen water to the ocean. During the summer of 1999, no icebergs were seen to enter North Atlantic shipping lanes, possibly because of warmer than usual ocean temperatures.

The balance between evaporation, precipitation, glaciers, and oceans, known as the hydrological cycle, is usually considered to be in rough equilibrium over the Earth, although there is evidence that the Greenland ice sheet shrank substantially during the mid-1990s. There is also evidence that the West Antarctic ice sheet thinned during the same period.

During times of global cooling and glacial advance, ocean water levels dropped as increasing amounts of water were stored in ice sheets, as was the case during the Pleistocene glaciation of Eurasia and America. At warmer times in the geological record, glaciers have had dramatic retreats, resulting in a worldwide rise in ocean levels. The change from a cold period to a warm period may occur rapidly over the course of a century. Since 1960, mid-latitude glaciers have receded dramatically. Glaciers in the Caucasus are estimated to have lost about half their mass, while the Tien Shan Mountains are estimated to have lost about 22 percent of their glacial ice between 1960 and 2000. The decade of the 1990s is believed by most scientists to have been the warmest in many millennia. It has been postulated that these changes foreshadow a more prolonged global warming period that may be partially attributable to anthropogenic alterations in atmospheric composition.

TROPICAL STORMS

Solar heating of tropical oceans warms the surface water, promoting evaporation. Where the equatorial surface waters are warmest and the northeast and southeast trade winds meet, a band of cirrostratus and cirrus clouds spreads out from convective precipitation regions. This area is known as the Intertropical Convergence Zone.

When tropical ocean surface water temperatures exceed 26°C near the edge of the Intertropical Convergence Zone, and air aloft is warm and moist, conditions are favorable for the development of large tropical cyclones. These storms begin as weak depressions or disturbances, most of which fail to develop into organized systems. When conditions favor storm development, pressures drop in the center and winds increase in a tight 30–60 km band around a central eye. Large storms are powered by the latent heat of condensation released as clouds form from the moisture-laden air.

When a tropical storm has winds in excess of 120 km/hr, it is officially classed as a hurricane. Large hurricanes have a highly organized rotary structure, with a central eye surrounded by tightly curving bands of clouds extending up to 2,000 km in diameter, although most important activity occurs within 100 km of the eye. Large hurricanes draw enormous amounts of moisture from the Intertropical Convergence Zone.

The most powerful hurricanes (called "Category 5") have sustained winds exceeding 248 km/hr. In general, hurricanes move slowly with the average wind speed of the troposphere. When these hurricanes strike land, they bring a devastating combination of high winds, torrential rain, and a storm surge. The storm surge is an uplifting of the water level resulting from an air pressure drop and wind-driven water; the most powerful hurricanes have a storm surge exceeding 18 feet (5.5 m). Hurricane Gilbert, a massive Category 5 hurricane in 1988, dominated about 20 percent of the entire global Intertropical Convergence Zone, causing the cloudiness in the zone outside the storm to dissipate. Hurricane Andrew, which devastated South Florida in 1992, was also a Category 5 hurricane.

As a hurricane travels over warm ocean water, it lowers the sea surface temperature by about 3°C in a 100 km swath. When a hurricane is stationary, this surface ocean cooling weakens the storm intensity. Hurricanes also rapidly lose strength when they move over cold water or land.

How is energy utilized in a hurricane? Hurricanes derive energy mainly from the release of latent heat, and lose energy, in part, through precipitation and frictional loss of the wind. For an average hurricane, the rate of release of latent heat has been estimated at 10^{14} watts. This is equivalent to the energy output of a 20-megaton bomb every fifteen minutes. An average hurricane with maximum winds of 50 m/s and a radius of 30 km dissipates wind energy at a rate of 3×10^{12} watts. Thus, it takes only about 3 percent of the input energy to maintain the destructive winds of an average hurricane.

THUNDERSTORMS

Thunderstorms (cumulonimbus clouds) come in many sizes and shapes, ranging from small "air-mass" thunderstorms to large "supercells." Thunderstorms are influenced by the surrounding atmosphere and nearby convective activity. Sometimes a thunderstorm is composed of a single, isolated cumulonimbus cloud. At other times, cumulonimbus clouds are so numerous that they form a continuous sheet, losing any separate identity.

The air-mass thunderstorm is the least severe of all thunderstorms. In its simplest form, an air-mass thunderstorm grows as a single cell when solar radiation heats the surface air in an unstable atmosphere. Its life cycle lasts around 30 minutes. Towering cumulus clouds are formed as in-cloud updrafts push moisture upward. The tower may reach a height about five times the diameter of the cloud base in the growth phase.

When water vapor is deep enough for continued convective activity, the thunderstorm reaches an active phase, in which the top of the cloud glaciates, often forming a distinctive anvil. Strong updraft and downdraft regions form within the cumulonimbus cloud. The change from a towering cumulus cloud to a cumulonimbus cloud is usually quite rapid as the top turns to ice, and lightning and heavy rain begin.

The final stage in the life of a cumulonimbus cloud is marked by dissipation. The lower regions of the cloud break up, while the upper anvil spreads out.

Mixing with the environment lowers vertical wind velocities by reducing the in-cloud temperatures through evaporation and mechanical mixing with the cooler surrounding air.

Most air mass thunderstorms form in groups, facilitating growth by the reduction of environmental mixing. These multicellular storms may occur as compact clusters of cells or, if there is some external organization, laterally aligned in squall lines.

Supercells have greater size, organization, and duration than air mass thunderstorms. A supercell rotates, with persistent updrafts and downdrafts, and lasts for many hours. Updrafts in supercells may exceed 140 km/hr. Supercells develop when there are large changes in wind velocity with height. Moist, warm air entering from the front side is lifted at the cold air gust front until condensation occurs, releasing latent energy. This air parcel then moves rapidly upward and, usually, out ahead of the storm at upper levels in the atmosphere. Dry air moves in from the back side of the supercell, is cooled by rain falling out of the rising air, and then descends as a downdraft. Several different arrangements of this flow are possible. Supercells frequently move slower than the mean winds aloft. These storms are notorious for their ability to spawn tornadoes; they may show a tornadic “hook echo” on radar displays. Supercells and regular cells can combine in a multicellular complex, which then exhibits some characteristics of both types of storms.

Thunderstorms arise from convective activity, driven by energy derived from the latent heat of condensation and sublimation of water vapor within cumuliform clouds. Buoyant air movements caused by surface heating, by orographic (relating to mountains) forcing, and by lifting of warm moist surface air along frontal zones, are some of the important mechanisms for initiating the upward transfer of energy.

TORNADOES

Calculations of tornado energy are difficult to make—the aftermath of a large destructive tornado sometimes resembles carpet bombing in a war situation, with buildings ripped off foundations, large numbers of trees uprooted, and asphalt stripped from roadways. Several reports describe the derailment of up to five train cars as tornadoes have apparently lifted cars off the tracks. Large building debris has been found at a distance of 20 km from its original location.

Meteorologists categorize tornadoes by their wind speeds as deduced subjectively from severity of the damage. Each tornado is given a Fujita F-scale class: F0 (light damage), 40–72 mph; F1 (moderate damage), 73–112 mph; F2 (considerable damage), 113–157 mph; F3 (severe damage), 158–206 mph; F4 (devastating damage), 207–260 mph; F5 (incredible damage), 261–318 mph. The highest reported toronto wind speed (reported from Doppler reader) was 318 mph in the F5 Oklahoma City tornado of May 3, 1999.

The F-scale classification is only a first approximation to tornado damage. Some buildings are wind-sensitive while others are wind-resistant. The lower pressure of the tornado core also weakens the integrity of the building. Mobile homes, wood-frame houses, buildings with sheet metal roofs, and those with unreinforced masonry walls are particularly sensitive, often damaged by winds less than 100 mph. In rural counties without building codes, wood-frame houses using nails to anchor walls to foundations can be blown off and destroyed by 80-mph winds. Structurally engineered buildings are seldom destroyed or even severely damaged. People are most often injured and killed by falling building materials and by projectiles in the debris suspended in the tornado.

Tornadoes form in several ways. The most common tornadoes form at the edge of thunderstorm cold air outflow, and they are called gustnadoes. Gustnadoes fall into the F0 or F1 class and only rarely inflict intense damage along a short, narrow path. Waterspouts and landspouts form in areas where a pre-existing surface circulation becomes entrained, stretched, and intensified as a thunderstorm updraft passes over. Waterspouts and landspouts may attain in F2 class, and several have been reported to inflict moderate damage to marinas or poorly constructed buiildings. The least frequent and most severe tornadoes form and descend from supercell thunderstorms, which may persist for many hours and spawn multiple tornadoes. However, the most severe tornadoes, although less frequent, are those that descend from supercells. Supercells may persist for many hours. A single supercell, moving over several hundred kilometers, has been observed to spawn a series or “family” of up to eight tornadoes along its route. These tornadoes are associated with rotating circulations called mesocyclones. The mesocyclone is 10 to 20 km in diameter, much bigger than a single tornado. It can sometimes be seen as a larger-scale rotation of the clouds at the bottom of the supercell. Rotation begins at an altitude between 4 km and 8 km and builds downward. Sometimes a mesocyclone produces more than one tornado at a time.

An average of 800 tornadoes is reported within the United States yearly, with possibly 2,000 small tornadoes going unreported. Tornadoes have been reported in every state, including Hawaii and Alaska. The Great Plains has the highest occurrence of damaging tornadoes. Occasionally, tornadoes occur in outbreaks. The super-outbreak of April 2 and 4, 1974, had 148 reported tornadoes in 13 states. Hurricanes can spawn tornadoes; in 1967, Hurricane Beulah generated 115 reported tornadoes. The majority of tornadoes occur in late afternoon and evening. However, tornadoes can form at any time of day or night. Nocturnal tornadoes are relatively common on the U.S. Gulf Coast.

Television news reporting gives the impression that people incur substantial risk in tornado-prone regions, but the likelihood of any particular building being hit is on the order of once every million years.

LIGHTNING

Within cumulonimbus clouds, precipitation processes and ambient physical conditions interact to produce regions of high electrical charge. The mechanisms by which charge separation occurs in cumulonimbus clouds are poorly understood by cloud physicists. Some researchers believe that electrical charges build in strength when ice pellets fall through a region of ice crystals and water droplets.

Lightning is the visible manifestation of a plasma channel. The plasma is very hot, with peak temperatures greater than 30,000°C, compared to 6,000°C for the sun. Although the peak current in a lightning stroke may be as high as 100 kiloamperes, charge transfer is limited by the brief duration of the flash. Movement within the plasma is limited; a typical electron in the lightning channel may move only two meters. Most of the charge transfer occurs by way of a continuing current between the strokes comprising the flash, and by relatively low amplitude currents following strokes. Usually lightning transfers negative charge to the ground. However, positively charged cloud-to-ground lightning also occurs.

After electrical potentials on the order of 300 to 400 kV/m are produced in discrete regions within the cloud, streamers extend their way forward along the cloud's charge gradient in a tree-like structure. When electrical potentials on the order of 1,000 kV/m develop, streamers become self-propagating. A plasma channel then moves toward regions of opposite charge within the cloud, neutralizing much of the electric charge within the cloud as it travels through diffusely charged regions. As the channel tip advances, it may branch in several directions simultaneously. If it penetrates into highly charged a region, a recoil streamer may flow along the channel to the initiating region.

About 80 percent of lightning channels begin and end in the cloud. The remaining 20 percent of streamers extend horizontally into the clear air outside the cloud. They propagate in a stepwise fashion called step leaders. Discharges ending in the clear air are usually highly branched, and generally quite weak. When a step leader approaches the earth's surface, an upward streamer propagates from the ground toward the channel tip aloft. These plasma channels intersect at an altitude of about 100 m above the ground. Completion of the circuit causes an upward rush of electrons called a return stroke, substantially increasing the brightness of the luminous plasma channel. Frequently, a second pulse of energy, the dart leader, moves smoothly down from the cloud, following the same path to the ground. Return strokes may follow the dart leader. Typically, cloud-to-ground flashes have four or more separate strokes.

Capturing electricity from a stray lightning flash is an intriguing but impractical idea. Presumably, Benjamin Franklin, in his famous kite experiment, transferred energy from a lightning flash to a Leyden jar, a primitive type of battery. A typical lightning flash has 25 coulombs of charge and 30,000 instantaneous amps. However, the stroke is very brief; 0.01 seconds. This is only enough energy to power one 100-watt bulb for a few months. One hundred thousand 1,000-ft towers would be needed to capture lightning energy equivalent to the output of a typical small power station.

Very powerful lightning discharges, known as superbolts, are about 100 times more powerful than the typical lightning stroke. Superbolts are most common in the wintertime off the coasts of Japan and the eastern United States. The radius of a superbolt channel is estimated to be 20 cm, compared to 2 cm of a typical lightning stroke. Because superbolts are thought to be rare over land, tapping energy from them is even more problematic than obtaining energy from a regular lightning strike.

High-altitude discharges above active thunderstorms have been studied with cameras that sense very low light levels. Several distinct, differently colored phenomena have been identified. Unusual cloud-to-air discharges from the anvil top upward to heights of 35 to 50 km (into the stratosphere) are called "blue jets." Blue jets propagate upward at about 100 km/sec in a narrow conical or trumpet shape. Red discharges extending 50 to 95 km upward above thunderstorm anvils are called "sprites." Sprites have widths ranging from 10 km to 50 km, and have been observed only above large (30,000 km^2) multicellular thunderstorms. "Elves" are very brief (1 msec) red halos that form at altitudes of 60 to 100 km.

WINDS

Winds arise through a complex interplay of forces. As Earth rotates around its axis every 24 hours, the atmosphere moves along with the earth. In the troposphere, large-scale weather systems, covering regions of around two million square kilometers, form an interlocking grid pattern over the globe. The growth and decay of these large systems produces day-to-day changes in weather conditions around the world. Large-scale weather systems develop quickly; they may double in intensity in a period of 12 to 48 hours. Once formed, these systems decay slowly, generally halving in intensity in four days.

Temperature swings following frontal passages are common in North America and Eurasia, but are rare in the tropics, where differences in cloudiness and precipitation arise from seasonal variability in thermal forcing. Poleward from the tropical regions, extratropical cyclones transform latitudinal temperature gradients into kinetic energy.

In coastal areas, temperature differences between the land and the water produce air pressure variations, creating sea and lake breezes that are superimposed on the normal winds. These winds vary diurnally and as a function of cloudiness. During the daytime, winds blow from the cool sea toward the warm land, while at night the land becomes cooler than the sea surface, and the winds blow from land to sea.

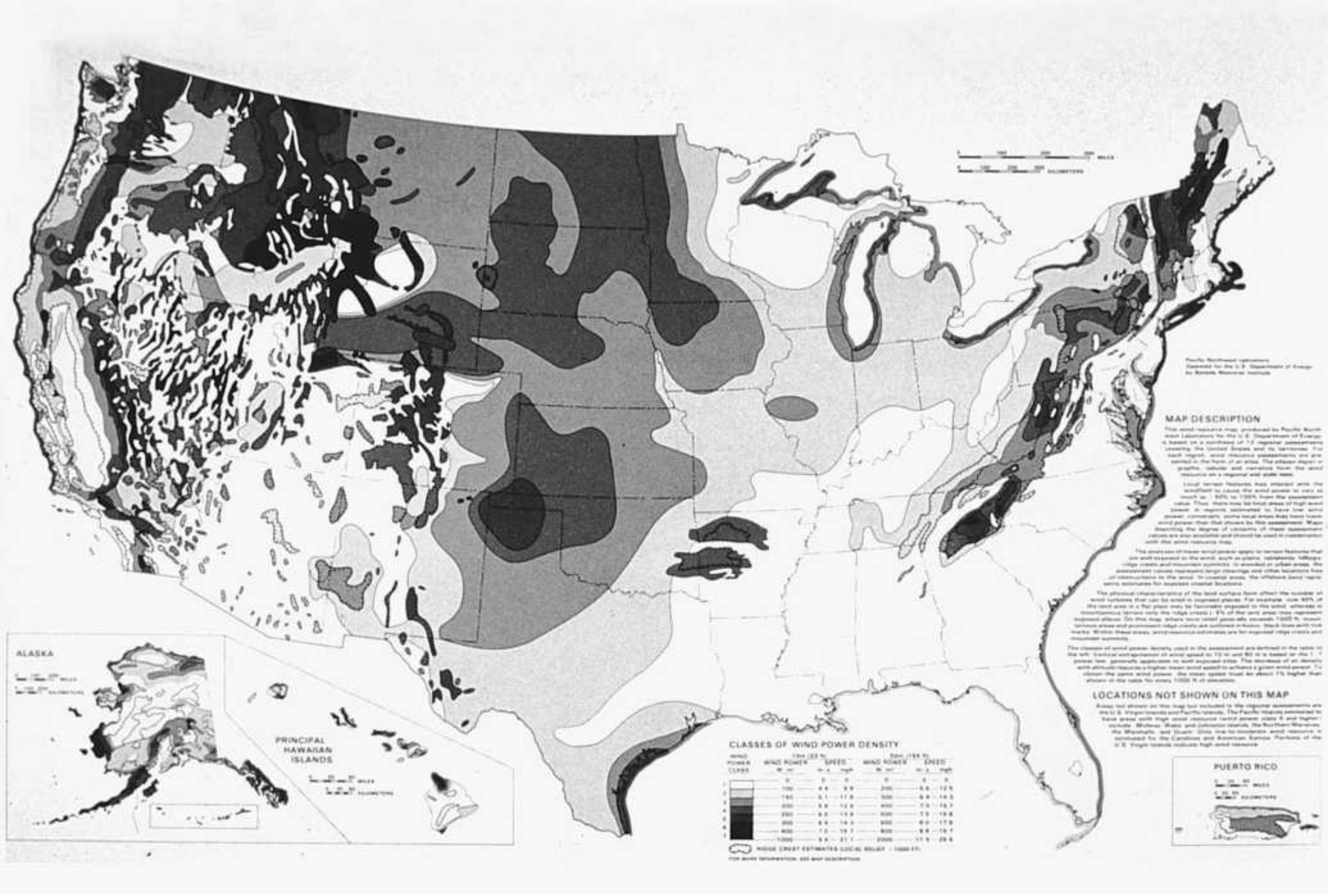
On a larger scale, continents produce flows, known as "monsoon winds," over wide areas between the surrounding seas and lands. These winds respond to seasonal forcing. The best example is the Indian monsoon. During the summer monsoon, from June through September, moist winds blow northward off the Indian Ocean. Convergence of this moist air with other air masses produces intense precipitation. The monsoon slowly moves northward in spring and summer, traveling about 5 km/day. From December through February, the Siberian high dominates Eurasian air circulation, and the general flow of air is reversed, with cold, dry air traveling from the continental land masses southward over the warmer surface waters of the Indian Ocean.

Topography can substantially change air flow. Local mountain winds form when surface heating causes winds to flow up the sides of the mountain: technically known as "anabatic flow." Anabatic winds are generally strongest in early afternoon. At night, winds flow down off hills or mountains, technically known as "katabatic flow." In hilly terrain, with slopes of about two degrees, winds on the order of 3 km/hr descend as the ground surface cools.

Mountains modify the velocity and direction of wind. The coastal mountains along western North, Central, and South America play a major role in determining regional winds on the eastern rim of the Pacific Ocean. Dynamically induced winds may attain substantial speeds in mountainous regions, sometimes exceeding 100 km/hr. Some orographic winds have been given names associated with a specific region, such as the "Santa Ana" winds that occur as dry continental air descends from the Sierra Nevada Mountains to Southern California coastal areas during spring and autumn. A strong, warm wind on the leeward side of a mountain range is called a chinook (North America) or föhn (Europe). Strong chinooks, with damaging winds reaching 160 km/hr, occur several times each winter along the Front Range of the Rocky Mountains.

WIND ENERGY EXPLOITATION

Wind turbines produce power by converting the force of the wind into torque. The power produced is a function of the wind energy flux (power), which, in turn, is a function of the air density multiplied by the wind velocity raised to the third power. Changes of air density with time at a particular site are negligible compared to the fluctuations in wind velocity. Meteorologists usually report wind speed as an average. To get the potential wind power, the average



Areas with annual average wind speeds of 13 mph or greater are found throughout the United States. Regions with Class 4 winds are considered attractive for wind turbine siting. (Developed by Battelle Pacific Northwest Laboratories for the U.S. Department of Energy)

wind speed is raised to the third power and then adjusted using a Weibull statistical distribution too account for the natural instantaneous wind variability.

Wind speed, and thus available wind power, at any given location is a function of several factors: global variations; local variations, especially around coast lines with sea or lake breezes and topography; and diurnal variations of wind speed from differences in the stability of the air next to the ground. Turbulence associated with unstable air during the afternoon, or on cloudy days, mixes higher velocity winds aloft with the winds slowed by friction at the surface. On clear nights the air is stable, and there is little transport of the high winds aloft to the ground. Thus, wind speeds near the ground are normally higher during the daytime than at night, with the highest wind speeds occurring in the afternoon, and minimum wind speeds around dusk and dawn. In general, gusts are greatest in the afternoon.

Wind speed varies with height above the ground. Because surface wind speeds are routinely measured at 10 m, winds turbine heights (usually higher than 10 m) must be estimated. The turbulence level of wind also varies. Forests, buildings, and other obstacles slow wind down, and increase turbulence levels. Long grass, shrubs, and crops can slow the wind considerably. Such variations can be corrected by use of “roughness classes” or “roughness lengths.” The sea is assigned a roughness class of zero, while a landscape with many trees and buildings has a class of three or four. (Sheep can keep the roughness down through grazing.) When no data are available for a site, a wind rose from the nearest observations may provide a rough estimate of wind speed. However, data availability frequently is sparse in areas with substantial wind generation potential.

During the late 1970s and early 1980s, there was considerable interest in harnessing wind energy in the

United States. During this time, efforts were made to determine the national wind energy potential. Maps were drawn using the "Batelle Wind Power Classes," ranking nominal wind energy at 10 m, 30 m, and 50 m elevations. These classes, which remain standard in mapping wind energy, are shown on the map. In general, Class 4 and higher winds are considered favorable for wind energy exploitation in the United States.

Because of its large population and the tacit assumption that its varied topography would be ideal for wind power exploitation, California conducted its own program to determine wind energy potential. This study demonstrated the meteorological difficulties in characterizing wind speeds in hilly terrain. Some wind turbines were constructed in areas thought to be ideal, but which proved to be quite marginal. Three California passes were identified as among the best wind energy sites in the world, with average wind speeds in excess of 8 m/s. Tehachapi and San Gorgonio have proven successful, and the Altamount Pass wind farm has over 7,500 wind turbines in operation.

Within the United States, some areas are especially suited for wind power generation, including North and South Dakota, Minnesota, Montana, Wyoming, the Front Range of the Rocky Mountains, the Cascade Mountains, the Great Lakes shoreline, and the ridge crests and peaks of the Appalachians. Close examination of specific geographical and topographical features may help wind power planners identify suitable sites. This has proven to be the case for Buffalo Ridge, a 100-km long ridge stretching from Spirit Lake, Iowa, through southwestern Minnesota north through Lake Benton to Sica Hollow, South Dakota. It has the potential to yield 3 Terawatt hours yearly.

Dennis G. Baker
Anita Baker-Blocker

See also: Acid Rain; Climatic Effects; Turbines, Wind.

BIBLIOGRAPHY

Church, C.; Burgess, D.; Doswell, C.; and Davies-Jones, R. (1993). *The Tornado: Its Structure, Dynamics, Prediction, and Hazards.* Washington, DC: American Geophysical Union.

DeHarpporte, D. (1983). *Northeast and Great Lakes Wind Atlas.* New York: Van Nostrand Reinhold.

DeHarpporte, D. (1984). *South and Southeast Wind Atlas.* New York: Van Nostrand Reinhold.

Eagleman, J. R. (1983). *Severe and Unusual Weather.* New York: Van Nostrand.

Eagleman, J. R.; Muirhead, V. U.; and Willems, N. (1975). *Thunderstorms, Tornadoes, and Building Damage.* Toronto: Lexington Books.

Emmanuel, K. A. (1999). *Weather*, Vol. 54: *The Power of a Hurricane.* Reading, UK: Royal Meteorological Society.

Gedzelman, S. D. (1980). *The Science and Wonders of the Atmosphere.* New York: John Wiley & Sons.

Gipe, P. (1995). *Wind Energy Comes of Age.* New York: John Wiley & Sons.

Jursa, A. S. (1985). *Handbook of Geophysics and the Space Environment.* Springfield VA: National Technical Information Service.

Kessler, E. (1986). *Thunderstorm Morphology and Dynamics*, 2nd ed. Norman, OK: University of Oklahoma Press.

Lane, F. W. (1965). *The Elements Rage.* Philadelphia: Chilton Books.

Linacre, E., and Geerts, B. (1997). *Climate and Weather Explained.* London: Routledge.

Lovelock, J. E. (1991). *GAIA, a New Look at Life on Earth.* New York: Oxford University Press.

Ludlum, F. H. (1980). *Clouds and Storms.* University Park: Pennsylvania State University.

Miller, K. B.; Matchett, J. M.; Purcell, C. W.; and Turner, M. H. (1995). *Strategies for Managing Ozone-Depleting Refrigerants, Confronting the Future.* Columbus, OH: Battelle Press.

Newton, C. W., and Holopainen, E. O. (1990). *Extratropical Cyclones.* Boston: American Meteorological Society.

Pacific Northwest Laboratories. (1991). *Wind Energy Resource Atlas of the United States.* Richland, WA: Author.

Pielke, R. A., Jr., and Pielke, R. A., Sr. (1997). *Hurricanes: Their Nature and Impacts on Society.* Chichester, England: John Wiley & Sons.

Uman, M A. (1984). *Lightning.* Minneola, NY: Dover.

Uman, M. A. (1986). *All About Lightning.* Minneola, NY: Dover.

Uman, M. A. (1987). *The Lightning Discharge.* Orlando, FL: Academic Press.

Viemeister, P. E. (1972). *The Nature of Lightning and How to Protect Yourself from It.* Cambridge, MA: MIT Press.

Williamson, S. J. (1973). *Fundamentals of Air Pollution.* Reading, MA: Addison-Wesley.

Yoshino, M. M. (1975). *Climate in a Small Area.* Tokyo: University of Tokyo Press.

ATOMIC BOMB

See: Nuclear Fission

AUDITING OF ENERGY USE

Energy conservation has always proceeded along two main avenues. One involves new technology—by continually improving the efficiency of appliances or the mileage of automobiles, the overall energy intensity (energy use per person) of society decreases. Both the government and private industry have spent large sums of research-and-development dollars on countless products that has lead to great improvements in the energy efficiency of products. The second major avenue is improvement of actual practice, whether at home, in a commercial building or on the factory floor. It is based on the assumption that through ignorance, poor operation, insufficient maintenance, priority conflicts or in some cases simply sloth, that energy is used less efficiently than the current state of technology allows.

One way of closing the gap between the current state of operations and what would be considered "best practice" is to formally examine energy use through energy audit. Using the term "audit" literally, one "counts" the energy consumed (and paid for) and matches that to necessary energy uses and practices that result in energy waste. Remedial actions are then planned to minimize the energy waste and save money. In reality, a one-for-one accounting of energy in versus energy out is rarely done. The term audit is often avoided because of negative connotations. Why are these negative connotations? Other terms used include "energy survey," "energy assessment," or "energy use analysis." All do essentially the same thing, namely examine how energy is consumed and try to identify areas where energy and money can be saved.

HOW THEY WORK

Energy audits are classified according to the client served, falling generally under the categories of residential, industrial, and commercial. Commercial audits include public and semipublic buildings like schools and hospitals and are sometimes referred to as institutional audits.

Audits are done by a variety of groups and agencies, again depending on the type of audit. Most utilities have residential auditing programs. State and community agencies run a number of auditing programs for institutions and low-income housing. The people actually doing the work are either employees of the funding organizations (from nonprofits and universities), or work at "for profit" energy service companies (ESCOs), which either do contract work for the funding sources or work directly for the client.

The scope of the audit also varies considerably. It can consist of anything from a brief walkthrough by an auditor who notes possible areas for improvement to a several month forty-person study at a major manufacturing operation. Nearly any type of audit is of some benefit. The simpler walkthrough type audits can be automated to point where a computer-printed report can be handed to the homeowner at the end of the audit. (Consequently, the costs are quite modest). In larger auditing efforts significant engineering analysis often is required to generate customized recommendations that quantify both the costs and benefits of a particular project.

A common phrase, which has been attributed to many different people, is that "you can't control what you can't measure." This is particularly true of energy use and energy waste. One way energy audits provide information to the client or end-user is by making measurements that show energy waste and allow its magnitude to be calculated. Therefore the toolbox the auditor carries is almost as important as the auditor him. The most important parameter to measure is temperature. A thermocouple can easily measure inside temperatures of the air and hot water systems, but often more is needed. One common tool is an infrared camera that is used to "map" the temperatures of walls and ceilings. This allows hot spots to be found where insulation is not functioning properly and excessive heat is escaping. Ceiling temperatures can reveal problems or suggest the installation of destratification fans to mix the air. Another important tool is a combustion analyzer. Most furnaces need to be tuned periodically to maximize performance. Typically furnace maintenance people adjust a flame by "eye," which cannot match the accuracy of measuring the composition of the flue gas with a combustion analyzer. Other important tools are electric power meters (which can measure for low power factor and line imbalance), flow meters (for fan and pump sizing), light meters and ultrasonic sound sensors (for picking up leaks in gas systems). For best results, measurements

A Southern California Edison employee logs test results at the company's electric meter site. (Corbis-Bettmann)

should be made over a significant period of time. To do this, data-loggers are available that provide inexpensive long-term measurements. Basically, the more sophisticated the audit, the greater the emphasis on measurements.

For some more advance audits, total building modeling can be done using any of several good software packages. One area where this is very important is in studies of ventilation. Excess ventilation wastes energy and some audits in the past involved measuring total infiltration into a building by putting the building under suction. Recently many of the old "rules of thumb" have come under scrutiny because insufficient ventilation can result in air quality problems. Changes in ventilation rates has led to more reliance on modeling.

An audit should also catalog the hardware used at a site. Older hardware can be operating well within its expected efficiency range and still be wasteful. The auditor needs to know about old devices and their performance specifications, as well as what is newly available. It is also essential that the auditor understand *why* the newer device works better. Many products that come on the market do not live up to their marketing hype. The auditor must filter through these and determine which are clearly indicated for their clients.

FINANCING

While some energy audits are paid for directly, most are either subsidized or leveraged in some way. Utilities in the United States are required to provide assistance in energy conservation through their demand-side management programs. Even with the deregulation of electricity, many states are requiring that all energy providers pay into a "public benefits" pool, and money from this pool be used in part to support energy auditing. Another way of leveraging the costs of audits is through "performance contracting." This process normally involves an ESCO that will provide energy audits and system upgrades for no direct costs to the client. The client agrees to pay some fraction of the savings from the system improvements to the ESCO for a contract period (normally 5-10 years). This type of contracting has worked best when energy improvements are clearly measurable, such as lighting upgrades. Other types of projects normally present a larger risk and are therefore less common.

Energy audits are also leveraged with a system commonly called “over the fence” energy. In this scheme, the end-user does not buy equipment, but purchases a commodity such as steam, heat, or compressed air from a third party. The third party owns, leases, or operates the equipment used to provide the energy service. It is then in the interest of that third party to ensure that energy audits are routinely carried out and unnecessary energy use is kept to a minimum.

During the 1990s, system commissioning started to become popular. Historically, builders, architects and owners have agreed on hardware. The owners are guaranteed that the building will be warm in the winter and cool in the summer. Owners now often ask for an additional guarantee—a guarantee of performance in terms of energy costs. A commissioning audit does not recommend improvements, but acts as a measuring device to ensure compliance, often of new buildings before owners take possession. In a commissioning audit, the auditor ensures that heating and air conditioning systems, energy management and control systems, and other complex systems are installed and operating properly.

TRENDS

The interest in energy audits by end-users has historically tracked with the price of energy. An additional motivation appeared in the 1990s with the concern about global warming. Energy use (the majority of which comes from the burning of fossil fuels) directly correlates with the emission of greenhouse gases. Saving energy now also reduces CO_2 emissions and the buildup of greenhouse gases in the atmosphere. For many who wish to be environmentally sensitive, auditing is a proactive step. Another change impacting energy auditing is energy deregulation. Deregulation is leading to open competition for both electricity and gas. For commercial and industrial customers, electricity prices are going down, probably decreasing interest in energy conservation. However, advances in electrical generation technologies are increasing the likelihood that businesses and institutions will be generating at least part of their own electricity needs. There have also been recent advances in engine-driven technologies to replace the electric motors in chillers, air compressors and pumps. Therefore the energy systems within a building or plant are becoming more complex. The way in which energy is purchased is also becoming more complicated. Consumers are being approached by different energy providers, some of whom are packaging technical services such as audits with their energy. Some fuel supplies are provided with “interruptible service” requiring backup fuels. Some energy is provided with “time-of-use” charges and “ratchets” that can result in twelve months of surcharges for the ill-advised use of energy at an inopportune time. This added complexity makes energy auditing an essential periodic check on operations.

Michael Muller

BIBLIOGRAPHY

Felters, J. L. (1998). *The Handbook of Lighting Surveys and Audits*. Boca Raton, FL: CRC Press.

Schueman, D., ed. (1992). *The Residential Engery Audit Manual*, 2nd ed. Lilburn, GA: Fairmont Press.

Thumann, A. (1998). *Handbook of Energy Audits*, 5th ed. Lilburn, GA : Fairmont Press.

AUTOMOBILE

See: Transportation, Evolution of Energy Use and

AUTOMOBILE PERFORMANCE

At the beginning of the twentieth century, the automobile was still a novelty. In the United States more cars were then powered by steam engines and battery-electric systems than by internal-combustion engines. By the end of the twentieth century, the automobile had become an integral part of the American lifestyle, with approximately one privately owned passenger vehicle in operation for every two people. Practically all of these vehicles were powered by internal-combustion engines burning a fuel derived from petroleum.

New cars are purchased on the basis of such qualities as performance, fuel economy, reliability, durability, ride quality, noise and vibration, comfort, convenience, maintenance requirements, styling, safety, environmental qualities, price, and resale value. Many of these attributes conflict. The conflict between performance and fuel economy serves as an example. The typical driver wants a vehicle with sufficient power to merge safely into high-speed freeway traffic, or to pass a slowly moving highway truck on an upgrade. To maintain performance when the vehicle is carrying a heavy load, or pulling a trailer, or operating at high altitude can call for even more installed engine power. That desire for performance potential persists with the typical consumer even though it is used only very occasionally in normal driving.

Installing a more powerful engine to meet performance expectations typically penalizes average fuel economy. To accommodate a range of expectations, manufacturers frequently offer a given car model with a choice of more than one engine. In a recent example, one car model is offered with its base engine and two engine options that exceed the base engine in rated power by 13 percent and 26 percent respectively. While offering greater performance, those optional engines decrease fuel economy in typical driving by 5 percent and 10 percent respectively. In striking a balance between performance and fuel economy, the typical United States consumer leans more heavily toward performance than his overseas counterpart because the pump price of gasoline in the United States is only about a third of what it is in many overseas nations. Moreover, tax policies in some overseas countries, discourage the use of large and powerful engines.

The petroleum from which gasoline is derived is a depletable resource. Passenger cars and light-duty trucks account for about 40 percent of national petroleum consumption. In recent years, imported petroleum has supplied an increasing proportion of U.S. needs. This adversely affects the U.S. balance of trade. Currently about half of the oil consumed is imported. Nearly a quarter of U.S. fuel consumed comes from Organization for Petroleum Exporting Countries (OPEC), a group organized to control the price of oil. About 10 percent of consumption is imported from the Persian Gulf segment of OPEC, where the potential for political instability is a national-security issue. Gasoline combustion also accounts for about 20 percent of the carbon dioxide generated in the combustion of fossil fuels in the United States. The growing concentration of carbon dioxide in the atmosphere threatens to increase world average temperature. For reasons such as these, the federal government has established minimum fuel-economy standards for light-duty vehicles.

Several different metrics are used to assess vehicle performance. Included are the time required to travel a specified distance from a standing start, the time required to accelerate from rest or from some initial speed to a specified final speed, and the speed that can be maintained on a specified upgrade without downshifting the transmission from its highest gear. In the United States, the most frequently used metric of this nature is the time required to reach 60 mph from a standing start.

This is not to imply that a full-throttle acceleration from rest to that speed is a maneuver frequently executed by the typical driver. The time to 60 mph is rather an easily measured parameter that serves as a surrogate for other performance metrics. A car that is slow from 0 to 60 mph will likely have slow response from 40 to 60 mph for freeway merging, or prove lethargic when climbing hills. Reflecting the market preference of the typical new-car buyer, for the average new U.S. passenger car, the acceleration time from 0 to 60 mph has decreased from about 14 seconds in 1975 to fewer than 11 seconds in 1995.

This gain has not come entirely at the expense of fuel economy, however. Over that same twenty-year span, the fuel economy of the average new U.S. car increased by 80 percent. These advances are attributable to lower vehicle weight, improved tires, reduced aerodynamic drag, improved transmissions, and gains in engine efficiency.

ACCELERATION PERFORMANCE

When a driver commands an increase in vehicle velocity, that vehicle obeys Newton's first law of motion, which states that when a force (F) acts on a body of mass (M) and initially at rest, that body will experience an acceleration (a). For an automobile, typical units for acceleration, which is the rate of change of velocity, would be miles per hour per second. Mass is further defined as the weight of the body (W) divided by the acceleration of gravity (g).

The product of force F and the rolling radius (R) of the tires on the drive wheels is the wheel torque (T). Power depends on both torque and rotational speed (N). By definition, power is given by $P = 2\pi NFR = 2\pi NT$. When driving at constant speed, the driver adjusts the accelerator pedal so the drive-wheel power exactly matches the *power required* (P_r) to overcome the resistance of the vehicle (discussed later in this article). To accelerate the vehicle, the driver further depresses the accelerator pedal so that the *power available* at the drive wheels (P_a) exceeds P_r.

When applying Newton's law to a moving automobile, acceleration depends on the excess of power over that required for constant-speed driving, namely $P_a - P_r$. From this it follows that the instantaneous acceleration (a) of the vehicle at a given road speed (V) is

$$a = g\,(P_a - P_r) / WV \qquad (1)$$

For maximum vehicle acceleration, the driver depresses the accelerator pedal to the floorboard and the engine operates with a wide-open throttle. The power required curve traces the power needed by the car as a function of vehicle velocity when it is operated at constant speed in still air on a level road. At any given speed, the difference between these curves, Pa–Pr in Equation 1, is available for accelerating and hill climbing.

In the above expression for vehicle acceleration, the proper weight to use is the effective vehicle weight, which is the actual weight plus an additional increment that accounts for the rotating inertia of the engine, drivetrain, and wheels. For a passenger car driven in high gear, the ratio used for normal driving, this increment amounts to about 10 percent of the vehicle weight. It is substantially higher in the lower gears, however, because at a given road speed, the rotational speed of the engine and part of the transmission is multiplied by the transmission gear ratio. This increase in rotational speed magnifies their influence on the effective weight of the vehicle. For illustrative purposes, however, this adjustment to vehicle weight for rotational inertia is ignored here.

POWER REQUIRED

The force required to move the vehicle forward is the sum of four components: rolling resistance, aerodynamic drag, acceleration force, and grade requirement. This required force is converted into power required by multiplying by the forward velocity of the vehicle with respect to the road.

Rolling resistance stems from the energy expended in deforming the tire and the surface of the road at the contact patch between tire and roadbed. The power required to overcome rolling resistance (P_r) depends on the rolling resistance coefficient (C_r), and the vehicle weight (W) and velocity (V_v). It is given by

$$P_r = C_r\,W\,V_v \qquad (2)$$

C_r tends to increase slightly with speed but is often considered constant over the normal speed range of an automobile.

The rolling resistance coefficient of a tire depends on the construction of the tire carcass, the elastic characteristics of the tire material, the tread design, and characteristics of the roadbed. It increases with decreasing wheel diameter, tire underinflation, and roadbed compliance. It decreases as the operating temperature of the tire rises.

Before the 1960s, the bias-ply tire exemplified standard construction. It had a typical rolling resistance coefficient of 0.015 on hard pavement. Since then, the radial-ply tire has emerged, offering a coefficient closer to 0.010. Coefficients as low as 0.008 to 0.009 have been claimed in tires suitable for use on passenger cars. Cutting the coefficient from 0.015 to 0.008 offers the opportunity for about a 10 percent reduction in fuel consumption.

Other qualities sought in a tire include ride quality, cornering ability, traction characteristics on both dry and slippery roads, tire noise, life, and cost. Addressing these qualities often opposes the objective of lower rolling resistance.

The force of aerodynamic drag opposing forward motion of the vehicle depends on its drag coefficient (C_d), its frontal area (A_f), the air density (ρ), and the velocity of the wind with respect to the vehicle. In still air, this velocity is simply the vehicle velocity (V_v). If driving into a headwind of velocity V_w, however, the wind velocity with respect to the vehicle is the sum of these two. Multiplying the aerodynamic drag force by vehicle velocity provides the aerodynamic power requirement (P_a).

$$P_a = \tfrac{1}{2}\,\rho A_f V_v\,(V_v + V_w)^2 \qquad (3)$$

The drag coefficient for an automobile body is typically estimated from wind-tunnel tests. In the wind tunnel, the drag force acting on a stationary model of the vehicle, or the vehicle itself, is measured as a stream of air is blown over it at the simulated vehicle speed. Drag coefficient depends primarily on the shape of the body, but in an actual vehicle is also influenced by other factors not always simulated in a test model.

For example, the windage loss associated with the rotating wheels increases drag. Covering the wheel wells can reduce this adverse effect. Although suitable for rear wheels, a body-mounted cover over the front wheel wells interferes with steering because during sharp turns, the front wheels extend beyond the plan-view profile of the car body.

A smooth underbody also would improve fuel economy by reducing drag but is rarely used because of more pressing demands. For example, a continuous smooth underbody interferes with engine-compartment ventilation, blocking the normal exit route for engine cooling airflow that has passed through the radiator. It also interferes with accessibility for routine engine servicing. Farther to the rear, it is desirable to expose the exterior surfaces of the hot running exhaust system, which includes a catalytic converter and muffler, to the flowing airstream for cooling. Therefore, covering them with an underbody panel creates problems.

Another source of increased drag involves the external-surface details. In this regard, stand-alone bumpers, externally protruding door handles, and running boards alongside the passenger compartment to facilitate entry, all of which were once universal, have disappeared. Flush rather than recessed side windows have come into recent use. On the other hand, external rear view mirrors, which increase drag, have been added as a safety measure.

Over the years, the flat vertical windshield of 1920 has given way to an increasingly raked windshield, first curved in two dimensions, but now in three. The once ubiquitous vertical flat radiator that fronted the engine compartment has disappeared into a compartment covered by a streamlined front body. Fenders and headlights, both of which were once free standing, are now incorporated into that body.

A major fraction of the aerodynamic drag in the modern streamlined car is caused by flow separation at the rear of the body. Alleviating that separation calls for a long afterbody that tapers to a point. Depending on driving conditions, this change could decrease fuel consumption by about 15 percent, with even greater improvement on the highway but less in the city. However, such a sharp tailpiece is useless for carrying passengers, of minimal utility in storing luggage, prohibitively dangerous for storing fuel, and impairs vehicle handling and parking.

The equation given above for the power required to overcome aerodynamic drag (P_a) is expressed for a vehicle driving into a headwind. If, instead, the wind direction is from the rear, the sign on the wind velocity changes from positive to negative, and the aerodynamic drag is reduced. In practice, however, the wind almost never blows directly from either the front or the rear. When the wind approaches from an angle oblique to the direction of travel, the wind velocity relative to the vehicle is the vector sum of the travel velocity and the wind velocity. The vehicle cross-sectional area encountered by this relative wind is greater than the frontal area A_f, and the vehicle shape was not designed for that oblique wind direction. As a result, the drag coefficient can increase as much as 50 percent above its value when the direction of the wind is aligned with the vehicle centerline.

The power expended in accelerating the vehicle represents an investment in kinetic energy that is stored in the vehicle by virtue of its motion. It can be partially recovered during vehicle coasting, but most of it is usually dissipated as heat in the brakes when the vehicle is decelerated or brought to a stop. The acceleration power (P_k) depends on vehicle weight (W), velocity (V_v), and the rate of change of velocity with respect to time, which is the instantaneous acceleration (a). The power for acceleration is given by

$$P_k = (W\ a\ V_v) / g \qquad (4)$$

where g is the gravitational constant. If the car is being powered in a decelerating mode, the sign on this acceleration term becomes negative.

The final term in the equation for required power is that which accounts for driving on a grade (P_h). The severity of a grade is normally defined as the ratio of its vertical rise (h) to its horizontal run (L), expressed as a percentage. On U.S. interstate highways, the grade is usually limited to 4 percent ($h/L = 0.04$). On public roads, grades as high as 12 percent may be encountered. For grades no steeper than that, P_h is closely approximated by

$$P_h = W\,V_v\,(h/L) \qquad (5)$$

If the vehicle is on a downgrade rather than an upgrade, the sign on this term becomes negative.

POWER AVAILABLE

The full-throttle capability of a typical gasoline engine, as delivered to the wheels by the drivetrain, was shown as a function of car speed for four transmission gear-ratios. In any given gear, engine full-throttle power rises with increasing engine speed, leveling off at maximum power and then falling again as engine speed is further increased. This characteristic is dominated by the influence of engine speed on mass airflow rate and on engine friction.

Mass airflow rate is the principal determinant of the maximum power that can be developed within the engine cylinders. As speed increases, the cylinders ingest more air. Therefore, one might expect the power developed in the cylinders to increase in proportion to engine speed. However, the aerodynamic losses in the air passing through the cylinders rise at an increasing rate as engine speed is increased. This causes the mass airflow rate, and with it the power developed in the cylinders, to reach a maximum at some high engine speed that typically lies beyond the normal operating range of the engine.

The power delivered by the engine crankshaft is less than that developed within the engine cylinders by the power expended in overcoming engine friction. Friction power also increases with speed at an increasing rate. As a consequence, the power output delivered at the engine crankshaft peaks at some speed less than that at which the cylinders achieve their maximum power.

The power developed on the crankshaft is further depreciated by the requirements of such accessories as the electric alternator, power steering pump, and air conditioner. The drivetrain that connects the crankshaft output to the vehicle drive wheels causes a further loss in power. Drivetrain efficiency generally falls within 80 to 95 percent. The remaining useful propulsive power varies with speed.

Design features in the engine and transmission can effect seemingly subtle changes to the shape of the power available curve that are important to vehicle drivability. Automotive engineers normally address this issue in terms of available torque, rather than power, as a function of speed. Torque, of course, is determined by dividing power by rotational speed. An objective is to produce high torque at low speeds in order to minimize the need for transmission shifting.

Before the 1980s, engine air was normally inducted through an air cleaner mounted directly on the carburetor. Replacing the carburetor with electronically controlled fuel injection just ahead of the intake valves has minimized large areas of intake-manifold walls wetted with liquid fuel, which impaired engine response to quick depression of the accelerator pedal. This virtual elimination of wetted walls in the air intake system has facilitated use of a carefully designed intake system that includes a plenum and long air pipes. The components of this system are carefully proportioned to enhance torque at low engine speeds.

Another method of advantageously reshaping the torque variation with engine speed is the incorporation of variable valve actuation. This includes the options of variable valve timing; variable valve lift; and in engines with two intake valves per cylinder, the deactivation of one of them at low engine speeds. Variable valve actuation may be used not only to enhance low-speed torque but also to improve other operating aspects of the engine.

In automatic transmissions, a torque converter has long been incorporated to increase delivered torque at low vehicle speeds, as during acceleration from a standing start. In the torque converter, the engine shaft is fastened to a centrifugal pump that converts the torque delivered by the engine into flow energy in a hydraulic fluid. An adjacent hydraulic turbine converts this flow energy back into shaft torque for delivery to the input side of the transmission gearbox. Because there is no mechanical connection between pump and turbine, turbine speed is independent of pump speed and allows the vehicle wheels to be stationary while the pump rotates at engine speed. Through the action of a row of stationary vanes located between the pump inlet and the turbine discharge, the torque converter is able to multiply engine output torque at the expense of gearbox input speed. This ability of the torque converter to multiply torque typically allows the gearbox of the automatic transmission to meet the needs of the vehicle with one less gear step than would be needed in an equivalent manual transmission.

When the vehicle is at a standstill, the torque converter delivers approximately twice the engine torque to the gearbox. This torque multiplication falls toward unity as the vehicle accelerates from rest and as the turbine speed catches up with the pump speed. At the coupling point, turbine speed nearly equals pump speed; torque multiplication ceases; and the stator is unlocked, allowing it to spin freely. The small speed difference, or slip, between pump and turbine then existing causes a minor loss in transmitted power. The 1970s witnessed increasing use of a torque-converter clutch that reduces this slip loss by joining the pump to the converter when the coupling point has been reached. (Actually, a slight slip in this clutch may be employed to minimize the transmission of engine vibrations to the driver.)

Because of the importance of mass airflow rate in establishing engine output power, power available is sensitive to ambient conditions. Full-throttle engine power varies approximately inversely with inlet-air absolute temperature, but more significantly, approximately directly with ambient pressure. Mountain passes exist on public roads in the United States that have altitudes of over 12,000 ft. The normal atmospheric pressure at such altitudes results in a one-third loss in power capability in the typical passenger-car engine.

POWER RESERVE

Power available at full throttle in top gear is again plotted against vehicle speed in Figure 1, along with power required curves for three different conditions. Figure 1a is for driving on a level road in still air, Figure 1b for driving on a level road into a 25 mp/h headwind, and Figure 1c for driving up a 6 percent grade in still air. In each case, the difference between power available and power required that is available for acceleration, termed the power reserve, is shown shaded. Clearly, the power reserve in top gear is diminished significantly by driving into a headwind or by hill climbing.

Vehicle maximum speed is indicated in Figure 1 by the intersections of the power available and power required curves. It is seen to fall from more than 120 mph to 90 mph in going from conditions of Figure 1a to Figure 1c. In the early days of the automobile, top speed was of greater importance than today. The Panhard Levassor of 1886 was capable of only 12 mph. In about 1900, cars with a top speed of about 40 mph had become available, which may have been adequate for existing roads. When the first concrete road appeared in 1909, the Olds Limited could reach 65 mph. With its 108 horsepower engine, the 1935 Cadillac attained 85 mph. An additional 25 hp enabled the 1951 Cadillac to top 98 mph.

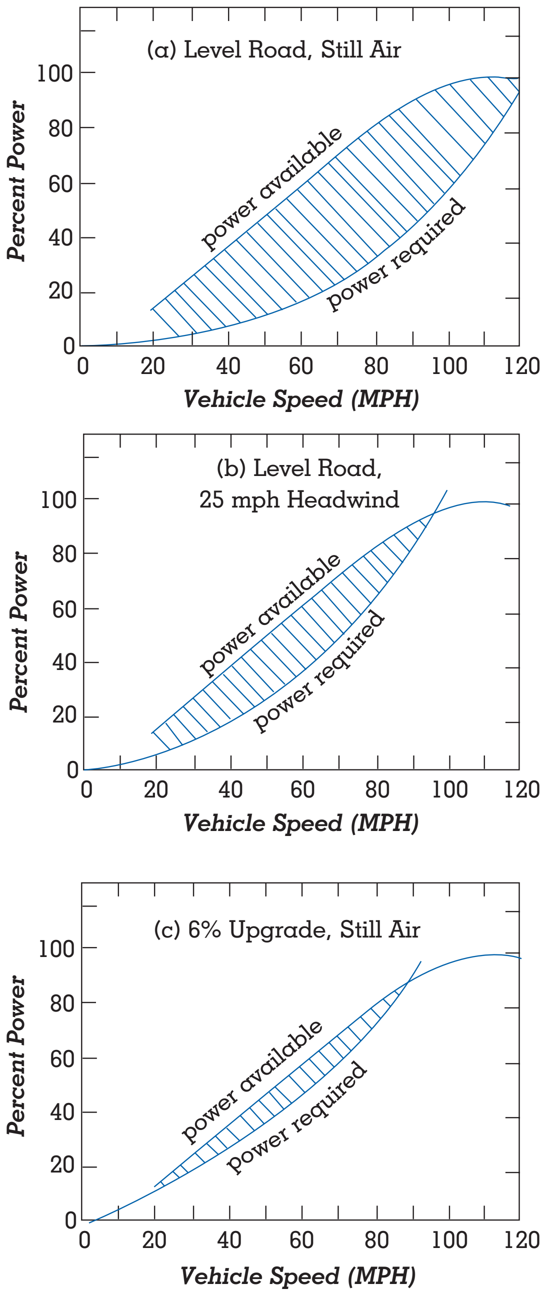


Figure 1.
Power available at wide-open throttle and power required on a level road in still air for car with a four-speed manual transmission.

That a modern automobile may be able to exceed 100 mph is of no direct consequence in U.S. driving

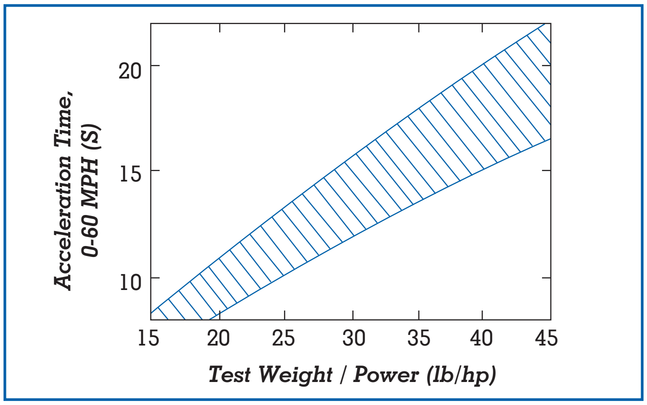


Figure 2.
Effect of vehicle weight/power ratio on acceleration time.

because these speeds are illegally high in nearly all locations. What accompanies the deterioration in top speed illustrated in Figure 1, however, is the obvious loss in power reserve at more commonly driven speeds. This reserve is so small—for example at 40 mph on a 6 percent grade—that in real driving, frequent gear shifting between fourth and third gear could be rendered necessary by minor disturbances such as wind gusts or on-off cycling of the air conditioner. Such shifting requirements, whether imposed on the driver by a manual transmission or done for the driver in an automatic transmission, are sources of driver dissatisfaction. In addition, fuel economy is poorer in the downshifted gear because the engine is forced to operate at a higher speed with lower thermal efficiency.

A corresponding situation occurs at high altitude, where one-third of the sea-level power available has been lost due to low atmospheric pressure. This low air density also reduces aerodynamic drag, but rolling resistance is unaffected by altitude. As a result, power reserve is seen to suffer. In fact, at this altitude, the power available in fourth gear is insufficient to operate the vehicle on a 6 percent grade at any speed without downshifting.

Past studies of passenger-car performance have shown that the acceleration time from 0 to 60 mph correlates well with vehicle weight/power ratio. Although a small difference was found between automatic and manual transmissions, most measured data falls within the band shown in Figure 2, where test weight corresponds to vehicle curb weight plus two 150-lb passengers or their weight equivalent. This suggests that performance can be improved either by increasing installed engine power, which tends to depreciate fuel economy, or by decreasing vehicle weight, which tends to improve fuel economy.

FUEL ECONOMY

Vehicle fuel economy is normally measured in miles per gallon. At any given instant, it depends on the energy content of a gallon of fuel (Q_f), the vehicle velocity (V_v) and power required (P_{req}), the thermal efficiency with which the engine converts fuel energy into useful output work (η_e), and the mechanical efficiency with which the driveline delivers that work to the vehicle wheels (η_d). Specifically,

$$\text{Miles/Gallon} = Q_f V_v \eta_e \eta_d / P_{req} \qquad (6)$$

On a transient driving schedule, this instantaneous fuel economy must be averaged over the distance driven.

In earlier times it was customary to eliminate vehicle velocity as a variable by measuring fuel economy while driving at a constant speed, typically 40 mph. Fuel economy at this speed generally falls within 10 percent of its maximum level-road value. Fuel economy falls off at both much lower and much higher speeds.

The shortcoming of a constant-speed test is that in traffic, nobody drives at constant speed. Consequently, with the onset of the federal fuel-economy standards that took effect in the United States in 1978, the Environmental Protection Agency (EPA) prescribed transient driving schedules deemed representative of both urban and highway driving.

The EPA urban driving schedule is diagramed in Figure 3a. Its 18 start-and-stop cycles cover 7.5 miles in 1,372 seconds, with an average speed of 19.5 mph and a peak speed of 56.7 mph. The Federal Test Procedure (FTP) used for measuring exhaust emissions is based on this schedule. It involves operating a car, which has been initially stabilized at room temperature, for the prescribed 7.5 miles, shutting off the engine for 10 minutes, then restarting and repeating the first five cycles. The FTP is performed on a chassis dynamometer, with the vehicle stationary and the drive wheels turning a roller in the floor that is connected to an electric generator. That generator is loaded to simulate the rolling resistance and aerodynamic drag that would be encountered when actually driving the vehicle on the road. The urban fuel economy (MPG_u) is determined from this test.

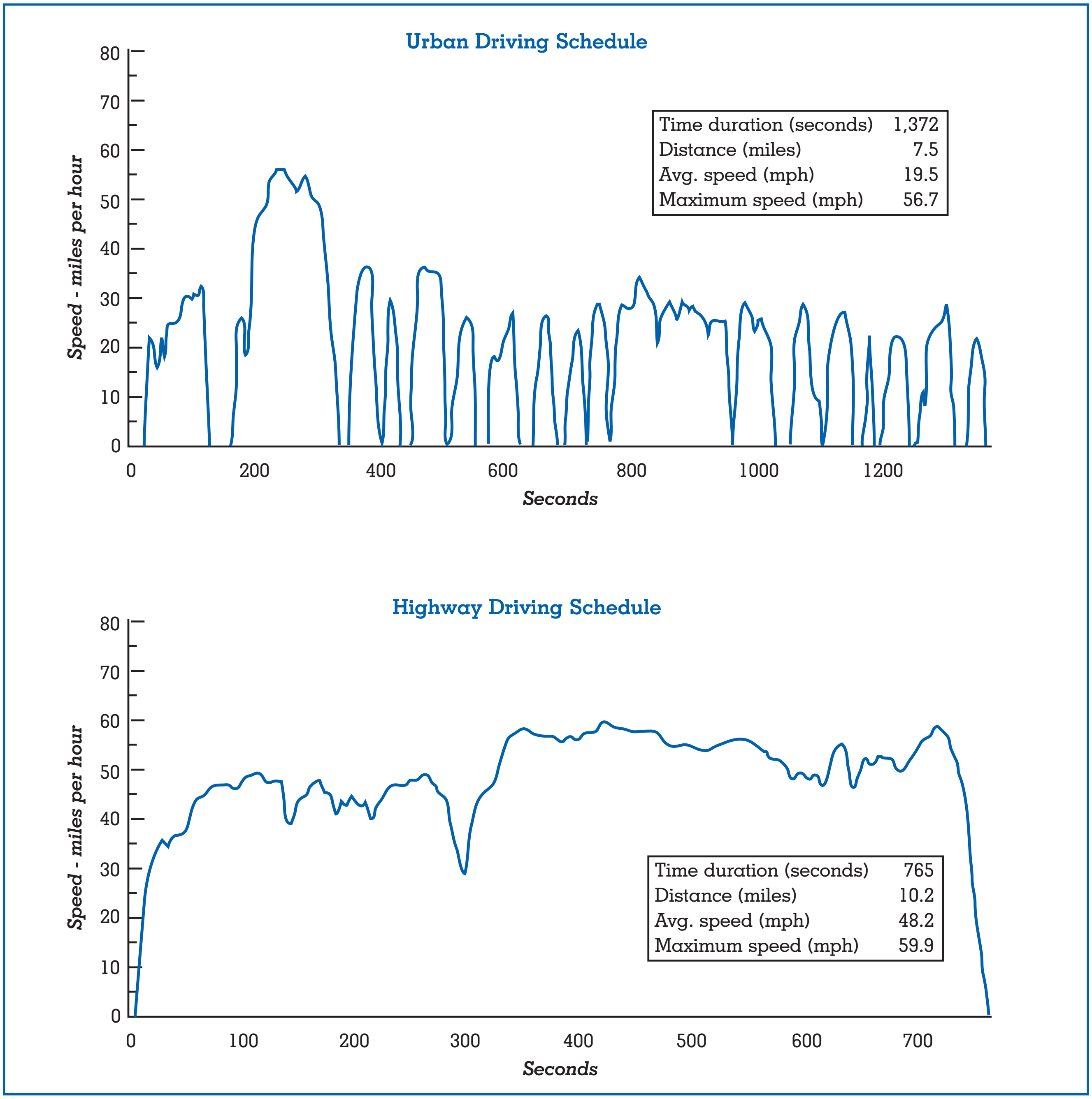


Figure 3.

Standard U.S. driving schedules: (a) urban and (b) highway.

SOURCE: *Code of Federal Regulations*, "Subpart B-Fuel Economy Regulations for 1978 and Later Model Year Automobiles-Test Procedures," July 1, 1988 ed., p. 676.

The schedule used to measure highway fuel economy (MPG$_h$), also driven on a chassis dynamometer, is diagramed in Figure 3b. It covers 10.2 miles in 765 seconds at an average speed of 48.2 mph and a peak speed of 59.9 mph. In contrast to the urban schedule, there are no stops and starts during the highway schedule.

For regulatory purposes the fuel economy assigned to the vehicle from these tests is based on the premise that the vehicle will accumulate 55 percent of its mileage on the urban schedule and 45 percent on the highway. The resulting composite fuel economy (MPG_c) is thus calculated according to:

$$1/MPG_c = 1/(0.55/MPG_u + 0.45/MPG_h) \qquad (7)$$

For a typical passenger car of the 1990s, the effect on fuel consumed over the urban and highway schedules, and also on composite fuel consumption, is illustrated in Figure 4 for independent 10 percent changes in vehicle weight, aerodynamic drag, and rolling resistance coefficient. On the urban schedule, vehicle weight has the greatest significance because of the dominant effect of acceleration power associated with the many start-and-stop cycles. This power, invested in kinetic energy during accelerations, is not recouped because the brakes are applied during the succeeding decelerations. On the highway schedule, which lacks frequent starts and stops, the high average speed makes aerodynamic drag the dominant factor. This reflects the exponential dependence of drag on speed.

At the bottom of Figure 4, the effect of fuel consumed during unpowered engine operation, which occurs during closed-throttle decelerations and engine idling at zero vehicle speed, also is shown. The fuel expended during unpowered operation is greatest on the urban schedule because of the time spent braking and standing during that schedule.

It is clear from Figure 4 that weight is the vehicle parameter of greatest significance to composite fuel economy. The power increments required to overcome rolling resistance, provide acceleration performance, and climb hills all vary directly with vehicle weight. The power required to overcome aerodynamic drag relates indirectly because vehicle frontal area, a factor in determining interior roominess, tends to correlate with vehicle weight. The importance of weight to performance was established in Figure 2.

Over the years, consumer desire for more interior roominess and greater luggage capacity has combined with mandated weight-increasing safety and emissions devices to complicate weight reduction. Despite this, the curb weight of the average new U.S. passenger car decreased 21 percent from 1975 to 1995, in which year it was 2,970 lb. Over those same years, however, the curb weight of the average light-duty truck increased 4 percent, to 3,920 lb in 1995. During those two decades, the category of light-duty trucks has been swollen by the popularity of light-duty vans and sports-utility vehicles included in this category. The sales ratio of these light-duty trucks to the lighter passenger cars climbed from about 20 percent in 1975 to near 40 percent in 1995. This increase in the proportion of truck sales meant less reduction in the weight of the fleet-average vehicle than might otherwise be expected.

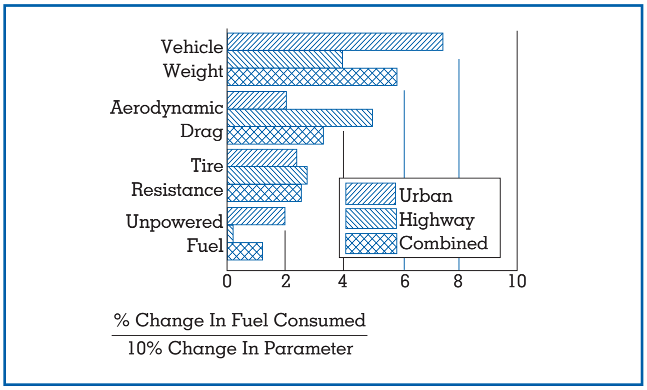


Figure 4.
Individual effects of design characteristics on fuel economy of a typical passenger car.

Much of the passenger-car weight reduction has resulted from basic design changes, but some came from materials substitution. From 1978 to 1997, the portion of domestic-car weight attributable to ferrous materials has dropped from 74 to 67 percent. Concurrently, lighter aluminum climbed from 3.2 to 6.3 percent of car weight, and plastics and composites from 5.0 to 7.5 percent. Magnesium, which is lighter than aluminum, is finding limited application. Research is intense on composites, which combine glass or carbon fibers with a plastic binding material. Such composites are light, strong, and can be formed into complex shapes, but their high cost and long manufacturing cycle times have impeded their acceptance in high-production automobiles.

HISTORICAL PERSPECTIVE

U.S. automotive history reveals a fairly continuous improvement in both performance and fuel economy, but the relative interest in each is influenced by externalities. When gasoline is plentiful and inexpensive, the consumer is more interested in performance and/or larger vehicles, both of which tend to decrease fuel economy. During the Arab oil shocks of the 1970s, when the gasoline supply was stifled and driv-

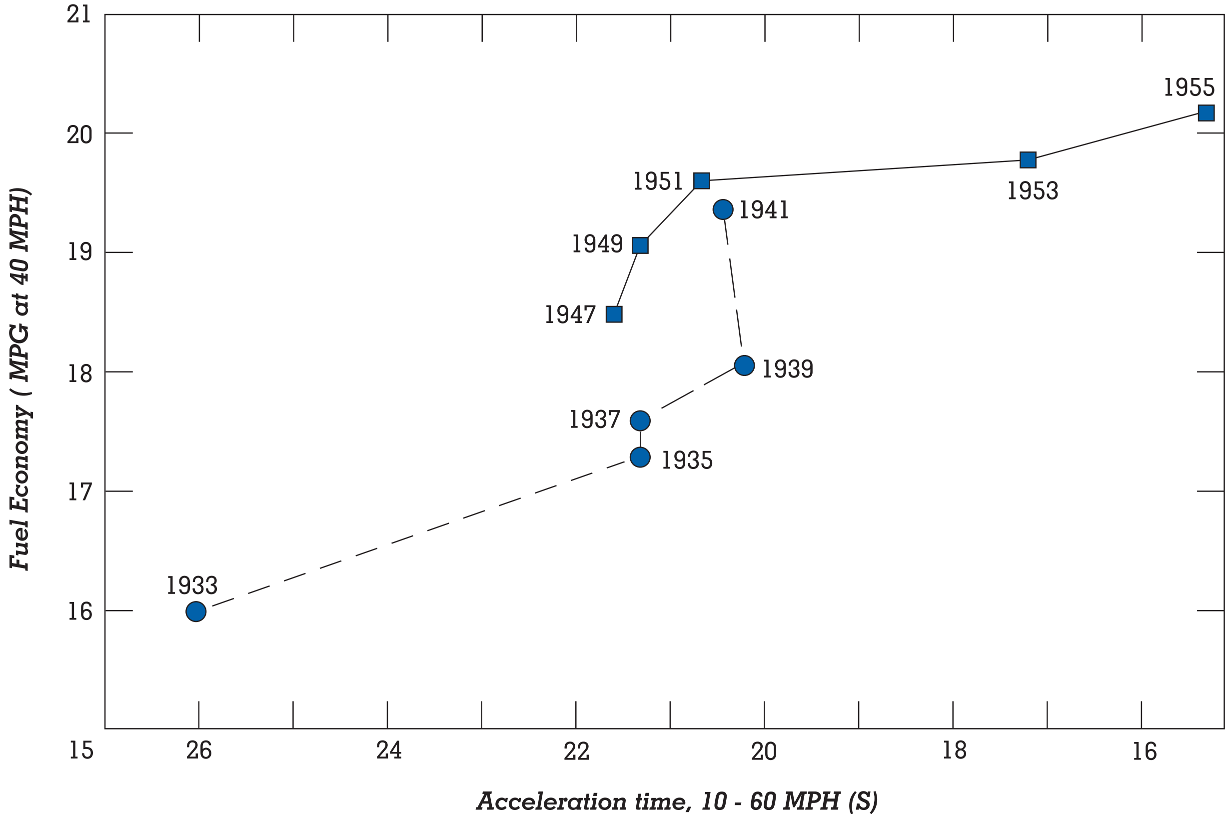


Figure 5.
Historical trend in trade-off between performance and fuel economy (average of thirty-eight car models, 1975–1995).

ers sometimes waited in line to fill their tanks, fuel economy became more important. Likewise, fuel economy received special emphasis from 1975 to 1980 as manufacturers scrambled to meet the newly imposed and rapidly rising fuel-economy standards.

In earlier days, performance was often measured in terms of acceleration time to 60 mph from an initial speed of 10 mph rather than from a standstill. Fuel economy at a steady 40 mh is plotted against that acceleration time in Figure 5. The acceleration scale is reversed so that the best-performing cars are on the right. The data are averaged over a fleet of thirty-eight models from eighteen U.S. manufacturers. The two branches of the plot represent fleets from 1933 to World War II, during which civilian car production was suspended, and from the postwar years until 1955. Over this twenty-two year span, fuel economy increased 25 percent and acceleration time was improved 40 percent.

More recent data, from 1975 through 1995, are presented in Figure 6. The EPA classifies passenger-cars according to interior roominess. The four classes represented in Figure 6—sub compact, compact, mid size sedan, and large sedan—accounted for from 74 percent of the new passenger-car fleet in 1975 to 94 percent in 1995. The average composite fuel economy at five-year intervals is plotted for each car class against the corresponding average acceleration performance, as estimated by the EPA from vehicle installed power and test weight. In these estimates, each car is assumed to be carrying a 300-lb load. Curves on the plot indicate trends for each of the five years sampled.

The trend shown in a given year has been to offer somewhat better performance in larger cars than in smaller ones. However, the large sedan is able to carry a heavier load than a sub compact. Loading the

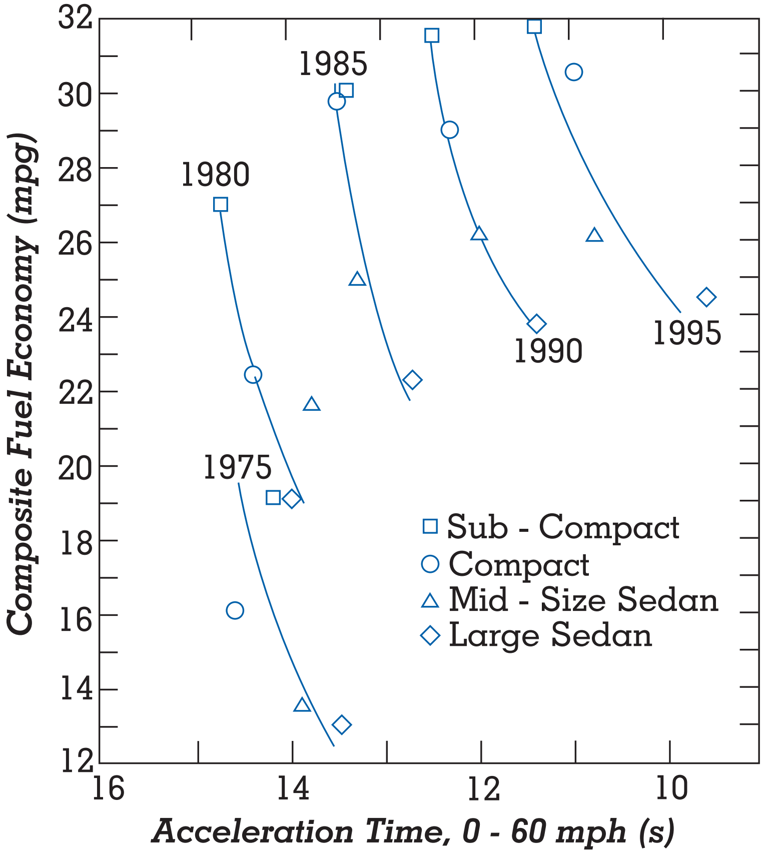


Figure 6.
Historical trend in trade-off between performance and fuel economy, segregated by car size (average for U.S. new-car fleet, 1975–1995).

large sedan more heavily, rather than with the same 300 lb as in the sub compact, would erode the apparent performance advantage of the large car.

It is seen from Figure 6 that in any given year, a fuel-economy spread of 6 to 8 mpg exists between the sub compact and the large sedan. A major contributor to this difference is the increased weight of the larger car. The average large sedan was 60 percent heavier than the average sub compact in 1955, with that spread decreasing to 30 percent in 1995.

Reduced weight has been a major contributor to the improvements in performance and fuel economy over the twenty-year span of Figure 6. From 1975 to 1980 the test weight of the average new passenger car fell nearly 25 percent. It has crept up slowly in the following fifteen years, but in 1995 was still 20 percent less than in 1975. The trend from rear-wheel to front-wheel drive aided in reducing weight. In 1975, front-wheel drive appeared in fewer than 10 percent of the new-car fleet, but its fleet penetration rose to 80 percent in 1995.

Although the average passenger-car engine produced about 10 percent more maximum power in 1995 than in 1975, it did so with approximately 40 percent less piston displacement as a result of advances in engine technology. The average 1995 engine also was lightened through greater use of non ferrous materials in its construction.

Typical passenger-car fuel economy was boosted slightly around 1980 with the entry of the light-duty diesel. A 25 percent increase in miles per gallon was claimed for that engine, about half of which resulted from the higher energy content of a gallon of diesel fuel relative to gasoline. Diesel-engine penetration of the new-passenger-car fleet peaked in 1981 at 5.9 percent, and a year later in the light-duty truck fleet at 9.3 percent. By 1985 the popularity of the diesel had waned, with penetration of the combined fleet being less than 1 percent as of the turn of the century.

In the drivetrain arena, the torque-converter automatic transmission has long dominated the U.S. passenger-car fleet, appearing in about 80 percent of new passenger cars in both 1975 and 1995. Its penetration in light-duty trucks has risen from 63 percent in 1975 to nearly 80 percent in 1995. However, the automatic transmission of 1995 was much improved over the 1975 model. In 1975, three-speed automatics and four-speed manual transmissions were the norm. By 1995 the more efficient four-speed automatic with torque-converter clutch had taken over, while manual-transmission vehicles enjoyed enhanced fuel economy through the addition of a fifth gear.

Looking back a century in the United States, when draft horses outnumbered automobiles and police enforced a speed limit of 10 mph on cars in some locales, the individual improvements that have since accrued to the automotive vehicle have transformed it into a necessity in the lives of many Americans. Gains in both performance and fuel economy have contributed importantly to its popularity. In the last quarter of the twentieth century, during which fuel economy regulation began in the United States, average new-car fuel economy has increased 80 percent at the same time that performance capability, as measured by acceleration time to 60 mph, has improved 25 percent. Both have benefited most significantly from a reduction in the weight of the average car, but over this time period the summation of effects from myriad other individual improvements has contributed even more to fuel economy. These individual improvements can be grouped into factors influencing engine efficiency, the efficiency of the drivetrain connecting the engine to the drive wheels of the vehicle, the rolling resistance of the tires, vehicle aerodynamic drag, and the power consumed by accessories either required for engine operation or desired for passenger comfort and convenience. Engine improvements can be further classified as to whether they improve the quality and control of the mixture inducted into the cylinders, increase the air capacity of the engine, improve the efficiency of the energy conversion process associated with combustion, or reduce the parasitic losses associated with engine friction and the pumping of gas through the engine cylinder. With multiple technological improvements possible in each of these classifications, their additive effects can become significant even though the improvement attached to each individually is small. Sometimes gains are realized through interaction among separate components. For example, an improved transmission may enable the engine to run at a more efficient point in its operating range, even though the engine itself experiences no change in its overall efficiency characteristics.

Charles A. Amann

BIBLIOGRAPHY

Robert Bosch GmbH. (1986). *Automotive Handbook*, 2nd ed. Stuttgart, Germany: Author.

Burke, C. E.; Nagler, L. H.; Campbell, E. C.; Zierer, W. E.; Welch, H. L.; Lundsrom, L. C.; Kosier, T. D.; and McConnell, W. A. (1957). “Where Does All the Power Go?” *SAE Transactions* 65:713-737.

Caris, D. F.; Mitchell, B. J.; McDuffie, A. D.; and Wyczalek, F. A. (1956). “Mechanical Octanes for Higher Efficiency.” *SAE Transactions* 64:77-96.

Gillespie, T. D. (1992). *Fundamentals of Vehicle Dynamics.* Warrendale, Pa: Society of Automotive Engineers.

Heavenrich, R. M., and Hellman, K. H. (1996). “Light-Duty Automotive Technology and Fuel Economy Trends Through 1996.” Technical Report EPA/AA/TDSG/96-01. Ann Arbor, MI: U.S. Environmental Protection Agency.

Hoerner, S. (1965). *Fluid-Dynamic Drag.* Midland Park, NJ: Author.

Malliaris, A. C.; Hsia, H.; and Gould, H. (1976). “Concise Description of Auto Fuel Economy and Performance in Recent Model Years.” SAE Paper 760045. Warrendale, PA: Society of Automotive Engineering.

McCuen, C. L. (1952). “Economic Relationship of Engine-Fuel Research.” *SAE Transactions* 6:291-303.

Sovran, G.; Morel, T.; and Mason, W. T., Jr. (1978). *Aerodynamic Drag Mechanisms of Bluff Bodies and Road Vehicles.* New York: Plenum Press.

AVAILABLE ENERGY

See: Reserves and Resources

AVIATION FUEL

Aviation fuel is the fuel used to power aircraft in flight. It must satisfy the unique requirements of both the engine and the airframe of the aircraft. Currently the great majority (more than 99%) of aviation fuel used in both civil and military aircraft is jet fuel. A small quantity of aviation gasoline is still used in small aircraft. Early aircraft used motor gasoline to power their spark ignition engines because the aviation and auto worlds shared the same early engines. In recognition of aviation's more stringent requirements compared to ground transportation, separate specifications for aviation gasoline were developed after World War I. Subsequent aircraft spark ignition engine developments as World War II approached identified the need for high octane in aviation fuel for improved performance. This resulted in the development of 100-octane aviation gasoline and the unique refinery processes necessary to produce it. Beginning in the mid-1930s, research was initiated in both Great Britain and Germany on the development of a gas turbine aircraft engine, which was radically different from the spark-ignition, reciprocating engines used since the days of the Wright brothers. The new jet engine was capable of markedly improved high-speed performance. During this development, illuminating kerosene used as a fuel for lamps, was chosen as the liquid fuel for the jet engine because it did not conflict with the very strong military demand for high-octane aviation gasoline. This use for jet engines of distillate-based fuels different in composition from high-octane gasoline has continued to this day. The first operational use of jet-engine-powered aircraft occurred in a military aircraft (the German Me-262) late in World War II, and its performance proved so superior to propeller-powered, piston-engine aircraft that subsequently all air forces changed to the use of jet aircraft. The development and rapid growth of higher-speed commercial transport aircraft using jet engines began in the late 1950s. As a result of the switch of both military and commercial aircraft to jet engines from spark ignition engines, jet fuel demand rose rapidly, and jet fuel over time displaced aviation gasoline as the dominant fuel for aviation use.

FUEL TYPES

Jet fuels in use today are essentially all kerosene-based but differ somewhat in their compositions. For civil fuels, Jet A is used primarily in the United States and Jet A-1 throughout most of the rest of the world. Jet A and Jet A-1 differ principally in their freezing point, which is the temperature at which solid wax crystals form in the liquid fuel as it cools. Commercial aircraft store their fuel primarily in wing tanks, and there is a concern that during long international flights through cold-weather conditions the formation of wax could interfere with the flow of fuel from the wing tanks into the engines. Thus all jet fuels specify a freezing point suitable for its intended flight use. The military fuel used by both the U.S. Air Force and NATO air forces is JP-8, which is similar in composition to commercial Jet A-1, but employs military-designated additives. The U.S. Navy uses JP-5, a jet fuel with a higher flash point (a measure of the fire hazard associated with the fuel) than Jet A, Jet A-1, or JP-8 because of concern about fire safety aboard aircraft carriers, particularly in combat operations. In the past the U.S. Air Force used a very low flash point fuel called JP-4, composed of a mixture of kerosene and lighter-boiling refinery streams, but switched to the higher-flash-point kerosene-based JP-8 fuel to reduce combat losses and post-crash fire and handling incidents. A commercial low-flash-point fuel designated Jet B, similar to military JP-4, is used only in very cold Arctic areas because of difficulties in starting engines with the more viscous kerosene-type fuels.

SPECIFICATIONS

Jet fuel requirements are defined by engine and airframe technical needs, which are balanced against the need for a widely available and low-cost fuel. These technical and economic requirements are translated into fuel specifications that define physical properties and chemical compositional limits and that also require the fuel to pass a number of unique performance tests designed to predict satisfactory use. Jet fuel is a tightly specified, high-technology commodity. A number of commercial and military jet fuel specifications are used throughout the world. Commercial specifications include ASTM D 1655, which is an industry consensus specification; Defense Standard 91/91, issued by U.K. Ministry of Defense for their Civil Aviation Authority; and the

International Air Transport Association (IATA) Guidance Material. At major airports, where fueling systems are operated by a number of different companies rather than a single company, a combination of the most stringent requirements of ASTM D 1655, Defense Standard 91/91, and the IATA Guidance Material called the "Check List" is often used. Attempts are under way to harmonize the major commercial Western specifications and to get non-Western countries to join in using common worldwide specification and test methods. In addition, the U.S. military as well as other governments write specifications for jet fuel.

PROPERTY REQUIREMENTS

Because the jet engine was free of the demanding need for high-octane fuel, in the early days of the jet-engine development it was thought that it could use practically any liquid fuel. However, subsequent experience proved this to be untrue, as a number of potential problem areas indicated that control of fuel properties, reflecting both bulk and trace components, were important for satisfactory use. Over the years these important property requirements were translated into specification requirements that put restrictions on what is acceptable as jet fuel.

During the early jet-engine development work, it was recognized that the combustion system of the engine would be a critical component. Fuel-combustion-related properties are controlled via limits on the total concentration of aromatic-type compounds as well as the concentration of condensed ring aromatic compounds (i.e., naphthalenes). Since these types of hydrocarbon compounds tend to burn with higher levels of flame radiation than do other hydrocarbon compound types, it makes combustor wall cooling more difficult. In addition, specific combustion performance prediction tests such, as the smoke point test and the luminometer number test, were developed and added to specification requirements. Other concerns are energy content, density, and volatility. The minimum energy content of the fuel is specified for range considerations. The density of the fuel controls the weight of fuel that can be carried in a given volume. The volatility or boiling range of the fuel is controlled because it impacts on a number of properties. Lower boiling fuels are easier to use, either when starting a cold engine or when attempting to relight an engine at altitude, but early attempts to use very light fuels encountered problems with fuel boiling off from the vented wing tanks at the low pressure at higher altitudes. Subsequent experience also demonstrated the increased safety risks inherent in low-flash-point fuels in either civil or combat military use. Specifications include controls on the boiling range of the fuel as well as a flash point test measurement. The potential for harmful wax crystal formation in aircraft wing tanks at low temperatures in flight is controlled via the inclusion of a freeze point test requirement in all specifications. The viscosity of the fuel at low temperatures is limited to ensure the proper operation of fuel injection nozzles during low-temperature start-up. Problems with the stability of the fuel in storage leading to unwanted gums and deposit formation were anticipated in early fuel development work and led to restrictions on the olefin (unsaturated hydrocarbon) content of the fuel in specifications. Subsequent operational experience also discovered stability problems in flight caused by the exposure of the jet fuel to hot metal surfaces where reactions of the fuel with the dissolved oxygen in the fuel led to deposit formation in critical components such as the fuel nozzle, heat exchanger surfaces, and narrow-tolerance moving components in fuel control units. These high-temperature-thermal stability problems led to the development of tests designed to simulate the high-temperature exposure of the fuel, and all specifications require the fuel to pass such a test.

The use of high-sulfur-content fuels could enhance undesirable carbon-forming tendencies in the engine combustion chamber as well as result in higher amounts of corrosive sulfur oxides in the combustion gases. Mercaptans (a type of sulfur compound) cause odor problems and can attack some fuel system elastomers. Both the concentration of total sulfur compounds as well as the concentration of mercaptan sulfur compounds are controlled in

Net Heat of Combustion, MJ/kg	42.8 Min
Boiling Range, 10% Recovered, ˚C	205 Max
Boiling Range, Final Boiling Point, ˚C	300 Max
Flash Point, ˚C	38 Min
Density, kg/m3	775 to 840
Freezing (wax appearance) Point, ˚C	-40 Max

Table 1.
Important Jet Fuel Properties

A Twin Otter turboprop airplane refuels at a cache before heading for the North Pole. (Corbis-Bettmann)

specifications. The corrosivity of the fuel toward metals caused by the presence of elemental sulfur or hydrogen sulfide is controlled by the use of tests such as the copper strip corrosive test. Acidic compounds present in the fuel, such as organic acids or phenols, are controlled by a total acidity test.

Another area of importance is contamination. Jet fuels are tested for the presence of heavier fuel contamination by use of an existent gum test, which detects the presence of heavier hydrocarbons from other products. Testing also is carried out to detect the presence of excessive levels of undissolved water and solids, as well as for surfactants that can adversely affect the ability of filters and coalescers to remove dirt and water from the fuel.

Additives also are used to enhance jet fuel quality in a manner similar to that of gasoline, but unlike gasoline, are tightly controlled. Only additives specifically cited in a specification can be used within allowed limits. The mandated or permitted use of additives varies somewhat in different specifications, with military fuels tending to the greater use of additives compared to civil fuels. A static dissipater additive is used in many fuels to enhance the rapid dissipation of any electrostatic charge in the fuel created by the microfiltration used for dirt removal. To prevent the formation of deleterious hydroperoxides during prolonged fuel storage, many specifications require that an antioxidant (a compound that slows down or prevents oxidation) additive be added to fuels that have been hydrotreated. This must be done because the natural antioxidants present in the fuel that were unavoidably removed.

Another additive used is a metal deactivator to chemically deactivate any catalytic metals such as copper accidentally dissolved in the fuel from metal surfaces. Uless they are chemically deactivated, dissolved metals cause the loss of good stability quality.

Corrosion inhibitor/lubricity improvement additives are used particularly in military fuel for the dual purpose of passivating metal surfaces and improving the lubricating properties of the fuel in equipment such as fuel pumps. The military also specifies the use of a fuel system icing inhibitor as an additive to prevent filter blocking by ice crystal formation, because military aircraft tend not to use fuel line filter heaters, which are standard equipment on civil aircraft.

COMPOSITION AND MANUFACTURING

Aviation turbine fuels are produced in refineries primarily using petroleum crude oil as the sole starting material. The exceptions are Canada, which uses some liquids produced from tar sands, and South Africa, which uses some liquids produced from coal. The overwhelming percentage of chemical compounds—present in jet fuel are hydrocarbon compounds, that is, compounds composed of carbon and hydrogen. These hydrocarbon compounds include branched and normal paraffins; single-ring and multiring cycloparaffins, which are also called naphthenes; and single-ring and multiring aromatics, hydroaromatics, and olefins. The distribution of hydrocarbon compound types varies considerably, primarily depending on crude source. Heteroatom compounds, which are hydrocarbon compounds that also contain sulfur, nitrogen, or oxygen, are present at trace levels, and are important because heteroatoms can have a disproportionate effect on fuel properties.

Much jet fuel is produced by simply distilling a kerosene fraction from the crude oil followed by some form of additional processing. The initial boiling points for the distillation are generally set to produce a jet fuel that meets the flash-point requirement, and the final boiling points are set to meet requirements such as freeze-point, smoke-point or naphthalene content. Jet fuel often is blended from a number of streams. In addition to simply distilled kerosene fractions, blend stocks are produced from heavier crude oil fractions or refinery product streams by breaking them down into lower-boiling fractions. The processing steps used to prepare blend stocks after distillation vary considerably, depending on factors such as crude oil type, refinery capabilities, and specification requirements. Crude oils, whose kerosene fractions are low in total sulfur content, can be chemically processed to reduce mercaptan sulfur or organic acid content. For example, a kerosene with a high organic acid content but a low total and mercaptan sulfur content can be simply treated with caustic (sodium hydroxide) to lower the acid level. Similarly, a jet fuel blend stock low in total sulfur but too high in mercaptan sulfur can be chemically treated in a so-called sweetening process, which converts the odorous mercaptan sulfur compounds into odor-free disulfide compounds. Chemical treatment is often followed by passage through both a salt drier to lower water levels, and a clay adsorption bed to remove any trace impurities still present.

Another type of processing employed is treatment in a catalytic unit with hydrogen at elevated temperatures and pressures. Catalytic processing is used, for example, when higher total sulfur levels require total sulfur removal, which cannot be achieved by chemical treatment. In addition, catalytic treatment with hydrogen, depending on the process conditions, can be used to break down heavier fractions into the kerosene range. More severe processing conditions such as higher pressures used to affect a boiling point reduction, will also generally extensively remove heteroatoms and markedly lower the level of olefins. Jet fuel normally sells at a premium compared to other distillates, and reflects the cost of crude oil.

Jet fuel is shipped in a highly complex system designed to prevent or eliminate excess water, particulates such as dirt and rust, microbial growths, and contamination from other products in the fuel being delivered into aircraft. Transportation may involve shipment in pipelines, railcars, barges, tankers, and/or trucks. Techniques employed include dedicated storage and transportation, the use of filters to remove particulates, and the use of coalescers and water-adsorbing media to remove water. The elimination of water will prevent microbial growth.

ENVIRONMENTAL ISSUES

As for all hydrocarbon fuels, the combustion of jet fuel produces carbon dioxide and water. Turbine engines are designed to be highly efficient and to produce low levels of unburned hydrocarbons and carbon monoxide. Nitrogen oxides and sulfur oxides also are emitted from the turbine engine. There is little organic nitrogen in the fuel, and oxides of nitrogen are produced from the nitrogen and oxygen in the air during the combustion process. As a result, control of nitrogen oxides emissions is essentially an engine combustor design issue. Sulfur oxides are produced from the low levels of sulfur compounds present in jet fuel during the combustion process, and thus control of sulfur oxides is essentially a fuel-related issue. Only a small fraction of all sulfur oxides emissions are produced by the combustion of jet fuel because of the relatively low use of jet fuel compared to total fossil-fuel combustion. However, jet-fuel-produced sulfur oxides emissions are unique because aircraft engines are the only source emitting these species directly into the upper troposphere and lower stratosphere an issue of growing interest to atmospheric and climate change researchers.

William F. Taylor

See also: Aircraft; Air Pollution; Air Travel; Climatic Effects; Gasoline and Additives; Gasoline Engines; Kerosene; Military Energy Use, Modern Aspects of; Transportation, Evaluation of Energy Use and.

BIBLIOGRAPHY

ASTM. (1998). "D1655 Standard Specification for Aviation Turbine Fuels." In *1998 Annual Book of ASTM Standards.* West Conshohocken, PA: Author.

Coordinating Research Council. (1983). *Handbook of Aviation Fuel Properties.* Atlanta, GA: Author.

Dukek, W. G. (1969). "Milestones in Aviation Fuels." AIAA Paper 69-779, July 14.

Dyroff, G. V., ed. (1993). "Aviation Fuels." In *Manual on Significance of Tests for Petroleum Products,* 6th ed. Philadelphia ASTM.

Smith, M. (1970). *Aviation Fuels.* Henley-on-Thames, Oxfordshire, Eng.: G. T. Foulis.

Taylor, W. F. (1997). "Jet Fuel Chemistry and Formulation." In *Aviation Fuels with Improved Fire Safety—National Research Council Proceedings.* Washington, DC: National Academy Press.

Waite, R., ed. (1995). *Manual of Aviation Fuel Quality Control Procedures.* West Conshohocken, PA: ASTM.

B

BATTERIES

A battery is a series of electrochemical cells. Electrochemical cells are devices that, whenever in use, can continuously and directly convert chemical energy into electrical energy.

The demand for energy from batteries is enormous. In the United States alone, battery sales exceeded $4.3 billion in 1997. Batteries have become an ubiquitous aspect of modern life, providing energy for almost every type of technology imaginable, from small-capacity devices such as watches, electronic toys, pocket radios, portable computers, cardiac pacemakers, hearing aids, and smart cards to the large-capacity applications such as motor vehicles, satellites, remote off-grid equipment, and uninterruptible backup power systems for hospitals.

Because of the ever-growing demand for a wide array of cordless electrical technology, and the markedly increased use of electronic devices and information technologies using integrated circuits, battery technology advanced remarkably in the latter half of the twentieth century. In particular, major advances in construction technology and new materials applications have resulted in smaller yet more powerful batteries. It has largely been the result of new combinations of more refined electrochemical materials that have made batteries lighter, thinner, and more efficient. Probably the most noteworthy advances have been alkaline batteries for consumer products, lithium rechargeable batteries for electronic technology, and nickel-hydrogen batteries for use in spacecraft.

These advances are likely to continue as industry and government liberally invest in battery research and development. Spending priorities of industry and government are far different, however. Whereas the majority of spending by industry is focused on improving battery technology for communication and computer technology, the majority of government-funded research is for military, spacecraft, transportation, and distributed energy applications.

BATTERY BASICS

All flashlight batteries, button batteries, compact rechargeable batteries and vehicle storage batteries operate under the same basic principles. An electrochemical cell is constructed of two chemicals with different electron-attracting capabilities. Called an electrochemical couple, these two chemicals, immersed in an electrolyte (material that carries the flow of energy between electrodes), are connected to each other through an external circuit.

The chemical process that produces an electrical current from chemical energy is called an oxidation-reduction reaction. The oxidation-reduction reaction in a battery involves the loss of electrons by one compound (oxidation) and the gain of electrons (reduction) by another compound. Electrons are released from one part of the battery and the external circuit allows the electrons to flow from that part to another part of the battery. In any battery, current flows from the anode to the cathode. The anode is the electrode where positive current enters the device, which means it releases electrons to the external circuit. The cathode, or “positive” terminal of the battery, is where positive current leaves the device, which means this is where external electrons are taken from the external circuit.

To generate electrical energy from chemical energy, the reactant molecules must be separated into oppositely charged ions (electrically charged atoms or groups of atoms) and the exothermic reaction (gives off heat) between the two reactants requires the proper conditions and hardware to drive an electric current through an external circuit. An electrolyte, which can be a free liquid, a paste, or even a rigid solid, is the ionic conductor serving as the charge transfer medium.

The ions move between electrodes in the electrolyte due to voltage potential gradients. The velocity of these chemical currents increases with temperature. Hence, electrolytic conductivity increases as temperature goes up. This is the opposite of electrical currents in metallic conductors, which increase as the temperature goes down.

The external circuit provides a path for the excess electrons at the negative electrode to move toward the positive electrode. Although the flow of electrons would seem to cancel out the two charges, chemical processes in the battery build the charges up as fast as the depletion rate. When the electrodes can no longer pass electrons between them, the battery "dies."

During charging, energy from electric current is stored on the plates of the cell. Substances on these plates are converted to new substances taking on a higher energy. Then when the battery is connected to any load with electrical resistance, the substances on the battery plates begin to retransform themselves back to the earlier state, producing electricity as a result of the electrochemical reactions.

The quantity of electric charge is measured in coulombs, and the unit of electric current—the number of coulombs per second that go past any point—is the ampere (A), named after French physicist André Marie Ampère:

$$1 \text{ ampere} = 1 \text{ coulomb/second}$$

Batteries are rated by ampere-hours, which is the total amount of charge that can be delivered. More specifically, it is a measure of the number of electrons that can be released at the anode and accepted at the cathode. It is also related to the energy content or capacity of the battery. If a battery has a capacity of 90 A-h, it can supply a current of 90 A for one hour, 30 A for three hours, or 1 A for 90 hours. The greater the current, the shorter the period the battery can supply it; likewise, the less the current, the longer the period the battery can supply it.

A battery produces a potential difference between its two electrodes that is brought about by a decrease in potential energy as one coulomb moves from the negative to positive electrode. Measured in volts, it is equal to the energy difference per unit charge. Voltage measures the "push" behind a current and is determined by the oxidation-reduction reactions that take place inside each cell. The better the oxidation-reduction characteristics, the higher the potential difference and the greater the current.

There is a huge range of cell capacities. At one extreme are miniature batteries providing milliamp-hours; at the other extreme are huge submarine batteries that can provide five million watt-hours. Battery cell voltages for any single cells fall within in a narrow range of one to two volts; cells are stacked in series to get higher voltages. The connection is always negative electrode to positive electrode so that each battery supplies more push to the electrons flowing through them. A typical golf cart uses three or four lead-acid batteries in series, the 1995 General Motors Impact electric car uses thirty-two lead-acid batteries in series, and the largest utility battery storage system, California Edison's substation at Chino, California connects hundreds of lead-acid batteries in series. In all, this substation houses 8,000 batteries capable of deep discharge (near-complete discharge without hampering the battery's capacity to be recharged), providing 40,000 kilowatt-hours of storage.

Another important concept for batteries is resistance. Expressed in ohms, resistance is what limits current. For example, if a 12V car battery were connected to a circuit with 4 ohms resistance, the current would be 3 A. If the battery had a capacity of 90 A-h, it would supply 3 A for 30 hours. (Resistance of 4 ohms here refers to external resistance of the battery and the headlight circuit.) The greater the resistance of the circuit and cell itself, the less the current for a given applied voltage.

The current from a cell is determined by the voltage level of the cell and the resistance of the total circuit, including that of the cell itself. If large amounts of current are needed to flow in the circuit, a low-resistance cell and circuit are required. The total area of the electrodes determine the maximum current. The amount of voltage that a battery can actually deliver is always less than the battery's advertised maximum deliverable voltage. That is because once current begins, cell voltage decreases because of the slowness of the chemical process at the electrodes and the voltage drop across the internal resistance of the cell.

Ohm's law, named after German physicist Georg Simon Ohm, explains the relationship between cur-

HOW BAD BATTERIES LEAD TO BETTER EFFICIENCY

The pace of innovations in batteries has not matched the remarkably fast evolution in the development of portable electronic technology. Batteries are widely regarded as the weak point slowing the rapid innovations in portable technology. Chief among consumer complaints are that batteries do not last long enough and that they do not last as long as the manufacturer claims they will.

Portable electronic manufacturers had the foresight to realize that battery technology was not keeping up with innovation in electronic technology, and it probably never would. To satisfy consumer demand for smaller and more powerful electronic devices that could go a longer time between charges, improvements in the energy efficiency of the devices themselves were required.

Consumer preference to "unplug" from the wall has been responsible for making all electronic equipment more energy-stingy. The growing demand for portable energy for ever-smaller phones, laptop computers, and DVD players has been the driving force behind the tremendous energy efficiency improvements in electronic equipment. Although the market for laptops, notebooks and palm-size computers is only a small fraction of what it is for desktops, much of the smart energy-saving electronics developed for these devices have been incorporated in desktop computers. In the United States, companies and consumers are saving millions of dollars in energy costs since new computers, mostly replacement systems, are being plugged in at a rate of over thirty million a year. Just as importantly, society receives the environmental benefit of reducing the need for additional electric power plants.

This rapid pace of innovation in electronics will continue. With each advance, electronic circuits become smaller, more sophisticated, more reliable, and ever more energy-stingy. In 2000, Transmeta Corporation introduced the Crusoe microprocessor that is specifically designed as a low-power option for notebooks and Internet appliances. Transmeta claims that a notebook equipped with a 500 to 700 MHz clock speed Crusoe chip requires slightly more than 1 watt of power on average—far less power than the average microprocessor's 6 to 10 watts. The chip garners additional savings since it emits less heat, eliminating the need for a cooling fan.

The Crusoe microprocessor and similar microprocessors coming from other chip manufacturers mark a revolutionary advance: all-day computing with full PC capabilities from a single smaller and lighter battery. As these more energy-stingy microprocessors migrate to the "plugged in" office world, it will help slow the rate of electricity demand for office equipment.

rent, potential difference, and resistance in a given metallic conductor: voltage = current × resistance (when voltage and current are constant). In other words, current depends on resistance and voltage applied. By simply altering the resistance of a battery, it can be designed for either fast current drain (low resistance) or slow current drain (high resistance) operation.

To ensure the safety and reliability of battery-using technology, the American National Standards Institute (ANSI), the International Electrotechnical Commission (IEC) and the International Standards Organization (ISO) have developed standards for battery sizes, voltages, and amperages. These standards are now accepted on a world-wide basis.

TYPES OF BATTERIES

There are two major types of electrochemical cells: primary batteries and secondary, or storage, batteries. Primary battery construction allows for only one continuous or intermittent discharge; secondary battery construction, on the other hand, allows for recharging as well. Since the charging process is the reverse of the discharge process, the electrode reactions in secondary batteries must be reversible.

Primary batteries include dry, wet, and solid electrolyte. The term "dry cell" is actually a misnomer, for dry cells contain a liquid electrolyte, yet the electrolyte is sealed in such a way that no free liquid is present. These cells are sealed to prevent damage to persons or materials that would result from seepage of the electrolyte or reaction products. Dry cells are typically found in products like flashlights, radios, and cameras. Wet cells use a free-flowing aqueous electrolyte. Because of the large capacity and moderately high currents of wet cells, the primary uses are for signal systems for highway, railway and marine applications. Solid electrolyte batteries use electrolytes of crystalline salts. Designed for very slow drain long-term operations, they are popular for certain kinds of electronic devices.

Secondary batteries consist of a series of electrochemical cells. The most popular types are the lead-acid type used for starting, lighting, and electrical systems in motor vehicles and the small rechargeable batteries used in laptops, camcorders, digital phones, and portable electronic appliances.

Because of the great diversity of applications requiring batteries, and the very distinct requirements for those applications, there is an array of batteries on the market in a wide variety of materials, sizes and capacities. Some estimates put the number of battery varieties at over 7,000. Decisions about batteries are not simple for battery manufacturers, designers of technology, or consumers. Many factors must be taken into account in choosing a battery. For any given application, important criteria may include longer life, high voltage, fast-draining, durability, lower weight, salvageability, recharging speed and control, number of charge-and-discharge cycles, and lower cost (material and construction costs).

Manufacturers often feel compelled to be very vague about how well a battery will perform because of the uncertainty about how the battery will be used. Technology manufacturers invest heavily in marketing research before deciding what type of battery to provide with a new product. Sometimes it is a clear-cut decision; other times there is no one best choice for all potential users. Perhaps the most important issue is whether a product will be used continuously or intermittently. For instance, selecting a battery for a DVD player is much easier than a cellular phone. The DVD player is likely to be used continuously by the vast majority of buyers; a cellular phone will be used very differently in a wide mix of continuous and intermittent uses.

The life span of a battery is also dependent on the power needs of a product. Some batteries produce a lot of power for a relatively short time before fully discharging, while others are designed to provide less peak power but more sustained power for a very low drain rate. Other important variables affecting battery lifespan are the design and efficiency of the device being used and the conditions of use. In particular, exposing a battery to excessive heat can dramatically curtail the length of a battery's life.

Zinc-Manganese Dioxide Primary Dry Cells

Zinc-manganese dioxide cells are the most prominent commercial battery. Over ten billion are manufactured each year for electronic equipment, camera flashes, flashlights, and toys. Initial voltage for such a cell starts in the range of 1.58 to 1.7 volts, and its useful life ends after voltage declines to about 0.8 volt. The three types are the zinc-carbon cell, the zinc chloride cell and the alkaline cell. The zinc-carbon cell was invented by the French engineer Georges Leclanché in 1866, and thus also referred to as the Leclanché cell. Because zinc and manganese were readily available at a low cost, the Leclanché cell immediately became a commercial success. It remains the least expensive general-purpose dry cell and still enjoys worldwide use.

Figure 1 shows a cross section of a Leclanché cell battery. The outer casing, or "cup," is the anode (negative electrode). It is primarily made of zinc, but also may include small amounts of lead, cadmium, and mercury in the alloy. Because of the environmental hazards of mercury, manufacturers in the late twentieth century dramatically reduced or eliminated mercury in the zinc alloy used in disposable batteries by developing better zinc alloys, reducing impurities in the electrolyte, and adding corrosion inhibitors. The moist paste electrolyte consists of ammonium chloride, zinc chloride, and manganese dioxide. Formed around the porous carbon cathode (positive electrode) rod is a pasty mixture of manganese dioxide and carbon black.

The zinc chloride cell, which was first patented in 1899, is actually an adaptation of the Leclanché cell. The major innovation was the development of plastic seals that permitted the replacement of ammonium chloride in the electrolyte.

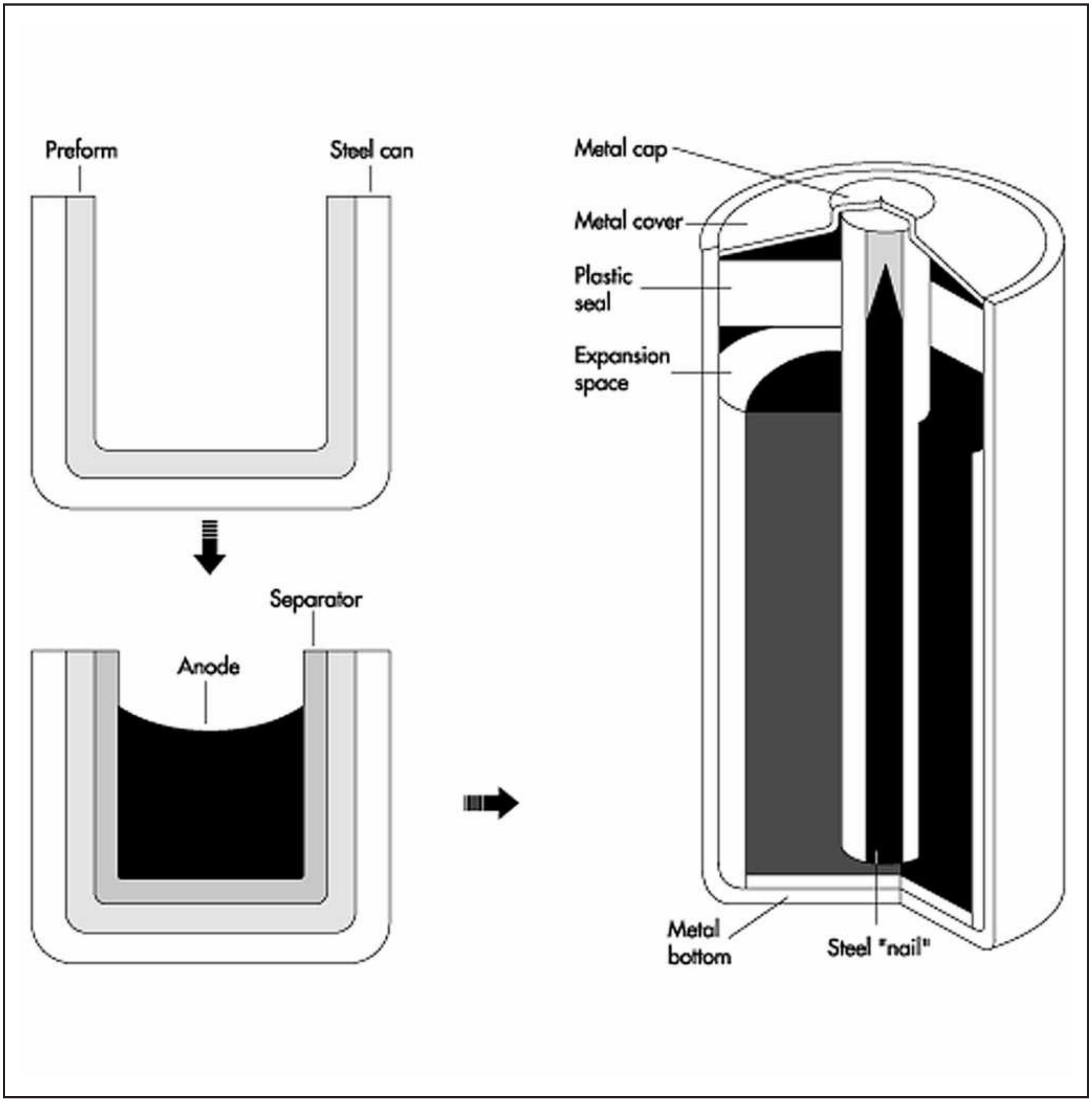


Figure 1.
Cross-section of a Leclanché cell battery. (Gale Group, Inc.)

The alkaline cell appeared commercially in the 1950s. It differs from the other zinc-manganese dioxide cells in that the electrolyte is alkaline, which primarily reduces corrosion and allows for a much longer battery life. Although these batteries look quite simple, manufacturing high-performance alkaline batteries is tedious. Many contain over twenty-five different components and materials and require up to forty manufacturing steps, which makes them more expensive than standard batteries.

Alkaline cells offer the highest energy density (more energy per given volume) of any zinc-manganese dioxide cell, and the manufacturers continue to improve on performance. In 1998, Duracell introduced the Duracell Ultra, which they claimed will last 50 percent longer than the standard Duracell. And Eveready claims the Energizer Advanced Formula of 1998 provides 100 percent better performance than the standard Energizer of 1996.

Compact Alkaline Rechargeable Batteries

A variety of different rechargeable batteries provides the power for camcorders, laptops, digital phones and other portable electronic equipment. The following are some of the most common batteries used for these applications.

Nickel-Cadmium Cells. Nickel (hydroxide)-cadmium, or Nicad, is the most common rechargeable battery. These batteries are very durable and work well in extreme temperatures. Because the sealed cells use “jelly roll” electrodes that minimize resistance, high current can be delivered efficiently, which allows for quick recharging and works well for technology requiring high current. Other advantages of Nicad batteries are the better tolerance to overcharging and overdischarging and the ability to withstand 700 to 750 charge-and-discharge cycles.

However, Nicad batteries have some major drawbacks. In comparison to primary batteries and lead-acid batteries, Nicad batteries are heavier and have a relatively low energy density. They also require special attention from users to ensure full discharge each cycle before recharging to prevent the capacity-lowering memory effect (hysteresis) that also shortens battery life. When recharging a Nicad battery that is only partially discharged, chemical compounds around the cadmium anode are created that interfere with recharging. If a three-hour Nicad battery is repeatedly recharged after operating only one hour, the memory effect will set in and the battery will last for only one hour before it needs to be recharged again. Sometimes the memory effect can be erased by fully discharging the battery before recharging. The best way to prevent the memory effect is to closely monitor battery use and carry a backup battery. This makes it possible to fully drain each Nicad battery before recharging, and operate the laptop or camcorder for extended periods.

Nickel-metal hydride is a popular alternative to Nicad batteries since they are capable of operating 75 percent longer after each charge, are less likely to suffer memory effects, and pose less of an environmental disposal problem. The difference between nickel-metal and Nicad batteries is that the negative plate in sealed Nicad cells is replaced in the nickel-metal battery with hydrogen absorbed in a metal alloy. Hydrogen-absorbing alloys, which provide a reversible hydrogen sink that can operate at ordinary temperatures and pressures, were first used in the late 1960s in hydrogen-nickel oxide secondary cells for communication equipment and satellites. Nickel-metal hydride batteries are generally more expensive and offer a more uncertain life expectancy of between 500 and 1,000 charge-and-discharge cycles.

Lithium-Ion Cells. Lithium-ion cells and the newer alternative, lithium-ion-polymer, can usually run much longer on a charge than comparable-size Nicad and nickel-metal hydride batteries. “Usually” is the key word here since it depends on the battery’s application. If the product using the battery requires low levels of sustained current, the lithium battery will perform very well; however, for high-power technology, lithium cells do not perform as well as Nicad or nickel-metal hydride batteries.

Because of lithium’s low density and high standard potential difference (good oxidation reduction characteristics), cells using lithium at the anode have a very high energy density relative to lead, nickel and even zinc. Its high cost limits use to the more sophisticated and expensive electronic equipment.

The lithium-ion-polymer battery, which uses a cathode that contains lithium instead of cobalt, is likely to eventually replace lithium-ion. Lithium-ion-polymer batteries boast a longer life expectancy (over 500 charge-and-discharge cycles as opposed to around 400), much more versatility (they are flat and flexible and can be cut to fit almost any shape), and better safety (far less likely to vent flames while recharging).

For “smart” cards, micro-robots and small precision instruments, thin laminated micro-cells are being developed. Some of these developmental thin-film devices—using an electrolyte of lithium, a copper cathode, and lithium again for the electrode—can charge and discharge up to 3 volts, and can be expected to tolerate up to 1,000 charge-and-discharge cycles.

Rechargeable Alkaline Manganese Cells. Rechargeable alkaline manganese cells began to reach the market in the 1990s. Compared to Nicad batteries, these batteries are less expensive, weigh less (because no heavy metals are used), and boast higher energy density. The major disadvantage is its more limited life, particularly if deeply discharged.

The major innovation that led to alkaline manganese rechargeable batteries was the ability to prevent discharge of manganese dioxide beyond the 1 electron level. This has been accomplished by either limiting discharge to 0.9 volt or by using a zinc-limited anode. Other innovations that made alkaline rechargeables possible were the development of a special porous zinc anode and the development of a laminated separator that cannot be penetrated by zinc.

Button Primary Cells

The button cells that provide the energy for watches, electronic calculators, hearing aids, and pacemakers are commonly alkaline systems of the silver oxide-zinc or mercuric oxide-zinc variety. These alkaline systems provide a very high energy density, approximately four times greater than that of the alkaline zinc-manganese dioxide battery.

Lithium batteries began to replace mercury-zinc batteries in the late 1970s since lithium is the lightest of all the metals and loses electrons easily. Before the introduction of lithium batteries, pacemakers usually failed within one to two years. Since the pulse generator could not be sealed tightly enough, body fluids leaking in caused an early short circuit, significantly shortening the life expectancy of five years. Small and lightweight lithium-iodine batteries, which remain most common, dramatically cut the weight of the pulse generator to about 1 ounce for battery and circuitry combined. This extended the generator's life expectancy to almost ten years.

Lithium primary cells have also been introduced up to the AA-size. These "voltage compatible" cells have a working voltage of 1.5 V, and can deliver up to four times the energy of a comparable alkaline cell.

Lead-Acid Secondary Cells

Lead-acid batteries have been a workhorse battery for nearly a century. Over 300 million units are manufactured each year. One of the early uses of lead-acid batteries was for electric cars early in the twentieth century. The development of the electric starter made it possible for battery power to replace hand crank power for starting automobiles in the 1920s. First-generation portable radios with vacuum tubes and discrete transistors, which were critical for military operations, used very large lead acid batteries. The development of very large-scale integrated circuits made it possible to vastly reduce the size of these large and bulky portable radios and use much smaller batteries, since energy needs also were significantly reduced.

SMART BATTERY TECHNOLOGY

Smart battery technology started appearing on the market in the mid-1990s, mostly in high-end electronic products such as laptop computers and camcorders. Sensors send temperature, current output, and battery voltage data to microprocessors, which prevent improper charging. These features are helpful to users who do not care to take the time to regularly monitor their batteries; for those who do, they offer a way to eliminate the guesswork from recharging. It also reduces anxiety for the user and lessens the need for emergency backup batteries. For those relying on off-grid solar or wind energy systems, smart battery technology also provides critical monitoring for the times when the owner of the equipment is away from the site.

Smart batteries have been aided by tremendous advances in smart electronic circuitry. Found in everything from laptops to pacemakers, this circuitry has improved the operating time of everything from pacemakers to laptops. In pacemakers, this circuitry not only senses a patient's own heartbeat, and fires only when necessary, but also senses other body signals such as rises in body temperature and respiratory rate from exercise and respond by increasing the heart rate to meet the body's changing needs. In laptop computers, smart electronic circuitry entails power saving functions that shut off the monitor and put the computer processor in a sleep mode when not being used. Eventually most electronic technology probably will come with specialized computer chips devoted to power management, monitoring charge levels and optimizing the charging process.

Lead-acid batteries remain popular because of their capability to service high and low current demand, produce high voltage, provide capacity up to 100 A-h, and recharge well. Moreover, the lead-acid battery has important material and construction advantages, such as simple fabrication of lead components, the low cost of materials (lead is abundant and much less expensive than nickel or lithium), and excellent salvageability because of lead's low melting point.

The lead-acid battery is constructed of lead plates or grids. The negative electrode grid is coated in lead and other additives such as calcium lignosulfate, and the positive electrode grid is coated in lead oxide. Between them is the electrolyte of sulfuric acid. During discharge, the lead dioxide at the positive electrode combines with the electrolyte to produce lead sulfate and water. Meanwhile, the lead at the negative electrode combines with the sulfuric acid to produce lead sulfate and hydrogen, which replaces the hydrogen consumed at the positive electrode. The reversible reaction:

$$Pb + PbO_2 + 2\ H_2SO_4 \rightleftarrows 2\ PbSO_4 + 2\ H_20 + \text{electrical energy}$$

When the above reaction is read from right to left, it reflects what happens when the cell is discharging; read from left to right, it represents what happens when the cell is charging.

A 12-V storage battery, the type in automobiles, consists of six 2-V cells connected in series. Although the theoretical potential of a standard lead-acid cell is 1.92 volts, the actual potential can be raised to 2 V by properly adjusting the concentrations of the reactants. These batteries provide high current, and are designed to tolerate thousands of shallow-depth discharges over a period of several years. Traditionally, for automobiles and trucks, three or six cells have been connected in series to make a 6- or 12-volt battery, respectively. Manufacturers are likely to begin equipping automobiles and trucks with higher voltage batteries to accommodate increased electricity requirements and to improve the efficiency of air conditioning systems.

Unlike the automobile-type battery that is quite portable, the stationary lead-acid batteries that provide uninterruptible power to hospitals and other important facilities are not. Some may weigh over several tons because of the much heavier grid structure and other features to extend life expectancy and improve deep discharge capabilities.

Weight is usually inconsequential for stationary applications but of foremost concern for electric vehicles. Vehicle weight is detrimental in transportation because the greater the vehicle weight, the greater the energy consumed overcoming resistance. In fact, reductions in vehicle weight was one of the major reasons for the doubling of the fuel efficiency for the U.S. vehicle fleet from 1973 to 1993. Electric vehicles also present other concerns for battery makers such as charge time and vehicle range on a single charge.

ADVANCED BATTERIES FOR ELECTRIC VEHICLES

To hasten development of batteries for electric vehicles, Chrysler, Ford, and General Motors formed the U.S. Advanced Battery Consortium (USABC). In 1991 USABC, battery manufacturers, the Electric Power Research Institute (EPRI), and the U.S. Department of Energy (DOE) launched a joint research effort to identify, develop and license promising battery technology for electric vehicles—vehicles with the range, performance and similar costs of gasoline-powered vehicles.

Whatever the battery construction and mix of materials being considered, the technical challenges facing developers are

1. Improve electrolyte material, using better conductors that are still chemically compatible with the electrodes.
2. Reduce the amount of change in the electrolyte and electrodes per charge-and-discharge cycle to extend life expectancy.
3. Improve the interface of the electrodes with the electrolyte by enlarging the effective surface area of the electrodes.
4. Identify electrode and electrolyte material that is inexpensive and readily available in order to achieve a low-cost battery.
5. Improve the energy density in terms of unit mass, unit volume, or both.

Advanced lead-acid batteries have resulted in higher energy density, longer life expectancy, and weight reductions. Energy density has improved by making use of the "starved" electrolyte concept in which a fiberglass and polyethylene separator immobilizes the electrolyte, allowing for the passage of oxygen from the anode to the cathode where it recombines with hydrogen to form water. Besides improving energy density, this concept also eliminated the need to add water and vent hydrogen. Weight reduction has been achieved by replacing the electrode grids of lead-antimony and lead-arsenic with lead-cadmium. Lead-cadmium also has extend the life expectancy.

To further reduce weight and improve energy density, several companies are developing thin lead film electrodes in a spiral-wound construction with glass fiber separators. Already on the market for cordless electric tools, this battery technology may eventually be used in electric vehicles.

However, since the chemical reactions for lead-acid batteries cannot sustain high current or voltage continuously during charge, fast charging remains a major problem. Charging has to be deliberate because the lead grids heat up quickly and cool slowly. When excessive heating or "gassing" (liquids inside the battery becoming a gas) occurs during charging, hydrogen is released from the battery's vent cap, which reduces battery performance and life.

Other Advanced Batteries

Despite the many virtues of lead-acid batteries, there are major limitations: range, long recharge times, and more frequent replacement than most potential electric vehicle buyers are likely to accept. Much research is taking place to develop alternatives that store more energy per given volume and mass, cost less to manufacture and operate, weigh less, and last longer. Some of the most promising materials include sodium-sulfur, zinc-bromine, nickel, and lithium.

Sodium-Sulfur Batteries. The sodium-sulfur battery consists of molten sodium at the anode, molten sulfur at the cathode, and a solid electrolyte of a material that allows for the passage of sodium only. For the solid electrolyte to be sufficiently conductive and to keep the sodium and sulfur in a liquid state, sodium-sulfur cells must operate at 300°C to 350°C (570°F to 660°F). There has been great interest in this technology because sodium and sulfur are widely available and inexpensive, and each cell can deliver up to 2.3 volts.

Though sodium-sulfur batteries have been under development for many years, major problems still exists with material stability. It is likely that the first commercial uses of this battery will not be for electric vehicles. Sodium-sulfur storage batteries may be more well-suited for hybrid electric vehicles or as part of a distributed energy resources system to provide power in remote areas or to help meet municipal peak power requirements.

Zinc-Bromide. Unlike sodium-sulfur batteries, zinc-bromide batteries operate at ordinary temperatures. Although they use low-cost, readily available components, these batteries are much more complex than lead-acid and sodium-sulfur batteries because they include pumps and reservoirs to store and circulate the zinc-bromide electrolyte within the cell.

Nickel-Hydrogen, Nickel-Iron, and Nickel-Metal Hydride. First developed for communication satellites in the early 1970s, nickel-hydrogen batteries are durable, require low maintenance, and have a long life expectancy. The major disadvantage is the high initial cost. For these batteries to be a viable option for electric vehicles, mass production techniques will have to be developed to reduce the cost.

A more appropriate battery for transportation applications is probably a nickel-iron or nickel-metal hydride battery. These batteries are not as susceptible to heat and gassing as lead-acid batteries, so they can better withstand high current or high voltage charges that can dramatically shorten charging time.

In the mid-1990s, Chrysler and EPRI developed engineering prototypes of an electric minivan called the TEVan using a nickel-iron battery. Compared to a lead-acid battery, this battery was lighter, stored more energy, and lasted twice as long, resulting in a vehicle range of about 60 to 100 miles and a top speed of 65 mph. However, nickel-iron are much more expensive than lead-acid batteries and about half as energy efficient due to higher internal resistance. (This is not just the battery, but the battery-charger-vehicle combination according to Model Year 1999 EPA Fuel Economy Guide for vehicles offering both battery types.) In warmer climates, nickel-metal hydride energy efficiency drops even farther since there is a need to expend battery capacity cooling the battery.

In the late 1990s, Toyota began leasing the RAV4-EV sports utility vehicle using twenty-four nickel metal hydride batteries. Rated at 288 volts, these vehicles achieved a top speed of 78 mph, and a combined city/highway driving range of around 125 miles. The electric RAV4 weighed over 500 pounds more than the internal combustion engine version because the batteries weighed over 900 pounds. Nickel-metal hydride batteries are also proving to be the technology of choice for hybrid vehicles such as the Honda Insight because high current and voltage can be delivered very quickly and efficiently for the initial drive-away and during acceleration.

Lithium-Ion and Lithium Polymer. Major research efforts are taking place in the United States, Japan, and Europe to develop lithium-ion technology for potential use in electric vehicles. Sony Corporation, which was the first electronics manufacturer to offer lithium batteries for top-end video cameras, teamed up with Nissan in the late 1990s to develop the lithium-ion batteries for the Altra electric vehicle. Lithium-ion batteries in electric vehicles offer a higher power-to-weight ratio for acceleration and a much better energy-to-weight ratio for extending range. Nissan claims these batteries have an energy density three times that of conventional lead-acid batteries. The Altra can attain a top speed of 75 mph and a between-charges range of 120 miles in combined city and highway driving. In the future, it is likely that the conventional non-liquid electrolyte will be replaced with polymer electrolytes to permit the fabrication of solid-state cells.

John Zumerchik

See also: Electric Vehicles; Fuel Cells; Fuel Cell Vehicles; Office Equipment; Storage; Storage Technology.

BIBLIOGRAPHY

Gray, F. (1991). *Solid Polymer Electrolytes: Fundamentals and Technological Applications.* Cambridge, MA: VCH.

Ingersoll, J. (1991). "Energy Storage Systems." In *The Energy Sourcebook*, ed. R. Howes and A. Fainberg. New York: AIP.

Kassakian, J. G.; Wolf, H.-C.; Miller, J. M.; and Hurton, C. (1996). "Automotive Electrical Systems Circa 2005." *IEEE Spectrum* 33 (8) :22–27.

Linden, D. (1995). *Handbook of Batteries*, 2nd ed. New York: McGraw-Hill.

Pistoia, G. (1994). *Lithium Batteries.* Amsterdam: Elsevier.

Scrosati, B., and Vincent, C. (1997). *Modern Batteries*, 2nd ed. London: Arnold.

BEARINGS

A world without bearings would look far different. Although rarely noticed, bearings are ubiquitous: computers have a few, electric appliances contain several, and an automobile has hundreds. Without bearings, much of the motion that is taken for granted would not be possible.

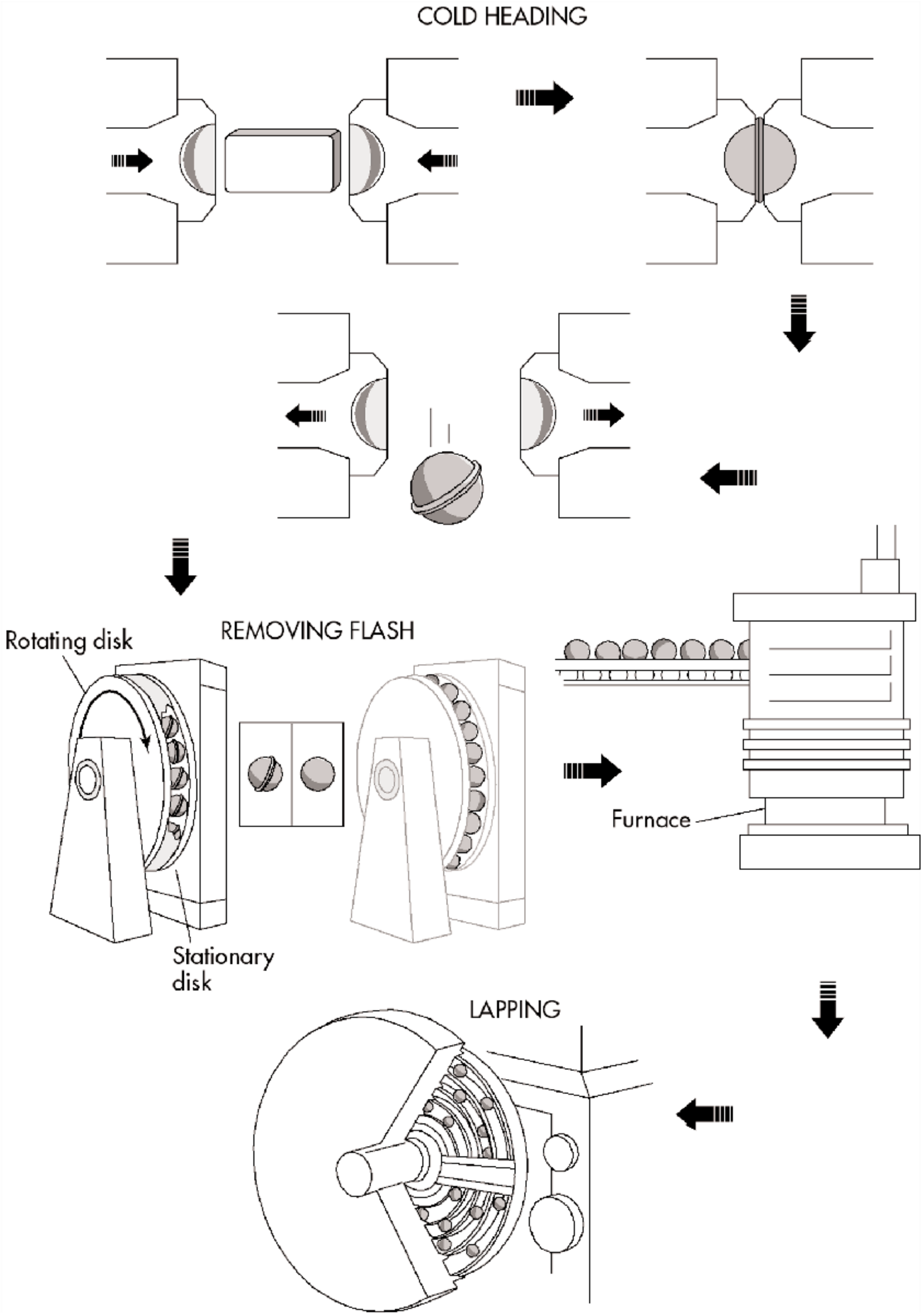
Since the first wheel was invented, people understood that it takes less effort to move an object on rollers than to simply slide it over a surface. People later discovered that lubrication also reduces the effort to side objects. Bearings combine these two basic findings to provide rolling motion necessary for things as simple as a tiny in-line skate wheel and as complex and large as a steam turbine. Bearings save energy, which is otherwise required to counteract friction arising from any elements related rotation, the better the bearing, the greater the energy savings.

TYPES OF BEARINGS

Bearings can be divided into rolling element bearings and sleeve (or plain) bearings. Sleeve (or plain) bearings consist of many sizes, shapes, and types, each of which functions essentially as a band (or sleeve) of close fitting material that encloses and supports a moving member. The sleeve is usually stationary and is called the bearing. The moving member is generally called the journal. Rolling element bearings are generally constructed with one or two rows of steel balls or rollers positioned between inner and outer rings. A retainer is used to equally space these rolling elements. Grooves, or raceways, are cut into the inner and outer rings to guide the rolling elements.

Rolling bearings have a lot of advantages compared with the sleeve bearings. Just to name a few: lower energy consumption, lower starting moment, lower friction at all speeds, and higher reliability. This is why there is a larger variety of rolling bearings than sleeve bearings. The variety of applications calls for the multitude of the rolling bearing types and designs. There are simple applications such as bicycles, in-line skates and electric motors. There are also complex applications such as aircraft gas turbines, rolling mills, dental drill assemblies, gyroscopes and power transmission products. Better automobile transmissions, containing hundred of bearings and delivering more mechanical energy to the wheels, have resulted in dramatic improvement in fuel economy and performance from 1975 to 2000. In delivering more mechanical energy to the wheels, and much of that improvement can be attributed to better bearings.

Rolling bearings can be further classified into ball and roller bearings. The following are some of the very common types of rolling bearings.



Components: (1) inner race, outer race, cage, and ball; (2) completed ball bearing. Manufacturing process: (1) cold heading—blank steel compressing between mold which forms ball with flash; (2) removing flash—balls are in machine with rotating disk and stationary disk, then balls enter furnace; (3) lapping—balls get polished. (Gale Group)

Bearing type	*Advantages*	*Cost/ Performance*	*Applications*
Deep grove ball bearing	high speed and high precision, average radial and thrust load	excellent	automobiles, cutting tools, water pumps, agricultural machinery
Self-aligning ball bearings	supporting radial and thrust load where shaft and housing are subject to misalignment	excellent	rubber mixers, oil field drilling rigs, vertical pumps
Angular contact ball bearings	average speed and support of radial and thrust load	fair	orbital sanders, food-processing machinery
Thrust ball bearings	support only thrust load	good	autombile clutches, gauges and instruments
Cylindrical roller bearings	low speed and heavy load, but only support radial load	excellent	tractors, machine tools, mid- and small-size motors
Needle roller bearings	support of radial load where radial dimension is limited	good	oil pumps, harvester combines
Spherical roller bearings	support of radial and thrust load, expecially when the shaft is long	excellent	paper mill machinery, air compressors, speed reducers, cranes

Table 1.
Comparison of Some Ball Bearing Types

Deep Groove Ball Bearings

Deep groove ball bearings, widely found in automobile applications, are the most popular of all rolling bearings. They are available in single and double row designs. The single row bearings are also available in a sealed version. They are simple in construction as well as easy to operate and maintain. They can run at high speeds and can support both radial and thrust loads imposed by rotating shaft and other moving objects. They are versatile, quiet, lubricated-for-life, and maintenance-free. The bearing cost/performance ratio for deep groove ball bearings is excellent. They are widely found in automobile applications.

Self-Aligning Ball Bearings

Commonly found in vertical pumps, self-aligning ball bearings have two rows of balls with a common spherical outer ring raceway. This feature gives the bearings their self-aligning property, permitting angular misalignment of the shaft with respect to the housing. Self-aligning ball bearings show very low vibration and noise level owing to the high accuracy of form and smoothness of the raceways.

Angular Contact Ball Bearings

Angular contact ball bearings are available in single and double row designs as well as four-point contact ball bearings. They are designed for a combined load and provide stiff bearing arrangements. Angular contact ball bearings have raceways in the inner and outer rings, which are displaced with respect to each other in the direction of the bearing axis. This means that they are particularly suitable for the accommodation of combined loads (i.e., simultaneously acting radial and thrust loads such as for orbital sanders). The benefits are high-load carrying capacity and speed capability, low operating temperatures, long relubrication intervals and quiet operation.

Thrust Ball Bearings

Thrust ball bearings are manufactured in single direction and double direction designs. They are only able to accept thrust loads but can be operated at relatively high speeds. Mounting is simple because the various bearing components (shaft washer, housing washer, ball and cage thrust assembly) can be installed separately. The benefits of using thrust ball bearings derive from their high running accuracy and high load carrying capacity, which is why they are used in automobile clutches and speed reducers.

Cylindrical Roller Bearings

Cylindrical roller bearings, found in tractor and machine tools, can carry heavy radial loads at high speeds because the rollers and raceway are in linear contact. Single row bearings have optimized internal geometry that increases their radial and thrust load carrying capacity, reduces their sensitivity to misalignment, and facilitates their lubrication. Full complement bearings incorporate the maximum number of rollers and have no cage, and are intended for very heavy loads and moderate speeds.

Needle Roller Bearings

Needle roller bearings can support heavy radial load such as clutches. A wide variety of designs, including bearings for combined radial and thrust loads, provide simple, compact and economic bearing arrangements. Their small sectional height makes them suitable for limited radial space of the housing.

Spherical Roller Bearings

Spherical roller bearings are robust, self-aligning bearings that are insensitive to angular misalignment. They offer high reliability and long life even under difficult operating conditions. They are mounted on an adapter assembly or withdrawal sleeve and housed in plummer blocks. They are also available with seals for maintenance-free operation.

Tapered Roller Bearings

Tapered roller bearings are designed for heavy combined loads or impact loading such as freight train locomotives and rail cars. Composed of a cup, a cone and a number of rollers, tapered roller bearings can do a much better job of withstanding sideward forces. These three components have tapered surfaces whose apexes converge at a common point on the bearing axis. Their excellent load carrying capacity/cross section ratios provide an economic bearing arrangement.

APPLICATIONS

High-carbon chrome bearing steel specified in SAE 52100 is used as a general material in bearing rings and rolling elements. The cages can be made of various materials, such as steel sheet, steel, copper alloy and synthetic resins. Once relegated to high-end applications, such as aircraft wing-flap actuators and precision instruments, hybrid bearings with ceramic balls are moving into the mainstream. Hybrid ceramic bearings offer many new options for demanding applications. Benefits include high speed, corrosion resistance, durability, reduced vibration, ability to operate with less lubricant, and electrical insulation.

Today's in-line skate market takes the advantage of the newly available technology in bearing design. For example, the ABEC (Annular Bearing Engineers' Committee) scale classifies different accuracy and tolerance ranges for bearings. A good in-line skate bearing typically adopts a high ABEC rating bearing, chromium steel for rings and balls, 100 percent synthetic speed oil for skate lube, ultrafast self-lubricating synthetic resin retainers, and unique and attractive packaging.

Jiang Long

See also: Materials; Tribology.

BIBLIOGRAPHY

Harris, T. A. (2000). *Rolling Bearing Analysis*, 4th ed. New York: John Wiley.

BECQUEREL, ALEXANDRE-EDMOND (1820–1891)

Edmond Becquerel was one of a family of scientists. His father, Antoine-César, was professor of physics at the Muséum d'Histoire Naturelle, and his son, [Antoine-]Henri Becquerel, also a physicist, discovered the phenomenon of radioactivity (for which he received the Nobel Prize in 1903).

The scientific work of Edmond began in 1838, at the very early age of eighteen. When the Chair of

Professor of Physics Applied to Natural History was created for his father at the Muséum d'Histoire Naturelle, Edmund had the dilemma of choosing to attend l'Ecole Normale, l'Ecole Polytechnique, or become an assistant to his father for the course that went with the professorship. He chose to assist his father, and their collaboration continued for decades. Thus, his title on the title page of the book published in 1855-1856 with his father is given as: "Professeur au Conservatoire impérial des Arts et Métiers, Aide-naturaliste au Muséum d'Histoire Naturelle, etc." After a short period as assistant at la Sorbonne, and then as Professor at the Institut Agronomique de Versailles, he became Professor at the Conservatoire des Arts et Métiers in 1852, where he worked for almost forty years. When his father died in 1878, Edmond succeeded him as director of the Muséum in addition to his professorship. He received a degree as Doctor of Science from the University of Paris in 1840, and was elected a member of l'Académie des Sciences in 1863.

He published a great number of scientific articles and a number of books: the three volume *Traité d'électricité et de magnétisme, et des applications de ces sciences á la chimie, á la physiologie et aux arts.* (1855-1856 with his father); *Recherches sur divers effets lumineux qui résultent de l'action de la lumiére sur les corps* (1859); and *La lumiére, ses causes et ses effets* 1867, in two volumes.

Electricity, magnetism, and light were the main subjects of his work. At the time, these subjects were "hot" topics. Hans Christian Ørsted had made his discovery that an electric current had an effect on a magnet in the year Edmond was born. Michael Faraday had just (in 1831) discovered the effect of induction, and Louis Daguerre invented the photographic plate in 1837, the year before Edmund began his scientific work. Edmond set out to study the chemical effect of light, and in 1839 he discovered a remarkable effect: electricity was emitted following the chemical actions due to the light—the photoelectric effect. He was thus led to the construction of the "actinometer," which allows the measurement of light intensities by measuring the electric current generated by the light. Using photographic plates, he examined the sunlight spectrum and discovered that the dark lines, observed by Fraunhofer in the visible part, continue into the violet and ultraviolet region, and that the plates, when exposed briefly to ultraviolet radiation, become sensitive to the red part as well, and can actually acquire an image without development of the plates.

Parallel to these investigations, he continued (with his father) to study electricity. He used the method of compensation to measure the resistivity of a large number of materials, including liquid solutions. The effect of the electrodes was, in the latter case, taken into account by using tubes in which it was possible to change the distance between the electrodes. Electrochemical effects and their practical applications were also a main concern. The second volume of the "*Traité ...*" is mainly concerned with the feasibility of extracting silver from minerals in Mexico by electrochemical methods as opposed to the methods then in use involving either charcoal or mercury. Mercury was expensive and charcoal was becoming increasingly expensive due to the shortage of wood. In the introduction to the "*Traité ...*" he mentions that if the consumption of wood in Mexico continued at the ten current rate, it would have severe effects, and that the Mexican government should be concerned. (It is at this point worth noting that Mexico had obtained independence in 1821, and the subject of French intervention in the internal wars that followed was a major political issue.) He studied extensively the electromotive force and internal resistance of a large number of batteries. He and his father used the thermoelectric effect to construct thermometers that could measure temperatures that were otherwise difficult to measure and at places that were hard to access by other means. The temperature in the ground was, for instance, measured throughout the year.

Another study was begun, in 1839, with Jean-Baptiste Biot (who had measured quantitatively the force that an electric current produces on a magnet, the effect that Ørsted had discovered qualitatively), namely, on phosphorescence, fluorescence and luminescence. To study the phenomenon that certain substances emit light after having been exposed to light, Becquerel devised an ingenious apparatus. The main idea was to have two discs with holes in them rotating about an axis parallel to the beam of light illuminating the sample placed between the discs. The sample receives light only when a hole passes in front of it; otherwise, the disc blocks the path of light. Likewise, the emitted light is observed only when a hole in the other disc passes the sample and can be examined at varying times after the exposure by changing either the relative positions of the holes or by changing the speed of rotation. Furthermore, a prism could be inserted in the path of the emitted light and spectral analysis performed. With this simple apparatus he was able to reduce to 1/40,000 of a second the time separating the luminous excitation and the observation. A number of important results were obtained; for instance, that fluorescence differed from phosphorescence only by its very short duration, and that the spectrum of the fluorescent light is characteristic for each substance (one of the first instances of nondestructive testing).

The discovery and detailed investigations of the phenomenon of fluorescence is generally considered the main contribution of Edmond Becquerel. It had the further impact of leading later to the discovery of radioactivity by his son Henri, as Henri continued these studies, including among the substances examined salts of uranium.

Edmond Becquerel was interested in and dedicated to science in general. He was a very careful and imaginative experimenter with an acute sense of the practical aspects of science. He put great effort and insight into exploring the practical uses of physics, especially the new phenomena of electricity and magnetism or, when combined, electromagnetism.

Stig Steenstrup

BIBLIOGRAPHY

Becquerel, H. (1892). “La chaire de physique du Muséum.” *Revue Scientifique* 49:674-678.

Harvey, E. N. (1957). *A History of Luminescence from the Earliest times Until 1900.* Philadelphia: American Philosophical Society.

Violle, J. (1892). “L’œuvre scientifique de M. Edmond Becquerel.” *Revue Scientifique* 49:353-360.

BEHAVIOR

Americans for years have been embracing new technology to make life more convenient, and at the same time grew accustomed to cheap, abundant energy to power that technology. Each year more than $500 billion is spent in the United States for energy to perform work and provide heat, and the energy needed for creating and building businesses, home-making, getting around, purchasing goods, and seeking pleasure. The decision-making exhibited while engaging in these daily activities varies widely, which results in diverse energy use behavior.

THE LONG-TERM IMPACTS OF DECISIONS

Energy-use behavior cannot be looked at just in the here and now or in isolation. Americans developed a greater reliance on heat, light, and power than any other nation, primarily due to decisions made by industry, government, and individuals of earlier generations that largely established the patterns for behavior today. Collective choices made long ago have behavioral consequences today, just as collective choices made today will have behavioral consequences for many more years to come.

The high-energy-consumption culture evolved primarily as the automobile became affordable in the 1920s and 1930s, offering tremendous mobility and the possibility of distancing home from work, school, and pleasure. It was then accelerated by the post-World War II federal policies of funding a vast network of highways and offering subsidies (home mortgage and property tax deductions) to make home ownership more affordable. Government policymakers at the time never realized the energy consequence of these nonenergy policies. Highways to everywhere, and home ownership subsidies promoted flight from city centers to suburbia and a high-energy-consumption economy largely based on the automobile. With these policies in place, the energy use patterns established a trend of greater and greater consumption. Moreover, the many benefactors of these policies amassed significant power and influence in the political system, discouraging any reversal.

Another factor contributing to high-consumption behavior is the ever-growing affluence of the population. In the 1950s, U.S. consumers spent around 25 percent of disposable income for food, which fell to less than 10 percent by the 1990s. For many, this translates into less than three hours of wages to cover the family’s weekly food bill. (For much of the developing world, it can take seven or eight hours of wages to feed a family of four for one day.) Moreover, six or seven hours of wages in the United States can cover the cost of a month’s worth of gasoline for the cars, and electricity and natural gas for the home.

This growing affluence allows a greater percentage of disposable income to go toward purchases of energy-using technology and ever-more-elective applications, rather than the cost of energy to power the technology. For example, while energy costs had remained flat or risen only slightly from 1980 through 2000, the average cost of an automobile tripled, going from around $7,000 to $21,000. Furthermore, as the United States became more affluent, the distinction between want and need—elective and essential—has blurred. Technology that was rare or nonexistent in the 1960s became a want in the 1970s and 1980s, and evolved into what many consider a need or necessity in the 1990s. Most households are now dependent on all sorts of energy-hungry technology that did not even exist in the 1960s—appliances such as lawn trimmers, personal computers, and microwave ovens, to name a few. Greater affluence, and an economy geared toward high consumption, has resulted in U.S. per capita energy use growing to more than thirty times that of developing world nations, and more than twice that of Western Europe or Japan.

One of the biggest fears with massive consumption and the high energy lifestyle is the high level of immigration that has accompanied it. Although the U.S. birthrate has been slightly more than self-sustaining since the 1970s, a liberal immigration policy had added more than 30 million people to the population by 1995, and is likely to push the population to more than 400 million by 2050. Millions of people from around the world try to immigrate, legally or illegally, to enjoy the freedom and the opportunities of the United States. Because most immigrants, regardless of culture and country of origin, tend to quickly assume the U.S. norm of high-energy-use behavior that increases carbon emissions, exacerbates air pollution (through a need for more power plants and automobile use), and worsens the dependence on imported crude oil, overpopulation and poverty is as much a United States problem as a world problem.

LIFESTYLES AND ENERGY USE

Energy is the ability to do work. The work can be done by man or machine; the fuel to do the work can come from food or fuel. Humans often have a choice of whether or not to continue to develop technology that replaces the work of man as well as do other types of work that man is not capable of doing. Historically, the choice has been almost always to embrace technology, driving the steady increase in demand for energy.

There is often a mental disconnect between technology and the energy needed to make technology work. Except for a small percentage of the population, people are not much interested in energy itself, but only what it can do for them. That is because energy is largely invisible. The units of energy—a kilowatt hour of electricity or a cubic foot of natural gas—cannot be seen. Thus, energy is primarily thought of in financial terms: the dollar amount of the electric or natural gas bill, or the cost of refueling the car.

Energy behavior can be categorized loosely as conserving (frugal), efficient, or high-energy (Table 1). It is not always fair to make such clear-cut distinctions because many behavior patterns exist within each income level that are beyond easy classification. For example, a few cross-classifications could be rich and frugal, rich and efficient, poor and frugal, and poor and wasteful. Ironically, the poor are often the most wasteful and inefficient because they own older cars and live in leaky homes.

The distinction between high-use versus conserving or efficient is well-known, but the difference between energy-conserving behavior and energy-efficient behavior is a more subtle distinction. Energy is used to perform important tasks such as getting to work. A vehicle will use a given amount of British thermal units (Btus) per mile. Multiplying the Btus per mile by the number of miles traveled, results in the total Btus used. The goal of energy conservation is to reduce the number of Btus used. Energy-conserving behavior is not strictly about forgoing consumption. The conserving person still needs to get to work; he just wants to get there by using less energy, say, by substituting the train, bus or carpool for driving alone. On a per passenger basis, the difference between the two could be 100 to 160 miles per gallon (mpg) for carpooling or mass transit, compared to 15 to 40 mpg for driving alone. Remembering to turn off the lights when leaving a room, or turning down the thermostat to 60°F, or shoveling the walk instead of buying a snow blower, are other examples of energy-conserving behavior.

Energy-efficient behavior usually involves embracing all technology to do work, but picking the most energy-efficient products to do that work. The energy-efficient choice is to stay in the single passenger automobile, but reduce the Btus per mile by buying an efficient model that gets 40 mpg rather than

	Conserving (Frugal) Lifestyle	***Efficiency Lifestyle***	***High Use Lifestyle***
Food	Eating a healthy diet and only the required 2,500 to 3,000 calories.	Prone to overeating, but tries to keep a healthy diet.	Overeating and an overabundance of meat and empty calories, highly processed junk foods.
Exercise	Commuting by walking or riding a bike.	Membership in fitness club to overeat yet remain thin.	Very little or none. Likely to become obese.
Recreational Activities	Walking and bicycling.	Running, swimming and aerobics.	Golfing with a cart, personal watercraft, all-terrain vehicles, and snow-mobiling.
Hobbies	Participatory activities.	Participatory activities.	Watching television and attending sporting events.
Home	1,500 sq. ft. apartment or condominium in a city high-rise.	5,000 sq. ft. energy-efficient home in the suburbs.	5,000 sq. ft., nonenergy-efficient home in the suburbs.
Heating	High-efficiency furnace and water heater. Bundles up with a sweater to keep thermostat at 65 degrees.	High-efficiency furnace and water heater. Keeps thermostat at 72 degrees, but programs it to 60 degrees for overnight hours.	Low-efficiency furnace and water heater. Keeps thermostat at 78 degrees at all times.
Air Conditioning	Only a few rooms, and only when occupied.	Central air that is programmed to come on an hour before arriving home.	Central air all the time, regardless if anybody is home.
Appliances	With limited space, fewer and smaller; only the necessities. The necessities must be energy-efficient. Buys food supplies daily on the way home from work; no need for a large refrigerator and freezer.	Desires all technology, but wants the most energy-efficient products.	Desires all technology and does not put a priority on energy efficiency in making buying decisions. Huge refrigerator and freezer for the convenience of limiting food shopping to once a week.
Vehicles	None	Two energy-efficient models.	Three or four automobiles and a recreational vehicle.
Commute	3 miles by walking, bicycling, or mass transit.	40 miles in an energy-efficient automobile. If there are high-occupancy vehicle lanes, will make effort to carpool.	60 miles in a 14 mpg sports utility vehicle, and a disdain for mass transit, carpooling and high-occupancy vehicle lanes.
Schools	Children walk to nearby school.	Bus to local school.	Drives children to school.
Children's activities	Nearby so that they can walk or take public transportation.	Effort to carpool and make it convenient for drop-off and pick-up on the way to and from work.	Little regard for proximity. The baby sitter will do driving.
Shopping	The majority is done locally and daily; accessible by walking or mass transit.	Drives to nearby shopping. Realizes that the additional gasoline cost of driving negates much of the discount achieved by giant retailer shopping.	Willing to drive many additional miles to get to malls and giant retailers.
Recycling	Yes	Yes, if there is a financial incentive.	No, unless mandated. Throw-away mentality.

Table 1.
Lifestyles for Three Affluent Four-Person Families

Note: This is an example for comparison only, to show how a high-energy-use lifestyle can result in energy consumption two or three times the level of an energy-efficient lifestyle, and over ten times that of an energy-conserving lifestyle. In reality, rarely do individuals exhibit behavior that is solely conserving, efficient or high-use.

one that gets 15 mpg, or replace a 75-watt incandescent light bulb with a 25 watt fluorescent tube. The amount of illumination is the same, but the amount of energy used is one-third less. In terms of total energy savings to a nation, energy conservation is more important than energy efficiency since there is less energy used when no light is on than when an efficient one is on.

A person exhibiting energy-conserving behavior is likely to exhibit energy-efficient behavior as well, but someone who exhibits energy-efficient behavior is not necessarily going to be into energy conservation. Energy conservation usually indicates a sacrifice in something, be it safety, aesthetics, comfort, convenience, or performance. To many energy-efficient people, this is an unacceptable compromise. Energy-efficient homeowners will take energy efficiency into account in making decisions about home lighting; yet, if they have a strong preference for art gallery quality lighting, they do not hesitate to spend more for electricity than a conserving neighbor whose only concern is a good reading light. Whereas energy-efficient behavior primarily entails an approach to making technology-buying decisions energy-conservation habits require a full-time, conscious effort to reduce the amount of technology used.

DECISION-MAKING THEORY

The reasons for the range of energy-use lifestyles, and the motivation behind energy decision-making, are varied and complex. Prior to the 1970s, there was little study of energy-use behavior. But energy scarcities and the growth of the environmental movement made the energy problem a social problem that economists, psychologists, sociologists, and anthropologists all began to address.

Economists primarily look at energy decision-making, like all other decision-making, as a function of price and utility: The individual is a rational utility maximizer who gathers and weighs all the relevant information to make cost-benefit evaluations to arrive at decisions. If an energy conservation or efficiency product proves beneficial, the purchase will be made. It is uncertain how much of the population realizes that often the greater initial out-of-pocket cost for energy-conservation measures and energy-efficient products will be made up by greater savings through the lifetime of the product. And even when aware of the life-cycle savings, the purchase of more energy-efficient products may be rejected because of an unacceptable compromise in safety, aesthetics, comfort, convenience, quality, or performance. For example, the five-passenger Mercedes might be preferred over the five-passenger Hyundai since the Hyundai's acceleration, reliability, and luxury shortcomings more than outweigh the benefits of its lower price and better fuel economy.

The current price of energy is a primary factor entering into decision-making, yet the expected future price of energy can be even more important since the product life cycles of most technology is over ten years (autos and appliances 14 or more years, real estate 30 or more years). After the Energy Crisis of 1973, people adopted energy-conserving and efficient behavior more because of the fear of where energy prices were headed than the price at that time. Almost all energy forecasters in the 1970s and early 1980s predicted that energy prices would soar and oil resources would soon disappear. *National Geographic* in 1981 predicted an increase to $80 per barrel by 1985—a quadrupling of gasoline prices. By the 1990s, when these forecasts turned out to be wildly erroneous, most people reverted back to their old lifestyles, and became much more skeptical of any proclaimed impending crisis.

Whereas the focus of economists is on price and utility, the focus of psychologists has been on attitudes and social norms. For example, the nonprice reasons for choosing an energy-conserving or efficient lifestyle have been for social conformity and compliance—a sense of patriotism, good citizenship, or because it was the environmentally friendly thing to do. The main theory is that attitudes determine behavior, and if attitudes can be changed, it is more likely that behavior will change too. However, there is also the opposite theory: Behavior causes attitudes to change. If persons assume an energy-conserving lifestyle, they will assume attitudes to support that action to avoid internal conflict or hypocrisy (Stern, 1992).

Another problem in studying attitudes and behavior toward energy is the pace of change. Even in times of energy shortage, attitudes and behavior tend to change slowly and are influenced as much by cultural and geographical factors as price. A family living in a 5,000 sq. ft. home can turn down the thermostat, yet still needs to heat a significant amount of space. And if the family lives 75 miles from work, they can switch to a more-fuel-efficient car, yet still must burn considerable gasoline commuting. Of course, this family can make a drastic lifestyle change and consume a fraction of the energy by moving into a 1,200 sq. ft. condominium a few blocks from work and school, but this choice is unlikely for many who have grown accustomed to the comfort, lifestyle and perceived safety of suburban life and feel it is well worth the inconvenience and higher energy costs.

THE COMMERCIAL SECTOR

Energy is a capital investment decision that is often neglected by corporations. When a corporation decides to build or rent office space, the energy needed to heat, air-condition, and illuminate the facility usually is not a top priority. In the 1990s, layout, ergonomics, aesthetics, corporate image, proximity of highways and mass transportation, and other productivity factors ranked much higher. Moreover, there remains skepticism about the ability of energy-efficient technologies and design strategies to simultaneously save energy and improve labor productivity. If the price of energy skyrocketed, lowering energy costs would quickly become a priority again. Many office facilities, not competitive on an energy-efficiency basis, could become obsolete well before their expected lifetime of 50 to 100 years.

THE INDUSTRIAL SECTOR

Lean production—the improvement of processes and operations—is something that all industrial executives encourage, and is largely why the U.S. industrial sector energy consumption dropped about 20 percent from 1973 to 1983. With much of energy efficiency gains already achieved, company executives of the 1990s found the payback from other investments more economically efficient, and consequently have put the emphasis elsewhere. A top priority in the 1990s was faster and cleaner processes that reduced manufacturing costs and reduced waste. If these processes also reduced energy use, it was an added bonus.

Besides operational energy use, corporate decisions have an impact throughout society. One reason that per capita energy consumption in North America is much higher than in the rest of the world was the decision of corporate leaders to expand and relocate away from city centers and to major beltway loops in the suburbs. This necessitated more trucking and a workforce reliant on the private automobile instead of mass transit. In an era of tremendous job insecurity, even the most energy-conserving person is hesitant to reside near work, or live without the mobility of the personal vehicle.

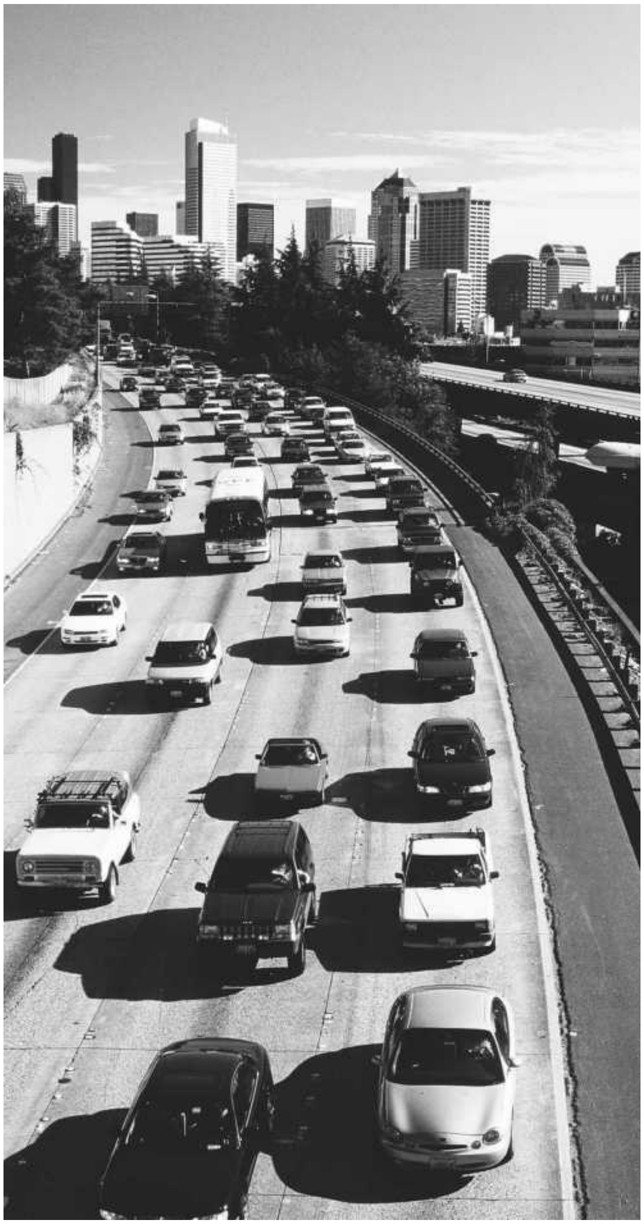
THE TRANSPORTATION SECTOR

Driving by personal vehicle is the most popular mode of transportation. And although there is a desire for a fuel-efficient automobile, fuel efficiency is a consideration well behind style, performance, comfort, durability, reliability, status, and safety. The weak demand for a 40 mpg automobile occurs for several reasons: It is not a status symbol (not stylistic), accelerates too slowly (smaller engine), cramps the driver and occupants (smaller interior), and often offers inadequate protection (too light) in case of an accident.

Another reason for weak demand is that fuel cost in 1999 were a much smaller fraction of the cost of owning and operating a vehicle than it was in 1975. While the average cost of owning and operating an automobile more than tripled from 1975 to 1999, the price of fuel increased only marginally, and average fuel economy improved from 15 to 27 miles per gallon. As fuel cost declines relative to all other costs of operating an automobile, the purchase of a fuel-efficient automobile increasingly becomes a more secondary financial consideration. Most motorists express a preference for greater size, luxury, and performance from the automobiles and trucks purchased, knowing that such attributes usually are detrimental to fuel economy.

The Private Auto Versus Public Transit

The freedom of private transportation is something most Americans have taken for granted and consider a necessity. Only a generation ago, an automobile was considered a luxury, and a generation before that, a rarity. The ascension of the automobile coincided with the decline in mass transit in the 1920s. This trend accelerated after World War II when American society decided to build a vast interstate highway system, emphasizing private automobiles at the expense of mass transit. It is not easy to undo the technological impetus of large infrastructure changes of this nature. According to the U.S. Census Bureau, by 1990, 73 percent of the population drove alone to work, up from 64 percent in 1980. In percentage terms, walking (3.9%), bicycling (0.4%), public transit (5.3%), and carpooling (13.4%) all declined between 1980 and 1990. For other travel (shopping, family business, and leisure), which was responsible for two-



Traffic crawls past the skyline of downtown Seattle on Interstate 5. (Corbis-Bettmann)

thirds of all travel, the share of collective modes and nonmotorized modes was even lower.

A major mass transit handicap is sprawling growth. Mass transit works most effectively in a hub and spoke manner, and the decentralization of urban areas makes it harder and harder to design effective mass transit. This mismatched segregation of home, work, and leisure—particularly in cities such as Houston, Atlanta, and Los Angeles—would require a long time to reverse.

Even where mass transit offers better speed and comparable comfort and convenience, the love of the automobile results in uneconomic decisions affecting energy consumption. For someone driving 30 miles to New York City, tolls, fuel, and parking can easily exceed $20 a day. In comparison, a monthly train pass for the same 30 mile trip runs less than $10 a day. This premium for the privilege of driving is even greater if the yearly $6,000 to $10,000 cost of leasing, insurance, and maintenance is included. Because of congestion, few can claim to drive for speed reasons. The roadway trip often takes more time than via mass transit. Getting more people out of single-passenger cars and into more-energy-efficient mass transit or car pools is going to take more than improving mass transit and the price of gasoline tripling. Americans, like most others in the industrial world, have an emotional attachment to the automobile.

Another reason for choosing the automobile is its development as a mobile office. Cellular phones, laptop computers, and satellite linkups to hand-held communication devices—all of which are getting smaller and smaller—have made it possible to be more productive from the roadways.

Aside from work, the automobile is central for shopping and pleasure. Most people who lived in the suburbs in the 1990s had grown up in the suburbs rather than in central cities. The automobile, and the mobility it affords its owner, is a central aspect of suburban life. Shopping centers and businesses, lured by cheaper land in the exurban areas, do not fear locating away from city centers because Americans show an eagerness to drive greater distances. The auto also caters to the needs of family life. Few suburban parents drive directly to and from work anymore. The growth in extracurricular sports and activities, for parents as well as children, requires the convenience of the automobile.

Higher incomes, higher automobile ownership, and a decline in the population and workplaces that can be served by mass transit has lead to the declining mass transit demand. Criticism of this shift toward the private automobile comes mainly because the individual driver receives the short-term benefits (privacy, comfort, speed, and convenience), while the negative social consequences (air pollution, traffic jams, and resource depletion) are shared by all. Moreover, if people drove less, and drove more-fuel-efficient vehicles, the positive national goal of less dependence on imported oil would be achieved.

Flight

The explosive growth in air travel from 1980 to 2000 occurred because deregulation reduced air fares, disposable incomes rose, and travelers desired to get places faster. People are willing to pay a large premium for speed by flying instead of driving. The premium is largely a reflection of the much greater energy costs of flying, yet it is not always a greater cost. On a passenger-miles-per-gallon basis, usually more energy can be conserved by flying a full aircraft rather than having each passenger drive solo to a given destination.

People make decisions to fly based on speed, the price of the flight, and an airline's on-time arrival. If these factors are the same, then secondary concerns such as comfort become a factor—fewer seats with more spacious seating, and more room to walk about. Since greater comfort means fewer paying passengers, the airlines' decision to cater to the desire for comfort will adversely affect fuel economy per passenger.

The widespread disfavor toward prop planes is another preference adversely affecting fuel economy. Prior to the late 1990s, the 30- to 50-seat plane market was dominated by the more-energy-efficient turboprop planes, yet regional airlines ordered jet engine replacements, citing strong customer preference. Airlines are turning to jets because people want to get places faster. Turboprop planes, which attain near-jet speeds and near-jet performance with less fuel consumption, are noisier and thought to be less safe, even though safety records do not warrant this belief.

The future demand for jet travel is very uncertain. Businesses have traditionally felt a need to travel for face-to-face meetings, but the new communications technology revolution might in the future make business travel less necessary. However, any dropoff in business travel is likely to be replaced by growth in leisure travel as more affluent Americans decide to take more but shorter vacations, and fly for more weekend getaways.

Freight

Individuals and companies make shipping decisions based on price, speed, and reliability. If reliability and speed are equivalent, the decision usually comes down to price. Because energy costs of shipping by air, railway, waterway, and truck vary tremendously, so does the price. Whereas commodities and basic materials are moved by rail, just-in-time components and finished product are mainly moved by truck, accounting for over 80 percent of all freight revenue. Rail is cheaper and a much more energy-efficient means to move freight, and has been approaching trucking for speed; yet businesses continue to pay a premium for trucking because rail historically has been less reliable.

THE RESIDENTIAL SECTOR

The American home is widely perceived as a good investment that appreciates. Any additional improvement to the home is considered wise for two reasons: the enjoyment of the improvement, and greater profit when the time comes to sell. This belief, in part, explains the preference for bigger new homes with higher ceilings over the large stock of older, smaller homes built from 1950 to 1970. The average new home grew from 1,500 square feet in 1970 to over 2,200 by 1997, and the inclusion of central air conditioning grew from 34 to 82 percent. These new homes are usually farther from city centers, and indicates a general willingness to endure the inconvenience and higher energy cost of longer commutes. The continuing trend of bigger suburban homes farther from city centers is attributable to the affordable automobile, the expansive highway system, and the pride of owning one's own home. Whereas a generation ago a family of eight felt comfortable sharing 2,000 square feet of living space, the generation of the 1990s located twice as far from work so that they could afford twice the space for a family of four. But home size alone can be a poor indicator of energy use. Energy consumption can often vary by a factor of two or three for similar families living in identical homes (Socolow, 1978).

Many people approach home ownership investments with only two major concerns: how much down and how much a month. They may be tuned into the present energy costs of a new home, but usually give little thought to how much energy will cost 10 or 20 years into the future. If energy-efficient features push up how much down and how much a month beyond what the buyer can afford, buyers must forgo energy-efficient features. However, selecting energy efficient features for a new home does not necessarily have to increase the down payment and monthly mortgage payment.

Aesthetics is often more important than energy efficiency to home buyers. Energy efficiency that is economically efficient is welcomed if it does not come at the expense of aesthetics. Tightening up a home with better insulation and caulking is fine, but solar collectors on the roof are thought by many to be an unsightly addition to a home. Prior to the 1960s, buildings in the sunbelt were usually built with white roofs. But as air conditioning became widespread, the "cooler" white roof grew in disfavor. Darker roofing shingles were perceived to be more attractive, and did a better job of concealing dirt and mold. Aesthetics won out over energy conservation. In a city like Los Angeles, replacing a dark roof with a white roof can save more than $40 in air-conditioning bills for the hot summer months. Eventually, through education, the aesthetic benefits of a dark roof might be deemed less important than its energy wastefulness.

The energy consequences of bigger homes filled with more power-hungry technology is a need for more energy. Because of federal standards for appliances and heating and air conditioning equipment, the new, larger homes have only incrementally increased consumption. However, the desire for ever-bigger refrigerators and freezers, and the continued introduction of more plug-in appliances, many of which run on standby mode, promises to keep increasing the residential demand for electricity.

By the 1990s, after decades of extremely reliable service, most customers have come to expect electricity to be available when they want it and how they want it, and feel the utility has the obligation to supply the power. If the customer's preference is to keep the air conditioning on all day, it is the obligation of the utility to supply the power for this preference. Customers have become so accustomed and reliant on electricity that the 1990s consumer felt service without interruption was a right. This was far different than a few decades earlier. New York City residents tolerated the major blackout of 1965 and the inconveniences it caused, yet the minor blackout of 1999 was widely believed to be inexcusable and that Con Edison (the local utility) should compensate its customers for damages, and it did.

During prolonged heat waves which can cause blackouts and brownouts, electric utilities always request that customers limit energy consumption by raising thermostats and turning off the air conditioning when not home. It is uncertain if these requests are heeded, and if they are, it is uncertain whether conformity is out of altruism (help the utilities) or self-interest (save money).

FOOD

Just as gasoline is the energy source for the automobile engine, food is the energy source for the human engine, yet it is not common for people to think of food in this way. It is more common to look at food as a means to satisfy a hunger craving. Even among those claiming to want to lose weight, 35 percent of men and 40 percent of women are not counting calories according to the American Medical Association.

The minimum average energy requirement for a sedentary adult to survive is around 2,000 kcal a day, which rises to 4,000 or more if the person is engaged in strenuous labor much of the day. Only a very small percentage of the U.S. population is involved in strenuous labor, yet much of the population consumes over 4,000 calories each day, resulting in an America that is 50 percent overweight and 20 percent obese.

Few people do manual labor anymore. Technology has modified behavior so that most people burn 2,500 to 3,000 calories a day, not 4,000 or more calories as was common up until the 1960s. The washboard was replaced by the washing machine; the manual push mower was replaced by the power mower; the snow shovel was replaced by the snow blower; the stairs were replaced by the elevator. Technology has made it easy to be inactive, yet few individuals will blame technology for their obesity.

The tremendous growth in health and fitness clubs since the 1960s caters to those Americans who would rather burn more calories than reduce calorie intake. By working out regularly, they can overindulge. However, only a small minority of the U.S. population exercises regularly; the majority overindulges and remains overweight.

If a sedentary person consumes more than 4,000 calories a day, obesity is likely to result; yet, among this group, a failure to match food energy input with activity energy output is seldom mentioned as the reason for obesity. Because much of the population cannot control the short-term pleasure of overindulging, and despite awareness of the long-term consequences to health and appearance, some politicians have proposed that health insurance, Medicare, and Medicaid should pay for surgical procedures (stomach staple) and diet pills (Redux, Fen-Phen) to combat what they call an "overeating addiction."

Discarded aluminum cans are disposed of in a red collection bin to be recycled and then reclaimed at Portsmouth Recycling Center. (Corbis Corporation)

MATERIALS

When given a choice among materials, determining the most energy-wise material seldom easy. Consider the energy needed to build an aluminum bicycle frame versus a steel tubular frame. The aluminum bicycle frame takes considerably more energy to build, and needs replacement more often. However, if the bicycle is meant to be a substitute for the automobile, the lighter aluminum frame material can dramatically lower the human power output needed to climb hills. The energy saved during the use of the aluminum frame will more than compensate for the greater energy needed to manufacture it.

The energy-wise choice of material for packaging—paper, plastic, tin, aluminum, or glass—can be equally tricky. Although it takes less energy to make glass and plastic bottles than aluminum cans, few buyers think of purchasing a favorite soda because it comes in glass or plastic rather than aluminum. Energy rarely enters into the buyer's decision. Price and quality concerns are foremost. Plastic wins out on price and glass on quality; rightly or wrongly, there is a perception that glass is better at not altering the taste of the product. Few consumers would be receptive to fine wines coming in plastic bottles or aluminum cans. Glass could someday compete with plastic at the lower end of the beverage market as well if the price of crude oil increased significantly. But glass must also overcome a higher freight cost (glass is much heavier than plastic).

The recycling of bottles and containers is almost universally viewed as an energy-wise and environmentally sound moral good—one of the best way to conserve energy, resources and landfill space. Usually this is true. However, few people realize that for some materials there are great energy savings, and for others very little. For instance, whereas the net energy savings to recycle aluminum cans is substantial, it can take nearly as much energy, and generate as much pollution and waste, to recycle tin cans as to produce new ones from raw materials.

A modest amount of recycling of metals occurs today. Much more can be done, and much more can be done to encourage energy-wise buying decisions as well (choosing products packaged with materials requiring less energy over those requiring more). Whether the American public is willing to invest more time and effort into recycling, and willing to alter product packaging choices solely for energy and environmental reasons, is highly questionable.

BEHAVIOR CHANGE

Energy-related behavior modification goes on continually. Advertisers market the latest household electrical technology to make life easier; tourism agencies promote exotic cruises and resorts; and governments mandate energy conservation and energy efficiency measures to lower carbon emissions and reduce energy imports. These efforts, either direct or indirect, subtle or overt, are all designed to modify behavior that affects energy consumption.

Government

Governments have taken an active role to alter behavior, to conserve energy, and to use it more efficiently. The rationale is that lower energy consumption reduces the need for additional fossil fuel power plants and the need for less crude oil imports. Since the Energy Crisis of 1973, the U.S. government's behavior modification efforts have been many and can be grouped into five categories: information campaigns, feedback, reinforcement, punishment, and reward.

Information campaigns refer to the broad range of brochures, flyers, billboards/signs, workshops, and television and radio advertisements designed to encourage energy-conserving behavior. The success of an information campaign depends on the audience paying attention and taking the message seriously; moreover, the intended audience must trust the government source, and receive confirmation from friends and associates (Stern, 1992).

A lack of trust and confirmation is one of the reasons that the energy-related informational campaigns of governments and environmental groups have faired so poorly. Despite millions being spent in the 1990s to promote conservation, energy efficiency, and renewable energy to combat global warming, the sales of energy-guzzling vehicles skyrocketed as economy car sales declined, the number of vehicles and average miles per vehicle increased, and the average home size and the number of electric appliances in each home kept rising. In a 1997 *New York Times* poll, when asked to rank environmental issues, only 7 percent ranked global warming first (47 percent said air and water pollution), and a CNN poll in that same year found 24 percent of Americans concerned about global warming, down from 35 percent in 1989. Since earlier informational campaigns to change behavior (for example, the supposed energy supply crisis of the 1970s was projected to only get worse in the 1980s) turned out to be erroneous, the American public viewed the government's global warming campaign much more skeptically.

Feedback modification efforts are targeted information programs that address the lack of awareness of people about the consequences, the belief being that if people are aware or educated about the negative consequences of such behavior, it is more likely that they can be convinced to engage in behavior more beneficial to the environment. For example, if the public better understood that all forms of energy consumption come with unfavorable environmental side effects, people would be more likely to conserve and use energy more efficiently. Once made aware of the deleterious environmental consequences, people would be more likely, in theory, to buy a more fuel-efficient minivan than a gas-guzzling sports utility vehicle that marketing research shows customers rarely or ever take off-road.

Reinforcement efforts include rewarding people for engaging in energy- and environmentally-beneficial behavior. High-occupancy vehicle lanes reinforce carpooling because the driver and passengers benefit with a faster trip for the ride-sharing sacrifice, and the EPA-DOE Energy Star label affixed to energy-efficient products is a sign to others that the purchaser uses energy in an environmentally friendly way. Punishments that have been tried include a fine or premium for energy use that causes pollution: the taxation of gas-guzzling automobiles. Rewards tried include tax credits for more environmentally friendly renewable energy and conservation measures, the construction of carpool park-and-ride lots, and government-mandated employer bonuses for employees who do not drive and therefore do not take advantage of subsidized parking (probably the most effective behavior modification program.

It is very difficult to determine the effectiveness of efforts to change behavior. There are those who feel a need to adopt an energy-conserving or efficient lifestyle for appearance sake, yet will assume high-energy behavior when no one is watching (Bell et al., 1990). Moreover, no universally effective method has been found for getting people to reduce energy consumption, and there is great debate as to whether it is a policy worth pursuing. First, many policy changes entail encroachments on freedoms. Since the freedom to choose is a major right that most are very reluctant to give up, passing policy that curtails freedom is controversial. Second, there is disagreement about whether human beings are rational about energy use, and actually modify behavior for any reason except self-interest (say, altruism). Third, sometimes the cost of the program exceeds the value of the energy saved, or the cost-to-benefit ratio is unfavorable. Finally, nonenergy policy has a greater impact on energy consumption than any energy policy itself. Government subsidies for home ownership, the public funding of highways, and policies encouraging suburban sprawl were far more responsible for per capita energy consumption being much greater in the United States than Japan or Europe than was any energy policy.

Corporate

From the corporate end, almost all of the behavior modification efforts directed at consumers are to get people to use more technology. It usually follows that greater technology use results in greater energy use.

To combat a corporate image of a single-mindedness toward greater consumer consumption, companies such as Dupont and 3M make great efforts at what they refer to as eco-efficiency—improving processes to reduce energy use, waste, and air pollution. The message to consumers is yes, we produce the technology you demand, but we do so in an energy-efficient way that is friendly to the environment. Dupont has been aggressively seeking to hold energy use flat, relying more on renewable sources such as wind and biomass, and is trying to reduce its 1990 levels of greenhouse gas emissions by 45 percent by 2001, and by 65 percent by 2010. The U.S. government encourages this behavior, and would like more corporations to no longer think of energy as just another cost, equal in importance with all other costs. Except for the EPA-administered punishments for pollution, most of the Federal effort is toward reinforcing and rewarding good corporate citizenship actions, not mandating them.

When questioned by electric utilities, a majority of residential customers show a willingness to consider paying a modest amount more per month for electricity powered from nonpolluting renewable energy sources, despite not knowing much about them. How many would actually choose to pay a premium for renewable energy is very uncertain. For those who are skeptical or ambivalent about the possibility of fossil fuel resources exhaustion, air pollution and global warming, the primary interest is the lowest price and best service. Early green-energy marketing efforts have shown promise in reaching those who are concerned about environmental issues. In 1999, Mountain Energy of Vermont signed up over 100,000 Pennsylvanians and Californians who will pay a 5 to 35 percent premium for electricity generation not involving nuclear power or coal. It is a surprisingly good start for Mountain Energy, who cannot actually get green power to the home, but instead must sell the concept—the green power the customer buys displaces the traditionally generated electricity in the area. Because many consumers buying the green energy are in a different region from where it is produced and therefore will not "receive" the cleaner air for which they paid a premium, it is uncertain whether green marketing programs will provide measurable environmental improvements. The result may be significant new renewable energy development, or it may just support the renewable base already in operation.

Food

Because of the huge food surplus in the United States, attributed to the high productivity of mechanized agriculture, there is probably no area of the energy consumption picture where more behavior modification takes place than food. The huge agricultural surplus keeps food products cheap, and would increase tenfold if Americans cut meat consumption in half and ate grains instead. Humans get from grain-fattened cattle only about 5 percent of the food energy they could get by eating the grain the cattle are fed.

There are thousands of special-interest groups trying to modify consumer behavior to eat more, or eat higher up the food chain. This would not be possible without a huge food surplus. It has never been easier to overeat. Even for those who lack the time or motivation to cook, the fast-food industry and microwaveable meals have made food more convenient and widely available than ever before. At the other extreme is the $33 billion a year weight loss industry promoting products to lose the weight gained "eating down" the food surplus. In the middle is the U.S. Department of Agriculture (USDA) with an inherent conflict of interest: One arm of the organization promotes the consumption of food while another arm publishes the highly political and ubiquitous food pyramid of good eating.

Since Americans enjoy the cheap food supply made possible from a huge agricultural surplus, and there seems to be no desire to reject technology and go back to a reliance on manual labor, the food surplus problem of an overweight and obese America is likely to remain for years to come.

OUTLOOK

Easy access to inexpensive energy has come to be viewed as a basic right. The American public goes about its daily life largely optimistic, feeling that any fossil-fuel shortage will be alleviated by new breakthroughs in developing supplemental sources, and that new end-use technology will be developed that can be powered by these new energy sources. Energy conservation and energy efficiency are lifestyle options. However, if the day ever comes when energy conservation or an accelerated adoption of energy-efficient products will need to be mandated, the American public would be more likely to choose the energy efficiency route because it does not necessarily entail sacrifice.

John Zumerchik

See also: Air Travel; Bicycling; Capital Investment Decisions; Communications and Energy; Economically Efficient Energy Choices; Economic Growth and Energy Consumption; Freight Movement; Green Energy; Government Agencies; Industry and Business, Productivity and Energy Efficiency in; Materials; Propulsion; Traffic Flow Management.

BIBLIOGRAPHY

Bell, P. A.; Baum, A.; Fisher, J. D.; and Greene, T. E. (1990). *Environmental Psychology,* 3rd ed. Fort Worth, TX: Holt, Rinehart and Winston Inc.

DeSimone, L. D., and Popoff, F. (1997). *Eco-Efficiency: The Business Link to Sustainable Development*. Cambridge, MA: MIT Press.

Energy Information Administration. (1997). *Household Vehicles Energy Consumption, 1994*. Washington, DC: U.S. Department of Energy.

Engel, J. F.; Blackwell, R. D.; and Miniard, P. W. (1998). *Consumer Behavior,* 8th ed. Hinsdale, IL: Dryden Press.

Farhar, B. C. (1994). "Trends: Public Opinion about Energy." *Public Opinion Quarterly* 58(4):603-632.

Farhar, B. C., and Houston, A. (1996, October). "Willingness to Pay for Electricity from Renewable Energy." National Renewable Energy Laboratory, NREL/TP-461-20813.

Gardner, G. T., and Stern, P. C. (1996). *Environmental Problems and Human Behavior*. Englewood Cliffs, NJ: Allyn & Bacon.

Kempton, W. (1995). *Environmental Values in American Culture*. Cambridge, MA: MIT Press.

Kempton, W., and Neiman, M., eds. (1987). *Energy Efficiency: Perspectives on Individual Behavior*. Washington, DC: American Council for an Energy Efficient Economy.

Nye, D. E. (1998). *Consuming Power: A Social History of American Energies*. Cambridge, MA: MIT Press.

Smil, V. (1999). *Energies: An Illustrated Guide to the Biosphere and Civilization*. Cambridge, MA: MIT Press.

Socolow, R. H. (1978). "The Twin Rivers Program on Energy Conservation in Housing: Highlights and Conclusions." In *Saving Energy in the Home*. Cambridge, MA: Ballinger Publishing Company.

Stern, P. C. (1992). "What Psychology Knows About Energy Conservation." *American Psychologist* 47:1224-1232.

U.S. Census Bureau. (1992). Census data, 1980 and 1990, Journey-to-Work and Migration Statistics Branch, Population Division, Washington, DC.

Veitch, R.; Arkkelin, D.; and Arkkelin, R. (1998). *Environmental Psychology: An Interdisciplinary Perspective*. Englewood Cliffs, NJ: Prentice-Hall.

BERNOULLI, DANIEL (1700–1782)

Daniel Bernoulli, the son of Johann Bernoulli, was born in Groningen while his father held a chair of mathematics at the university. He was born into a dynasty of mathematicians who were prone to bitter rivalry. His father tried to map out Daniel's life by selecting a wife and a career for him. By the time Daniel was thirteen, his father was reconciled to the fact that his son would never be a merchant, but absolutely refused to allow him to take up mathematics, decreeing that Daniel would become a doctor. Daniel gained his baccalaureate in 1715 and master's degree in 1716 at Basle University, but, while studying philosophy at Basle, he began learning about the calculus from his father and his older brother Nikolas. He studied medicine at Heidelberg in 1718, Strasbourg in 1719, and then returned to Basle in 1720 to complete his doctorate. About this time, he was attracted to the work of William Harvey, *On the Movement of Heat and Blood in Animals,* which combined his interests in mathematics and fluids. By 1720 his father had introduced him to what would later be called "conservation of energy," which he applied in his medical studies, writing his doctoral dissertation on the mechanics of breathing.

After completing his medical studies in 1721, he applied for a chair at Basle, but like his father before him, he lost out in a lottery. Disappointed with his lack of success, he accepted an invitation from Catherine I, Empress of Russia, to become Professor of Mathematics at the Imperial Academy in St. Petersburg in 1725. Catherine was so desperate to secure Daniel that she agreed to offer a second chair to his brother, Nikolas. Unfortunately, Nikolas died of tuberculosis shortly after arriving in Russia. Despondent over his death, Daniel thought of returning home, but stayed when his father suggested that one of his own students, Leonard Euler, would make an able assistant.

Bernoulli and Euler dominated the mechanics of flexible and elastic bodies for many years. They also investigated the flow of fluids. In particular, they wanted to know about the relationship between the speed at which blood flows and its pressure. Bernoulli experimented by puncturing the wall of a pipe with a small, open-ended straw, and noted that as the fluid passed through the tube the height to which the fluid rose up the straw was related to fluid's pressure. Soon physicians all over Europe were measuring patients' blood pressure by sticking pointed-ended glass tubes directly into their arteries. (It was not until 1896 that an Italian doctor discovered a less painful method that is still in widespread use.) However, Bernoulli's method of measuring air pressure is still used today to measure the airspeed of airplanes. Around the same time, he made yet another fundamental discovery when he showed that the movements of strings of musical instruments are composed of an infinite number of harmonic vibrations, all superimposed on the string.

Another major contribution that Bernoulli made while in Russia was the discovery that whereas a moving body traded its kinetic energy for potential energy when it gained height, a moving fluid traded its kinetic energy for pressure. In terms of mathematical symbols, the law of conservation of energy becomes:

$$P + \rho v^2 = \text{constant},$$

where P is pressure, ρ is the density of the fluid and v is its velocity. A consequence of this law is that if the pressure falls, then the velocity or the density must increase, and conversely. This explains how an airplane wing can generate lift: the air above a wing travels faster than that below it, creating a pressure difference.

By 1730 Bernoulli longed to return to Basle, but despite numerous attempts, he lost out in ballots for academic positions until 1732. However, in 1734 the French Academy of Sciences awarded a joint prize to Daniel and his father in recognition of their work. Johann found it difficult to admit that his son was at least his equal, and once again the house of Bernoulli was divided.

Of all the work that Bernoulli carried out in Russia, perhaps the most important was in hydrodynamics, a draft account of which was completed in 1734. The final version appeared in 1738 with the frontispiece "*Hydrodynamica*, by Daniel Bernoulli, Son of Johann." It is thought that Daniel identified himself in this humble fashion in an attempt to mend the conflict between himself and his father. *Hydrodynamica* contains much discussion on the principle of conservation of energy, which he had studied with his father since 1720. In addition, it gives the basic laws for the theory of gases and gave, although not in full detail, the equation of state discovered by Johannes Van der Waals a century later. A year later, his father published his own work, *Hydraulics*, which appeared to have a lot in common with that of his son, and the talk was of blatant plagiarism.

Hydrodynamica marked the beginning of fluid dynamics—the study of the way fluids and gases behave. Each particle in a gas obeys Isaac Newton's laws of motion, but instead of simple planetary motion, a much richer variety of behavior can be observed. In the third century B.C.E., Archimedes of Syracuse studied fluids at rest, hydrostatics, but it was nearly 2,000 years before Daniel Bernoulli took the next step. Using calculus, he combined Archimedes' idea of pressure with Newton's laws of motion. Fluid dynamics is a vast area of study that can be used to describe many phenomena, from the study of simple fluids such as water, to the behavior of the plasma in the interior of stars, and even interstellar gases.

After the dispute with his father in 1734, Daniel Bernoulli lost much of his drive to study mathematics and turned his attention to medicine and physiology. Finally, in 1750, Daniel was appointed chair of physics at Basle, where he taught until his death on March 17, 1782.

Douglas Quinney

BIBLIOGRAPHY

Bell, E. T. (1965). *Men of Mathematics*. London: Penguin.

Cannon, J. T., and Dostrovsky, S. (1981). *The Evolution of Dynamics: Vibration Theory from 1687 to 1742*. New York: Springes.

Fauvel, J., and Gray, J. (1987). *The History of Mathematics*. Houndmills, United Kingdom: Macmillan.

Hollingdale, S. (1983). *Makers of Mathematics*. London: Pelican.

BETHE, HANS ALBRECHT (1906–)

Hans Bethe, an only child, was born on July 2, 1906, in Strasbourg, when Alsace was part of the Wilhelminian empire. His father was a widely respected physiologist who accepted a professorship in Frankfurt when Hans was nine years old; his mother was a gifted musician who was raised in Strasbourg where her father had been a professor of medicine. The high school Bethe attended in Frankfurt was a traditional Humanistisches

Hans Albrecht Bethe. (Library of Congress)

Gymnasium with a heavy emphasis on Greek and Latin. While there, he learned Latin and Greek, read Kant, Goethe, and Schiller, and also learned French and English and a good deal of science. Classes were from 8 A.M. to 1 P.M., six days a week, with much homework assigned daily.

Bethe's talents, particularly his numerical and mathematical abilities, manifested themselves early. By the time he had finished Gymnasium he knew he wanted to be a scientist and his poor manual dexterity steered him first into mathematics, and then into theoretical physics. In the fall of 1926, after completing two years of studies at the University in Frankfurt, Bethe went to Arnold Sommerfeld's seminar in Munich. Sommerfeld was a forceful and charismatic figure, and among his students were many of the outstanding theorists of their generation: Peter Debye, Paul Epstein, Paul Ewald, Max von Laue, Wolfgang Pauli, Werner Heisenberg, Gregor Wentzel, and Fritz London. In Munich, Bethe discovered his exceptional talents and his extraordinary proficiency in physics and Sommerfeld gave him indications that he was among the very best students who had studied with him.

Bethe obtained his doctorate in 1928 summa cum laude and became Paul Ewald's assistant in Stuttgart. Ewald—whose wife was the niece of a famous and influential reform rabbi—opened his home to the young Bethe and he became a frequent visitor. Ten years later he married Rose, one of Ella and Paul Ewald's daughters. After his brief stay in Stuttgart, Bethe returned to Munich to do his Habilitation with Sommerfeld.

During the academic year 1930-1931 Bethe was a Rockefeller fellow at the Cavendish in Cambridge and in Rome in Enrico Fermi's Institute. In 1932 he again spent six months in Rome working with Fermi. Fermi and Sommerfeld were the great formative influences on Bethe. Bethe's craftsmanship is an amalgam of what he learned from these two great physicists and teachers, combining the best of both: the thoroughness and rigor of Sommerfeld with the clarity and simplicity of Fermi. This craftsmanship is displayed in full force in the many "reviews" that Bethe has written.

By 1933 Bethe was recognized as one of the outstanding theorists of his generation. His book length *Handbuch der Physik* articles on the quantum theory of one- and two-electron systems and on the quantum theory of solids became classics as soon as they were published. In April 1933, after Adolf Hitler's accession to power, he was removed from his position in Tübingen because he had two Jewish grandparents. He went to England, and in the fall of 1934 he accepted a position at Cornell University and remained there for the rest of his career. At Cornell Bethe built a school of physics where he trained and influenced some of the outstanding theoretical physicists of their generation including Emil Konopinski, Morris Rose, Robert Marshak, Richard Feynman, Freeman Dyson, Richard Dalitz, Edwin Salpeter, Geoffrey Goldstone, Robert Brout, David Thouless, Peter Carruthers, Roman Jackiw, and John Negele.

In 1938 Bethe formulated the mechanism for energy generation in stars. This research grew out of his participation at the third Washington conference on theoretical physics in April 1938. The reaction

$$^1H + {}^1H \rightarrow {}^2H + \beta^+ + \text{neutrino}$$

had earlier been suggested by Carl von Weizsäcker as a possibility for energy generation and the production of deuterium in stars. The rate of this reaction in stars was calculated by Bethe and Charles L. Critchfield before the conference. Their conclusion was that the rate of such a reaction under the conditions in stellar interiors would be enough to account for the radiation of the sun, though for stars much brighter than the sun, other more effective sources of energy would be required. Until Bethe tackled the problem nucleosynthesis was conflated with the problem of energy generation. Bethe, on the other hand, *separated* the two problems. He advanced two sets of reactions—the proton-proton and the carbon cycle—that were to account for energy production in stars like the sun. The second depended on the presence of carbon in the star. At that time there was no way to account for the abundance of carbon in stars, that is, it was not at all clear what nuclear reactions in stars between elements lighter than carbon could produce this element. However, the presence of carbon in stars had been corroborated by their spectral lines in stellar atmospheres. Bethe accepted this fact and proceeded to compute the characteristics of stars nourished by the two cycles, and found that the carbon-nitrogen cycle gives about the correct energy production in the sun.

During World War II, Bethe worked on armor penetration, radar, and helped design atomic weaponry. He was a member of the Radiation Laboratory at the Massachusetts Institute of Technology from 1942 till the spring of 1943 when he joined Oppenheimer at Los Alamos and became the head of the theoretical division. Bethe is the supreme example why theoretical physicists proved to be so valuable in the war effort. It was his ability to translate his understanding of the microscopic world—that is, the world of nuclei, atoms, molecules—into an understanding of the macroscopic properties and behavior of materials, and into the design of macroscopic devices that rendered his services so valuable at Los Alamos and later on to industry. Bethe's mastery of quantum mechanics and statistical mechanics allowed him to infer the properties of materials at the extreme temperatures and pressures that would exist in an atomic bomb. Bethe, Fermi, and the other physicists on the Manhattan Project converted their knowledge of the interaction of neutrons with nuclei into diffusion equations, and the solutions of the latter into reactors and bombs.

After the war, Bethe became deeply involved in the peaceful applications of nuclear power, in investigating the feasibility of developing fusion bombs and ballistic missiles, and in helping to design them. He served on numerous advisory committees to the government including the President's Science Advisory Committee (PSAC) and was influential in getting the United States and the Soviet Union to sign a Nuclear Test Ban Treaty in 1963. In 1967 he won the Nobel Prize for his 1938 theoretical investigations explaining the mechanism of energy production in stars. Beginning in the mid-1970s Bethe collaborated with G. E. Brown, and this association resulted in exceptional productivity. He contributed importantly to the elucidation of supernovae explosions and to the solar neutrino problem. His most recent researches were concerned with the life cycle of supernovas and the properties of the neutrinos involved in the fusion processes in the sun. He has been and continued into his mid-nineties (at the time this article was written) to be an enormously productive scientist.

Silvan S. Schweber

See also: Nuclear Energy; Nuclear Energy, Historical Evolution of the Use of; Nuclear Fission; Nuclear Fusion.

BIBLIOGRAPHY

Bernstein, J. (1979). *Hans Bethe, Prophet of Energy.* New York: Basic Books.

Rhodes, R. (1987). *The Making of the Atomic Bomb.* New York: Simon and Schuster.

Schweber, S. S. (2000). *In the Shadow of the Bomb.* Princeton, NJ: Princeton University Press.

BICYCLING

Bicycling is a simple, affordable, and energy-efficient means of transportation. Of all human-powered locomotion, it is the fastest and least energy-demanding.

The bicycle has stunning efficiency advantages over other vehicles for several reasons:

- It weighs roughly one-fifth of its payload weight. (By comparison, even a small motorcycle weighs more than the rider.)
- With the exception of avid sport and competitive use, it is typically operated at low speeds that do not cause high aerodynamic drag.
- Its internal mechanical efficiency can be nearly perfect—a 1999 test at Johns Hopkins University

Mode	*Vehicle Miles Per Gallon of Gasoline or Food Equivalent*	*Energy Use (BTU's) Per Pasenger Mile*
Bicycle	1560	80
Auto-high economy	50	600 (4 pass.)
Motorcycle	60	2100
Bus-Intercity	5	600 (45 passengers)
Subway Train		900 (1000 passengers)
747 Jet plane	0.1	3,440 (360 passengers)

Table 1.
Energy Use of Various Forms of Transportation

> showed chain efficiencies as high as 98.6 percent—and energy losses due to rolling resistance are far less than for other vehicles.

The light weight and mechanical efficiency not only allow the bicycle to be powered by a nonathletic human, but it can be walked over extreme terrain, or laden with heavy cargo, or picked up and carried. These options make the bicycle more versatile than any other vehicle, and allow a bicycle user door-to-door, on-demand transport.

For any given speed, the energy demands are close to half that of running, and 15 to 20 percent that of ice skating or roller skating. Moreover, better bikes, better techniques, and better athletes continually help set new speed and endurance records. The record speeds accomplished with recumbent two-wheeled bikes enclosed in an aerodynamic shield have surpassed 110 kilometers per hour (68 miles per hour). The record distance that cyclists can cover in an hour on an unfaired (no separate process to make-it more aerodynamic), upright bike has risen from 35 kilometers (21.7 miles) in the 1890s to over 51 kilometers (31.7 miles) in the 1990s.

The bicycle's energy efficiency superiority extends beyond human locomotion and beyond all other forms of transportation. By converting food into the energy equivalent of gasoline, the kilocalories of food energy needed by a human to pedal a bicycle is only a fraction of that needed to propel planes, trains, and automobiles (see Table 1).

HISTORY

Bicycles have been around since the early part of the nineteenth century. In 1817 Karl Von Drais invented a walking machine to get around the royal gardens faster. Made entirely of wood, the rider straddled two same-size, in-line wheels, steering with the front wheel and locomoting by pushing against the ground. This steerable hobby horse, which could surpass runners and horse-drawn carriages, never became a viable transportation option because of the need for smooth pathways, which were rare at that time.

A second major effort at self-propelled transportation came when Pierre Michaux invented in 1861 the velocipede ("fast foot") that applied pedals directly to the front wheel. To achieve greater speed with every pedal revolution, designers tried larger and larger front wheels, with some reaching almost two meters in diameter. Despite garnering interest from hobbyists, the velocipede had three major deficiencies as transportation: First, lacking gears, it was difficult to climb even a modest grade; second, because the construction was entirely of wood, with metal tires coming slightly later, the cobblestone roads of the day made for an extremely uncomfortable ride; third, the big front wheel created problems. Riding was extremely dangerous and inaccessible to most women and children.

A fresh start and the true beginning of bicycles becoming a popular means of transportation can be traced to around 1886 and the efforts of John Kemp Starley and William Sutton. With equal-sized wheels, tubular steel diamond-shaped frame geometry, and a chain-and-sprocket chain drive to the rear wheel, the "safety bike" looked much like the modern version.

During the late 1890s, bicycles were the worldwide focus of invention and technical innovation, much as biotech engineering and computers are today. We owe many of today's industrial manufacturing processes, designs for bearings, axles, and gearing mechanisms, and the knowledge of lightweight structures, to the explosion of inventions that bicycles produced.

In the United States, England, and other major nations, patents were awarded at the rate of about 5,000 per year per nation. In one peak year, bicycle-related patents comprised close to one-third of all patent-writing activity. Many of these patents were decades ahead of the technology to manufacture them; for example, suspension systems invented a century ago became viable only with modern-day elastic materials and manufacturing technology.

The two most important single inventions of this long-ago era were Starley's tension-spoked wheel and John Dunlop's pneumatic tire. The tension-spoked wheel was and is a marvel of lightweight structures; it allows four ounces of spokes, on a wheel weighing a total of three or four pounds, to support a 200-pound rider. (Today's carbon fiber wheels have yet to show a clear advantage over the wheel made from humble carbon steel spokes.) The pneumatic tire, which Dunlop invented in 1888, vastly improved the bike's comfort, and it also shielded the lightweight working mechanisms from excess vibration and fatigue. The coaster brake appeared in 1889, and it has been a staple of children's bikes ever since.

Because a bicycle uses a low-power engine (the rider) and because that rider can only apply power only over a small rpm range, gearing is essential to match the rider's output to the riding conditions. The first patent for bicycle gears was granted in France in 1868; the rider pedaled forward for one gear and backward for the other gear.

An 1869 patent by France's Barberon and Meunier foresaw today's derailleurs. It described a mechanism that would shift a belt or chain sideways among three sprockets or discs. That same year, Barberon and Meunier also patented a primitive gear hub.

These technical innovations dramatically improved performance. Bicycles of that era were hand-made in cottage industries, and were highly sought after and expensive. So impassioned were their owners that the League of American Wheelmen, founded in 1880, was for several years around the turn of the century the strongest political lobby in the United States, with a membership in the hundreds of thousands. The League's "Good Roads Movement" was the first political movement to lobby for a network of high-quality, paved roads throughout the nation.

Between the development of the chain-drive bicycle in the 1880s, and before Henry Ford popularized the automobile in the 1910s and 1920s, bicycling was extremely popular as an efficient means to quickly get around. It was over three times as fast as walking, and more convenient and cheaper than having a horse-drawn carriage at one's disposal. Unfortunately, the quickly improving performance, reliability and affordability of the automobile made it a formidable competitor to the bicycle.

The automobile ascended to dominance in North America starting in the 1920s, and in Europe by the 1950s. After World War II, most nonauto ground transport in the United States rapidly disappeared. The decades following World War II saw vast reductions in train travel, bus service, and public transportation systems in major cities.

The shift to personal autos was slower in many European countries, where many of the population could not afford autos, and where the governments placed very high taxes on gasoline and automobiles. Bicycles, motorcycles and public transportation continued to be widely used in these countries. In addition to the economic factor, there was a cultural reason for Europe's slower embrace of the automobile. Europeans have long lived with high population density within finite borders. The United States of the 1940s was a far more rural nation, with sprawling farmland inside the borders of major cities. That autos took up lots of space in a city was an obvious drawback to the European mind, but irrelevant in Texas.

The poorer nations of the planet had no choice, and used bicycles and public transportation exclusively. For example, when China was under Mao Tse Tung's rule, the number of private automobiles was only in the hundreds.

Beginning in the 1970s, the bicycle saw a resurgence of interest in North America. Reasons included environmental concerns with the internal combustion engine, the popularity of bicycling as a multifaceted sport (racing, touring, mountain-bike riding, family cycling), the desire for fitness, and the need for alternative ways of commuting in crowded cities.

Bicycling to work is viewed as the most environmentally friendly means of travel, the best way to avoid congested roadways (in some cases, it turns out to be quicker than mass transit or driving), and a means of turning commuting time to exercise time. The bicycle is far more efficient than the automobile in making good use of city streets. Traffic counts taken during the 1980 New York City transit strike showed that a single street lane devoted to bicycles

This Malaysian street scene shows two ways in which bicycles are used worldwide. (Cory Langley)

could carry six to eight times the number of people per hour than as a lane for auto traffic.

During these past few decades, the bicycle itself has vastly improved. The bicycles of the early 1970s usually had mechanical idiosyncrasies. Competition among manufacturers, led by the Japanese companies that entered the U.S. market, resulted in vastly improved quality control. The mountain bike, first made available on a widespread basis in 1983, offered a delightful alternative that mushroomed in popularity, and a decade later, traditional "road" bicycles had all but disappeared from stores. The mountain bike has become most people's vehicle of choice for city riding as well as recreational trail riding.

It has also helped that the bicycle buyer is continually offered more bicycle for less money. No longer are most bikes steel. While steel continues to be a fine material, the dominant material reported by bike enthusiast is now aluminum, which, in the hands of most designers, yields a lighter but more rigid frame. Well-heeled bicyclists opt for titanium or carbon-fiber composites.

There has also been a huge improvement in the value of today's mountain bikes. A mountain bike in the $300 to $500 price range typically has a suspension fork and aluminum frame that would have made it a $1,500 bike a decade earlier. The $800 dual-suspension bike of 2000 far outclasses the early 1990s $2,500 offering.

THE HUMAN ENGINE

The bicycle's advantages as the world's most mechanically efficient means of transportation are clouded by the limitations of the human engine. To put it in power output terms, the human body can produce sustained power only at modest levels. For most people, 100 watts would be too much, and for an elite athlete, 400 watts is the approximate ceiling. (The athlete may manage a brief burst of 1.1 kilowatts.)

The lower power output is inevitable because a body cannot long produce more power than it can simultaneously convert from the chemical mixing of blood glucose and oxygen (aerobic exercise). The higher brief bursts of power do not rely on real-time glucose/oxygen consumption, which is why the athlete is out of breath for minutes afterwards. The athlete's muscles have "borrowed" the results of future oxygen and glucose consumption in anaerobic exercise. For long-distance travel, yet another limitation appears: the body stores about a two-hour supply of glucose. After that glucose is exhausted, the body has to revert to the far less desirable mechanism of burning fat. The muscles produce less power with fat than they do with glucose, and only a trained endurance athlete can comfortably exercise beyond the glucose barrier into the fat-burning zone.

When the automobile became preeminent in the early twentieth century, it did so with good reason. Whether the energy to power a bicycle is anaerobic or aerobic in nature, it is still minuscule in comparison to what an automobile's internal combustion engine can deliver. In the United States, almost all subcompact cars are equipped with engines that can generate 100 or more horsepower (74,600 watts), and can sustain this output all day long.

The human engine cannot match this power output, yet the mechanical efficiency of the bicycle helps tremendously because a very small amount of horsepower can generate great speed. For example, 0.4 horsepower (298 watts) of output can result in 25 mph (40 kph) speeds or better. One set of calculations shows that if a cyclist rode on level ground, with no rolling resistance, and aided by a 25 mph tailwind, it would require only around 0.2 horsepower (150 watts) to sustain a 25 mph pace.

Gravitational Resistance

If not for the need to climb steep hills, bicycling at a 15 mph (24 kph) clip would never be a strenuous exercise. It takes approximately 82 watts, or 0.11 horsepower, on an efficient bicycle, to ride 15 mph on flat ground. But ground is seldom flat. That same 82 watts achieves only 8 mph climbing a barely discernible two percent grade. A five percent grade slows one down to just over 4 mph. Most riders don't want to slow down that much, so they work harder to maintain some speed. On descents, they work less hard, while going still faster.

Any weight reduction helps. Gravitational forces do not discriminate between bike and rider mass, but the human body does. If one spends an extra thousand dollars to shave ten pounds off a bike, there will be a 10-pound advantage, but if one sheds the 10 pounds from the belly, the body will not have to nurture those ten pounds of living tissue, and one will be a more efficient engine.

Air Resistance

Air resistance is a greater factor than most people realize. Even at 10 mph on flat ground, almost half the rider's energy goes to overcoming wind resistance. Rolling resistance is almost nil at the speed. At 15 mph, two thirds of the energy is need to overcome wind resistance. At 25 mph, about 85 percent of the rider's energy is devoted to overcoming wind resistance, with the remainder overcoming rolling resistance and the tiny frictional losses within the bicycle itself. So sensitive is the wind resistance to the rider's aerodynamic profile that riders who race time trials at these speeds feel the bike slow down dramatically if they sit up to take a drink from their water bottles.

For these reasons, most riders cannot, or will not, increase their speed much above their personal comfort levels, even with lots of training. The additional speed just costs the rider too much energy. A rider going 15 mph must double his power output to ride 20 mph. Why these dramatic numbers? Wind resistance varies with the square of the rider's airspeed, but the energy to overcome wind resistance increases with the cube of the rider's airspeed.

Of course, the air is rarely still, just as the ground is rarely flat. Tailwinds do speed the rider, but headwinds have a direct effect on the rider's speed, too: A rider traveling at 20 mph into a 15 mph (24 kph) headwind encounters as much air resistance as a rider traveling at 35 mph (56 kph) under windless conditions. A headwind will slow one down by half its own speed. If one normally rides 15 mph, and then steers into a 10 mph headwind, the resulting speed will be 10 mph.

Because of the significant increase in drag a rider encounters as speed is increased, small adjustments of bike and body contours can significantly alter energy expenditures. These alterations can be to clothing, frame design, handlebars, wheels (spokes), rider profile, and the race strategy of drafting. Of the two resistance factors, the rider and the bike, the rider accounts for approximately 70 percent of the wind resistance encountered, while the bicycle accounts for only 30 percent. Unlike a sleek automobile, the high upright bike and rider is a very inefficient aerodynamic profile, so encircling a rider and bike fully or partially in a streamlined fairing can drop the drag coefficient by 0.25, resulting in a top speed increase from 30 to 36 mph (48 to 58 kph). However, this gain could be somewhat negated on hot days because of the body overheating, or from instability caused by gusty crosswinds.

Fully faired bicycles and faired adult tricycles are today the stuff of cutting edge inventors and hobbyists. Virtually all of them are recumbent bikes, because it makes sense to start out with a smaller frontal area to begin with. Some are reasonably practical for daily use, with lights, radios and ventilation systems; others are pure race machines.

Rolling Resistance

Rolling resistance is almost directly proportional to the total weight on the tires. It is the sum of the deformation of the wheel, tire, and road surface at the point of contact. Energy loss occurs when the three do not return all of the energy to the cycle.

Rolling resistance varies tremendously by tire. Greater air pressure and less contact area is the reason the rolling resistance that a tops of-the-line racing tire encounters on smooth pavement is half or one-third that of a heavily-knobbed mountain bike tire.

But unlike air resistance, rolling resistance varies directly with speed, which means that as speed increases, the rolling resistance factor becomes less important relative to the air resistance. A bike going 20 mph has twice the rolling resistance of a bike going 10 mph. If the bike has good-quality tires with proper inflation pressure, neither rolling resistance number is high enough to be particularly significant.

Rolling resistance declines with smoother and harder road surfaces, larger diameter wheels, higher tire pressures, smoother and thinner tread, and narrower tires. In the case of rough, pot-holed roads, energy is lost in bounce. For soft surfaces such as gravel or sand, energy is robbed and absorbed by the surface. Anyone who has ridden with severely underinflated tires or through mud can attest to the extremely wasteful loss of human energy.

Over rough surfaces, an opposite effect, not easily measured in the laboratory, becomes apparent to the rider. A tire inflated to very high pressures (for example, 120 pounds) bounces off the peaks of the road surface, making the bike harder to control, and negating any theoretical decrease in rolling resistance. For that reason, top racers often use moderate inflation pressures (85 to 90 pounds). Studies have found that superinflatable tires (120 to 140 pounds) offer no noticeable advantage over high-inflation tires because they do not appreciably decrease rolling resistance.

Wheel size can have as dramatic an effect on rolling resistance as tire inflation. On paper, a smaller wheel size has more rolling resistance, a rougher ride, and poorer handling over bumps than a larger wheel size. Rolling resistance is inversely proportional to the radius of the cylinder, that is, given the same conditions, smaller-wheel bikes experience more resistance to motion than larger-wheel bikes.

Fortunately for the makers of small-wheel folding bikes, several factors can mitigate these shortcomings, such as by using wider tires to compensate for the smaller diameter. The use of improved modern tire technology and the use of suspension in combination with small wheels also help. In modern times, the father of small-wheel suspension bikes, Alex Moulton, began designing these bikes in the 1950s, and his most recent designs have taken this bicycle type to a new level. Hot on the heels of the Moulton bicycle are bicycles such as Germany's sophisticated dual-suspension Birde. Many riders believe these bikes completely negate the alleged disadvantages of small-diameter wheels.

Pedal Cadence

Numerous physiological studies have addressed the optimum pedal cadence, but these studies usually miss the point because they focus on the seemingly important factor of efficiency.

Efficiency is measured in power output per oxygen consumed. However, the rider's supply of oxygen is, for all practical purposes, unlimited, and most riders do not pedal at an effort level that leaves them constantly breathless. Even if they did, they would not want to follow the results of efficiency studies. These studies consistently show that recreational cyclists produce the best sustained performances (lowest metabolic rate and highest efficiency) when the seat is raised 4 or 5 centimeters above the normal height, the pedal cranks are slightly longer to make it possible to use higher gears, and the pedal cadence is in the 40 to 70 rpm range. However, experienced riders, whether fast or slow, virtually always choose to ride differently than the studies recommend.

Why? Efficiency does not matter if one has unlimited oxygen. What does matter is long-term comfort.

Pedal power has not only been found to be the least energy-demanding and fastest form of ground-based human locomotion, but pedal power excels in the water and air as well. In 1987, a water craft pedaled like a bicycle, called the Flying Fish II, reached a speed of 6.5 meters per second (14.5 miles per hour, which is slightly faster than the top speed of a single rower), and a pedal-powered plane, the Massachusetts Institute of Technology's Monarch B, completed a 1,500-meter triangular course with average speed of almost 10 meters per second (22 miles per hour).

Pedaling at the rate shown to be most efficient uses up blood glucose the fastest. This leaves the rider more susceptible to that sudden loss of energy known as "the bonk," and it also tends to leave more lactic acid and other waste products in the muscles, increasing discomfort and extending recovery time.

By contrast, pedaling faster (90 to 100 rpm) in a lower gear at a lower effort level allows the body to burn some fat along with the glucose, thereby extending the glucose reserves. The more rapid leg motion promotes blood circulation, the better to remove waste products from the muscles. This faster cadence is undeniably less efficient, because the body uses energy just to spin the legs around, but it results in increased long-term comfort.

Novice cyclists will often prefer the "more efficient" slower cadence and higher saddle because this configuration uses the leg muscles in a manner more similar to the way walking uses those muscles. In addition, the novice cyclist is often unable to benefit from a higher pedal cadence because of an inability to apply force perpendicular to the crank, resulting in excessive body motion. A bike rider needs to gain some experience with the "less efficient" faster cadence so his/her muscles develop the coordination to function smoothly in this new discipline. Having done that, the rider is unlikely to go back to the slower, less efficient way.

Cleated shoes and toeclips are also advocated on efficiency grounds. Every world-class cyclist uses toeclips today because studies have shown significant aerobic and anaerobic benefits. Toe clips often give elite riders a false sense of power production during the stroke recovery phase. Elite riders feel that toeclips double their deliverable pedal power—a coordinated push-pull effort, exerting an upward force, with the trailing leg alongside the downward force of the forward leg. In reality, however, the cadence is too fast to create a pulling-up force. The importance of cleated shoes and toe clips is in their ability to stabilize the foot and more effectively generate a pushing force rather than in generating a pulling force during the recovery stage. Recreational cyclists benefit from this foot stabilization as much as elite cyclists, do but many of them are uncomfortable with toe clips. Moreover, toe clips are undeniably clumsy in the frequent start/stop environment of crowded city traffic.

Alternative Propulsion Systems

Throughout the history of the bicycle, inventors have questioned whether the century-old circular sprocket design is the most efficient. Many inventors have built elliptical chainwheels, usually to have a "higher gear" during the power stroke and a "lower gear" during the dead spots at the very top and bottom of the pedal stroke. Numerous studies show that even elite riders are unable to apply propulsive force to the pedals during these portions of the power stroke, much as they are unable to lift on the backside of the power stroke.

These elliptical chainwheels have never been widely popular. Sophisticated cyclists tend to shun them because cyclist develop a riding rhythm and a comfort pattern from years of experience. Novices do not even know they exist. Most bike designers fear that elliptical chainwheels would tend to make novice cyclists less inclined to develop a smooth pedaling style.

During the 1980s, Japan's Shimano Corporation invented a radically different alternative to round chainwheels. The development of Shimano Biopace chainwheels began with a very sophisticated study of the biomechanical performance of bicyclists, and Shimano discovered two flaws that it wanted to correct. The first flaw was that the leg was speeding downward as the foot approached the 6 o'clock (bottom) position, and this downward momentum of the leg mass was not being harnessed and converted into forward motion. The second flaw occured during the upstroke phase of the pedal path. Shimano discovered that at this point in the pedal path, the knee joint was switching from flexion to extension, and the switch was so fast that both sets of muscles were being energized at once—so that the body was fighting itself.

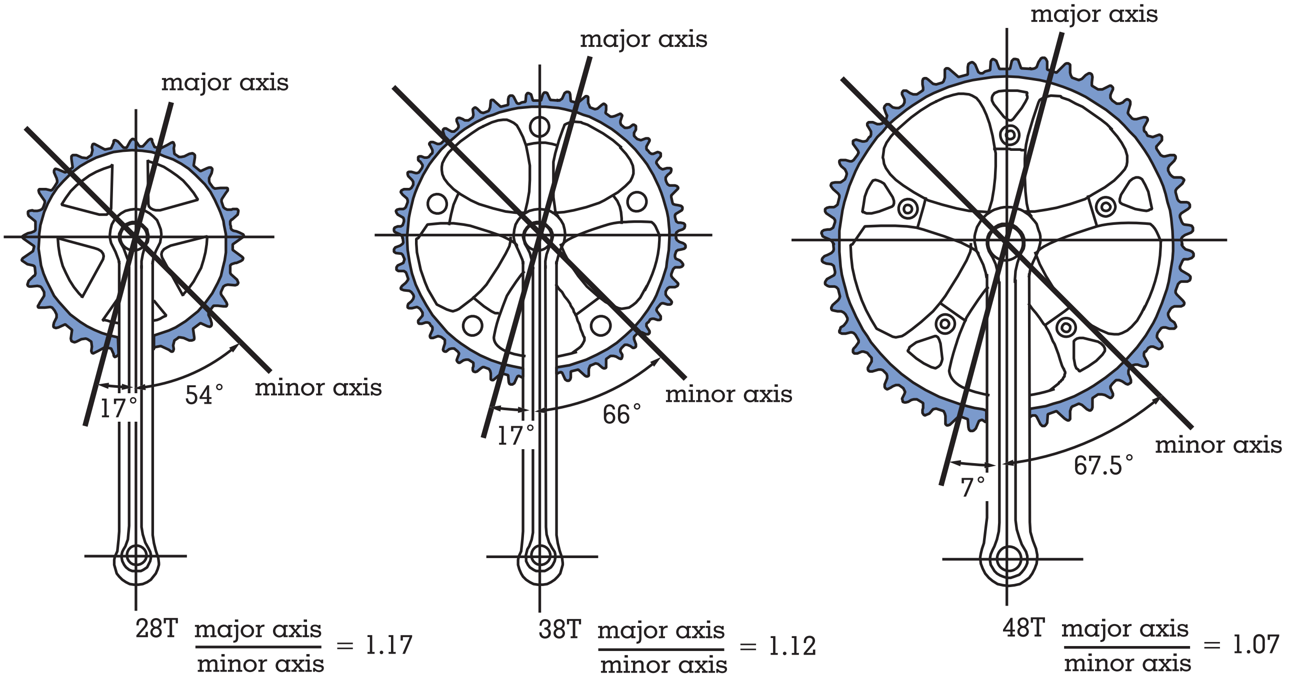


Figure 1.
Biopace chainwheels.

Shimano addressed the first problem with a design that seemed counterintuitive. The chainwheel was shaped so that the rider experienced the feeling of a lower gear during the power stroke. Then, at the bottom of the pedal stroke, the gear got higher, to absorb the energy from the leg mass's downward momentum (see Figure 1). The second problem was addressed by a change in the chainwheel shape, which slowed down that portion of the pedal stroke, giving the leg a few additional hundredths of a second to switch between flexion and extension.

As good as Shimano's research was, the product bombed in the notoriously conservative bicycle marketplace, and Biopace is history today. It's doubtful that bicycles will ever come with nonround chainwheels. The round ones do too good a job, and are too easy to make and market.

Lever propulsion—requiring an up-and-down stair-climbing motion—is another propulsion system long proposed, on efficiency grounds, as a replacement for the standard pedal-and-crank system. Though promising in theory, studies have shown that the muscle efficiency for pedaling a chain wheel is not inferior to that associated with stepping and steep-grade walking. Lever systems maximize the problem of harnessing the leg mass's downward momentum, and work against the smoothness that experienced riders have come to enjoy so much.

There is no basis for the theory that only when pushing the whole stroke vertically do the muscles work efficiently, and that the backward-and-forward foot movement over the top and bottom wastes energy. Certainly, there is some efficiency loss, but it is minimal. Toe clips and better variable gear systems have further minimized "top-dead-center problems" associated with the standard circular sprocket design.

THE BICYCLE AS TRANSPORTATION

Bicycles are the number-one mode of transportation in the world. More than 100 million new bicycles enter the market each year and, in Southeast Asia alone, around 700 million bicycles are used daily as a means of transportation.

China is the world's biggest producer and user of bicycles. Since 1985, China has been manufacturing more than 30 million bicycles each year. In Shanghai, more than 4 million people bicycle to work and school every day. Bicycles far outnumber automobiles. Whereas there is one automobile for every 1.7 people in the United States, in China there is only about one automobile for every 680 people. However, the developing nations seem determined to become overdependent on the automobile as well. In the quest for greater status and mobility, motorized vehicles are quickly becoming the most sought-after possessions in Asia and other developing nations. This pro-auto movement quickens with economic growth that creates more capital to purchase the vehicles. Paying for the energy to run the vehicle is a much smaller concern. As a larger segment of the Asian population rapidly becomes affluent and develops an addiction to the automobile—similar to that is the United States and other developed nations—world petroleum consumption will soar. The U.S. Department of Energy projects Asian transportation petroleum demand will nearly triple from 1995 to 2020, going from 9.6 to 26 quadrillion Btu's.

Although the bicycle is primarily used for recreation in North America, some urban areas (e.g., San Francisco, New York, and Toronto) are experiencing an upswing in the number of workers commuting by bicycle instead of car. Widespread use of bicycles for commuting would have a profound effect on energy supplies and society. If half of the U.S. labor force biked to work instead of driving, the United States could significantly curtail petroleum imports; moreover, the exercise attained by pedaling to work could shrink the obesity and overweight population, and thereby dramatically improve the overall health of many. Despite these benefits, a massive switch to bicycling is highly unlikely for one major reason: It is impractical. There are too many inconveniences involved that are simply unacceptable to the majority of an increasingly affluent population, particularly when the alternative is a climate-controlled, high-speed drive to work in a very comfortable automobile.

The primary inconvenience is the relative slowness and lack of comfort, especially during days of inclement weather. Many millions of Americans live great distances from where they work. Some might consider bicycling five miles to work each day, but few live that near to work. The typical 20-or-30-mile commute makes bicycling an unrealistic option for most people. And for those who do live close to work, few are willing to brave the elements to bicycle year around. Rain, snow, ice, high winds, extreme cold, and extreme heat that are minor inconveniences in a vehicle become major inconveniences on a bicycle. Thus, for backup, bicycle riders usually must own a vehicle for bad weather days, or have access to convenient mass transit.

By contrast, the bicycle survives as basic transportation in the Netherlands and Germany because those nations have a social infrastructure built to make it possible. Those countries have a neighborhood-centered way of life, and the trip distances on a bicycle are often two or three kilometers or less. People cycle slowly out of politeness to others on the crowded streets. Public transit is also far better than in the United States. It is amusing to see a German commuter train station with virtually no auto parking available, but hundreds of bike parking spots.

Intermodal commuting with a bicycle may someday be a way around congested roadways. One of the biggest problems facing mass transit developers is trying to provide service to a sprawling, residential community that needs to commute to equally decentralized work sites. The majority of the population resides a driving distance from mass transit stops, and mass transit service ends too far from a work site to make walking practical. If buses and trains were equipped to carry bicycles, or folding bicycles gained wider acceptance, the bicycle might someday serve as the intermodal link. The bike could be pedaled to the train, loaded for mass transit, and then pedaled the final leg of the commute. It is uncertain whether this alternative means of commuting will ever become attractive enought to overcome the drawbacks of weather, bicycle storage, and safety concerns.

Some people believe that bicyclists require dedicated bike lanes and trails. This is a half-truth. Certainly, these facilities are widespread in Germany and the Netherlands, but they are far less common in other cycling countries such as the United Kingdom, France, and Cuba. In those countries, bicyclists have long known that they can share the road with auto-

Bicycle use on Fifth Avenue in New York City helps to decongest the already crowded streets. (Corbis Corporation)

mobiles as long as all road users respect the same set of rules.

Dedicated facilities are often cited for safety purposes, but this, too, is not a simple truth. Most accidents occur at intersections, and dedicated facilities make intersections far more complex. When the Netherlands allowed their moped riders to travel in the auto lanes instead of the bike lanes, the moped accident rate fell by an astounding 70 percent. The bicycle accident rate would be higher if bicyclists tried to ride at brisk speed in these separated bike lanes. The sub-10-mph speeds that are considered polite in these countries largely allow bicyclists to compensate for the facilities' shortcomings, at the expense of travel time.

In the United States, advocates for bicycling are divided on the question of special facilities. It is doubtful that Dutch-style facilities would create much greater ridership in most U.S. locales, because the trip distances are too long. The best reading of accident statistics shows that adult riders well-schooled in sharing the road with automobiles and respecting the same rules ("vehicular cycling") have about one-fifth the accident rate of a control group of otherwise-well-educated adults.

Aside from the comfort and convenience disadvantages of bicycling, perhaps the biggest obstacle is the price of energy. Energy is far too cheap to make pedal power worthwhile, except as a hobby. For most people, the cost of the one or two gallons of gasoline used commuting each day is paid for in less than 10 minutes of work time. Of course, the pedal power equivalent of electrical energy is even less. To travel at 15 mph (24 kph) on flat ground requires only around 82 watts of power. If one ran on electrical energy instead of food energy, the cost of the electricity to generate that 82 watts to travel 14 miles to work is about 0.7 cent.

Sweating is another problem. Few people want to arrive at work with a sweat-soaked body. There are usually no showers at work, and few want the additional burden of lugging work clothes that would need pressing on arrival. For most riders, sweating is inevitable, particularly when climbing hills. The human body is about one-third efficient in converting the energy of glucose into muscle movement. The other two-thirds is waste heat, and sweat is how the body gets rid of excess heat.

Because of the desire to avoid perspiring, and the limited energy available to propel a bicycle, there is growing interest in electric motor-assisted bicycles that can make that 15-mile commute to work a near-effortless experience. The electric motor allows the rider to maintain a reasonable speed without exerting enough energy to perspire. The U.S. market saw at least a dozen brands of electric-assist bicycle during the 1990s. The more sophisticated offer regenerative braking 15 to 20 miles of range, and recharging cost of 2 to 3 cents for a 15-mile distance (at 10 cents kWh).

These bikes could be a wonderful transportation option, combining the efficiency, quiet and nonpollution of a bicycle with the ease of riding a small motorcycle. For these reasons, electric-assist bikes sell at the rate of over 200,000 per year in Japan. But in the United States, these bikes run up against a cultural barrier. Bicycle enthusiasts, who are less likely to object to their four-figure price tags, do not want a motor and do not want a bike that weighs 60 pounds. Others do not want to be seen riding to work. So U.S. sales have been poor.

As the number of motor scooters and motorcycles in Asia has grown at an alarming rate, policies to encourage the use of electricity-assisted bicycles are seriously being considered as a way to curtail vehicle emissions. In Bangkok, Thailand alone, over one million people needed treatment for smog-related respiratory problems in 1990. Greater affluence and the demand for faster transportation in China have resulted in yearly sales of motorcycles increasing from less than half a million sold in 1991 to over 10 million by 1996, further contributing to air pollution that is already some of the worst in the world. Unless zero-emission transportation, such as electric bicycles, gains in popularity, the number and severity of smog-related respiratory problems are certain to worsen in the most densely populated urban areas of Asia.

John Schubert
John Zunerchik

See also: Aerodynamics; Biological Energy Use, Cellular Processes of; Culture and Energy Usage; Flywheels.

BIBLIOGRAPHY

Allen, J. S. (1981). *The Complete Book of Bicycle Commuting.* Emmaus, PA: Rodale Press.

Cross, K. D. (1977). *A Study of Bicycle/Motor-Vehicle Accidents: Identification of Problem Types and Countermeasure Approaches.* Santa Barbara, CA: Anacapa Sciences.

Faria, I., and Cavanaugh, P. (1978). *The Physiology and Biomechanics of Cycling.* New York: Wiley.

Forester, J. (1994). *Bicycle Transportation.* Cambridge, MA: MIT Press.

Hunt, R. (1989). "Bicycles in the Physics Lab." *Physics Teacher* 27:160–165.

Romer, R. (1976). *The Energy Fact Book.* Amherst, MA: Department of Physics, Amherst College.

Schubert, J. (1988). *Cycling for Fitness.* New York: Random House.

Whitt, F., and Wilson, D. (1974). *Bicycle Science.* Cambridge, MA: MIT Press.

BIG BANG THEORY

The Big Bang Theory is the prevailing theory of the origin of the universe, and it is based on astronomical observations. According to this theory, about 15 billion years ago all the matter and energy in the visible universe was concentrated in a small, hot, dense region, which flew apart in a gigantic explosion.

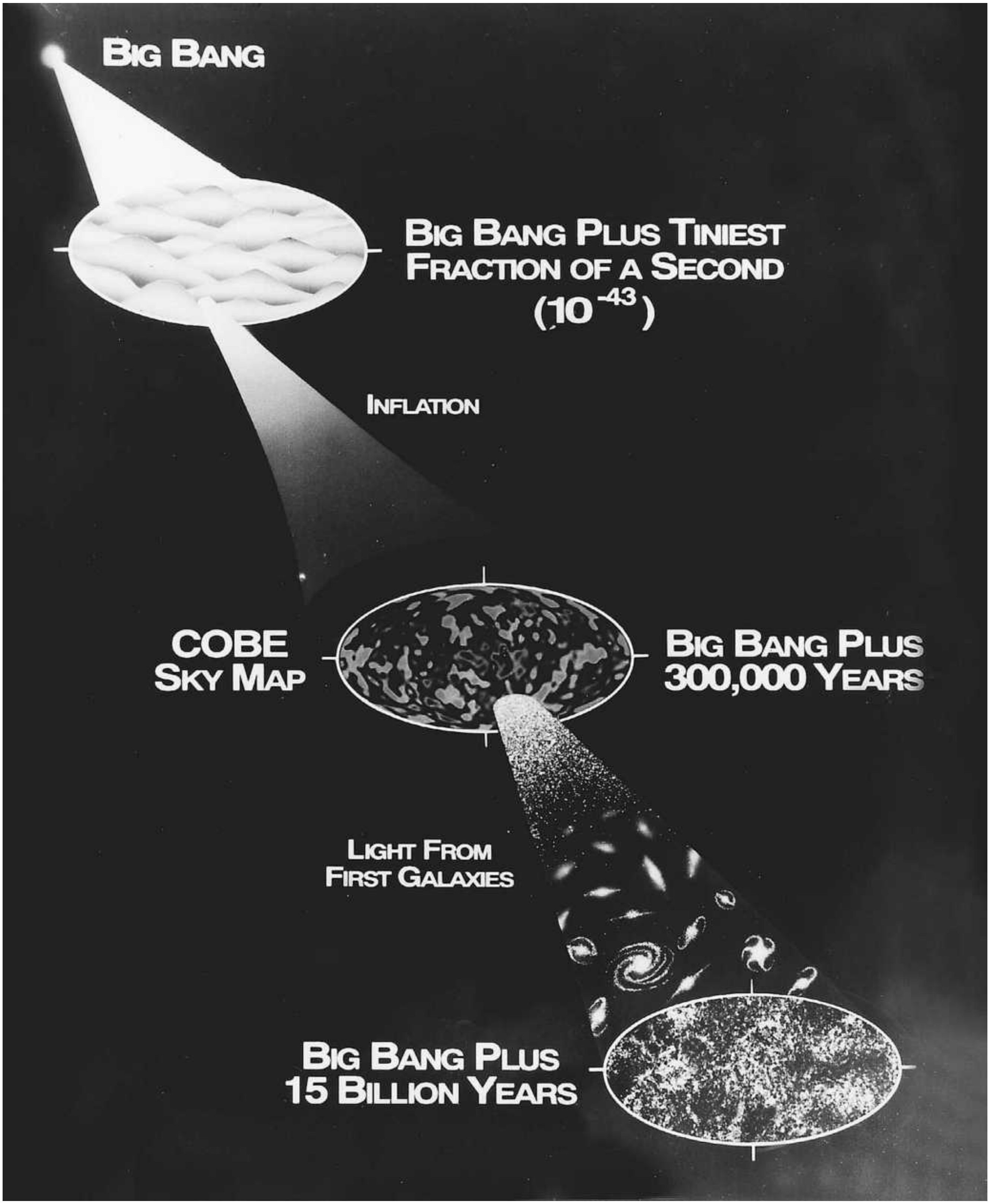
Before the twentieth century, most scientists believed the universe was static in the sense that it was neither growing nor shrinking as a whole, although individual stars and planets were moving. In 1915 Albert Einstein proposed the general theory of relativity, which is a theory of gravity that has superseded Isaac Newton's theory of gravity for very massive objects. Since general relativity was invented, its equations have been used to describe the possible ways in which the universe might change as time goes on. Einstein, like others before him, thought the universe was static, but the equations of general relativity do not allow for such a thing; according to the equations, the universe has to grow or shrink. In 1917, in order to allow for a static universe, Einstein changed the equations of general relativity by adding a term called "the cosmological constant."

AN EXPANDING UNIVERSE

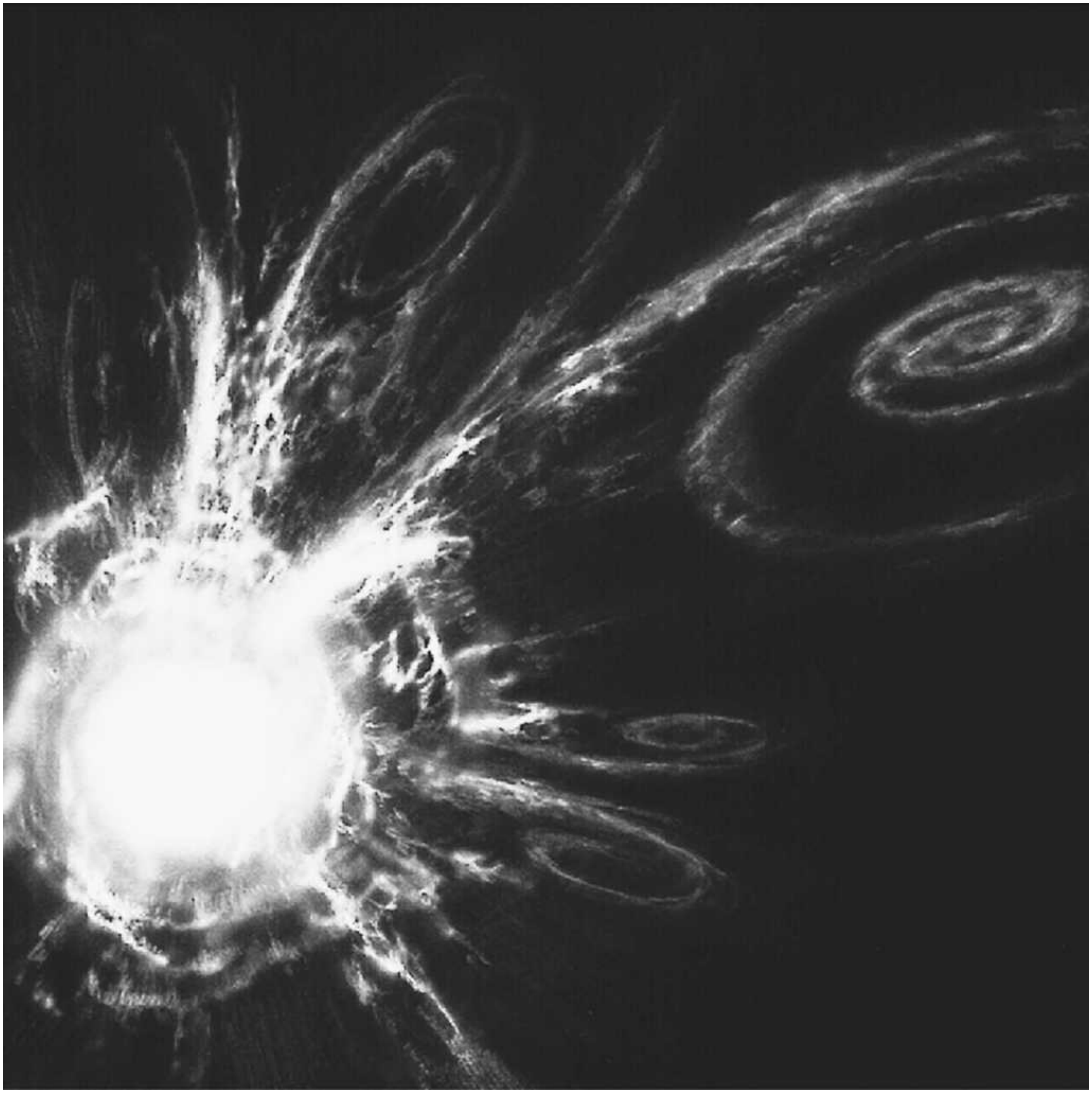
In the 1920s, cosmologists examined Einstein's original equations without the cosmological constant and found solutions corresponding to an expanding universe. Among those cosmologists was the Belgian Georges Lemaitre, who proposed that the universe began in a hot, dense state, and has been expanding ever since. This proposal came before there was any substantial evidence of an expanding universe.

Nearly all stars in the visible universe are in large clusters called galaxies. The Milky Way galaxy, the galaxy containing the sun and about 100 billion other stars, is one of about 50 billion galaxies that exist in the visible universe. In 1929 the astronomer Edwin Hubble, after making observations with a powerful telescope, discovered that distant galaxies are moving away from the earth and the Milky Way (and from one another). The farther these galaxies are from the earth, the faster they are moving, their speed being approximately proportional to their distance. Galaxies at the same distance from the earth appear to be moving away from us at the same speed, no matter in what direction in the sky the astronomers look. These observations do not mean that the earth is at the center of the universe; astronomers believe that if they made observations from any part of the visible universe that they would find the same general result.

If the galaxies are moving away from each other, then in the past they were closer to one another than they are now. Furthermore, it can be calculated from



Big Bang Theory, as conceptualized by NASA, 1992. (AP/Wide World Photos)



An artist's impression of galaxies being formed in the aftermath of the Big Bang. The spiral clouds of gas have already started condensing into the shapes of future galaxies. (Photo Researchers Inc.)

the present speeds and distances of the galaxies, that about 15 billion years in the past, all the matter and energy in the visible universe must have been in the same place. That is when the Big Bang happened. Scientists do not know what the universe was like "before" the Big Bang or even whether the concept of earlier time makes sense. The galaxies were formed out of the original matter and energy perhaps a billion years or more after the Big Bang.

THE THEORY GAINS ACCEPTANCE

Fred Hoyle, an astronomer and cosmologist who had a rival "steady state" theory of the universe, coined the name "Big Bang" in order to make fun of the theory in which the universe began in an explosion. The name stuck. Today, nearly all scientists prefer the Big Bang Theory because it can account for more observed properties of the universe than the steady state theory can. In particular, the observed microwave background radiation that appears everywhere in the sky is a remnant of the Big Bang. This radiation cannot be accounted for in a natural way by the steady state theory.

According to present theory, the galaxies are not flying apart into empty space, but space itself is growing larger. Another way of putting this is to say that the universe itself is expanding. Although the universe is expanding, one should not think that everything in the universe is expanding with it. Individual galaxies are not expanding, because their stars are prevented from flying apart by their mutual gravitational attractive forces. Likewise, other forces of nature keep the stars, the sun, Earth, and objects on Earth—down to atoms and nuclei—from expanding along with the universe.

THE FATE OF THE UNIVERSE

What will be the ultimate fate of the universe? Will the expansion go on forever or will gravity slow and then reverse the expansion into a collapse? According to general relativity, whether or not the universe will continue to expand or eventually collapse depends on the amount of matter and energy in the universe. If this matter and energy together are greater than a certain critical amount, their mutual gravitational attraction will reverse the expansion, and the universe will end with what astronomers call the "Big Crunch." If the sum of the matter and energy is below the critical amount, then, although gravity will slow the expansion, the universe will continue to expand forever. At the present time, most observations seem to favor a universe that will expand forever, but the uncertainties are large.

Astronomical observations made in the late 1990s, which are still preliminary, indicate that the expansion of the universe is not slowing down, as required by the attractive gravitational force of general relativity, but is speeding up. One way to account for this speeding up is to put back the cosmological constant into the equations of general relativity. If the cosmological constant has a certain value, general relativity allows for the speeding up that astronomers think they are seeing. When Einstein first learned that the universe was expanding, he abandoned the cosmological constant, calling it his greatest mistake. If he were alive today, what would he think about the possibility that his constant might be needed after all, but for an entirely different reason? In any case, astronomers continue to make better and better observations with their telescopes and are hoping to obtain more definite answers about the universe during the first decades of the twenty-first century. However, based on the recent history of discoveries in astronomy, it is probable that more surprises are in store.

Don Lichtenberg

See also: Matter and Energy; Particle Accelerators.

BIBLIOGRAPHY

Guth, A. (1997). *The Inflationary Universe*. Reading, MA: Addison-Wesley.

Rees, M. (1997) *Before the Beginning*. Reading, MA: Addison-Wesley.

Weinberg, S. (1977). *The First Three Minutes*. New York: Basic Books.

BIOFUELS

Biofuels are biomass (organic matter) or biomass products used for energy production. Energy created from the use of biofuels is often termed bioenergy. Biomass crops grown for the primary purpose of use in biofuels are called energy crops. Biofuels include wood and wood wastes, domestic wastes, agricultural crops and wastes, animal wastes, peat, and aquatic plants. Almost any type of combustible organic matter can potentially be used as an energy source.

Plants store solar energy by photosynthesis. During photosynthesis, carbon dioxide (CO_2) and water (H_2O) in the presence of light are converted into glucose ($C_6H_{12}O_6$) by the following chemical equation:

$$6CO_2 + 6H_2O + \text{light} \rightarrow C_6H_{12}O_6 + 6O_2.$$

Further processes in the plant make more complex molecules from the glucose. The exact makeup of biomass varies with type, but in general it has the chemical formula of $(CH_2O)_n$ and on average is about

75 percent carbohydrates or sugars and 25 percent lignin, a polymer that holds plant fibers together.

Biofuels are used to create a wide variety of energy sources. Ever since the harnessing of fire, biomass has been used for heating and cooking. Residential burning of biomass continues to be a primary source of fuel in less industrialized nations, but also has been used as fuel for electricity generation, and converted to liquid transportation fuels.

CURRENT USE OF BIOFUELS

Despite the fact that the world's biomass reserves are declining due to competing land use and deforestation, worldwide there remains more energy stored in biomass than there is in the known reserves of fossil fuels. Trees account for the largest amount of biomass. Currently biomass is the source of about percent of the energy used worldwide, primarily wood and animal dung used for residential heating and cooking. In developing countries, where electricity and motor vehicles are more scarce, use of biofuels is significantly higher (approximately 35 percent on average). At the higher end are countries such as India, where about 55 percent of the energy supply comes from biomass. Geography also is a determining factor; in some industrialized countries that have large sources of natural biomass forests near urban cities, such as Finland, Sweden, and Austria, there is a relatively high utilization of bioenergy (18, 16, and 13 percent, respectively). Municipal waste, which can be incinerated for energy production, also can be a large source of biomass for developed regions. France, Denmark and Switzerland recover 40, 60, and 80 percent of their municipal waste respectively.

At the low end is the United States, where biomass energy accounted for only about 3 percent (2.7 quadrillion Btus) of the total energy consumption in 1997. However, biomass use had been rising over the previous five years at an average rate of about 1 to 2 percent per year, but fell in 1997 due to a warmer-than-average heating season. Bioenergy produced in the United States is primarily from wood and wood waste and municipal solid waste.

These divergent energy production patterns between the developing world and the United States are understandable. Heating and cooking are the major uses of biomass in the developing world because of affordability, availability, and convenience. In the United States, where clean and convenient natural gas, propane, and electricity are widely available and affordable, biomass use has limited potential. Nevertheless, U.S. biomass energy production has been increasing because of technological advances for new and improved biomass applications for electricity generation, gasification, and liquid fuels.

The sources of biofuels and the methods for bioenergy production are too numerous for an exhaustive list to be described in detail here. Instead, electricity production using direct combustion, gasification, pyrolysis, and digester gas, and two transportation biofuels, ethanol and biodiesel, are discussed below.

ELECTRICITY GENERATION

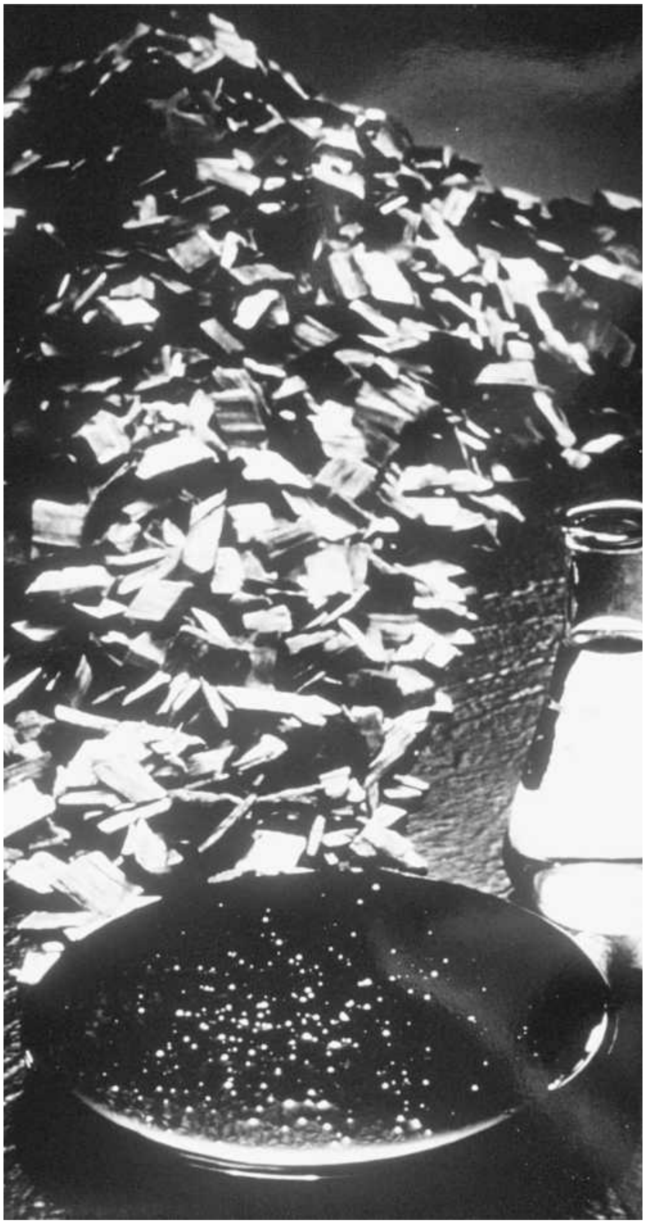
In the United States about 3 percent of all electricity produced comes from renewable sources; of this a little more than half comes from biomass. Most biomass energy generation comes from the lumber and paper industries from their conversion of mill residues to in-house energy. Municipal solid waste also is an important fuel for electricity production; approximately 16 percent of all municipal solid waste is disposed of by combustion. Converting industrial and municipal waste into bioenergy also decreases the necessity for landfill space.

These applications avoid the major obstacles for using biomass for electricity generation: fluctuation in the supply, and the type of biomass available. Seasonal variations and differing quality of feedstock are the biggest barriers to more widespread use. This is especially true for biomass wastes.

COMBUSTION

Combustion is the burning of fuels to produce heat. To produce energy, the heat from the combustion process is used to create steam, which in turn drives turbines to produce electricity.

Most electricity from biofuels is generated by direct combustion. Wood fuels are burned in stoker boilers, and mill waste lignin is combusted in special burners. Plants are generally small, being less than 50 MW in capacity. There is considerable interest in combustion of biomass in a process called cofiring, when biomass is added to traditional fuels for electricity production. Cofiring is usually done by adding biomass to coal, but biomass also can be cofired with



Wood chips, coal, and water—almost any type of combustible organic matter—can potentially be used as an energy source. (U.S. Department of Energy)

oil. There are several biomass cofiring plants in commercial operation in the eastern United States. The U.S. Department of Energy estimates that by 2020 the capacity for biomass cofiring could reach 20 to 30 GW. Cofiring has the advantage of requiring very little capital cost since most boilers can accommodate approximately 5 to 10 percent of biomass without modifications.

Estimates for delivery fuel costs for woody biomass range between $1.25 and $3.90 per million Btus compared to $0.90 to $1.35 per million Btus for coal. The cost associated with biomass electricity depends largely on the proximity of the plant to the biomass source and whether the feed is a waste material. At 10,000 Btu/kWh generation heat rate, each $1 per million Btus translates to 1 cent per kWh electrical cost. Thus biomass electricity costs can range from competitive with coal to several cents per kWh more expensive.

Cofiring biomass has environmental benefits in addition to lowering greenhouse gases. Since biomass has little or no sulfur, sulfur dioxide (SO_2) emissions are less when biomass fuels are used. In the United States, power plants have allowable sulfur dioxide levels for each gigawatt of power produced. If they produce less than the allowable amount of sulfur dioxide, they receive credits with which they can trade on the open market. The price for these sulfur dioxide credits is about $70 to $200 per ton.

Biomass also has lower levels of nitrogen than fossil fuels, leading to lower nitrogen oxide formation. The high water content in biomass also lowers the combustion temperature, decreasing the formation of thermal nitrogen oxides. In some cases this can lead to nonlinear reductions; for example, in one study when 7 percent wood was cofired with coal, nitrogen oxides emissions decreased by 15 percent. However, such reductions are not seen in all cases. Reburning is possible when using most biomass feedstocks and also can lower emissions.

Use of some biomass feedstocks can increase potential environmental risks. Municipal solid waste can contain toxic materials that can produce dioxins and other poisons in the flue gas, and these should not be burned without special emission controls. Demolition wood can contain lead from paint, other heavy metals, creosote, and halides used in preservative treatments. Sewage sludge has a high amount of sulfur, and sulfur dioxide emission can increase if sewage sludge is used as a feedstock.

GASIFICATION

Gasification of biofuels, which is in the early developmental stage, has been the focus of much recent research, since it has the potential of providing high conversion. During gasification, biomass is converted to a combustible gas by heating with a substoichiometric amount of oxygen. The biomass can be heated either directly or with an inert material such as sand. In some cases steam is added. The product gas consists of carbon monoxide, methane and other hydrocarbons, hydrogen, and noncombustible species such as carbon dioxide, nitrogen, and water; the relative amount of each depends on the type of biomass and the operating conditions. Generally the product gas has an energy content about one-half to one-quarter that of natural gas. The gas is cleaned by removing tars, volatile alkali, ash, and other unwanted materials. The gas is then sent to a steam boiler or combustion turbine for electricity production by a Rankine cycle or a combined cycle (IGCC). Use of gasification technology with an IGCC can double the efficiency of average biomass electricity production using advanced turbine technology.

The capital cost of an IGCC plant for biomass or coal is in the range of $1,500 to $2,000 per installed kW. A comparable natural gas fire facility costs about $750 to $1,000. The economics of biomass electricity based on IGCC technology depend on the relative cost of natural gas and biomass fuels. Biomass must be lower in cost than gas to pay back the additional capital cost of gas production and cleaning. A 1999 estimate suggestes that the biomass would have to be $3 per million Btus cheaper than natural gas for biomass to be economical.

PRYOLYSIS

Another emerging area in biofuels is pyrolysis, which is the decomposition of biomass into other more usable fuels using a high-temperature anaerobic process. Pyrolysis converts biomass into charcoal and a liquid called biocrude. This liquid has a high energy density and is cheaper to transport and store than the unconverted biomass. Biocrude can be burned in boilers or used in a gas turbine. Biocrude also can be chemical by altered into other fuels or chemicals. Use of pyrolysis may make bioenergy more feasible in regions not near biomass sources. Biocrude is about two to four times more expensive than petroleum crude.

BIOGAS PRODUCTION

Biogas is composed primarily of methane (CH_4) and carbon dioxide. Biogas is a by-product from anaerobic bacteria breaking down organic material. Large amounts of biogas can be released from areas such as livestock waste lagoons, sewage treatment plants, and landfills. Since biogas is primarily methane, it is similar to natural gas and can be used for energy generation, especially electricity using stationary engine-generators. The goals of capturing biogas are often to prevent these greenhouse gases from being released into the atmosphere, to control odor, and to produce fertilizer; energy production is secondary. Methane is a potent greenhouse gas, with twenty-one times the global warming potential of carbon dioxide. However, when methane is burned, it produces less carbon dioxide per Btu than any other hydrocarbon fuel.

Economics for generating electricity from biogas can be favorable. Landfill gas from municipal solid waste can supply about 4 percent of the energy consumed in the United States. In 1997, a total of 90 trillion Btus were generated by landfill gas, about 3 percent of total biomass energy consumption.

TRANSPORTATION FUELS

Although biomass used directly for heating and cooking is the thermodynamically most efficient use, followed by use for electricity generation, the economics are much more favorable to convert to a liquid fuel. Economic considerations outweigh thermodynamics; as an electricity generator, biomass must compete with relatively low-priced coal, but as a liquid fuel the competition is higher-priced oil.

Transportation fuels are the largest consumers of crude oil. Petroleum-based transportation fuels are responsible for 35 percent of greenhouse gas emissions in the United States. Only percent of transportation fuels comes from renewable nonpetroleum-based sources, primarily from the use of corn-based ethanol blended with gasoline to make gasohol. Increased use of biofuels could lower some of the pollution caused by the use of transportation fuels.

ETHANOL

The chemical formula for ethanol is CH_3CH_2OH. Ethanol is less toxic and more biodegradable than gasoline. For its octane boosting capability ethanol can be use as a fuel additive when blended with gasoline.

Demand for gasoline is 125 billion gals (473 billion l) per year according to 1998 estimates. The Clean Air Act Amendment of 1990 mandates the use of oxygenated fuels such as ethanol blends with up to 3.5 percent oxygen by weight in gasoline (E-10 or

Corn can be used to produce ethanol. (U.S. Department of Energy)

gasohol). Reformulated gasoline (RFG) is required year-round in areas that are not in compliance for ozone, and oxyfuels are required in the winter in areas that are not in compliance for carbon monoxide. These "program gasolines" total about 40 billion gals (151 billion l) per year.

In 1997 a total 1.3 billion gals of ethanol fuel was produced in the United States. Proposed new low sulfur conventional gasoline standards could greatly increase the demand for ethanol since desulfurization may lower gasoline octane. Almost all fuel ethanol is used as gasohol, but some is used to make E-85 (85% ethanol and 15% gasoline). E-85 can be used in flexible-fuel vehicles (FFVs) which can operate on gasoline or ethanol blends of to 85 percent ethanol.

Eighty-seven percent of the ethanol produced in the United States comes from corn. The remainder comes from milo, wheat, food wastes, and a small amount from wood waste. In Brazil, the largest producer of transportation biofuels, sugar cane is converted into ethanol at the rate of 16 billion l per year. There are 3.6 million cars in Brazil that run on 100 percent ethanol.

Ethanol is more costly to produce than gasoline. The cost of production of ethanol from corn ranges from about $0.80 per gal ($0.21 per l) for large depreciated wet mills to $1.20 per gal ($0.32 per l) for new dry mills. Better engineering designs, the development of new coproducts, and better uses for existing coproducts will help to lower the production cost. For example, recovering the corn germ in dry mills, which is currently in the development stage could lower ethanol production costs by $0.07 to $0.20 per gal ($0.02 to $0.05 per l). However, ethanol currently used for fuel is not competitive with gasoline without a federal excise tax exemption.

While the corn-to-ethanol industry is mature, conversion of energy crops to ethanol is in the commercial development stage. Engineering studies in 2000 estimate the cost of production per gallon for biomass ethanol at $1.22 per gal ($ 0.32 per l). The U.S. Department of Energy projects that technical advances can lower the cost to $0.60 per gallon. This would make ethanol competitive (without a tax exemption) on an energy basis with gasoline when petroleum is $25 per barrel.

To use biomass material, ethanol needs to be produced from the cellulose portion of the biomass, not just from the starch or sugars. Cellulose is more resistant to breakdown than starch or sugars, so different production methods are required. Acid-catalyzed reactions can be used for the breakdown of cellulose into products that can be converted into alcohol. This process, however, is expensive, and there are problems with the environmental disposal of dilute acid streams. Research for the development of an enzyme to break down cellulose began after World War II. It was discovered that a specific microbe, *Trichoderma reesei,* was responsible for the decomposition of canvas (cellulose) tents in tropical areas. Research on this microbe and others is being conducted. Using genetic engineering, new enzymes are being produced with the primary goal to increase efficiency of alcohol production from cellulose.

BIODIESEL

Biodiesel is diesel fuel produced from vegetable oils and other renewable resources. Many different types of oils can be used, including animal fats, used cooking oils, and soybean oil. Biodiesel is miscible with petroleum diesels and can be used in biodiesel-diesel blends. Most often blends are 20 percent biodiesel and 80 percent traditional diesel. Soy diesel can be used neat (100%), but many other types of biodiesel are too viscous, especially in winter, and must be used in blends to remain fluid. The properties of the fuel will vary depending on the raw material used. Typical values for biodiesel are shown in Table 1.

Density (@298 K), kg/m^3	860-900
Net heating value, MJ/kg	38-40
Viscosity @ 40 ˚C mm^2/s (cSt)	3.5-5.0
Cold Filter Plugging Point, K	269-293
Flash Point, K	390-440
Cetane Number	46-62

Table 1.
Typical Values for Biodiesel

Biodiesel does not present any special safety concerns. Pure biodiesel or biodiesel and petroleum diesel blends have a higher flash point than conventional diesel, making them safer to store and handle. Problems can occur with biodiesels in cold weather due to their high viscosity. Biodiesel has a higher degree of unsaturation in the fuel, which can make it vulnerable to oxidation during storage.

To produce biodiesel, the oil is transformed using a process of transesterification; agricultural oil reacts with methanol in the presence of a catalyst to form esters and glycerol. These monoalkyl esters, otherwise known as biodiesel, can operate in traditional diesel combustion-ignition engines. Glycerol from the transesterification process can be sold as a coproduct. Low petroleum prices continue to make petroleum-based diesel a more economical choice for use in diesel engines.

Current consumption of transportation diesel fuel in the United States is 25 billion gal (94.6 billion l) per year. The total production of all agricultural oils in the United States is about 2 billion gal (7.6 billion l) per year of which 75 percent is from soybeans. Total commodity waste oils total about 1 billion gal (3.8 billion l) per year. The amount of other truly waste greases cannot be quantified. Sewage trap greases consist of primarily free fatty acids and are disposed of for a fee. Trap greases might amount to 300 million gal (1.1 billion l) per year of biodiesel feedstock. The production of biodiesel esters in the United States in 1998 was about 30 million gal (114 million l). The most common oil used is soybean oil, accounting for 75 percent of oil production used for most biodiesel work. Rapeseed oil is the most common starting oil for biodiesel in Europe.

Production costs for biodiesel from soybean oil exceeds $2.00 per gal ($0.53 per l), compared to $0.55 to $0.65 per gal ($0.15 to $0.17 per l) for conventional diesel. The main cost in biodiesel is in the raw material. It takes about 7.7 lb (3.5 kg) of soybean oil valued at about $0.25 per lb (0.36 per kg) to make 1 gal (3.8l) of biodiesel. Waste oils, valued at $1 per gal ($3.79 per l) or less, have the potential to provide low feedstock cost. However, much “waste oil” is currently collected, reprocessed as yellow and white greases, and used for industrial purposes and as an animal feed supplement. Production of biodiesel from less expensive feedstocks such as commodity waste oil still costs more than petroleum diesel. Research has been done to develop fast-growing high-lipid microalgae plants for use in biodiesel production. These microalgae plants require high amounts of solar radiation and could be grown in the southwestern United States.

In addition to greenhouse benefits, biodiesels offer environmental advantages over conventional diesel. Biodiesels produce similar NO_x emissions to conventional diesel, fuel but less particulate matter. Biodiesel is more biodegradable that conventional diesel making any spills less damaging in sensitive areas. In general biodiesel provides more lubrication to the fuel system than low-sulfur diesel.

ENERGY INPUT-ENERGY OUTPUT OF BIOFUELS

Since the Sun, through photosynthesis, provides most of the energy in biomass production, energy recovered from biofuels can be substantially larger than the nonsolar energy used for the harvest and production. Estimates on conversion efficiency (energy out to non-solar energy in) of ethanol can be controversial and vary widely depending on the assumptions for type of crop grown and farming and production methods used. Net energy gain estimates for converting corn to ethanol vary between 21 and 38 percent. Conversion efficiencies can be increased if corn stover (leaves and stocks) is also used and converted to ethanol. Research is being conducted on converting other crops into ethanol. Switchgrass, a perennial, is one of the most promising alternatives. It has a net energy gain as high as 330 percent since it only has to be replanted about every ten years and because there are low chemical and fertilizer requirements. Net energy gains for the production of biodiesel are also high, with estimates ranging between 320 and 370 percent.

FUTURE USE OF BIOFUELS

One of the main benefits from future use of biofuels would be the reduction of greenhouse gases compared to the use of fossil fuels. Carbon dioxide, a greenhouse gas that contributes to global warming, is released into the air from combustion. Twenty-four percent of worldwide energy-related carbon emissions in 1997 were from the United States. Carbon and due to rising energy consumption, are expected to increase 1.3 percent per year through 2015.

When plants grow, they adsorb carbon dioxide from the atmosphere. If these plants are used for biofuels, the carbon dioxide released into the atmosphere during combustion is that which was adsorbed from the atmosphere while they were growing. Therefore the net balance of carbon dioxide from the use of biofuels is near zero. Since some fossil fuel use is required in both the planting and the production of bioenergy, there are some net carbon dioxide and other greenhouse gases released into the atmosphere. In determining the net carbon dioxide balance, important variables include growth rates, type of biomass, efficiency of biomass conversion, and the type of fossil fuel used for production. The amount of carbon accumulated in the soil and the amount of fertilizers used also have a large effect on the carbon balance. In particular, nitrous oxide (N_2O), a powerful greenhouse gas, can be released as a result of fertilizer application. Estimates for the amount of greenhouse emissions recycled using biomass for energy production range from a low of 20 to a high of 95 percent. Wood and perennial crops have higher greenhouse gas reduction potential than annual crops. Using biomass to replace energy intensive materials also can increase the carbon balance in favor of energy crops. It is estimated that the nation's annual carbon dioxide emissions could be reduced by 6 percent if 34.6 million acres were used to grow energy crops.

There is some greenhouse gas benefit from planting forests or other biomass and leaving the carbon stored in the plants by not harvesting. However, over the long term, increased carbon dioxide benefits are realized by using land that is not currently forested for growing some energy crops such as fast-growing poplar. The added benefits come from the displacing fossil fuels by the use of biofuels, since energy crops can be repeatedly harvested over the same land.

In the calculation of greenhouse gas benefits of planting energy crops, many assumptions are made. Among them is that the land will be well managed, appropriate crops for the region will be used, there will be careful use of fertilizers and other resources, and efficient production methods will be employed to get the maximum amount of energy from the biomass. Most importantly, it is assumed that biomass is grown in a sustainable manner. Harvested biomass that is not replanted increases greenhouse gas emissions in two ways: Carbon dioxide that had been previously stored in trees is released in the atmosphere, and future carbon fixation is stopped.

To comply with carbon reduction goals, some countries impose taxes on carbon dioxide emissions. Since biofuels have lower full-cycle carbon dioxide emissions than fossil fuels, biofuels are more cost-competitive with fossil fuels in regions where these taxes are imposed.

Another advantage to using biomass as an energy source is a possible increase in energy security for countries that import fossil fuels. More than two-thirds of the oil reserves are in the Middle East. More than half of the oil consumed in the United States is imported and oil accounts for approximately 40 percent of the trade deficit of the United Sates. A substantial biofuels program could help to the increase energy independence of importing nations and lessen the impact of an energy crisis.

There are some disadvantages with the use of biofuels as well. Some of the high-yield energy crops also have significant removal rates of nutrients from the soil. Each year the cultivation of row crops causes a loss of 2.7 million metric tons of soil organic matter in the United States. However, there are exceptions: Through the use of good farming practices, Brazilian sugarcane fields have had minimal deterioration from the repeated planting of sugarcane. Moreover, using switchgrass and other grasses increases soil organic matter and thus can help in reducing the soil erosion caused by the cultivation of rowcrops. Research is being conducted into improving sustainable crop yield with a minimal of fertilizer application. Possible solutions include coplanting energy crops with nitrogen-fixing crops to maintain nitrogen levels in the soil.

It is estimated that biomass is cultivated at a rate of 220 billion dry tons per year worldwide. This is about ten times worldwide energy consumption. Advocates suggest that by 2050, better use of cultivated biomass could lead to biomass providing 38 percent of the world's direct fuel and 17 percent of electricity generation. However, a large increase in bioenergy seems unlikely. When the U.S. Energy Information Administration (EIA) does not include any new greenhouse gas legislation into its energy utilization projections, only limited growth for renewable energy is predicted. The EIA estimates an average increase of 0.8 percent per year for fuels through 2020 and an average increase of 0.5 percent for renewable electrical generation without new legislation. Most of the increase comes from wind, municipal solid waste, and other biomass. The reason for low expected growth in biofuels is that natural gas and petroleum prices are expected to remain relatively low over the next few decades; in 2020 the average crude oil price is projected to be $22.73 a barrel (in 1997 dollars). The average wellhead price for natural gas is projected to increase from $2.23 per thousand cu ft ($2.17 per million Btus to $2.68 per thousand cu ft ($2.61 per million Btus in 2020 (prices in 1997 dollars). Low fossil fuel prices make it difficult for alternative fuels to compete. Projections for the amount of biomass energy use do rise, however, if it is assumed that the Kyoto protocols limiting greenhouse gases will be adopted, since biofuels contribute fewer greenhouse emissions than do fossil fuels. In the case where greenhouse gas emissions are kept to 1990 levels, renewable energy could account for as much as 22 percent of electricity generation in 2020. Even under this scenario, the biggest change in greenhouse gas emissions comes from a decrease in coal use and an increase in natural gas use.

While considerable amounts of biomass exist as wastes, the costs of collection, storage, transportation, and preparation are high. The largest obstacle for the wider us of biofuels is economics, but niche opportunities exist. Strategies to improve economics include extracting high-valued coproducts from the cellulosic matrix, offsetting disposal costs and mitigating environmental problems by using the waste.

Agricultural wastes such as corn stover (stalks, leaves, etc.) have been proposed as bioenergy sources. The annual planted corn acreage is near 80 million acres, and up to 1.5 tons of stover per acre could be collected. In many farm locations stover has a competitive use as animal feed, but in areas where higher-valued uses do not exist, it may be collected and used as an industrial feedstock. In California, rice straw presents a disposal problem, since burning has been disallowed, and the rice straw could be used for ethanol production. Alfalfa growers in Minnesota are developing a technology to separate stems from the protein-containing leaves. Since protein sources are economically valued on a ton-of-protein basis, the stems are available at essentially no cost for electricity generation. Diversion of demolition wood collected in urban areas from landfills also could yield low-cost fuels. However, if biomass is to become a large component of U.S. energy use, it will have to be grown commercially as an energy crop.

Because the energy density of biomass is much

Ethanol-powered snowplow in Hennepin County, Minnesota. (U.S. Department of Energy)

lower than that of fossil fuels, most cost analyses suggest that in order for conversion of biomass to fuels to be economical, the biomass source needs to be close to the processing facility, usually within fifty miles. Lower energy density also means that storage costs can be higher than with fossil fuels, and unlike fossil fuels, it is wholly important that storage time is minimized because weather and bacteria can lower the energy quality of the biomass.

The U.S. Department of Agriculture reports that in 1997 there were 432 million acres of cropland in the United States, of which 353 million acres were planted. Idled cropland accounted for 79 million acres, of which 33 million acres were in the Conservation Reserve Program (CRP). Some planted cropland as well as some or all the idled cropland may be available for energy crops depending on the ability of energy crops to compete economically with traditional crops and on public policy related to the use of CRP land. A 1999 study from University of Tennessee's Agricultural Policy Analysis Center and Oak Ridge National Laboratory used the POLYSYS (Policy Analysis System) model to estimate the amount of land that might be used for energy crops in 2008 based on two different scenarios. Under both scenarios it is assumed that producers are allowed to keep 75 percent of the rental rate paid by the U.S. government for CRP acreage. In both cases, switchgrass was the energy crop with the most economic potential. In the first scenario, it is assumed that the price for energy crops is $30 per dry ton ($2 per million Btus) and there are strict management practices in the CRP; in this case it is estimated that switchgrass would be competitive on 7.4 million acres. In the second scenario, it is assumed that the price for energy crops is $40 per dry ton ($2.70/per million Btus) and that there are lenient management practices in the CRP; under this scenario it is estimated that switchgrass would be competitive on 41.9 million acres. This would result in an increased annual ethanol production on the order of 4 billion to 21 billion gal (15 billion to 79 billion l) compared to the current corn ethanol production of about 1.5 billion gal (5.7 billion l) per year, or sufficient fuel for 6,000 to 36,000 MW of electrical generating capacity. Such a program could provide additional benefit to farmers by reducing the supply of commodity crops and in turn raising crop prices.

With dedicated feedstock supply systems, energy crops are grown with the primary purpose of energy generation. This means that fuel processors and growers will need to enter into long-term fuel supply contracts that provide early incentives to growers to tie up land. Woody species require four to seven years from planting to harvest. Switchgrass crops require approximately two years from planting to first harvest. High-growth species of poplar, sycamore, eucalyptus, silver maple, and willow are all being tested as energy crops. Hybrid species are being developed for pest and disease resistance. Willows have the advantage that common farm equipment can be modified for harvesting. Selection of biomass depends on many factors including climate, soil, and water availability.

Research is being done in the United States and worldwide to lower some of the barriers to biofuels. Researchers hope to develop high-yield, fast-growing feedstocks for reliable biomass fuel supplies. Research is also being done to improve the efficiency of energy conversion technologies so that more of the biomass is utilized.

Deborah L. Mowery

See also: Agriculture; Biological Energy Use, Cellular Processes of; Biological Energy Use, Ecosystem Functioning of; Diesel Fuel; Environmental Economics; Environmental Problems and Energy Use; Fossil Fuels; Gasoline and Additives; Geography and Energy Use; Green Energy; Hydrogen; Methane; Nitrogen Cycle; Renewable Energy; Reserves and Resources; Residual Fuels; Waste-to-Energy Technology.

BIBLIOGRAPHY

Bhattacharya, S. C. (1998). “State of the Art of Biomass Combustion.” *Energy Sources* 20:113-135.

Bridgwater, A. V., and. Double, J. M. (1994). “Production Costs of Liquid Fuels from Biomass.” *International Journal of Energy Research* 18:79-95.

Hinman, N. D. (1997). “The Benefits of Biofuels.” *Solar Today* 11:28-30.

Hohenstein, W. G., and Wright, L. L. (1994). “Biomass Energy Production in the United States: An Overview.” *Biomass and Bioenergy* 6:161-173.

Johansson, T. B.; Kelly, H.; Reddy, A. K. N.; Williams, R. H.; and Burnham, L. (1993). *Renewable Energy Sources for Fuels and Electricity*. Washington, DC: Island Press.

Kendall, A.; McDonald, A.; and Williams, A. (1997). “The Power of Biomass.” *Chemistry and Industry* 5:342-345.

Oritz-Canavate, J. V. (1994). “Characteristics of Different Types of Gaseous and Liquid Biofuels and Their Energy Balance.” *Journal of Agricultural Engineering Research* 59:231-238.

Sampson, R. N.; Wright, L. L.; Winjum, J. K.; Kinsman, J. D.; Benneman, J.; Kürsten, E.; Scurlock, J. M. O. (1993). “Biomass Management and Energy.” *Water, Air, and Soil Pollution* 70:139-159.

Schlamdinger, B., and Marland, G. (1996). “The Role of Forest and Bioenegy Strategies in the Global Carbon Cycle.” *Biomass and Bioenergy* 10(5/6):275-300.

Scholz, V.; Berg, W.; and Kaulfuβ , P. (1998) “Energy Balance of Solid Biofuels.” *Journal of Agricultural Engineering Research* 71:263-272.

Wright, L. L., and Hughes, E. E. (1993). “U.S. Carbon Offset Potential Using Biomass Energy Systems.” *Water, Air, and Soil Pollution* 70:483-497.

BIOLOGICAL ENERGY USE, CELLULAR PROCESSES OF

Just as an internal combustion engine requires fuel to do work, animals need fuel to power their body processes. Animals take in complex molecules as food and break them down to release the energy they contain. This process is called “catabolism.” Animals use the energy of catabolism to do work and to assemble complex molecules of their own from simple building blocks, a process called “anabolism.” The sum of anabolism and catabolism is “metabolism,” a broad term that includes all chemical reactions in the body.

LIVING SYSTEMS FOLLOW THE RULES OF THERMODYNAMICS

Living organisms are extremely complex. Perhaps this is the reason we often forget that all animals, including people, are made up entirely of chemicals and that these chemicals react with each other according to the same rules that govern chemical reactions in test tubes. Indeed, as recently as the 1800s some scientists believed that living organisms contained a "vital force" not found in inanimate objects that was necessary for life and controlled life processes. This idea, known as vitalism, is now rejected by science because this vital force has never been found and we can explain the chemical reactions in the body without resorting to the mystical thinking inherent in vitalism.

Two energy laws that apply to both living and non-living systems are the first and second laws of thermodynamics. The first law states that energy can be neither created nor destroyed, but can only be changed from one form to another. The second law states that the "entropy" (randomness, given the symbol S) of a closed system will increase spontaneously over time. At first glance, this second law would seem to make life itself impossible because living organisms increase in order and complexity (negative S) as they develop, and then maintain this order throughout adulthood. However, living organisms are not closed systems. They are able to maintain and even decrease their entropy through the input of energy from food (animals) or sunlight (plants).

The amount of energy contained in the bonds of a chemical is called the "free energy" of that chemical (given the symbol G). To understand how free energy and entropy are related, consider the following chemical reaction:

$$AB \rightarrow A + B. \tag{1}$$

The complex substrate molecule (AB) is broken down to simpler product molecules (A and B). The substrate and the products each have both a free energy and an entropy. For this, and for all chemical reactions the following relationship applies:

$$\Delta G = \Delta H - T\Delta S \tag{2}$$

where H stands for heat given off or taken up during the reaction and T stands for the absolute temperature of the reaction. G and S stand for free energy and entropy, and the Δ symbol means "change in" the variable during the reaction. Thus, Equation (2) can be stated as follows: for any chemical reaction, the change in free energy between the substrates and the products (ΔG = free energy of products - free energy of substrates) is equal to the amount of heat given off or taken up during the reaction (ΔH) minus the product of the reaction temperature times the change in entropy between the substrates and the products (ΔS = entropy of products - entropy of substrates). Reactions that give off heat are called "exothermic" reactions and have negative ΔH values. Reactions that take up heat are called "endothermic" reactions and have positive ΔH values.

According to the second law of thermodynamics, for a reaction to proceed spontaneously it must produce an increase in entropy ($\Delta S > 0$). Because most spontaneous chemical reactions in the body are exothermic ($\Delta H < 0$), most spontaneous chemical reactions will have ΔG values less than zero as well. This means that if, in the reaction shown in Equation (1) above, we begin with equal amounts of substrates and products ($[AB] = [A] \times [B]$), the reaction will proceed spontaneously (AB will be converted spontaneously to A and B) because the free energy contained in the bonds of AB is greater than the free energy contained in the bonds of A and B ($\Delta G < 0$). The more negative the value of ΔG, the greater the fraction of the available AB that will be converted to A and B.

In a practical sense, we can make use of Equation (2) above to understand this process in the following way. When a large complex molecule is broken down to smaller simpler molecules, energy is released because the smaller molecules contain less energy in their chemical bonds than the complex molecule ($\Delta G < 0$). Assuming the reaction is exothermic ($\Delta H < 0$), this energy will be released partially as heat and partially as an increased randomness in the chemical system ($\Delta S > 0$) and the reaction will occur spontaneously.

COUPLED REACTIONS CAPTURE ENERGY AND DRIVE "UNFAVORABLE" PROCESSES

While heat is vital to the human body, the reader may (quite correctly) suspect that the main reason we are interested in the energy released in chemical reactions such as Equation (1) is that this energy can also be captured, stored and used later to do useful work in the body. The energy of a chemical reaction is captured when an energy-releasing reaction ($\Delta G < 0$) is "coupled" to a reaction that requires energy. Consider the coupled set of reactions below:

$$\begin{gathered} AB \rightarrow A + B + \text{energy} \\ ADP + P_i + \text{energy} \rightarrow ATP. \end{gathered} \tag{3}$$

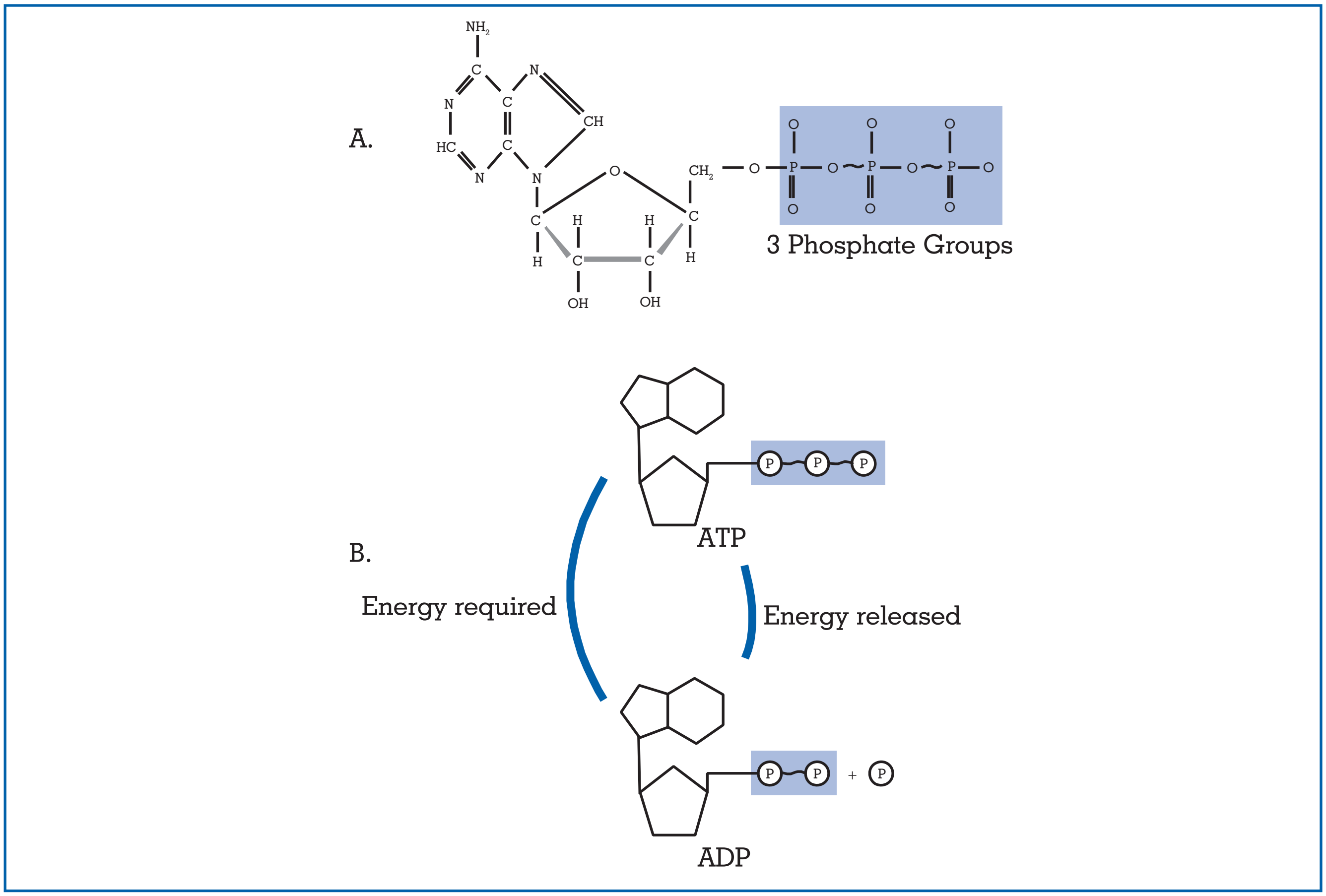


Figure 1.
The structure and formation of ATP. (A) The chemical structure of adenosine triphosphate (ATP). "C" indicates carbon, "N" nitrogen, "O" oxygen, "H" hydrogen and "P" phosphorus. Note the negative charges on the phosphate groups (PO_3^-). (B) ATP can be formed from adenosine diphosphate (ADP).

In these simultaneous reactions, the energy released when the complex molecule AB is broken down is immediately used to build a molecule of adenosine triphosphate (ATP) from a molecule of adenosine diphosphate (ADP) and an inorganic phosphate (P_i). ATP is a high energy compound. It is called the "energy currency" of the body because once it is formed, it provides energy that the body can "spend" later to drive vital reactions in cells (Figure 1).

ADP consists of an adenosine group bound to two phosphates, while ATP is the same molecule with a third phosphate bound. The phosphate groups have negative charges and repel each other as two magnets would if their negative poles were placed close together. Thus there is considerable energy in the bond that holds the second and third phosphates of ATP together. In the coupled reactions shown in Equation (3), the energy released from the breakdown of the chemical bonds in AB is transferred to the high-energy bond between the second and third phosphate groups of ATP.

Once a molecule of ATP is formed, it can be used by a cell to do work or to build complex molecules. Let us say that the cells require a complex molecule (XY). This molecule can be formed from its simpler parts (X and Y) in the reaction below:

$$X + Y \rightarrow XY. \tag{4}$$

However the formation of XY will not proceed spontaneously because the free energy of the product (XY) exceeds the free energy of the substrates (X and Y). We refer to the formation of XY as being an "unfavorable" process because, for Equation (4), $\Delta G > 0$. Cells can form the XY they need only by coupling its formation to a reaction, such as the breakdown of ATP, that provides the energy required to build the chemical bonds that hold X and Y together. This process is shown in the coupled reaction below:

$$\begin{gathered} \text{ATP} \rightarrow \text{ADP} + \text{Pi} + \text{energy} \\ \text{X} + \text{Y} + \text{energy} \rightarrow \text{XY} \end{gathered} \tag{5}$$

The energy released from the breakdown of ATP has been used to drive an unfavorable process. A reaction (the formation of XY) that would not have occurred spontaneously has taken place. Of course, the amount of energy required for the formation of one molecule of XY must be less than the amount released when one ATP is broken down, otherwise the system would have gained total energy during the coupled reaction, and violated the first law of thermodynamics.

ENZYMES INCREASE THE RATE OF CHEMICAL REACTIONS

Thus far we have discussed whether a chemical reaction will occur spontaneously or only with the addition of energy. We have said nothing about the rate of chemical reactions—how fast they occur. If we need to release the energy stored in our food to power the pumping of our heart and allow us to move, we need to release that energy rapidly. We cannot afford to wait hours or days for the energy-releasing reactions to occur.

Enzymes are complex molecules, usually proteins, that speed up chemical reactions. Figure 2 illustrates in graphic form how enzymes function. To fully understand Figure 2, imagine a chemical reaction in which a part of one compound is transferred to another compound:

$$\text{C} + \text{DE} \rightarrow \text{CD} + \text{E} \tag{6}$$

This reaction occurs spontaneously ($\Delta G < 0$), however it will only occur when a molecule of DE collides with a molecule of C with sufficient energy. The amount of energy that must be present in this collision is called the activation energy (E_a) of the reaction. If it is unusual for a molecule of DE to collide with a molecule of C with sufficient energy, the reaction will proceed very slowly. However, if an enzyme is present that binds both C and DE, the substrates will be brought closely together, reducing the activation energy. The rate at which products are formed will increase and the reaction will proceed more quickly in the presence of the enzyme. Note that enzymes have active sites with very specific shapes that bind substrate molecules and are not used up or altered during the reaction.

To understand activation energy, consider a boulder at the top of a hill. Imagine that the boulder has both the potential energy imparted by its position at the top of the hill and additional kinetic energy that causes it to "jiggle" randomly back and forth around its location. The potential energy of the boulder is analogous to the free energy of a chemical while the random motion is analogous to the random thermal motion that all molecules have. Just as chemical reactions with $\Delta G < 0$ proceed spontaneously, the boulder will have a natural tendency to roll down the hill because its potential energy state is lower at the bottom of the hill than at the top. However, if there is a small rise between the boulder and the slope of the hill, the boulder must have enough "activation" energy from its random motion to get over that rise. We could increase the likelihood that the boulder will roll down the hill either by adding more kinetic energy (giving it a push) or by lowering the rise. Enzymes work by lowering the rise (activation energy).

In thermodynamic terms, a spontaneous reaction ($\Delta G < 0$) may proceed only slowly without enzymes because of a large activation energy (E_a). Adding enzymes to the system does not change the free energy of either the substrates or products (and thus does not alter the ΔG of the reaction) but it does lower the activation energy and increase the rate of the reaction.

ANIMAL CELLS EFFICIENTLY CAPTURE THE ENERGY RELEASED DURING CATABOLISM

Animal cells obtain much of their energy from the breakdown (catabolism) of the six-carbon sugar glucose ($C_6H_{12}O_6$). The overall reaction for the catabolism of glucose is:

$$C_6H_{12}O_6 + 6\ O_2 \rightarrow 6\ CO_2 + 6\ H_2O + \text{energy} \tag{7}$$

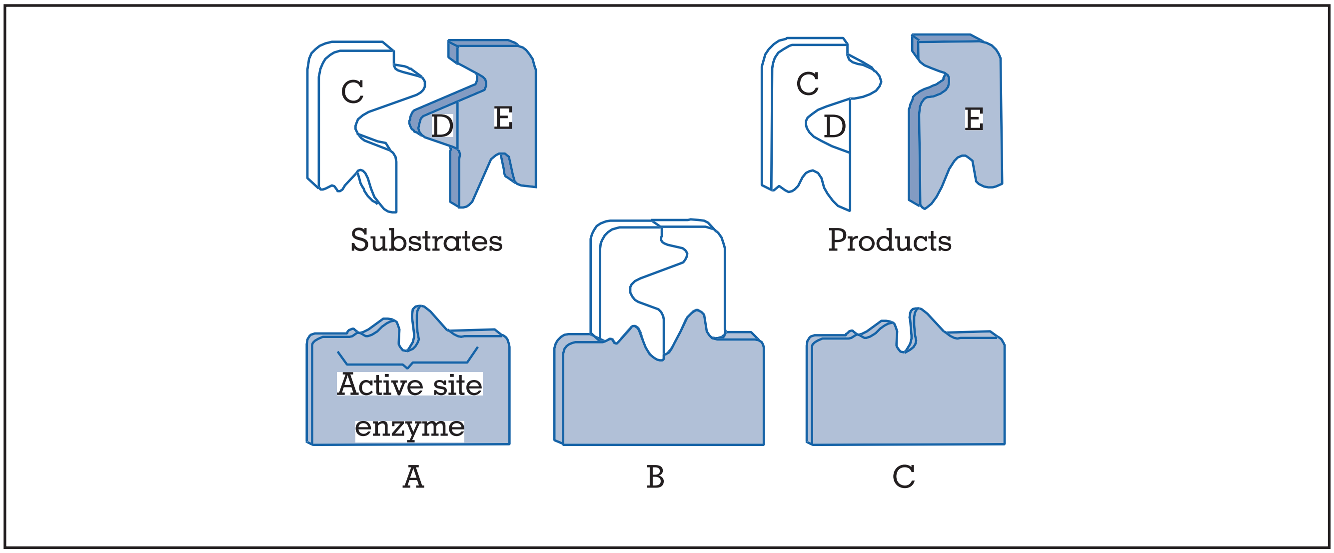


Figure 2.
The action of an enzyme. (A) An enzyme with substrate molecules C and DE. Note that the specific shape of the enzyme's active site matches the shape of the substrate. (B) The enzyme with the substrate molecules bound. (C) The enzyme, unchanged from its original form with the product molecules (CD and E).

In the presence of oxygen (O_2), glucose is broken down to carbon dioxide (CO_2) and water (H_2O). Energy is released because the free energy in the chemical bonds of the products is less than the free energy in the bonds of the glucose. It might seem simplest to couple the energy-liberating breakdown of glucose directly to each energy-requiring process in the body, much as the two chemical reactions in Equation (3) are coupled. However this is not practical. When glucose is broken down in a single step (such as by burning) a large amount of energy is released from every glucose molecule. If the catabolism of a glucose molecule were coupled directly to a process that required only a small amount of energy, the extra energy released from the glucose would be lost as heat. Thus, for efficiency, animal cells break glucose down by a multistep process. Cells release the energy in the bonds of the glucose molecule in a controlled way and capture this energy by using it to produce ATP. The breakdown of ATP, which releases energy in smaller amounts, is then coupled to energy-requiring reactions as in Equation (4).

The first segment of glucose catabolism is called "glycolysis." This process begins when glucose is transported into a cell. In a series of reactions within the cell, each of which requires a specific enzyme, a single six-carbon glucose molecule is converted to two molecules of pyruvic acid (three carbons each). For each molecule of glucose that undergoes glycolysis, two molecules of ADP are converted to ATP, and two molecules of nicotinamide adenine dinucleotide (NAD) accept a hydrogen atom and become NADH. The overall reaction of glycolysis is:

$$\text{Glucose} + 2\ \text{NAD} + 2\ \text{ADP} \rightarrow 2\ \text{Pyruvic Acid} + 2\ \text{NADH} + 2\ \text{ATP}. \quad (8)$$

The discerning reader will recognize that, while Equation (8) is written correctly, it does not explain one very interesting aspect of glycolysis. In the first two steps of glycolysis, phosphate groups are donated by ATP to glucose. This may seem odd because the goal of glucose catabolism is to liberate energy in the form of ATP but these first steps actually consume ATP! These steps have an important function, however. By adding a charged (polar) phosphate group to the glucose, they make this energy-rich molecule very insoluble in the lipid (nonpolar) cell membrane trapping the glucose inside the cell.

The next steps of glucose catabolism are called the "citric acid cycle." The pyruvic acid formed in glycolysis is transported into the mitochondria, which are subcellular organelles with double (inner and outer) membranes. They are referred to as the "powerhouses" of the cell because they produce most of the ATP. Inside the mitochondria, each three-carbon pyruvic acid molecule is converted to a two-carbon molecule of acetyl-coenzyme-A (acetyl CoA). A molecule of CO_2 is released and a molecule of NADH is generated. The acetyl CoA combines with a four-carbon molecule of oxaloacetic acid, forming the six-carbon molecule citric acid. Then, via a complex set of reactions, each of which requires its own enzyme, the citric acid is reconverted to oxaloacetic acid. Additional molecules of CO_2, NADH, $FADH_2$ (another hydrogen atom acceptor) and ATP are formed in the process. The overall reaction of the citric acid cycle is:

$$\text{Pyruvic Acid} + 4\,\text{NAD} + \text{FAD} + \text{ADP} \rightarrow 3\,CO_2 + 4\,\text{NADH} + FADH_2 + \text{ATP} \qquad (9)$$

The CO_2 generated when pyruvic acid is consumed in this cycle is the CO_2 product seen in Equation (7).

Thus far glucose catabolism has generated only a modest amount of ATP. It has, however, added a substantial number of hydrogen atoms to the hydrogen acceptor molecules NAD and FAD. The NADH and $FADH_2$ that result now pass their hydrogen atoms to a series of proteins in the mitochondrial membrane called the "electron transport system." This system splits the hydrogen atoms into a hydrogen ion (H^+) and an electron. The electron is passed from one protein to the next down the electron transport system. With each transfer, the electron gives up some energy and the protein of the transport system uses this energy to pump hydrogen ions from inside the mitochondrion to the space between the inner and outer membranes. These hydrogen ions then reenter the inner mitochondria through special hydrogen ion channels that capture the energy released in this hydrogen ion movement and use it to convert ADP to ATP. In the final step of glucose catabolism, the hydrogen ions and electrons are combined with oxygen to form water. These are the oxygen and water molecules seen in Equation (7).

Each time NADH gives up an electron to the electron transport system enough H^+ is pumped from the mitochondria to generate three molecules of ATP. However the energy of one ATP must be used to transport the NADH produced during glycolysis into the mitochondria, so this NADH generates a net gain of only two ATP for the cell. For each $FADH_2$ produced, an additional two ATP are generated. Thus, we find that for each molecule of glucose broken down the cell obtains:

2 ATP produced directly in glycolysis,
4 ATP from the 2 NADH produced in glycolysis (1 NADH per pyruvic acid),
24 ATP from the 8 NADH produced in the citric acid cycle (4 NADH per pyruvic acid),
2 ATP produced directly in the citric acid cycle, and
4 ATP from the 2 $FADH_2$ produced in the citric acid cycle (1 per pyruvic acid).

This yields a total of 36 ATP molecules produced per molecule of glucose consumed. The reader can now appreciate why it is vital that cells release the energy of glucose slowly in a multistep process rather than all at once in a single step. If glucose were broken down in a single reaction, the cell could never couple this breakdown to so many ATP-producing reactions and much energy would be lost. We can also see now why oxygen is vital to cells. In the absence of oxygen, glycolysis can proceed because the pyruvic acid generated is converted to lactic acid. (It is this lactic acid that makes muscles "burn" during heavy exercise.) This generates a small amount of ATP. However, in the absences of oxygen the electron transport system of the mitochondria backs up because it has no oxygen to accept the electrons and form water. Thus, lack of oxygen greatly reduces the ATP available to the cell and can result in cell death.

CARBOHYDRATE, FAT AND PROTEIN METABOLISM ARE CONNECTED

Both fats and proteins can also be catabolized for energy (i.e., ATP production). Dietary fats enter the bloodstream primarily as triglycerides (a three-carbon glycerol backbone and three fatty acids with sixteen or eighteen carbons in each). In cells, the fatty acids are split from the glycerol. The glycerol is converted to pyruvic acid and enters the mitochondria where it takes part in the citric acid cycle. The fatty acids undergo a process known as "beta-oxidation." During beta-oxidation, which occurs in the mitochondria, a molecule called coenzyme A is attached to the end carbon of the fatty acid chain. Then the last two carbons of the chain are cleaved with the CoA attached, producing a molecule of acetyl CoA,

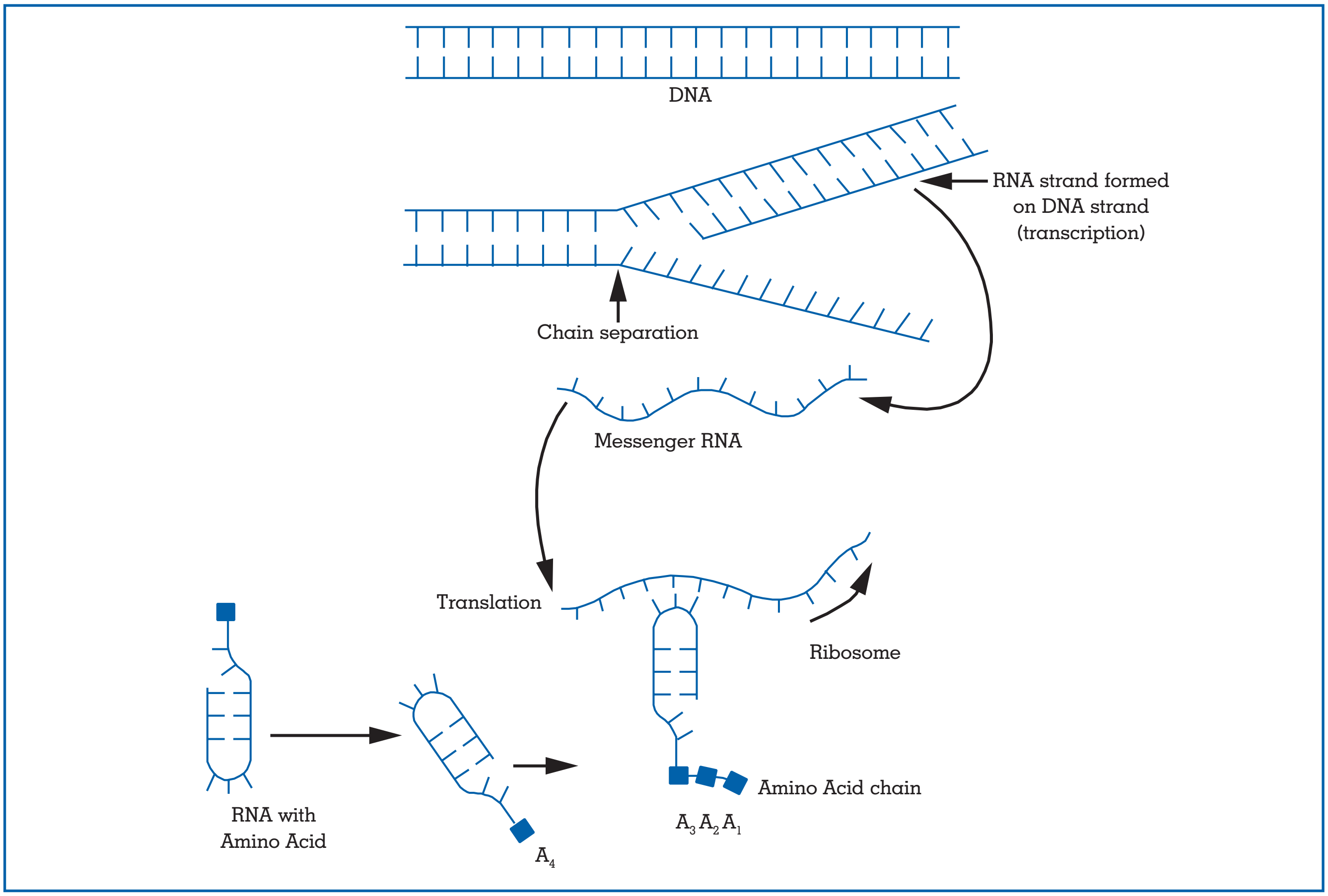


Figure 3.
A chemical reaction that will occur spontaneously because the energy level of the products (P) is less than the energy level of the substrate (S). (a) In the absence of an enzyme, activation energy is high. Few molecules have sufficient energy to overcome this barrier and the reaction proceeds slowly if at all. (b) In the presence of an enzyme, activation energy is lower and the reaction proceeds more quickly.

and a shortened fatty acid chain. This process is repeated until the entire fatty acid chain has been converted to acetyl CoA. The acetyl CoA enters the citric acid cycle. The catabolism of a single triglyceride molecule with three eighteen-carbon fatty acids yields over 450 molecules of ATP.

Dietary proteins are absorbed into the blood as amino acids, small molecules made up of a carbon backbone and a nitrogen-containing amino group (NH_2). Because protein is relatively rare and difficult to obtain, it is reasonable that the body should metabolize amino acids for energy primarily when other sources (such as sugars and fats) are unavailable. In times of great need, body cells can remove the amino group from amino acids, converting them to a form that can enter the citric acid cycle. Because muscle is made up primarily of protein, "crash" dieting causes the body to digest muscle tissue for energy.

ANIMAL CELLS USE ATP TO BUILD COMPLEX MOLECULES

Proteins are complex molecules that give cells structure and act as both enzymes and "motors" within cells. Proteins are long strings of amino acids folded in specific three-dimensional formations. There are twenty different animo acids in our bodies. DNA, the genetic material located in the cell nucleus, carries information for the order of the amino acids in each protein. Indeed, in the simplest sense, a "gene" is the section of DNA that carries the information for the construction of a single protein.

We have twenty-three pairs of chromosomes in our cells. Each chromosome is made up of a single huge molecule of DNA and contains many thousands of genes. The process by which the information in a gene instructs the cell in the formation of a protein is illustrated in Figure 3. DNA has the shape of a ladder. The ladder rungs are made up of four different molecules called "nucleotides." The information that the DNA carries is coded in the order of the ladder rungs. When a cell needs a particular protein it begins by "unzipping" the DNA ladder at the gene for that protein, exposing the information on the rungs. Then the cell makes a "messenger RNA" molecule (mRNA) that carries the same information as the gene. This process, called "transcription," requires that the cell build the messenger RNA from nucleotides. The mRNA then leaves the cell nucleus for the cell cytoplasm.

In the cytoplasm, the mRNA attaches to a ribosome and acts as a template for the construction of a protein with the proper amino acid sequence (a process known as "translation"). Single amino acids are brought to the ribosome by "transfer RNA" molecules (tRNA) and added to the growing amino acid chain in the order instructed by the mRNA. Each time a nucleotide is added to the growing RNA strand, one molecule of ATP is broken down to ADP. Each time a tRNA binds an amino acid and each time the amino acid is added to the protein, additional ATP is broken down to ADP. Because proteins can contain many hundreds of amino acids, the cell must expend the energy in 1,000 or more ATP molecules to build each protein molecule.

ANIMAL CELLS USE ATP TO DO WORK

Muscles can exert a force over a distance (i.e., do work). Thus, muscle contraction must use energy. The contractile machinery of muscle is made up of thin filaments that contain the protein "actin" and thick filaments that contain the protein "myosin" (Figure 4a). The myosin molecules have extensions known as "crossbridges" that protrude from the thick filaments. When muscle contracts, these crossbridges attach to the thin filaments at a 90-degree angle, and undergo a shape change to a 45-degree angle (power stroke) that draws the thin filaments across the thick filament. The crossbridge heads then detach, recock to 90 degrees, reattach to the thin filament and repeat the process. This entire process of myosin interaction with actin is known as the crossbridge cycle (Figure 4b).

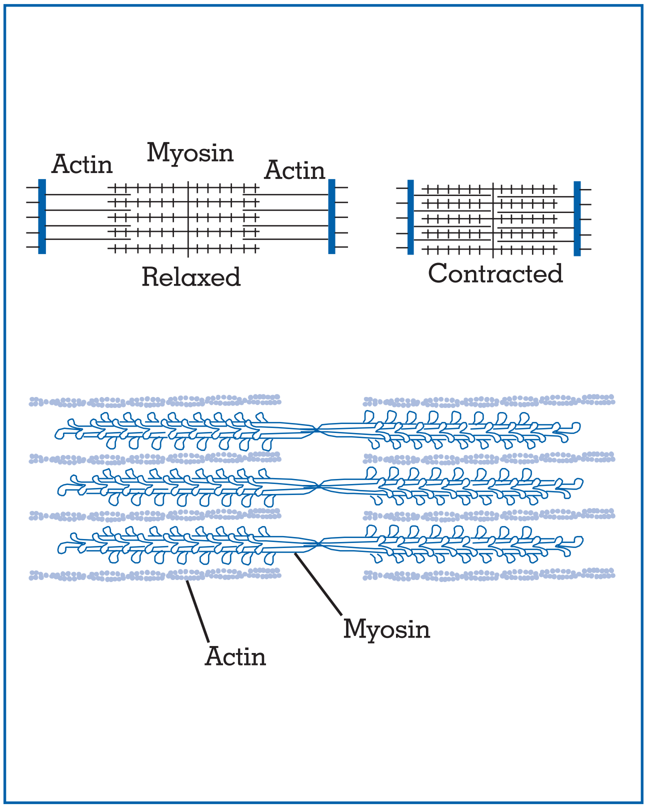


Figure 4a.
The structure and arrangement of the actin and myosin filaments in muscle. During muscle contraction the cyclic interaction of myosin crossbridges with actin filaments draws the actin filaments across the myosin filaments.

Each time a myosin crossbridge goes through its cycle it breaks down one molecule of ATP to ADP and uses the energy released to do work. It would be easier to understand this process if the energy release of ATP breakdown occurred simultaneously with the work performing step—the power stroke; however, a careful examination of Figure 4b reveals that this is not the case. The binding of ATP to myosin allows the myosin crossbridge to detach from the actin-containing thin filament. The breakdown of ATP to ADP with its energy release occurs when the crossbridge is detached and recocks the crossbridge, readying it for another power stroke.

"Efficiency" is the ratio of work done divided by

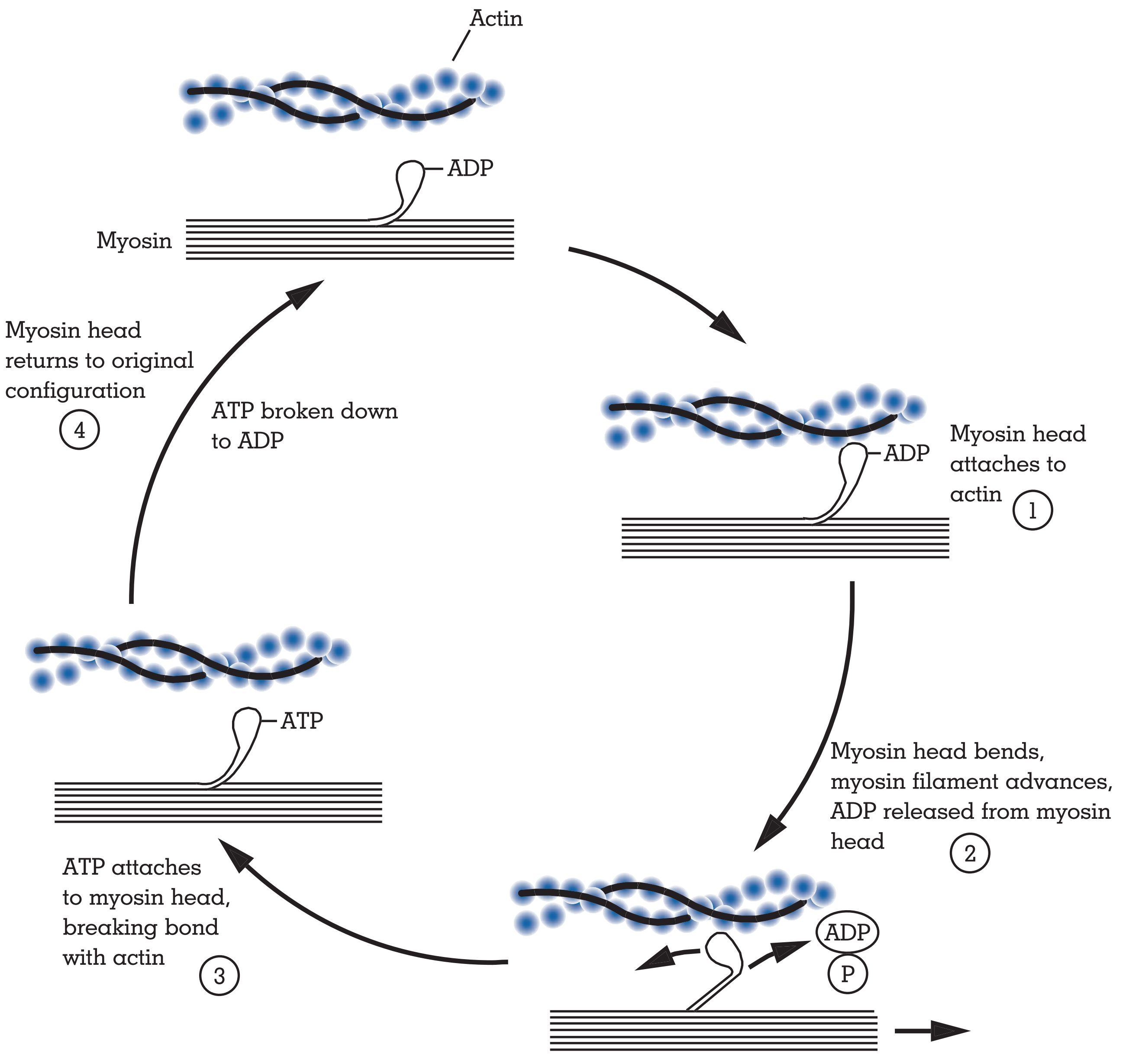


Figure 4b.
The crossbridge cycle in muscle. Myosin crossbridges interact cyclically with binding sites on actin filaments. Note that the energy release step—when ATP is broken down to ADP—recocks the crossbridge head.

energy expended. The efficiency of muscle's conversion of the chemical energy from ATP into mechanical work depends upon the rate of muscle contraction. Imagine an isolated muscle in a laboratory contracting against a weight. If the weight is too heavy for the muscle to lift, the muscle uses energy to develop force but does no work because it cannot move the weight. (Recall that work is equal to force times distance.) Thus, for contractions in which the muscle develops force but does not move a weight (isometric contractions), the muscle has an efficiency of zero. When a muscle applies a constant force to lift a weight through a distance (isotonic contractions), energy use efficiency is greatest (about 50 percent) when the muscle is contracting at one-third its maximum velocity and falls to lower levels when the muscle contracts either faster or more slowly than this. This may seem like a great waste of energy. However, much of the energy that does not do work ultimately appears as heat. This heat may not add to the strict definition of efficiency, but it is not wasted in a biological sense because it serves to maintain our body temperature.

THE METABOLIC RATE IS THE ENERGY OUTPUT OF THE ENTIRE BODY

A "calorie" is the amount of heat energy needed to raise the temperature of one gram of water by 1 degree celsius. Because this is a very small unit compared to the energy needs of the body, we use the kilocalorie, or dietary calorie (1 kilocalorie = 1,000 calories), when discussing total body metabolism. The term "Calorie" (capitalized) refers to kilocalories. The energy output of the entire body is called the "metabolic rate." This rate, expressed as Calories expended per unit time, has several components: 1) the energy needed to maintain life at rest—the basal metabolic rate or BMR, 2) the additional energy needed to digest food, and 3) any additional energy expended to perform exercise and work.

The basal metabolic rate for adults is 1 to 1.2 Calories/minute or 60 to 72 Calories/hour. This energy powers the movement of the chest during respiration and the beating of the heart—processes that are obviously necessary for life. However, a surprisingly large fraction of the BMR is used by cells to maintain ionic gradients between their interior and the fluid that surrounds them (the interstitial fluid or tissue fluid).

The interior of all body cells has a high concentration of potassium ions (K^+) and a low concentration of sodium ions (Na^+). The interstitial fluid and the blood plasma have a high Na^+ concentration and a low K^+ concentration. When electrical signals, known as "action potentials," pass along nerves, protein channels or gates in the nerve cell membrane open and allow sodium to enter the nerve cell and potassium to leave. It is the current carried by these ionic movements that is responsible for the action potential. Once the action potential has passed, the sodium that entered the nerve cell must be pumped back out and the potassium that left must be pumped back in, both against a concentration gradient. Another protein, known as the "sodium-potassium pump" does this pumping, at substantial energy cost. The work of the sodium-potassium pump comprises a significant part of the two-fifths of the BMR resulting from activity of the brain and spinal cord.

In addition to the BMR, the body uses energy to digest and store food. Digestion requires muscular contraction for the motion of the stomach and intestines, as well as the production of digestive enzymes, many of them proteins. The storage of food energy in the form of large molecules also requires energy. For example, glucose subunits are combined and stored as the large molecule glycogen in the liver and muscle. The production of glycogen from glucose requires energy input in the form of ATP.

The energy expenditure needed to produce glycogen is worthwhile for the body because glycogen serves as a ready source of glucose during periods of low food intake and high energy output. Glycogen can be broken down to glucose 6-phosphate (glucose with a phosphate group attached, see Figure 5). In muscle, glucose 6-phosphate is broken down to pyruvic acid through glycolysis and then enters the citric acid cycle. This process provides ATP for muscle contraction (Figure 4b). In the liver, the glucose 6-phosphate is converted to glucose. Without its charged phosphate group, glucose can leave the liver cells and provide for the energy requirements of other tissues, including the brain.

Body activity also adds to the metabolic rate. In general, the more strenuous the activity, the more work is done and the greater the increase in metabolic rate. For an adult male of average size, the BMR (measured lying down) accounts for 1,500-1,600 Calories per day. If this subject sat still but upright in a chair, he would use over 2,000 Calories per day, and if he engaged in prolonged strenuous activity he might expend as much as 10,000 Calories per day. Young people generally have higher metabolic rates than do elderly individuals, partially because younger people have, on average, more muscle mass than the elderly.

The metabolic rate is increased by several hormones including thyroid hormone, adrenalin and male sex hormones. The increase in metabolic rate caused by male sex hormones explains why males have slightly higher average metabolic rates than females of the same size and age. Living in a cold climate increases the metabolic rate because the cold stimulates thyroid hormone production and this hormone increases heat output of the body, while living in a warm climate causes the metabolic rate to decrease.

Training increases the body's ability to perform physical activity. The basic structure of muscle and crossbridge cycle are not altered by training. However, performing strength exercises makes muscle stronger by adding more thick and thin filaments (and thus more crossbridges) in parallel with those that already exist. Cardiovascular training increases the number and size of the blood vessels that supply oxygen to muscles, strengthens the heart and lungs, and even increases the ability of muscle cells to produce ATP by increasing the number of mitochondria they contain. As a result, a trained athlete can achieve a much higher metabolic rate and perform far more work when they exercise than can an untrained individual.

Most physical activity includes moving the body through a distance. In general, larger animals expend less energy to move each gram of body tissue at fixed velocity than do small animals. This difference probably results from the fact that small animals need a faster rate of muscle shortening (and therefore a faster crossbridge cycle) to achieve a given velocity of motion, and means that small animals are inherently less efficient in their locomotion than are large animals. On the other hand, because small animals have less inertia and experience less drag, they can accelerate to maximum speed more quickly and with less energy expenditure than larger animals.

Different animals employ different forms of locomotion, and these forms also differ in efficiency. Most swimming animals are at or near neutral buoyancy in water, and thus do not need to expend energy working against gravity. For this reason, swimming is inherently more efficient than flying even though the swimming animal must move through a medium (water) that is much denser than air. Running is the least efficient form of locomotion because running animals (including people) move their body mass up and down against gravity with every stride.

The metabolic rate can be measured in several ways. When no external work is being performed, the metabolic rate equals the heat output of the body. This heat output can be measured by a process called direct calorimetry. In this process, the subject is placed in an insulated chamber that is surrounded by a water jacket. Water flows through the jacket at constant input temperature. The heat from the subject's body warms the air of the chamber and is then removed by the water flowing through the jacketing. By measuring the difference between the inflow and outflow water temperatures and the volume of the water heated, it is possible to calculate the subject's heat output, and thus the metabolic rate, in calories.

Another method of measuring the metabolic rate, and one that allows measurements while the subject is performing external work, is indirect calorimetry. In this process, the subject breathes in and out of a collapsible chamber containing oxygen, while the carbon dioxide in the subject's exhaled air is absorbed by a chemical reaction. The volume decrease of the chamber, equivalent to the amount of oxygen used, is recorded. Because we know the total amount of energy released in the catabolism of glucose, and the amount of oxygen required for this process (Equation 7) it is possible to calculate the metabolic rate once the total oxygen consumption for a period of time is known. Of course, our bodies are not just breaking down glucose; and other nutrients (fats and proteins) require different amounts of oxygen per Calorie liberated than does glucose. For this reason, indirect calometric measurements are adjusted for the diet of the subject.

THE BODY CAN STORE ENERGY-RICH SUBSTANCES

Our bodies must have energy available as ATP to power chemical reactions. We must also store energy for use during periods of prolonged energy consumption. When we exercise, ATP powers the myosin crossbridge cycle of muscle contraction (Figure 4b). However, our muscle cells have only enough ATP for about one second of strenuous activity. Muscle also contains a second high-energy

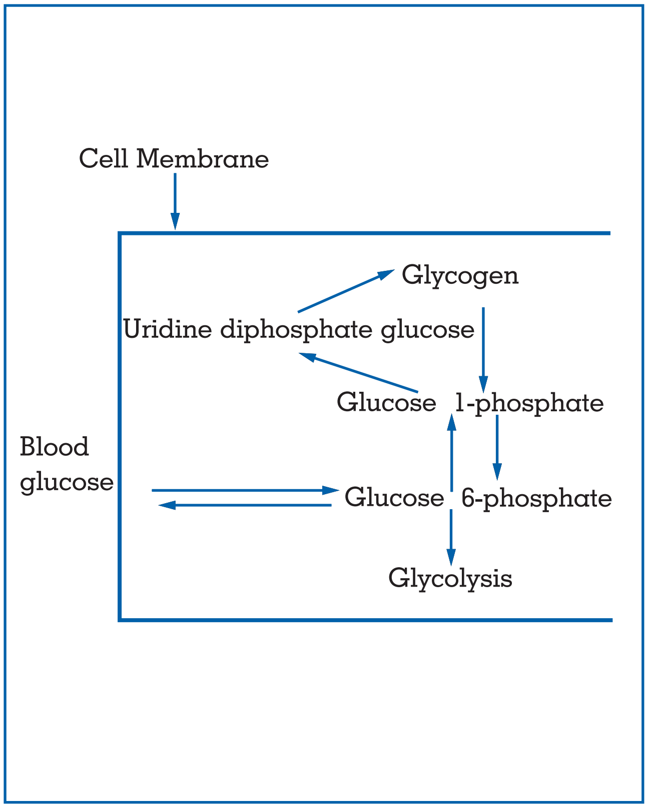


Figure 5.
The breakdown of glycogen to glucose. The glucose then enters glycolysis and the citric acid cycle, providing energy in the form of ATP.

compound called "creatine phosphate" that can give up energy to reconvert ADP to ATP by means of the coupled reactions below:

$$\begin{gathered}\text{creatine phosphate} \rightarrow \text{creatine} + \text{inorganic phosphate} \\ \text{ADP} + \text{inorganic phosphate} \rightarrow \text{ATP}\end{gathered} \tag{10}$$

Muscles contain enough creatine phosphate to power contraction for about ten seconds.

For muscle contraction to continue beyond this brief period, we must rely on the stored energy reserves of glycogen and fats. Glycogen, the storage form of carbohydrate (Figure 5), is present in muscle and liver. Even in the absence of sufficient oxygen, muscle glycogen can be broken down through glycolysis to provide enough energy for an additional five minutes of muscle contraction. When oxygen is present, the glycogen in muscle and liver provide enough energy (via the citric acid cycle) to power muscle contraction for three hours or more. The "carbo loading" athletes engage in before a marathon race (often including a large pasta dinner) is an attempt to "top off" their glycogen stores, and a depletion of glycogen may contribute to the phenomenon of "hitting the wall" in which athletic performance declines substantially after several hours of intense exercise. When glycogen stores are depleted, we must rely on fats to power muscle contraction.

In a prolonged fasting state, liver glycogen can supply glucose to the blood, and ultimately to tissues such as the brain that preferentially use glucose, for only about twelve hours, even at rest. Thereafter fat and protein stores are broken down for energy. Even people of normal weight have an average of 15 percent (for men) to 21 percent (for women) body fat. So, as long as their fluid intake is sufficient, a healthy person may survive as much as two months of fasting. However fasting has significant negative effects. Fats are broken down to glycerol and fatty acids. Glycerol is converted to glucose by the liver in a process called "gluconeogenesis," and the glucose is released into the blood to provide energy for the brain. The fatty acids are metabolized by a variety of tissues, leaving keto acids that cause the acid level in the blood to rise (fall in pH). Some protein is broken down to amino acids and these, like glycerol, takes part in gluconeogenesis. The metabolic rate falls as the body attempts to conserve energy. This reduction in metabolic rate with fasting is one reason crash dieting is so ineffective. When fat stores are used up, protein catabolism accelerates. Body muscle is digested and the person develops the stick-thin extremities characteristicsc of starving children and those with anorexia nervosa. If they do not receive nourishment, these people will soon die.

Those with a normal diet take in food in the forms of carbohydrates, fats and proteins. Because it has a low water content and produces so many ATP molecules, fat yields 9.3 Calories per gram while carbohydrates and proteins yield less than half as much (4.1 and 4.3 calories per gram respectively). Thus, we get a huge number of calories from a small quantity of fat eaten. The average person in the United States has a diet with 50 percent of the calories in the form of carbohydrates, 35 percent in the form of fat and 15 percent in the form of protein. We need about 1 gram of protein per kilogram of body weight per day to replace body proteins that are broken down. A 70 kg person on an average 5,000-Calorie per day diet receives over twice this amount (5,000 Calories per day × 0.15 of Calories as protein/ 4.3 Calories per gram of protein = 174 grams of protein per day). Thus, most of us are in little danger of protein deficiency. Most of us would probably be healthier if we ate less fat as well. A high fat diet is a known risk factor for diseases of the heart and blood vessels, as well as for colon cancer. Because autopsies on young otherwise healthy soldiers killed in combat indicate that fatty deposits in arteries can be well established by twenty years of age, it is never too early to begin reducing fat intake. A modest goal might to have no more than 30 percent of dietary calories from fat. However, studies from cultures where people consume little red meat indicate that it is possible, and almost certainly healthy, to reduce our fat intake far more than this.

We have seen that energy flow in the body's chemical reactions follows the same basic rules as does energy change in nonliving systems. Energy is taken in as food, then either stored as fat or glycogen, or released in an orderly manner through a multistep enzyme-controlled process and converted to ATP. The ATP is then used to synthesize large molecules needed by the body and to power body processes that do work. The body's overall metabolic rate can be measured and is affected by a variety of internal and external factors. Diet affects the body's energy stores, and insufficient or excess food intake influences metabolic processes. We can use an understanding of metabolism to match our food intake to our body needs and in so doing to maximize our health.

David E. Harris

See also: Biological Energy Use, Ecosystem Functioning of.

BIBLIOGRAPHY

Berne, R. M., and Levy, M. N. (1998). *Physiology*, 4th ed. Boston, MA: Mosby Year Book.

Eckert, R.; Randall, D.; and Augustine, G. (1988). *Animal Physiology*, 3rd ed. New York, NY: W. H. Freeman and Company.

Guyton, A. C., and Hall, J. E. (1996). *Textbook of Medical Physiology*, 9th ed. Philadelphia, PA: W. B. Saunders Company.

Marieb, E. N. (1999). *Human Anatomy & Physiology*, 5th ed. Reading, MA: Benjamin/Cummings Publishing Company.

Martini, F. M. (1998). *Fundamentals of Anatomy & Physiology*, 4th ed. Upper Saddle River, NJ: Prentice-Hall.

Schmidt-Nielsen, K. (1997). *Animal Physiology*, 5th ed. Cambridge, England: Cambridge University Press.

Segel, I. H. (1976). *Biochemical Calculations*, 2nd ed. New York, NY: John Wiley & Sons.

Woledge, R. C.; Curtin, N. A.; and Homsher, E. (1985). *Energetic Aspects of Muscle Contraction*. Monographs of the Physiology Society, No. 41. Boston, MA: Academic Press.

BIOLOGICAL ENERGY USE, ECOSYSTEM FUNCTIONING OF

THE NEED FOR ENERGY

The bald eagle has been the national symbol of the United States since 1782, representing freedom, power, and majesty. The eagle's impressively large beak and talons, combined with its ability to detect details at great distances, make this bird an imposing predator. Interestingly, the most important factor determining what this bird looks like, how it behaves, the prey it seeks, how it interacts with the environment, and the number of bald eagles that are supported by the environment, is energy. In fact, energy is probably the most important concept in all of biology.

Energy is the ability to do work, and work is done when energy has been transferred from one body or system to another, resulting in a change in those systems. Heat, motion, light, chemical processes, and electricity are all different forms of energy. Energy can either be be transferred or converted among these forms. For example, an engine can change energy from fuel into heat energy. It then converts the heat energy into mechanical energy that can be used to do work. Likewise, the chemicals in food help the bald eagle to do the work of flying; heat allows a stove to do the work of cooking; and light does the work of illuminating a room. Biological work includes processes such as growing, moving, thinking, reproducing, digesting food, and repairing damaged tissues. These are all actions that require energy.

The first law of thermodynamics states that ener-

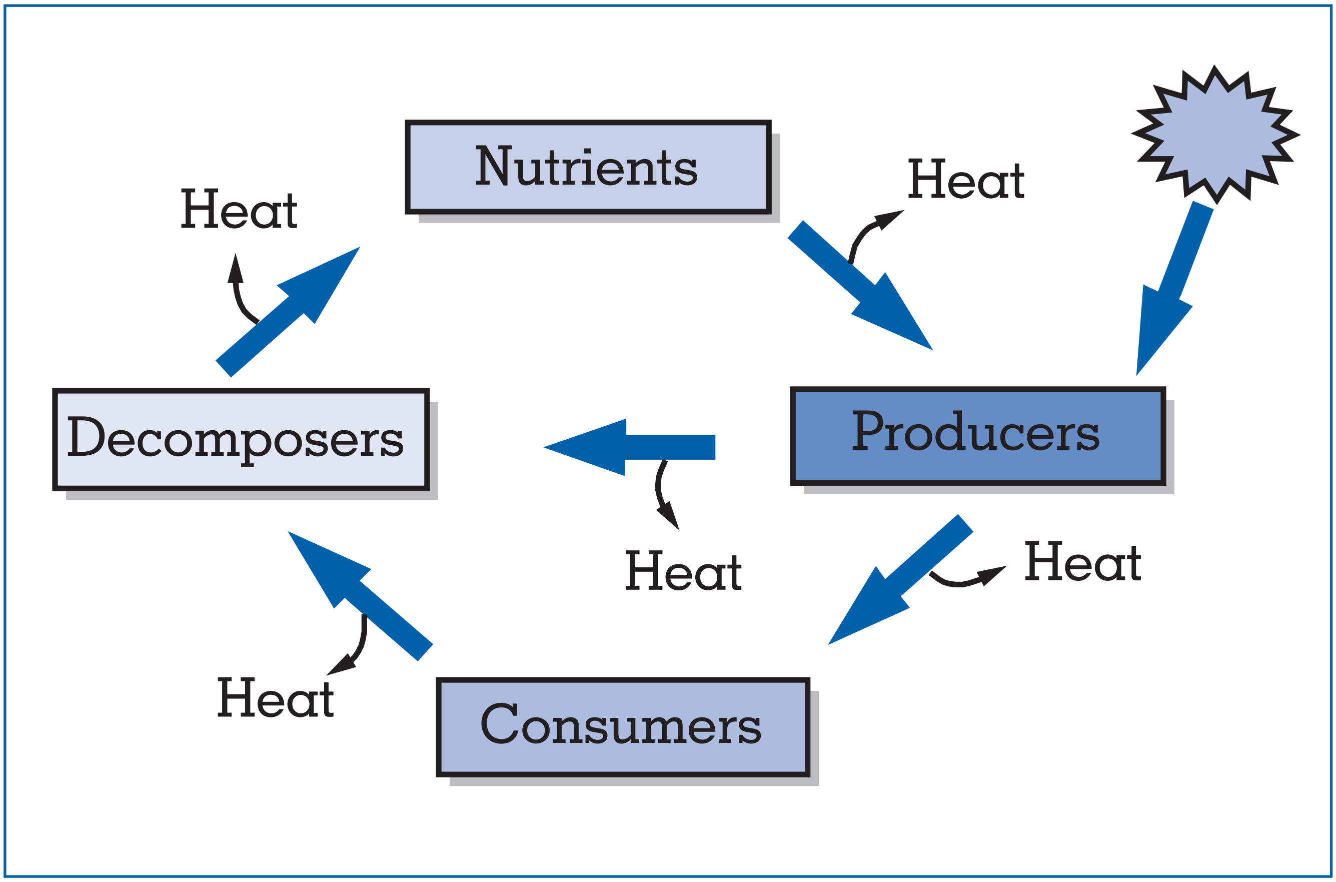


Figure 1.
Energy from the Sun enters the food chain through plants (the producers). At each stage of the food chain, energy is lost as heat.

gy can neither be created nor destroyed. This seems to imply that there is an abundance of energy: If energy cannot be destroyed, then there must be plenty available to do biological work. But biological work requires high-quality, organized energy.

Energy can be converted from one form to another. According to the second law of thermodynamics, each time energy is converted, some useful energy is degraded into a lower-quality form—usually heat that disperses into the surroundings. This second law is very interesting in terms of biology because it states that every time energy is used, energy quality is lost: The more energy we use, the less there is to do useful work. As we shall see, this principle influences every biological event, from interactions between predators and their prey to how many species can live in a habitat.

Plants and Animals Get Energy from the Sun

Living systems are masters of energy transformation. During the course of everyday life, organisms are constantly transforming energy into the energy of motion, chemical energy, electrical energy, or even to light energy. Consequently, as dictated by the laws of thermodynamics, each living system is steadily losing energy to its surroundings, and so must regularly replenish its supply. This single fact explains why animals must eat and plants must harvest light energy through photosynthesis.

Where does this supply of energy come from? Most life on Earth depends on radiant energy from the Sun, which delivers about 13×10^{23} calories to Earth each year. Living organisms take up less than 1 percent of this energy; Earth absorbs or reflects most of the rest. Absorbed energy is converted to heat, while energy is

The population size of a particular species that can be supported in any given ecosystem depends on the resource needs—ultimately, the energy needs—of that species.

Before the 1800s, the North American prairie was able to provide sufficient energy and other resources to support an incredible 30 million to 60 million bison. During the late 1800s all but a few hundred of these extraordinary animals were killed for their skins, meat, tongues, for sport, or to impact the Plains' Native American populations. An intensive breeding program was undertaken—this was among the first times that zoos took an active role to save a species from becoming extinct in the wild—and captive-bred bison were released back into the prairie. But most of the prairie has since been converted to farmland or extensively grazed by livestock. Therefore it would be impossible for bison herds to ever again reach the large numbers seen in the past, simply because there is not enough energy available in the system.

Similarly, widespread control of prairie dogs has been a major factor in the near-extinction of the black-footed ferret.

reflected as light (Figure 1). Solar energy helps create the different habitats where organisms live, is responsible for global weather patterns, and helps drive the biogeochemical cycles that recycle carbon, oxygen, other chemicals, and water. Clearly, solar energy profoundly influences all aspects of life.

Solar energy also is the source of the energy used by organisms; but to do biological work, this energy must first be converted. Most biological work is accomplished by chemical reactions, which are energy transformations that rely on making or breaking chemical bonds. Chemical energy is organized, high-quality energy that can do a great deal of work, and is the form of energy most useful to plants and animals. Solar energy can be used to create these chemical bonds because light energy that is absorbed by a molecule can boost that molecule's electrons to a higher energy level, making that molecule extremely reactive. The most important reactions involve the transferring of electrons, or oxidation-reduction (REDOX) reactions because, in general, removing electrons from a molecule (oxidation) corresponds to a release of energy. In other words, organisms can do work by oxidizing carbohydrates, converting the energy stored in the chemical bonds to other forms. This chemical energy is converted from solar energy during the process of photosynthesis, which occurs only in plants, some algae, and some bacteria.

How Plants Get Energy: Photosynthesis

A pigment is a material that absorbs light. Biologically important pigments absorb light in the violet, blue, and red wavelengths. Higher-energy wavelengths disrupt the structure and function of molecules, whereas longer (lower-energy) wavelengths do not contain enough energy to change electron energy levels. Photosynthesis is restricted to those organisms that contain the appropriate pigment combined with the appropriate structures such that the light energy can trigger useful chemical reactions. If an organism cannot perform photosynthesis, it cannot use solar energy. The Sun's energy enters the food chain through photosynthesis. Plants are the most important photosynthesizers in terrestrial systems, while photosynthesis in aquatic systems generally occurs in algae.

In plants, chlorophyll is the pigment that absorbs radiant energy from the Sun. This allows the transfer of electrons from water to carbon dioxide, creating the products glucose and oxygen. The equation for photosynthesis is:

$$6CO_2 + 6H_2O + \text{light} \rightarrow C_6H_{12}O_6 + 6O_2$$

Photosynthesis takes atmospheric carbon dioxide and incorporates it into organic molecules—the carbon dioxide is "fixed" into the carbohydrate. These molecules are then either converted into chemical energy or used as structural molecules. The first powers living systems; the second is what living systems are composed of.

To release energy, the electrons can be removed from glucose and used to create ATP, a molecule that supplies a cell's short-term energy needs. This latter occurs in a series of reactions known as respiration. (Body heat is a by-product of these reactions.) The most efficient respiration reactions are those that use oxygen to accept the electrons removed from glucose. Thus respiration is the reverse of photosynthesis: Organic compounds are oxidized in the presence of oxygen to produce water and carbon dioxide. Photosynthesis captures energy, and both products of photosynthesis are required in the energy-releasing reactions of respiration.

By converting radiant energy from the Sun into stored chemical energy, plants essentially make their own food: All that a plant needs to survive is water, carbon dioxide, sunlight, and nutrients. Plants need not rely on any other organism for their energy needs. In contrast, the only forms of life in terrestrial systems that do not depend on plants are a few kinds of bacteria and protists. Every other living thing must eat other organisms to capture the chemical energy produced by plants and other photosynthesizers.

How Animals Get Energy: Energy Flow Through an Ecosystem

A heterotroph is an organism that relies on outside sources of organic molecules for both energy and structural building blocks. This includes all animals and fungi and most single-celled organisms. An autotroph can synthesize organic molecules from inorganic elements. Most autotrophs are photosynthetic; a few, occurring in very restricted areas such as deep ocean trenches, are chemosynthetic.

Consequently, each organism depends in some way upon other organisms; organisms must interact with each other to survive. For example, a heterotroph must eat other organisms to obtain energy and structural compounds, while many important plant nutrients originate from animal wastes or the decay of dead animals. The study of these interactions is the subject of the science of ecology. Interactions that involve energy transfers between organisms create food chains. A food chain portrays the flow of energy from one organism to another.

Three categories can describe organisms in a community based on their position in the food chain: producers, consumers, and decomposers. These categories are also known as trophic levels. Plants “produce” energy through the process of photosynthesis. Consumers get this energy either directly or indirectly. Primary consumers eat plants directly, whereas secondary consumers get their energy from plants indirectly by eating the primary consumers. Energy enters the animal kingdom through the actions of herbivores, animals that eat plants. Animals that eat other animals are carnivores, and omnivores are animals that eat both plants and animals. Decomposers get their energy from consuming nonliving things; in the process releasing inorganic nutrients that are then available for reuse by plants.

A North American prairie food chain begins with grass as the producer. The grass is eaten by a prairie dog, a primary consumer. The prairie dog falls prey to a secondary consumer, such as a black-footed ferret. Some food chains contain a fourth level, the tertiary consumer. In this example, the ferret could be eaten by a golden eagle, a top predator. Because organisms tend to eat a variety of things, food chains are generally linked to form overlapping networks called a food web. In addition to the ferret, golden eagles eat ground squirrels and rabbits, each of which are primary consumers in separate food chains. A prairie food web would depict all of the interactions among each of these food chains.

Not all of the energy produced by photosynthesis in the grass is available to the prairie dog. The grass requires energy to grow and produce structural compounds. Similarly, the prairie dog eats food for the energy it needs for the work of everyday life: It needs to find and digest the grass; to detect and avoid predators; to find a mate and reproduce. Each time one of these processes uses energy, some energy is lost to the environment as heat. By the time the predator eats the prairie dog, very little of the original energy from the grass is passed on (Figure 1).

The fact that energy is lost at each stage of the food chain has tremendous influence on the numbers of producers and consumers that can be supported by any given habitat. This can best be illustrated graphically by ecological pyramids. Figure 2 is an example of an energy pyramid. Two other types of ecological pyramids are a pyramid of numbers and a pyramid of biomass. Each section of these pyramids represents a trophic level for the represented community—the producers form the base of the pyramid, the primary consumers the second level, and the secondary and tertiary consumers the third and fourth levels, respectively. Decomposers are also often included. The size of each section represents either the number of organisms, the amount of energy, or the amount of biomass.

Biomass refers to the total weight of organisms in the ecosystem. The small amount of solar radiation incorporated into living systems translates into the production of huge amounts of biomass: On a worldwide basis, 120 billion metric tons of organic matter are produced by photosynthesis each year. However,

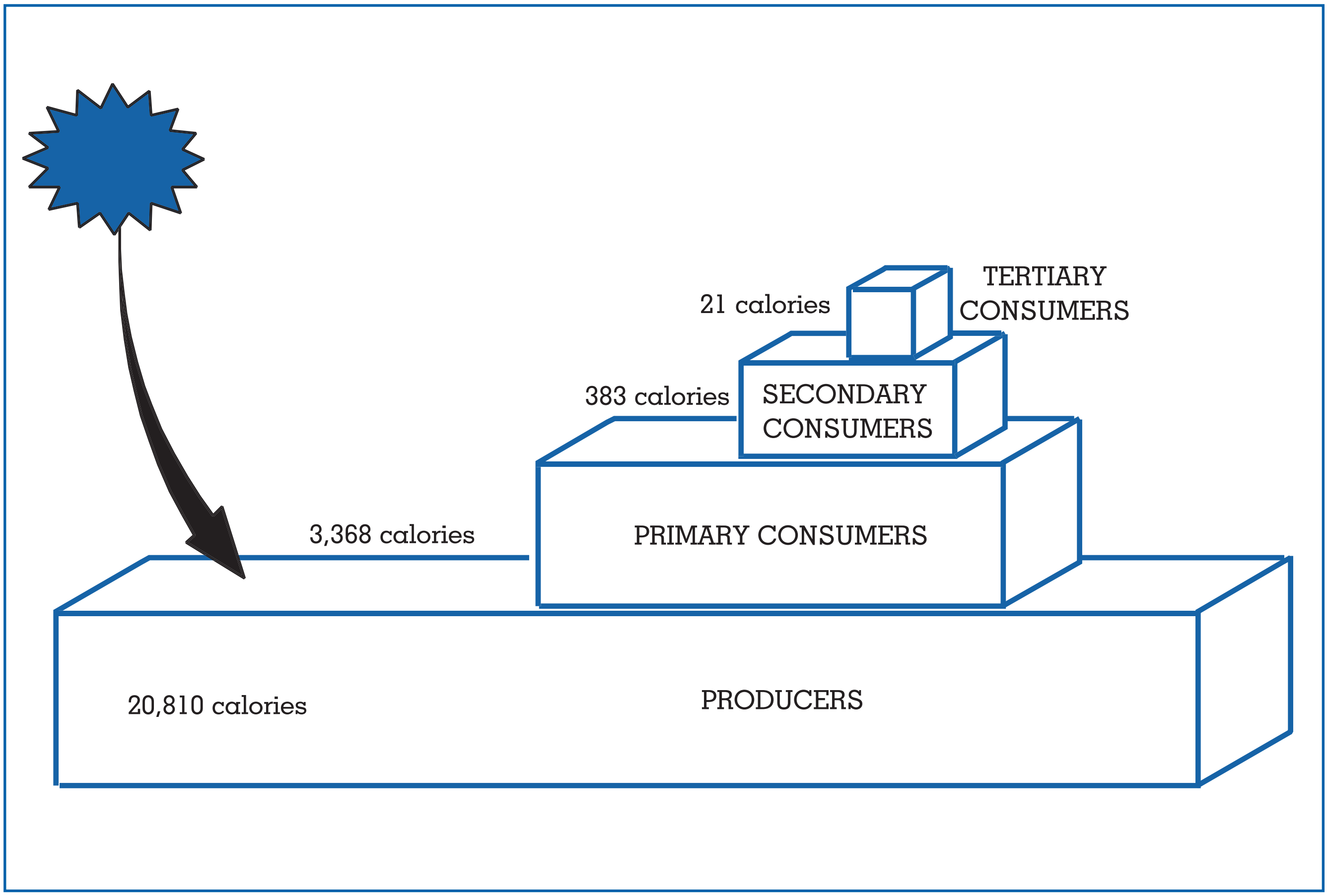


Figure 2.
Pyramids are useful tools for wildlife conservationists in determining the sustainable populations of particular species in any given ecosystem.

a plant uses only a fraction of the energy from photosynthesis to create this biomass; similarly, only a fraction of the energy consumed by an animal is converted to biomass. The rest is used up for metabolic processes and daily activities. The amount of biomass, the amount of energy, and the number of species present at each level of the pyramid are therefore less than in the trophic level before it. On the Serengeti Plain in Africa, there are many more blades of grass than there are zebras, and many more zebras than there are lions.

As shown in Figure 2, when a herbivore eats a plant, only about 10 percent of the energy stored in that plant is converted to animal biomass; the rest is used up in everyday activities. The same is true for each succeeding trophic level. Note that the commonly cited "10 percent energy transfer" figure is only a rough average based on many studies of different ecosystems. Scientists have found that actual transfer rates vary from 1 to 20 percent.

In a temperate forest ecosystem on Isle Royale, Michigan, ecologists found that it takes 762 pounds (346 kg) of plant food to support every 59 pounds (27 kg) of moose, and that 59 pounds of moose are required to support every one pound (0.45 kg) of wolf. The basic point is that massive amounts of energy do not flow from one trophic level to the next: energy is lost at each stage of the food chain, so there are more plants than herbivores and more herbivores than carnivores.

As links are added to the food chain, the amount of energy becomes more and more limited; this ultimately limits the total number of links in the food chain. Most habitats can support food chains with three to four trophic levels, with five being the usual limit. One study of 102 top predators demonstrated that there are usually only three links (four levels) in a food chain.

These factors lie behind the idea that adopting a vegetarian diet is a strategy in line with a sustainable lifestyle. Take the example of the plant → moose → wolf food chain given above. More energy is available to the wolf if it eats the plant rather than the moose. If a person is stuck on a desert island with a bag of grain and ten chickens, it makes little energetic sense to first feed the grain to the chickens, then eat the chickens. More energy is conserved (because fewer energy conversions are required) if the person first eats the chickens, then the grain. Similarly, if everybody were to adopt a vegetarian diet, much more food would be available for human consumption.

Thirty to 40 percent of the calories in a typical American diet come from animal products. If every person in the world consumed this much meat, all the agricultural systems in the world would be able to support only 2.5 billion people. In 1999 these same systems supported 6 billion people, primarily because the majority of the people living in less developed countries consume fewer animal products.

The energy flow through an ecosystem is the most important factor determining the numbers, the types, and the interactions of the plants and animals in that ecosystem.

Where an Animal's Energy Goes

During an animal's lifetime, about 50 percent of the energy it consumes will go to general maintenance (everyday metabolic processes), 25 percent to growth; and 25 percent to reproduction.

Animal nutritionists have developed formulas to guide them in recommending the amount of food to feed animals in captive situations such as in zoos. First, the number of calories needed to maintain the animal while at rest is determined—this is called the basal metabolic rate (BMR). In general, a reptile's BMR is only 15 percent that of a placental mammal, while a bird's is quite a bit higher than both a reptile's and a mammal's. For all animals, the number of calories they should receive on a maintenance diet is twice that used at the basal metabolic rate. A growing animal should receive three times the number of calories at the BMR, while an animal in the reproductive phase should receive four to six times the BMR.

Recall that during respiration, animals gain energy from glucose by oxidizing it—that is, by transferring the electrons from the glucose to oxygen. Because molecules of fat contain more hydrogen atoms (and therefore electrons) than either glucose or proteins, the oxidation of fat yields almost twice the calories as that of carbohydrates or proteins. Each gram of carbohydrate yields four calories, a gram of protein yields five calories, while a gram of fat yields nine calories.

THE EFFECTS OF ENERGY

Adaptations

Adaptations are features an organism has or actions it does that help it survive and reproduce in its habitat. Each organism has basic needs, including energy, water, and shelter; adaptations allow an organism to obtain its basic needs from its habitat. Body structures, behavioral characteristics, reproductive strategies, and physiological features are all examples of adaptations. Different habitats pose different challenges to organisms, and adaptations can be thought of as solutions to these challenges. For example, important adaptations in a desert habitat would include those that allow an animal to conserve water and store energy.

Energy affects adaptations in two fundamental ways. First, plants and animals need energy to survive, and adaptations may allow them to obtain, use, and, in many cases, conserve energy. Not only do plants and animals need a great deal of energy to fuel the chemical reactions necessary for survival, but also this energy can be in limited supply, particularly to consumers. Accordingly, many adaptations seen in different species revolve around the conservation of energy.

Second, solar radiation can result in striking differences among habitats. A rain forest is fundamentally different from a tundra first and foremost because there is a much greater amount of energy available to a rain forest. The average daylight hours in Barrow, Alaska, vary from about one hour in January to nearly twenty-four hours in June. The reduced growing season limits the amount of energy that plants can produce. Contrast this with Uaupés, Brazil, which receives twelve hours of sunlight each month of the year. The year-long growing season contributes to the fact that rain forests are the most productive habitats on Earth.

Regulating Body Temperature. Metabolism refers to all the chemical reactions that take place within an organism. To occur at rates that can sustain life, metabolic reactions have strict temperature requirements, which vary from species to species. There are two ways by which an organism can achieve the appropriate temperature for metabolism to occur. The first is to capture and utilize the heat generated by various energy conversions; the second is to rely on external energy sources such as direct sunlight. In other words, body temperature can be regulated either internally or externally.

An endothermic animal generates its own body temperature, while an ectothermic animal does not. In general, endothermic animals have constant body temperatures that are typically greater than that of the surrounding environment, while ectothermic animals have variable temperatures. Ectotherms rely on behavioral temperature regulation—a snake will move from sun to shade until it finds a suitable microclimate that is close to its optimal body temperature. When exposed to direct sunlight, an ectotherm can increase its body temperature as much as 1°C (32.8°F) per minute.

Endothermic animals can achieve and sustain levels of activity even when temperatures plummet or vary widely. This can be a huge advantage over ectothermy, especially in northern latitudes, at night, or during the winter. In colder climates, an ectothermic predator such as a snake will tend to be more sluggish and less successful than an endothermic predator. There are no reptiles or insects in the polar regions.

However, endothermy is a costly adaptation. An actively foraging mouse uses up to thirty times more energy than a foraging lizard of the same weight; at rest, an endothermic animal's metabolism is five to ten times greater than that of a comparably sized ectotherm. In certain habitats, this translates to a substantial advantage to ectothermy. Because of their greater energy economy and lower food requirements, tropical ectotherms outnumber endotherms in both number of species and number of individuals.

Birds and mammals are endothermic vertebrates. Not coincidentally, they are the only vertebrates with unique external body coverings—feathers and hair, respectively. For both groups, these body coverings evolved as an adaptation to reduce heat loss. A bird's feathers were originally adaptive because they helped keep the animal warm, not because they helped it to fly.

The Energetics of Body Size. Larger animals have lower energy requirements than smaller ones. Gram for gram, a harvest mouse has twenty times the energy requirements of an elephant. Part of the advantage of size probably stems from the fact that a larger animal has proportionately less surface area than a smaller one. When heat leaves the body, it does so through the body surface. It is more difficult for a smaller object to maintain a constant body temperature because it has a greater amount of surface area relative to its volume. One consequence of this relationship is that a whale, because of its lower metabolic rate, can hold its breath and thus remain underwater for longer periods of time than a water shrew.

The Costs of Locomotion. Because oxygen is required for energy-producing metabolic reactions (respiration), there is a direct correlation between the amount of oxygen consumed and the metabolic rate. Not surprisingly, metabolic rates increase with activity. During exercise, a person will consume fifteen to twenty times more oxygen than when at rest.

A comparison of different species reveals that a larger animal uses less energy than a smaller one traveling at the same velocity. This seems to be related to the amount of drag encountered while moving. Small animals have relatively large surface-area-to-volume ratios and therefore encounter relatively greater amounts of drag. Adaptations that reduce drag include streamlined body shapes. However, because a larger body must first overcome a greater amount of inertia, there is a greater cost of acceleration. Therefore, small animals tend to be able to start and stop abruptly, whereas larger animals have longer start-up and slow-down periods.

Interestingly, two different species of similar body size have similar energetic costs when performing the same type of locomotion. Differences in energy expenditure are seen when comparing the type of locomotion being performed rather than the species of animal. For a given body size moving at a given velocity, swimming is the most energetically efficient mode of locomotion, flying is of intermediate cost, and running is the most energetically expensive. An animal swimming at neutral buoyancy expends less effort than an animal trying to stay aloft while flying; running costs the most because of the way the muscles are used.

Nonetheless, birds have higher metabolic rates than mammals of similar size. Most small mammals reduce energy costs by seeking protected environments; birds spend much of their time exposed. Also, because fat is heavy, the need to fly restricts a bird's ability to store energy. Even with a high-protein diet, a bird must eat as much as 30 percent of its body weight each day. This factor alone probably accounts for some birds' summer migratory journey from the tropics to northern latitudes, which, because of their longer days, allow a bird more daylight hours in which to feed itself and its young.

The Energetics of Mating. Sexual reproduction requires a considerable amount of energy, including the energy invested while competing for mates; mating itself (including the production of gametes); and caring for the offspring. Three main mating systems are found in the animal kingdom: polygyny (one male, more than one female), monogamy (one male, one female), and polyandry (one female, more than one male. Each system can be defined in terms of the relative energetic investments of each sex. In a monogamous pair bond, both sexes invest approximately equal amounts of energy; consequently, courtship behaviors tend to be rather involved, and competition is equal between the sexes. Males in a polygynous system spend a great deal of energy competing for mates, while females invest more heavily in parental care, and therefore tend to be very particular with whom they mate. A polyandrous system is the opposite: Females invest in competition and mating, while males invest in parental care.

Adaptations to Habitats. Because of Earth's geometry and the position of its axis, the equator receives more solar energy per unit area than the polar regions. Because Earth's axis is tilted relative to the plane of Earth's orbit around the Sun, this angle of incident radiation varies seasonally. These factors, combined with Earth's rotation, establish the major patterns of temperature, air circulation, and precipitation.

A habitat, or biome, is made up of interacting living and nonliving elements. The major terrestrial habitats include deserts, temperate forests, grasslands, rain forests, tundra, and various types of wetlands. The boundaries of these different habitats are determined mainly by climatological factors such as temperature, precipitation, and the length of the growing season. These conditions are created by the influence of solar radiation, often in conjunction with local factors such as topography.

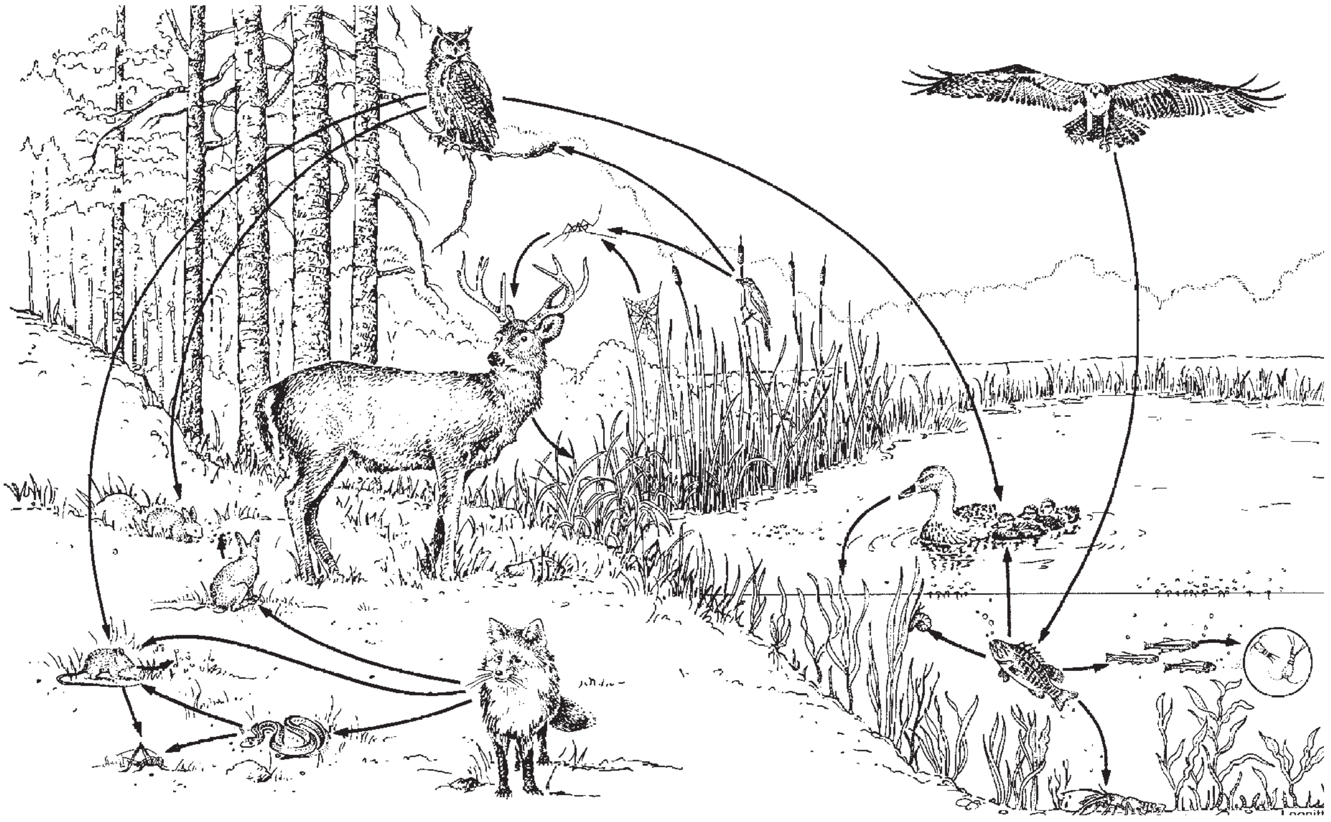
The amount of biomass produced in a habitat—the productivity of the habitat—is determined by the types of plants (some species are more efficient photosynthesizers than others), the intensity and duration of solar radiation, the amount of nutrients available, and climatic factors such as temperature and precipitation. Aquatic habitats tend to be less productive than terrestrial ones, largely because there is less sunlight available at depth and there is a scarcity of mineral nutrients. Tropical rain forests have conditions that favor high productivity; one result is that they also have the highest biodiversity of any habitat.

Cold Habitats. Because of considerations of surface area relative to body mass, animals that live in cold habitats tend to have larger body sizes and smaller extremities (especially ears and legs) compared to their counterparts in warmer habitats. Animals that live in cold habitats also have a greater amount of insulation, such as fat, fur, or feathers. Behavioral adaptations include gathering in groups, which effectively decreases the exposed surface area of each individual.

If energy resources are seasonally low, some animals adopt the strategy of migrating to areas with greater resources. A bird that is insectivorous is more likely to be migratory than one that is a seed-eater. Hibernation is another adaptation in response to seasonal energy shortages. The body temperature of a true hibernator closely matches that of its surroundings. The heart slows (a ground squirrel's heartbeat drops from 200 to 400 beats per minute to 7 to10 beats per minute), and metabolism is reduced to 20 to 100 times below normal. Hibernators tend to have much longer lifespans than non-hibernators.

Differences also can be seen in human populations living in cold habitats. Among the Inuit, the body maintains a high temperature by burning large amounts of fat and protein. Increased blood flow to the arms, legs, fingers, and toes helps prevent frostbite. An Australian Aborigine may sleep in below-freezing temperatures with little clothing or shelter, yet conserves energy by allowing the temperature in the legs and feet to drop. Heat is maintained in the trunk, where it is needed most.

Adaptations to Warm Habitats. When water evaporates into the surroundings, the vaporized molecules carry a great deal of heat away with them. One of the best ways to cool an animal's body is to evaporate water from its surface. Adaptations that take advantage of this property include sweating, panting, and licking the body. But water often is a limited resource in warm habitats such as deserts, so many desert animals have adaptations that reduce the amount of water that evaporates from the body. Most



A depiction of how food chains become interconnected to form food webs. (Gale Group, Inc.)

small desert animals avoid the heat and reduce water loss by being nocturnal and living in burrows. Large extremities, particularly ears, help to bring heat away from the body and dissipate it to the surroundings.

Other adaptations are perhaps best exemplified by examining the camel, which is able to conserve water by excreting a very concentrated urine. Also, the upper lip is split, so moisture dripping from the nose reenters the body through the mouth. More importantly, the camel can tolerate dehydration: It can lose an impressive 25 percent of its body weight in water with no ill effects. Its internal body temperature can fluctuate by as much as 6°C (10.8°F). By increasing its temperature during the day and dropping it at night, it can more closely track changes in external temperatures. This also helps to reduce water loss by as much as five liters of water per day. The fat stored in the camel's hump represents an important energy supply in the sparsely vegetated habitat in which it lives.

ENERGY AND HUMANS

The Costs of Technology

Paradoxically, organisms must use energy to get energy: A lion must hunt to eat, while a zebra must sometimes move long distances to find food. Most organisms can gain 2 to 20 calories in food energy for each calorie they use to obtain that energy. This holds for both a hummingbird, whose metabolic rate is 330 calories per minute, and a damselfly, which uses less than 1 calorie per day.

This is also true for human hunter-gatherer societies. Without technology, people use about 1 calorie to gain 5 to 10 calories. The energy return increases to 20 calories through the use of shifting agricultural practices.

Ironically, the cost of getting our food has increased with technological advances. In 1900 we gained a calorie for each calorie we used, while in 1995, for each calorie we invested we got only 0.1

calorie in return. Some of the energy costs associated with food include human labor, the cost of fertilizer, the cost of fuel for the farm machinery, and the cost of transportation of the food.

More developed countries rely heavily on burning fossil fuels to meet energy needs. Fossil fuels are the remains of plants that, over millions of years, have been transformed into coal, petroleum, and natural gas. Just like the natural systems examined in this article, our energy ultimately comes from the Sun and photosynthesis. Although more developed countries have less than 20 percent of the world's population, they use more than 80 percent of the world's energy.

People in less developed countries burn wood, plant residues, or animal dung to fuel stoves and lanterns. About 2 billion people rely on wood to cook their daily meals. Typically, four to five hours per day are spent gathering wood fuels from the surrounding habitat. Because these countries tend to have high population growth rates, there has been an ever-increasing demand for more wood, resulting in a significant amount of habitat degradation in many areas.

The Effects of Human Energy Use

The two major ways by which humans get energy is to either burn fossil fuels or to burn wood for fuel. Both contribute substantially to air pollution, and both can have serious effects on (1) the health of plants and animals and (2) the workings of Earth's atmosphere.

Burning fossil fuels can release air pollutants such as carbon dioxide, sulfur oxides, nitrogen oxides, ozone, and particulate matter. Sulfur and nitrogen oxides contribute to acid rain; ozone is a component of urban smog, and particulate matter affects respiratory health. In fact, several studies have documented a disturbing correlation between suspended particulate levels and human mortality. It is estimated that air pollution may help cause 500,000 premature deaths and millions of new respiratory illnesses each year.

Physiological effects of air pollution are dependent on dosage, the ability of the exposed organism to metabolize and excrete the pollution, and the type of pollutant. Many pollutants affect the functioning of the respiratory tract; some change the structure and function of molecules; others can enter the nucleus and turn genes on or off; and some cause chromosomal aberrations or mutations that result in cancer. For example, exposure to the air toxin benzene can increase the risk of getting myelogenous leukemia or aplastic anemia, while exposure to ground-level ozone can cause a 15 to 20 percent decrease in lung capacity in some healthy adults.

Air pollution affects plant health as well. Acid rain and ozone can directly damage a plant's leaves and bark, interfering with photosynthesis and plant growth. More serious effects can occur if soil nutrients are leached away and heavy metals are mobilized in the soils upon which plants depend. Without proper nutrients, plants become susceptible to a variety of diseases. The overall result is a decrease in the amount of energy produced by plants. Acid rain affects over 345,960 square miles (900,000 sq. km) of Eastern Europe, where it has taken a significant toll on cities, forest, lakes, and streams (Kaufman and Franz, 1993). Moreover, air pollution is reducing U.S. food production by 5 to 10 percent, costing an estimated $2 billion to 5 billion per year (Smith, 1992).

Burning fossil fuel releases carbon into the atmosphere—more than 6.3 billion tons in 1998 alone. Significant amounts of carbon also come from burning of live wood and deadwood. Such fires are often deliberately set to clear land for crops and pastures. In 1988 the smoke from fires set in the Amazon Basin covered 1,044,000 square miles. By far the most serious implication of this is the significant threat to Earth's ecosystems by global climate change.

Like all matter, carbon can neither be created nor destroyed; it can just be moved from one place to another. The carbon cycle depicts the various places where carbon can be found. Carbon occurs in the atmosphere, in the ocean, in plants and animals, and in fossil fuels. Carbon can be moved from the atmosphere into either producers (through the process of photosynthesis) or the ocean (through the process of diffusion). Some producers will become fossil fuels, and some will be eaten by either consumers or decomposers. The carbon is returned to the atmosphere when consumers respire, when fossil fuels are burned, and when plants are burned in a fire. The amount of carbon in the atmosphere can be changed by increasing or decreasing rates of photosynthesis, use of fossil fuels, and number of fires.

Scientists have been able to compare the seasonal changes in atmospheric carbon dioxide to the seasonal changes in photosynthesis in the Northern Hemisphere. Plants take up more carbon dioxide in the summer, and animals continue to respire carbon dioxide in the winter, when many plants are dormant. Correspondingly, atmospheric carbon dioxide increases in the winter and decreases in the summer.

Atmospheric carbon dioxide, water vapor, methane, and ozone are all "greenhouse gases." When solar energy is reflected from Earth's surface, the longer wavelengths are trapped in the troposphere by these greenhouse gases. This trapped radiation warms Earth. In fact, without greenhouse gases, which have been present for several billion years, Earth would be too cold to support life.

Habitat destruction is another important contributing factor to increased atmospheric carbon dioxide levels. The world's forests are being cut, burned, or degraded at an astounding rate: More than half of the world's tropical rain forests have been lost in the past one hundred years. The forests supply fuel wood for energy; land for crops or pastures; and the wood demands of the global economy. Burning the forests releases carbon dioxide into the atmosphere; cutting or degrading the forests results in fewer plants available to take carbon dioxide out of the atmosphere.

Although scientists have been able to measure increasing levels of carbon dioxide, it is difficult to predict what the effects will be. For example, some models predict that warmer temperatures and the greater availability of atmospheric carbon dioxide will stimulate productivity, which in turn will remove carbon dioxide from the atmosphere, thus neutralizing the problem. However, the availability of carbon dioxide is not what generally limits plant growth. Rather, plant productivity tends to be restricted by the availability of resources such as nitrogen, water, or sunlight. Therefore, increasing the amount of carbon dioxide available to plants probably will have little effect on productivity, especially because it likely will result in greater evaporation rates and changed weather patterns. Although increased levels of evaporation may actually increase rain in some parts of the world, it may not be in those places currently containing rain forests.

A 1999 study by the Institute of Terrestrial Ecology predicts that tropical rain forests will be able to continue to absorb carbon dioxide at the current rate of 2 billion tons per year until global temperatures rise by 8°F (4.5°C). At this point, evaporation rates will be high enough to decrease rainfall for the forests, leading to the collapse of tropical ecosystems. This collapse will decrease the amount of carbon dioxide leaving the atmosphere and have dire consequences for all life.

Allison Brody

See also: Acid Rain; Agriculture; Air Pollution; Atmosphere; Biological Energy Use, Cellular Processes of; Climatic Effects; Environmental Problems and Energy Use; Green Energy; Thermodynamics.

BIBLIOGRAPHY

Brown, L. R. (1999). *Vital Signs*. New York: W. W. Norton.

Curtis, H., and Barnes, N. S. (1989). *Biology*, 5th ed. New York: Worth.

Eckert, R., and Randall, D. (1983). *Animal Physiology*, 2nd ed. San Francisco: W. H. Freeman.

Kaufman, D., and Franz, C. (1993). *Biosphere 2000: Protecting Our Global Environment*. New York: Harper Collins.

Miller, G. T., Jr. (1996). *Living in the Environment*, 9th ed. Belmont, CA: Wadsworth.

Raven, P. H.; Berg, L. R.; and Johnson, G. B. (1995). *Environment*. Fort Worth, TX: Saunders.

Smith, Z. (1992). *The Environmental Policy Paradox*. Englewood Cliffs, NJ: Prentice-Hall.

Wildlife Conservation Society. (1995). *Voyage from the Sun*. New York: Author.

BLACK, JOSEPH (1728–1799)

Joseph Black was born in Bordeaux, France, the fourth child of parents of Scottish extraction. His father was a native of Belfast engaged in the Bordeaux wine trade; his mother was a daughter of an Aberdeen man who had settled in Bordeaux. In all, Black's parents had twelve children. At the age of twelve Black was sent to school in Belfast, and around 1744 proceeded to the University of Glasgow. Black followed the standard curriculum until pressed by his father to choose a profession. He opted for medicine. Black began to study anatomy and chemistry. William Cullen had recently inaugurated lectures in chemistry that were to have a decisive influence on Black's career. Recognizing Black's aptitude, Cullen employed Black as his laboratory assistant.

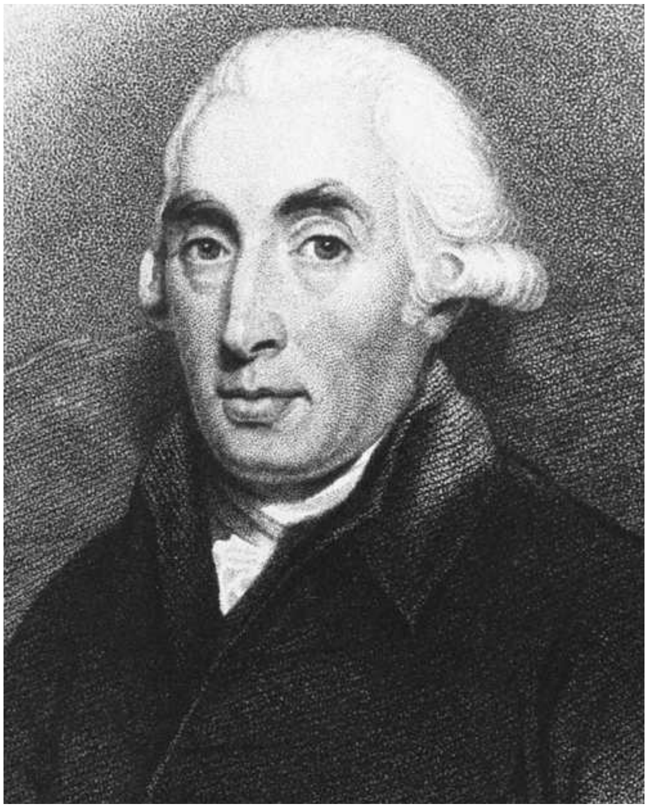
In 1752 Black transferred to Edinburgh University to continue his medical training, receiving his M.D. in 1754. In 1756 Cullen arrived at Edinburgh as Professor of Chemistry, and Black at the age of twenty-eight was appointed to the Chair of Anatomy and Chemistry at Glasgow. He became dissatisfied with his qualifications for that position and exchanged duties with the Professor of Medicine. During the ensuing ten years, Black lectured on the subject of medicine and also carried out an active and demanding medical practise. In 1766 Cullen vacated the Chair of Chemistry to become Professor of the Institutes of Medicine, and Black took over Cullen's Chair. Black remained at Edinburgh, but now limited his medical practice to a few close friends.

Joseph Black suffered all his life with ill health and had something of a reputation for parsimony, being as methodical in his financial affairs as in his science. Black never married but was no recluse. A prominent member of Edinburgh intellectual society, he regularly frequented the Royal Society of Edinburgh and the dining clubs for which Edinburgh was then famous. Due to worsening health Black gave his last lectures in 1795-1796, and died in December 1799 in his seventy-second year.

Black was the founder of the measurement of quantities of heat or calorimetry. During his period of tenure at Glasgow, Black carried out experimental research into the nature of heat and discovered what he termed "latent heat." In 1754, Cullen communicated to Black his observations concerning the intense cold produced by the evaporation of volatile substances such as ether. Black was also aware of Fahrenheit's observations concerning the freezing of water. Black reflected that solidification or evaporation required the transfer of quantities of heat not detectable by the thermometer—heat that Black termed "latent." Black concluded that the heat in water at 32°F is equal to the heat in ice at 32°F plus the latent heat of liquefaction. Not satisfied with demonstrating that there was such a thing as latent heat, Black proposed to measure it. He had to first prove the reliability of the thermometer as a measuring tool. After carrying out a number of ingenious but simple experiments involving the mixing of hot and cold water, Black concluded that the expansion of mercury was a reliable indicator of the temperatures of heat.

Black recognized a distinction between quantity of heat and temperature. Although not the first to make this distinction, he was the first to sense its fundamental importance and make use of it: a thermometer can be used to measure temperature but how can the quantity of heat be measured? In conducting his experiments Black was probably guided by Newton's experiments and law of cooling. In the winter of 1761 Black carried out experiments to determine the latent heat of fusion. A mass of water was cooled to 33°F and the time taken to raise the temperature 1°F was noted. As a comparison, Black also determined the time required to melt an identical mass of ice. Conversely, Black determined the time required to lower the temperature of a mass of water and compared it with the time it took to freeze it. From this Black obtained a value for the latent heat.

Black soon realized that latent heat must also play a part in vaporizing water. In 1762 Black carried out a series of investigative experiments. The time to heat water from 50°F to the boiling point was compared with the time it took the water to boil away. From these experiments Black calculated that the amount of heat required to evaporate water was equal to that required to raise the water to 810°F, were this to be possible. Black went on to make a second but closely



Joseph Black. (Library of Congress)

related discovery: Different substances have different heat capacities. In doing so Black was building up the work of Fahrenheit and George Martine, who had made observations concerning the different rates of temperature rise when heat was applied to equal quantities of water and mercury. Black concluded, since mercury increased in temperature almost twice as fast as water, it must have a smaller store of heat than water. He realized the significance of this effect, but did not initially pursue the matter.

James Watt, then a mathematical instrument maker, was called upon by Black to make items he needed for his experiments. Watt, in turn, was engaged in his own experiments concerning the Newcomen steam engine. Large quantities of water were required to condense the steam in the engine cylinder, and he turned to Black for an explanation. Black in turn explained to Watt his ideas about latent heat. Watt repeated Black's experiments with a smaller improved still. James Watt went on to develop his separate condenser, and later insisted his discovery had not been suggested by the doctrine of latent heat. Watt did, however, readily credit Black with clarifying the problems he had encountered and with teaching him to "reason and experiment with natural philosophy."

The problems raised by Watt revised Black's interest in heat. With his assistant, William Irvine, he set out to determine a more accurate value for the latent heat of steam using a common laboratory still as a water calorimeter. In 1764 it occurred to Black that his knowledge of the latent heat of fusion of ice could be used to measure the latent heat of steam. Plans to put this idea to the test were abandoned, however, when James Watt began to obtain values that Black thought sufficiently precise. Watt was the first to investigate specific heat experimentally and was responsible for drawing Black's attention to the significance and practical importance of such research. Black and his assistant investigated the specific heats of various solids by determining the heats communicated to water, using the method of mixtures. Black ceased his investigations into heat when he transferred to Edinburgh University in 1766. For reasons, that remain unexplained, Black was reluctant to publish his ideas and details of his experiments, his lecture notes were only printed after his death, by John Robinson.

Robert Sier

See also: Water Heating; Watt, James.

BIBLIOGRAPHY

Guerlac, H. (1970). "Joseph Black." In *Dictionary of Scientific Biography*, ed. C. C. Gillispie. New York: Scribners.

Ramsey, W. (1918). *Life and Letters of Joseph Black, M.D.* London: Constable.

Robinson, J., ed. (1803). *Lectures on the Elements of Chemistry, delivered in the University of Edinburgh*, 2 vols. Edinburgh: W. Creech.

BOILER

See: Heat and Heating; Steam Engines

BOMBS

See: Explosives and Propellants

BRITISH THERMAL UNITS

See: Units of Energy

BROOKHAVEN NATIONAL LABORATORY

See: National Energy Laboratories

BUILDING DESIGN, COMMERCIAL

More than 3.3 billion square feet of new commercial buildings were constructed from 1988 to 1998, with 170 percent more expected by the year 2030. Because this new stock is expected to have a lifetime of fifty to one hundred years, this has a dramatic impact on energy consumption not only today, but also for many years to come. This article will discuss the history of commercial building design, how technology has impacted commercial building design, and what impact this has had on building energy use.

HISTORY

The first commercial buildings, built in about 2000 B.C.E., were simple structures that represented the beginnings of architecture—a series of columns, walls, and roofs. Columns represented the upright human stance, walls represented human territoriality, and roofs both kept the rain out and created a crown, or head, for the structure. Walls also represented a separation between the plant and animal world and the human world. The walls of a courtyard formed a human space that became the city. Although the form of buildings has evolved over time, buildings today fundamentally provide these same basic human functions: artistic expression, separation, definition, and shelter.

Modern buildings are fundamentally defined by the mechanical principles that drive their utility. Technology has defined both the form of the buildings and energy use. Electrical lighting, mechanical ventilation, curtain-wall systems, air conditioning, and office equipment all contribute to a modern building's energy consumption. Two of these technologies that have had the most influence on the energy consumed by commercial buildings have been the lightbulb and the air conditioner.

Energy use in buildings has risen dramatically since 1900 because of technologies that enabled the creation of man-made indoor environments. It began with the invention and proliferation of the electric lightbulb in the 1880s. Fluorescent lighting became popular in the late 1930s, and by 1950 had largely replaced incandescent lamps in commercial buildings. However, incandescent lamps still are used in approximately 17 percent of the pre-1970 building stock.

Tasks such as health care, office work, manufacturing, studying, and other tasks requiring visual acuity all benefited from electric lighting. It also meant that workers didn't need access to a window to be productive. Designers could use artificial illumination to light tasks away from windows. Interior spaces that didn't require skylights or cleristories were now possible in commercial buildings.

The next major advance was elevators in the early 1900s followed by air conditioning in 1920s. Air conditioning changed where we lived and the way we lived, worked, and spent our leisure time. Prior to air conditioning, commercial buildings required natural ventilation. This defined the shape of the building, as each office or room required an operable window T-, H-, and L-shaped floor plans, which allowed the maximum number of windows to provide natural light and ventilation, are still visible in New York, Chicago, Boston, and Denver. During the pre-air-conditioning era, large cities developed in northern latitudes because workers could remain productive performing office work for most of the year.

The New York Stock Exchange was the first "laboratory" for air conditioning, in 1901. The American public first experienced air conditioning in the Olympia Theater in Miami, Florida, in 1920. Department stores and movie theaters realized the potential for air conditioning to increase their business. And the first high-rise office building to be completely air-conditioned was the Milam Building, built in 1928 in San Antonio, Texas this building demonstrated that office workers were also more productive in a temperature-controlled environment. For manufacturers, air conditioning benefited the manufacturing of textiles, printing, and tobacco by allowing precise temperature, humidity, and filtration controls.

This new technology also allowed greater architectural freedom. Post-World War II construction incorporated sealed aluminum and glass facades, or curtain walls. The United Nations Building was the first major post-World War II project to be designed with a curtain wall system and air conditioning. Larger floor cross sections were possible, and interior offices could be created with man-made environments.

The structure and energy use of post-World War II commercial buildings was re defined as a direct result of air conditioning.

Modern buildings are fundamentally defined by the mechanical principles that drive their utility. Technology has defined both the form of the buildings and energy use. (Archive Photos, Inc.)

Electrical energy use in buildings rose from the late 1800s through the 1970s while natural gas and fuel oil use declined. Pre-1970 buildings represent 30 billion square feet of floor area, and have an average size of roughly 12,000 square feet. Approximately 70 percent of these buildings are air conditioned and are illuminated with fluorescent lighting. Overall energy intensity declined from about 1910 to 1940, but increased after 1950 as air conditioning became more common. More than 70 percent of the energy used by pre-1970 buildings came from electricity or natural gas.

THE ENERGY CRISIS

In the late 1960s a host of new technologies allowed architects to expand the horizons of their designs. The boundless promise of low-cost energy and a real-estate boom created a certain freedom of expression. This expression also created some of the most energy-intensive buildings ever built, as technology could overcome the shortcomings of a building design that failed to properly shelter occupants from heat, cold, sun, and glare.

The Arab oil embargo and ensuing “energy crisis” in late 1973 began to change the way commercial buildings were designed and operated. The first reaction to the energy crisis was to conserve energy. Conservation of energy means using less energy regardless of the impact that has on the levels of amenities the building provides. This impacted three areas of building comfort; thermal comfort, visual comfort, and ventilation air

The combination of lower thermostat settings, reduced ventilation, and lower lighting levels created environmental quality problems in many buildings. Even today, some building owners associate energy conservation with poor indoor environmental conditions.

Starting in the early 1980s, Energy efficiency gradually replaced energy conservation as the mainstream approach to saving energy. Energy efficiency relies on three key principles:

1. Life-cycle cost analysis. Energy-efficient buildings are typically designed to be cheaper, on a life-cycle basis, than wasteful buildings.
2. Comprehensive design. Often, energy-efficient buildings are cheaper on a first-cost basis as well as on a life-cycle cost basis. This frequently results from approaching the design of all of the energy-consuming systems of a building comprehensively and finding synergies among the various energy efficiency measures. For example, introducing daylight and energy-efficient artificial lighting can reduce the internal heat loads in a building, which reduces the size of the heating, cooling and ventilating ducts. This in turn reduces the floor-to-floor height and the cost of the elevators, building skin, etc., and enables the developer to add more rentable floor space in a given building volume (which is often constrained by zoning restrictions).
3. Environmental quality. In contrast to being dark and cold in the winter and hot in the summer, well-designed, energy-efficient buildings usually are more enjoyable to inhabit. This is particularly true of well-lit buildings, with low glare, balanced luminance, and visual cues from the lighting as to how to most comfortably inhabit

the space. Well-designed efficient buildings also allow easy and balanced temperature control, avoiding problems such as some zones being too hot while others are too cold.

The energy crisis also created a rush to develop new technologies and practices to reduce the energy consumption of buildings. One popular technology was solar energy. Commercial building solar technologies focused on harnessing the Sun's energy to cool the building through absorption cooling, provide hot water, or illuminate building interiors.

The Frenchman's Reef Hotel in St. Thomas, U.S. Virgin Islands, was the first large-scale commercial demonstration of solar absorption cooling. Other popular solar applications in the 1970s and early 1980s included swimming pool heating, commercial laundry hot water heating, and the use of natural daylight to offset electrical lighting in buildings.

It took several years before real improvements in commercial building energy efficiency occurred. This was because first-cost decisions drive a large segment of the new construction market. If a construction project begins to run over budget, energy-efficient features may be the first items cut, since they are less visible to the building tenant. In the 1980s, marketplace dynamics helped to change the market for energy efficiency because of the following factors:

- Utilities offered financial incentives to customers to reduce energy consumption because it was less expensive than constructing new power plants to meet the growing demand for electricity.
- Manufacturers incorporated energy-efficient design features into their product lines, making new products and services available to commercial building owners.
- Widespread adoption of building energy codes requiring minimum levels of energy efficiency to be included in new buildings.

The commercial building market began to respond to these dynamics by significantly reducing energy use.

CURRENT COMMERCIAL BUILDING DESIGN

A recent study conducted by California utilities identified the efficiency of buildings constructed in the 1990s with respect to what was required by the state building code. The study showed that most newer buildings use 10 percent to 30 percent less energy than buildings barely meeting the code. This finding is noteworthy in part because California has one of the most stringnet building codes in the country. The technologies and practices that created this result will be discussed in more detail below.

The building envelope in a commercial building plays a very different role than in a residential building. Figure 1 shows the complex interactions of heat flow in a modern commercial building. Energy (Q) enters the building from direct solar gain and heat from people, equipment, and lights. Energy leaves the building through the walls, roof, and windows and by heating ventilation air. An efficient building minimizes energy entering the building and balances that with the energy leaving the building.

The Commercial Building Envelope

The building envelope is one of the keys to both building energy use and thermal comfort. A high-performance building envelope will require a smaller mechanical system, provide natural lighting, and shelter occupants from heat and glare. Building envelopes in a modern building have four key elements that impact energy use. These elements are

- thermal performance
- building orientation
- permeability (air and moisture)
- daylighting.

Envelope construction characteristics for a modern building are shown in Table 1. The predominant stock of buildings is small masonry or metal buildings. The building envelope is an integral part of the way a building is illuminated. Windows and skylights can deliver a significant portion of the lighting needed for building occupants to be productive. Proper daylighting requires that glare and direct sun be minimized while maximizing the use of diffuse light. Architecture and glass property selection accomplish this. Architectural features include aperture areas, overhangs, fins, and light shelves or horizontal surfaces that redirect light entering windows deep into a space. Glass properties include a low solar heat gain coefficient and a high visible light transmittance. When properly designed, daylighting can deliver improved learning rates in schools and higher sales in retail stores.

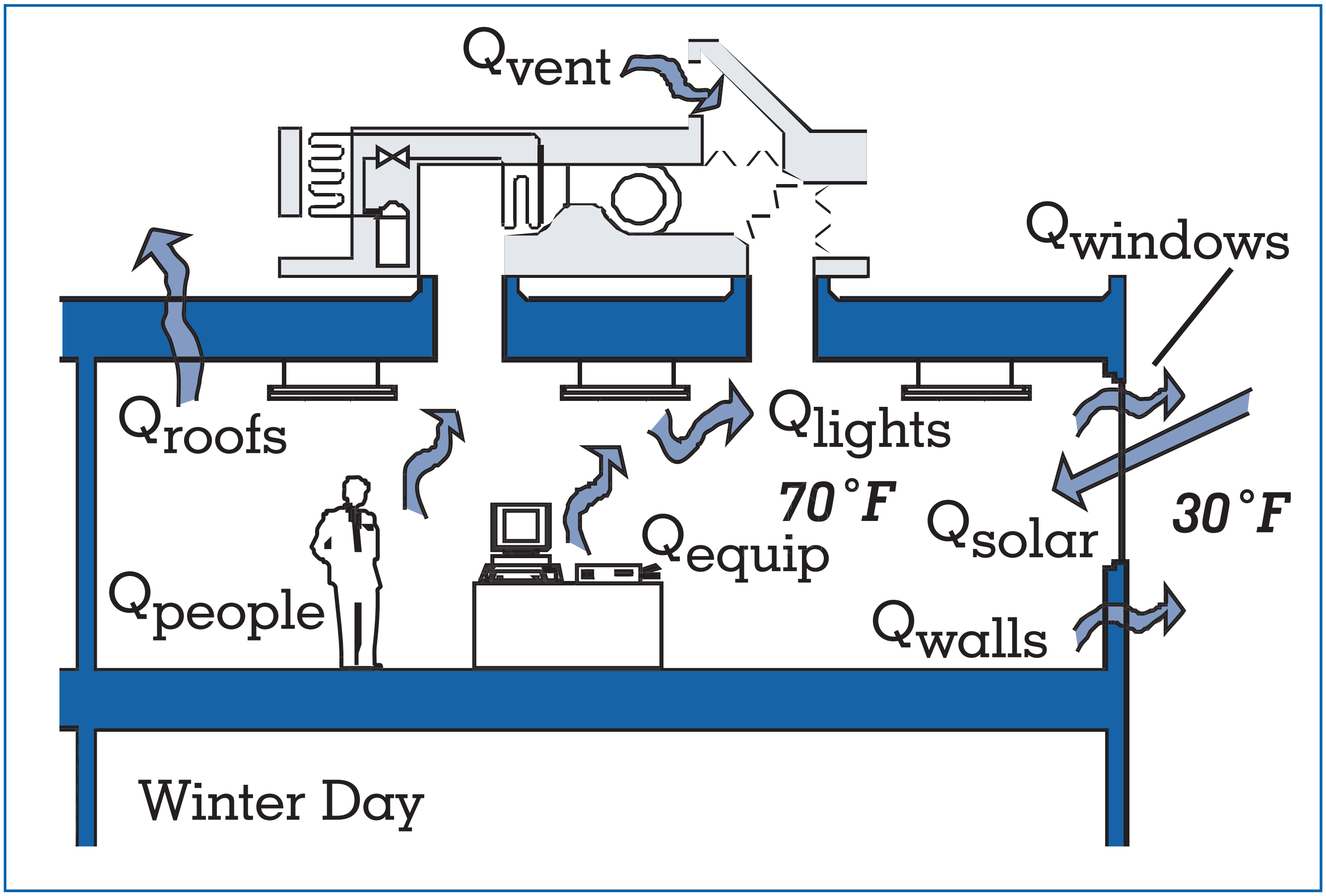


Figure 1.
Heat flow in a modern building.

Mechanical Systems

Building comfort is one of the key elements of a successful commercial building. The thermal comfort of a building is often as compelling as the aesthetics of the design. Modern building mechanical systems have two primary functions: maintain spaces within a predefined comfort range and deliver outdoor air to each space to assure proper ventilation. They have to do this in a quiet and efficient manner.

Heating, ventilating, and air-conditioning (HVAC) systems in modern buildings fall into two general categories: single-zone systems and multiple-zone systems. Both systems use air as the primary transfer mechanism to deliver heating and cooling to a space.

Single-zone systems deliver conditioned air to a single thermal zone. These systems are popular in small buildings (fewer than 10,000 square feet) and in single-story larger buildings. Usually they are vapor compression systems that cool air before it is delivered to a space. Single-zone systems serve 57 percent of post-1980 buildings.

Multiple-zone systems deliver conditioned air to more than one thermal zone. These systems typically have a direct expansion compressor or cold-water chiller to deliver cooling to the system. A central boiler or warm-air furnace at the system level, or electric heating elements at each zone, provide heating capabilities. This collection of components is controlled to maximize comfort and minimize energy use by computer-based controls. Roughly half of these buildings have an energy management control system that is operated by a trained energy manager.

A multiple-zone system must be able to deliver heating to a perimeter thermal zone in the winter while cooling an interior zone to offset internal

Building Characteristic	*Predominant Construction*	*Percent of Post-1980 Buildings*
Size	10,000 sq. ft. or less	74%
Exterior Walls	Concrete Masonry	54%
	Metal Siding	27%
Roofing	Metal Roof	40%
	Non-wood Shingles	30%
Insulation	Roof	79%
	Walls	71%
Windows	More than One Pane	51%
	Tinted Glass	36%

Table 1.
Post-1980 Envelope Construction Characteristics

gains. To accomplish this, modern multiple-zone systems use a technique called variable air volume (VAV).

A VAV system controls the temperature in a thermal zone by the volume of air delivered to that zone. If a zone thermostat demands heat, the system would first reduce the volume of air to the zone to the minimum required to meet outdoor air ventilation requirements and then begin to heat the air. This helps reduce energy use in two ways: first, the volume of air moving through the system is reduced to a minimum, lowering fan energy use, and second, reheating previously cooled air is minimized. Variable-speed drives are the most efficient method of controlling VAV fans and are used in 8 percent of post-1980 buildings.

The advantages of VAV are that temperature can be controlled while minimizing energy consumption. Also, the system can be smaller because the maximum demand for cooling never occurs simultaneously in all spaces. The disadvantages are the additional space required for the air-handling plants and ductwork.

Another energy-saving feature of modern HVAC systems is the ability to cool the building using outdoor air. This is accomplished through a control and damper arrangement called an outdoor air economizer. An economizer varies the outdoor air supply from the minimum setting to up to 100 percent outdoor air, provided the outside air temperature is less than the supply air temperature required to maintain comfort conditions. Outdoor air economizers are present in 85 percent of post-1980 buildings.

Lighting Systems

Lighting has a significant impact on building occupants, for better or for worse. Lighting also is a significant energy user and is rich with potential energy savings. For some time there has been a good deal of attention and effort invested in mining the energy savings from lighting systems. Preliminary studies into the ancillary benefits of energy-efficient lighting show that quality lighting can have positive effects such as improved productivity, reduced health complaints, and reduced absenteeism.

Figure 2 shows how buildings built after 1980 use energy. Lighting is the most significant energy expenditure in a modern building. The key qualities of an effective lighting system are

- energy efficiency
- room surface brightness
- reduction of glare
- adequate task illumination
- uniform light distribution
- good color lamps
- visual interest
- lighting controls.

Key technologies that are used in modern lighting include electronic ballasts, more efficient tubular fluorescent lamps, compact fluorescent lamps, and lighting controls. Fluorescent lighting is the predominant lighting system installed in post-1980 buildings and is used in 71 percent of floor space. Specialty retail stores use a combination of fluorescent and

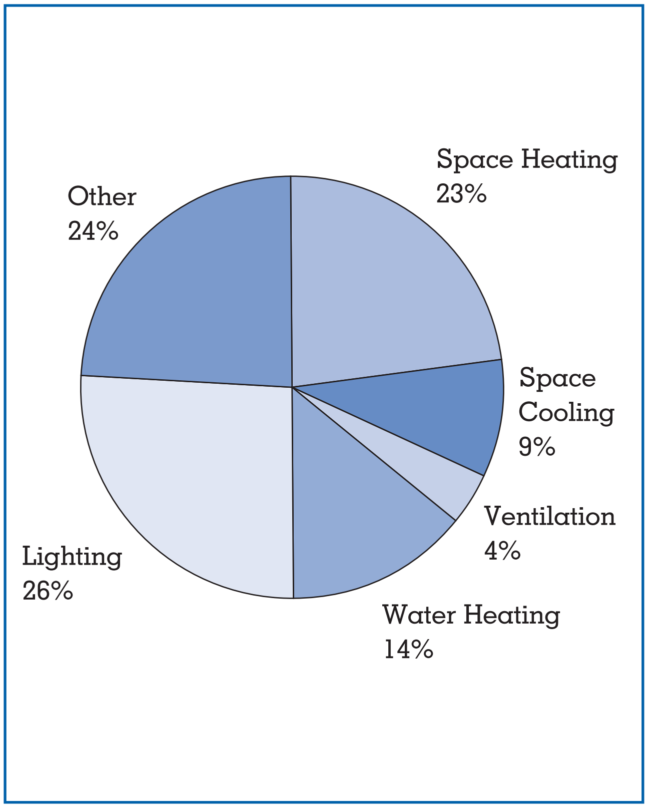


Figure 2.
Post-1980 energy end uses.

incandescent, with halogen lights accounting for 2 percent of floor space.

Lighting controls used in post-1980 buildings include automatic shutoff controls through an energy management system (6%) and an occupancy sensor (2%). Daylighting automatically dims artificial lighting in areas where light entering windows and skylights provides adequate illumination. Daylighting, while a promising technology, has not received widespread application.

Office Equipment

All of the advances in modern building design have not dramatically reduced the energy intensity of modern buildings. In fact, they have become more energy-intensive during the 1990s. Buildings constructed from 1990 to 1995 typically used 15 percent more energy per square foot than buildings constructed during the 1980s.

In 1995 an estimated 43 million PCs and computer terminals were used in commercial buildings. More than half of the 4.6 million buildings in the United States had at least one PC or computer terminal. The more PCs and computer terminals used in a given building, the greater the impact on the building's energy consumption. The proliferation of personal computers, printers, copiers, and other types of "plug loads" is the main cause of the rise in energy intensity in recent years.

THE FUTURE OF COMMERCIAL BUILDINGS

The workplace of the future has captured the imagination of researchers and designers internationally and across a wide range of disciplines, from computer science and furniture design to organizational systems. Noticeably missing from these discussions and visions is the energy research community. With more and more businesses operating under tighter and tighter margins, building owners and occupants are going to demand increased attention to the delivery and management of energy in buildings as other changes take place.

Trends in the Workplace

We are moving in a direction in which the information technologies associated with work processes are almost totally decoupled from energy and architectural technologies and systems. Not only is the workplace of the future going to demand more flexibility and rapid reconfiguration of space, it also is increasingly moving toward a radically different use of facilities. In addition, in the interest of meeting shareholder and organizational concerns for increasing profit margins, the workstation footprint is shrinking rapidly. All of these trends have implications for energy.

For example, current trends show telecommuting increasing 20 percent annually. According to one estimate, New York City will have an additional 180 million square feet of empty office space as a direct result of telecommuting. This could shift patterns in energy demand as well as require the development and deployment of new technologies that are suited to renovating facilities rather than for use in new buildings. Furthermore, although many telecommuting advocates see working at home as a way to reduce transportation impacts, there is some indication that the demand for transportation will actually increase. This is because many of the workers who have home offices are independent consultants who spend a good deal of their time in their cars visiting clients and developing work. Further, the introduction of new information technologies is opening up new markets in many areas, which, in turn, are associated with rapid delivery of products, which increase the demand for vehicles.

Many organizations are moving to open-planned, high-communication, team-centered layouts with little understanding of the organizational and performance implications of this fad, and with even less understanding of the implications for the design and delivery of energy where it is needed, when it is needed. Furthermore, the workstation "footprint" is becoming ever smaller as furniture manufacturers reduce the size of the cubicle setting in response to demand from businesses to put more people into smaller areas.

Energy Technology Development Implications

These trends have an impact on both energy use and the technologies that are developed to deliver thermal and visual comfort to the workplace. For example, how does an existing building with a fixed comfort delivery system meet the needs of these new and varying layouts? Current practice suggests that increased airflow is the answer, resulting in higher levels of reheating and recooling of air in a typical office system. Higher occupant density also translates into greater demand for outdoor air ventilation. Higher outdoor air rates can increase total building energy consumption from 20 percent to 28 percent in buildings built prior to 1989. These trends could have the following impacts leading to greater use in the workplace of the future:

- increased air movement to accommodate high occupant densities resulting in increased fan energy consumption;
- individuals bringing in desk lamps, resulting in the use of low-efficacy light sources (i.e., incandescent lamps) in addition to the high, efficacy light sources installed in the buildings;
- electric resistance heaters to warm areas of high air flow (too much cold air) and low air flow (not enough warm air) resulting from high-density loads with controls designed without a high enough degree of zone resolution;
- personal fans to create air movement that is impacted by typical partitions;
- increased ventilation requirements resulting in greater heating and cooling energy demand.

To avoid significantly higher energy consumption, building designers of the future must integrate the knowledge and expertise of the energy-efficiency community. There are several areas of integration, including:

- enhanced personal and team control over ambient conditions, including lighting, temperatures, ventilation, and acoustics;
- greater integration of energy technologies into the design of furnishings;
- glazing materials and window technologies that promote views and natural ventilation while reducing heat gain and glare;
- greater attention to understanding the energy demands that will be required to support the high-technology office of the future;
- development of ways to humanize windowless and underground spaces through features such as sensory variability, borrowed daylight through light tubes, and simulated windows;
- intelligent building systems that can identify and correct potential problems before they become large and more difficult to manage.

Jeffrey A. Johnson
William Steven Taber, Jr.

See also: Air Quality, Indoor; Building Design, Energy Codes and; Building Design, Residential; Cool Communities; Economically Efficient Energy Choices; Efficiency of Energy Use; Lighting; Office Equipment; Solar Energy; Windows.

BUILDING DESIGN, ENERGY CODES AND

Energy efficiency codes and standards for new buildings are the most cost-effective ways to improve the efficiency of a new building. Improvements can be made when a building is being designed and constructed at a fraction of the cost that it would take to make similar improvements once it is built. Energy codes require all new buildings to be built to a minimum level of energy efficiency that is cost-effective and technically feasible. Considering the average lifespan of a building, investments in energy saving technology ultimately pay for themselves many times over.

Building codes play a role in supporting the welfare of the community. They protect life, health and safety. They also protect the investment that has been made by bankers, insurance companies, businesses, and individuals. Finally, building codes promote economic development by protecting the value of the built environment.

This section will discuss the history of building construction regulations, the role of energy in building construction regulations, how these regulations have evolved into our current system, and the future of building construction regulations and energy.

BUILDING CONSTRUCTION REGULATIONS

The need to develop standards, codes, and other rules for the design, construction, operation, and maintenance of buildings has been driven over the years by a number of factors. These factors include comfort, fire safety catastrophic events, egress, sanitation, and health, among others. The first recorded building code was that of the Amorite king, Hammurabi, in 1763 B.C.E. His Code of Laws, one of the first codes of law in world history, contained 282 rules including the principles of "an eye for an eye" and "let the buyer beware." The Code of Laws stated: "If a builder build a house for some one, and does not construct it properly, and the house which he built fall in and kill its owner, then that builder shall be put to death."

Many of the building codes used initially in North America were imported from Europe by the early settlers. In 1630, the City of Boston building code stated: "No man shall build his chimney with wood nor cover his roof with thatch." In 1865, the City of New Orleans adopted an ordinance requiring the inspection of buildings for public use. In 1905 the National Board of Fire Underwriters published the first national building code in the United States, the Recommended National Building Code.

Some notable evolutionary processes occurred over the past during the twentieth century. While local government has generally retained the authority to enforce building construction regulations, the development of the codes and standards they adopt has shifted away from "home grown" criteria of various state and local agencies towards voluntary sector activities. In 1927 the Pacific Coast Building Officials (a precursor to the International Conference of Building Officials) conference published the Uniform Building Code. Then, in 1945, the Southern Building Code Congress International published the Standard Building Code.

MILLER SQA

The 290,000-square-foot Miller SQA building was designed by William McDonough, FAIA to be a state-of-the art "green" building. Miller SQA, a wholly owned subsidiary of Herman Miller, Inc., is a remanufacturer, manufacturer, and vendor of office furniture that provides "just in time" furniture products for small businesses and nonprofit institutions. The building is a manufacturing plant, warehouse, and headquarters housing approximately 600 workers in a manufacturing plant and 100 workers in the office portion. The SQA building also has a lunchroom; rest areas at each end of the manufacturing area; and a fitness center, including a full-size basketball court.

Energy-efficient aspects of the building include large-scale use of energy-efficient lighting, daylight controls, and state-of-the art digital HVAC controls, including sensors, controllers, and data loggers. Green components include environmentally sensitive materials throughout the building, minimally invasive site utilization (including a wetlands and use of natural field vegetation rather than planted and mowed grasses), enhanced indoor air quality, and extensive recycling. In addition, building materials were obtained locally whenever possible to reduce transportation costs and energy.

Studies showed that the new SQA building was associated overall with a higher quality of work life than the old building. For example, 16 percent of the office workers said they had headaches often or always in the old building, while only 7 percent did in the new building. In addition, there were small (typically less than 2%) increases in worker performance, can be significant to an organization in a competitive market. Finally, the new building uses 18 percent less energy than the old building.

Federal Legislation	*Scope*	*Date Enacted*
Energy Conservation and Production Act	Develop performance standards for all new buildings	8/14/76
New Buildings Act	HUD shall promulgate, implement and enforce energy performance standards	5/4/77
Department of Energy Organization Act	Transfer authority from HUD to DOE	8/4/77
Housing and Community Development Act of 1980	DOE shall promulgate interim standards that apply only to Federal buildings	10/8/80
Omnibus Budget Reconciliation Act	Standards to be developed through private sector.	8/13/81
Cranston-Gonzales National Affordable Housing Act	Energy efficiency standards for public housing	11/28/90

Table 1.
Codes and Standards Legislative History

Groups such as the American Society of Mechanical Engineers (ASME), National Fire Protection Association (NFPA), American Society of Heating, Refrigeration, and Air Conditioning Engineers (ASHRAE) and Underwriters Laboratories (UL) develop codes and standards for use by industry members. Model codes organizations including Building Officials and Code Administrators International (BOCA), International Conference of Building Officials (ICBO), and Southern Building Code Congress International (SBCCI) are code official organizations that develop model codes specific to their region of the country.

FEDERAL LEGISLATIVE HISTORY

Energy codes and standards stem from policy directives in the early 1970s that were hastened by the Arab oil embargo. The Secretary of the Department of Housing and Urban Development was directed in 1971 to reduce maximum permissible energy loss by about one-third for a typical home and to revise the insulation standards for apartments. The resulting thermal envelope criteria in HUD's Minimum Property Standards became the first energy code.

The Arab oil embargo and ensuing "energy crisis" in late 1973 began to change the scope of building construction regulations. The public became quickly aware of the true cost of energy as gasoline prices skyrocketed. The oil embargo created a ripple effect that demanded a larger public policy role to reduce wasteful energy use.

Table 1 describes the ensuing legislation that was promulgated to promote energy efficient construction as a result of the Arab oil embargo.

Development of standards, codes, and other regulations to address energy in buildings began in earnest in the early 1970s. The Omnibus Budget Reconciliation Act prohibited the government from promulgating regulations that would apply to nonpublic construction. This left the task of energy codes and standards development up to two key processes. They are the model code processes set up by the building regulation community and the consensus processes set up by the American Society of Heating, Refrigeration and Air-conditioning Engineers (ASHRAE) and their co-sponsor, the Illuminating Engineering Society of North America (IESNA).

ASHRAE Consensus Process

Some states recognized the need and the role that they could play if energy criteria were added to their building codes. Through the National Conference of States on Building Codes and Standards (NCSBCS) a request was made by the states for the National Bureau of Standards (NBS) to develop some criteria that could be used to address energy issues through building codes. The NBS released suggested energy

Title	*Scope*	*Description*
ASHRAE/IES Standard 90A-1980	All Buildings	Updated Sections 1 through 9 of ASHRAE/IES Standard 90-1975 with more stringent provisions.
ASHRAE/IES Standard 90.1-1989	Commercial and High Rise Residential	Updated entire commercial and high-rise residential standard with more stringent provisions.
ASHRAE Standard 90.2-1993	Low Rise Residential	Updated low-rise residential standard with more stringent provisions.
ASHRAE Energy Code for Commercial and High Rise Residential Buildings	Commercial and High Rise Residential	Standard 90.1-1989 in mandatory language for use by model code organizations.
ASHRAE/IESNA Standard 90.1-1999	Commercial and High Rise Residential	Updated entire commercial and high-rise residential standard with more stringent provisions

Table 2.
History of ASHRAE/IES Standards

requirements for new building designs in early 1974. Since the NBS criteria had not received widespread input from and agreement by the building community the states looked to a voluntary sector standards development organization to develop a standard for building energy design. Using the NBS criteria as a starting point the ASHRAE established a Standard 90 project and committee and began the process of writing a standard in cooperation with the IESNA. Released in August of 1975, ASHRAE/IES Standard 90-75 addressed energy conservation in new building design.

The DOE sponsored two major projects with ASHRAE as a result of the Omnibus Reconciliation Act; Special Project 41 and Special Project 53. These projects led to a series of changes and has continued to be amended from time to time. Table 2 outlines the evolution of the ASHRAE/IES Standard through 1999.

Model Codes Development Process

Because the ASHRAE/IES standard contained both mandatory requirements and recommendations, it was difficult to adopt and implement as a uniform building code. Such codes must clearly state minimum requirements that can be uniformly interpreted and applied. For this reason it was determined that a model code was needed that contained only the mandatory requirements of Standard 90-75 and placed them in enforceable code language that would fit within a building code. Working with NCSBCS the three U.S. model code organizations (Building Officials and Code Administrators, International Conference of Building Officials, and Southern Building Code Congress) developed and released a model energy code for new buildings in late 1997: Model Conservation Energy Code (MCEC) 1977. Table 3 shows the evolution of the model codes from the MCEC to the International Energy Conservation Code (IECC).

RECENT TRENDS IN CODE ADOPTION AND ENFORCEMENT

In 1975, the Energy Conservation and Production Act tied the availability of federal funds to states for their energy conservation efforts to the adoption by states of an energy code for new building construction. Beginning in the mid-1970s states began to adopt energy codes in earnest. This was accomplished in three ways:

1. Through individual state legislation, wherein the legislature adopted by reference a particular energy code or wrote the energy code requirements directly in the legislation.

Title	*Scope*	*Description*
Model Energy Code (MEC) 1983	All Buildings	Technically equivalent to ASHRAE/IES Standard 90A-1980
MEC 1986	All Buildings	Low-rise residential thermal envelope provisions updated to improve stringency.
MEC 1988	All Buildings	Minor upgrades.
MEC 1992	All Buildings	Low-rise residential thermal envelope provisions updated to improve stringency.
MEC 1993	All Buildings	Reference to ASHRAE Energy Code incorporated into Chapter 6.
MEC 1995	All Buildings	Window provisions added to reference National Fenestration Rating Council.
International Energy Conservation Code 1998	All Buildings	Simplified commercial provisions added as Chapter 7 and low-rise residential Solar Heat Gain Coefficient requirement.

Table 3.
Evolution of the Model Energy Code

2. To convey the authority to adopt an energy code to a state regulatory agency such as the state energy office or state office with authority for a state building code.
3. Take no action; leaving the decision up to local government.

Ultimately, the responsibility for properly implementing these legislative actions fell upon the local governments of counties and incorporated cities, towns, boroughs, and so on. In total there are over 2,000 counties in the United States and over 40,000 independent units of local government that have some ability to adopt, implement, and enforce energy codes.

Since 1975 numerous states have adopted energy codes through legislative or regulatory mechanisms. They range from those that apply to all new buildings to those that only apply to state-owned buildings, non-residential buildings, or only buildings in localities that have adopted a building code. Where the state has taken no action or only partial action on a selected building type, such as state-owned buildings, then local government is free to take action if it so chooses. In some cases state and local government are prohibited or preempted from taking any action.

Other notable adopters of energy codes are the U.S. Department of Housing and Urban Development, Department of Defense, and public utilities.

Once adopted there are many ways to implement and enforce energy codes. Where a state or local building code exists, there is already an adopted code and enforcement mechanism in place to ensure compliance with the energy code. Where no such building code infrastructure exists, then implementation in the absence of such support must rely on the following mechanisms:

- Builder certification as a condition for utility connection
- Warranty with penalties if non-compliance is verified
- Contractor licensing tied to code compliance
- Energy rating schemes that pull the market
- Professional architect and engineer certification
- Third party certification and inspection

Over time more states have secured authority to adopt and implement building codes and over time those codes have tended to be more uniform. Current activities are likely to keep the focus on national uniformity in the absence of Federal preemptive authority.

MEC Version or State Code that's Equivalent	*States Adopted*
98 IECC, For state-owned and stated-funded buildings	1 State (NE)
Exceeds 95 MEC, Statewide adoption/equivalence	4 States (CA, FL, OR, MN)
Exceeds 95 MEC, Partial adoption or equivalence (i.e. only state funded bldgs, dependent on local jurisdiction, etc.)	2 States (AK, WA[1])
95 MEC, Mandatory statewide adoption/equivalence	13 States (CT, GA, MD, MA, NC, NH, RI, SC, OH, VA, VT, UT, WI)
95 MEC, Partial adoption/equivalence	3 States (OK, LA[2], HI[2])
93 MEC, Mandatory statewide adoption/equivalence	1 State (DE)
93 MEC, Partial adoption/equivalence	5 States (TX[3], ND, MT, AL, KS)
92 MEC, Mandatory statewide adoption/equivalence	7 States (AR, IN, IA, KY, NM, TN, NY)

[1] Code exceeds 95 MEC for electrically heated buildings, but is less stringent for non-electrically heated buildings.
[2] LA and HI have 95 MEC adopted for multi-family low rise only.
[3] TX is listed twice because 93 MEC is mandatory only for state funded low-rise bldgs. Local jurisdictions are adopting 92, 93, 95 MEC on their own.

Table 4.
Status of State Adoption as of August 1999

CURRENT CODE POLICY AND STATE ADOPTION

The Energy Policy Act of 1992 (the Act) was a major policy action to promote the improved use of the nation's energy resources. The Act includes both improving the supply of energy resources and promoting the efficient use of those resources. One aspect of the Act focuses on improving the efficiency of new buildings through upgrading and adoption of energy efficiency codes and standards. The Act, combined with technical and financial support, provides states with an unprecedented level of federal support to improve the efficiency of new buildings.

State and locally adopted energy codes are supported by the Federal government in three ways:

- Federal law requires that States act to review and upgrade their codes.
- Technical support is provided by DOE in the form of tools, training programs, code user support through a toll-free number, and analysis directed by the states.
- DOE provides financial support in the form of over $4 million in special projects funding annually.

Since the adoption of the Act in 1992, state-of-the-art energy codes have been extended to cover an additional 39 percent of residential construction and 26 percent of commercial construction. Two-thirds of new U.S. residential construction (1 million homes annually) fall under federal, state, and local energy codes that meet or exceed the 1995 version of the Model Energy Code (MEC). An additional 975 million square feet of commercial construction falls under codes that meet or exceed ASHRAE/IES Standard 90.1-1989. Table 4 shows the status of state residential code adoption as of August 1999.

On December 9, 1994, the International Code

Council (ICC) was established as a nonprofit organization dedicated to developing a single set of comprehensive and coordinated national codes. The ICC founders—the Building Officials and Code Administrators (BOCA), the International Conference of Building (ICBO), and the Southern Building Code Congress International (SBCCI)—created the ICC in response to technical disparities among the three sets of model codes now in use in the United States.

Since 1972, the Council of American Building Officials (CABO) has served as the umbrella organization for BOCA, ICBO, and SBCCI. In November 1997, it was agreed to incorporate. CABO into the ICC. Responsibility for developing and maintaining the Model Energy Code (MEC) was transferred from CABO to the ICC in order to provide proper interface with the International Codes.

The first version of the International Energy Conservation Code (IECC) was published in 1998.

THE FUTURE OF ENERGY CODES

As with any market, the new construction market has a wide range of efficiencies within its participants. In theory, energy codes can shift the average efficiency of the market by eliminating the option of building to an efficiency level lower than that mandated by the code. This effect can produce significant savings even when the code minimum is set at the market "average" efficiency level.

Evidence from numerous code evaluations suggests that energy codes have transformed markets in three ways:

1. In areas where codes are well enforced, the stock of poor performing new buildings has been reduced to a minimum, and
2. In areas where utility incentive programs were successful, the overall efficiency of a typical building exceeds code.
3. Codes have brought more efficiency technologies into widespread use in the market (e.g., vinyl window frames, T-8 fluorescent lamps).

The focus of codes and standards will shift from specifying the installation of prescriptive measures to the actual performance of the final building. This shift is consistent with an overall emphasis on objective or performance-based codes within the building code community. The next generation of energy codes will

- Incorporate new technologies and practices into the standard that replace less efficient technologies and practices,
- Assure proper performance of measures once specified and installed thereby assuring that energy and environmental benefits to energy codes are realized by building owners and occupants, and
- Improve enforcement through partnerships and support of innovative enforcement practices by local governments.

Jeffrey A. Johnson
David Conover

See also: Building Design, Commercial; Building Design, Residential.

BIBLIOGRAPHY

American Society of Heating, Refrigerating, and Air Conditioning Engineers (ASHRAE). (1999). *ASHRAE/IESNA Standard 90.1-1999 Energy Standard for Building Except Low Rise Residential Buildings.* Washington, DC: Author.

Building Codes Assistance Project.*BCAP—Building Codes Assistance Project.* Washington, DC: Author. <http://www.bcap-energy.org/>.

International Energy Conservation Code 2000. (2000). Falls Church, VA: International Code Council.

U.S. Department of Energy. (July 20, 2000). *Building Standards & Guidelines Program (BSGP).* Washington, DC: Author. <http://www.energycodes.org/>.

BUILDING DESIGN, RESIDENTIAL

From early cave dwellings to the well-appointed suburban homes of today, the most fundamental reason for shelter remains to provide protection from weather and other possible dangers. Beyond this most basic purpose, modern culture has forged a number of other expectations for homes, including affordability, comfort, health, durability, and peace of mind. This article explains the concepts behind energy-efficient homes—how they meet our basic needs and expectations; gives an overview of the history of energy-efficient innovations in American housing; and provides projections for the future.

HISTORICAL DEVELOPMENT

The history of energy technology in homes from the turn of the twentieth century until the 1970s is dominated by developments that contributed to energy consumption, including modern heating and cooling systems and the proliferation of appliances and lighting systems. The trend to increased energy efficiency began after World War II with the development of individual technologies that, only in recent years, have been integrated into systems solutions.

The evolution of energy-efficient homes began in the 1830s with the advent of wood-framed construction, which is still the dominant building technique. At about the turn of the twentieth century, advances in glass manufacturing allowed builders to add windows in increasing size and quantity to both residential and commercial buildings. During the fuel shortages in World War II, there was a demand for increased insulation and some consideration given to passive solar design. But low oil prices following that war and the development of central heating and cooling systems curtailed the development of energy-efficient designs and led to increased reliance on mechanical systems to create home comfort.

During the OPEC oil embargo of the 1970s, energy efficiency became a national priority. Levels of insulation increased, and double-glazed windows became standard in colder climates. As a result of the oil crisis, there was a push to develop new energy-efficient technologies. Once wall insulation had improved and the building envelope air-sealed, losses due to windows, equipment inefficiency, and system design became more critical. New window technology featured low-E coatings and insulating gases between the two panes. These windows have became increasingly popular since the mid-1980s. Concurrent improvements in heating and cooling equipment included condensing furnaces and water heaters, heat pumps, and dramatic increases in air-conditioner efficiency.

Energy efficiency experienced some important setbacks in the mid-to late 1980s due to a number of prominent technology failures. For example, early pulse combustion gas boilers, compact fluorescent lighting, and some triple-glazed windows experienced a variety of performance problems because products were introduced before all technical issues were resolved. In some cases, energy efficiency just suffered from bad press. For instance, air-sealed homes were wrongly associated with bad indoor air quality.

By the early 1990s, the results of research and technology improvements began to be more effectively passed down to builders. In addition, other technology improvements were introduced or increased market penetration, including advanced wall system alternatives to stick framing (e.g., structural insulated panels and insulated concrete forms), geothermal heat pumps, combined space and water heating systems, mechanical ventilation with heat and recovery, and a new generation of compact fluorescent lights. Another critical development was advanced computer technology, which made it easy to model the complex dynamics of residential energy use. Energy audits and modeling have become widely available to assess the most cost-effective improvement measures. Finally, diagnostic procedures such as duct and infiltration testing and infrared imaging have enabled building scientists to refine their concept of the house as a system.

RESIDENTIAL ENERGY EFFICIENCY

Each aspect of energy use in a home influences the comfort, health, and safety of the occupants as well as their utility bills. An energy-efficient home properly utilizes systems solutions to effectively reduce energy use while improving the quality of life for its occupants. A systems solutions approach examines the interactive effects of all of the components within the house. While there is no definitive set of features, the building science community is converging on eight common elements to reduce home energy use when properly incorporated into the systems solution: air sealing, insulation, windows, duct sealing, heating and cooling, lighting and appliances, mechanical ventilation, and diagnostic testing.

Air Sealing

If outdoor air can easily leak into and through homes, both comfort and energy-efficient performance will be difficult to maintain. Today, off-the-shelf technologies that contribute to airtight construction include a variety of house wraps, sealants, foams, and tapes. In energy-efficient homes, builders use these tools to seal the myriad of cracks

Location of Common Air Leakage Paths

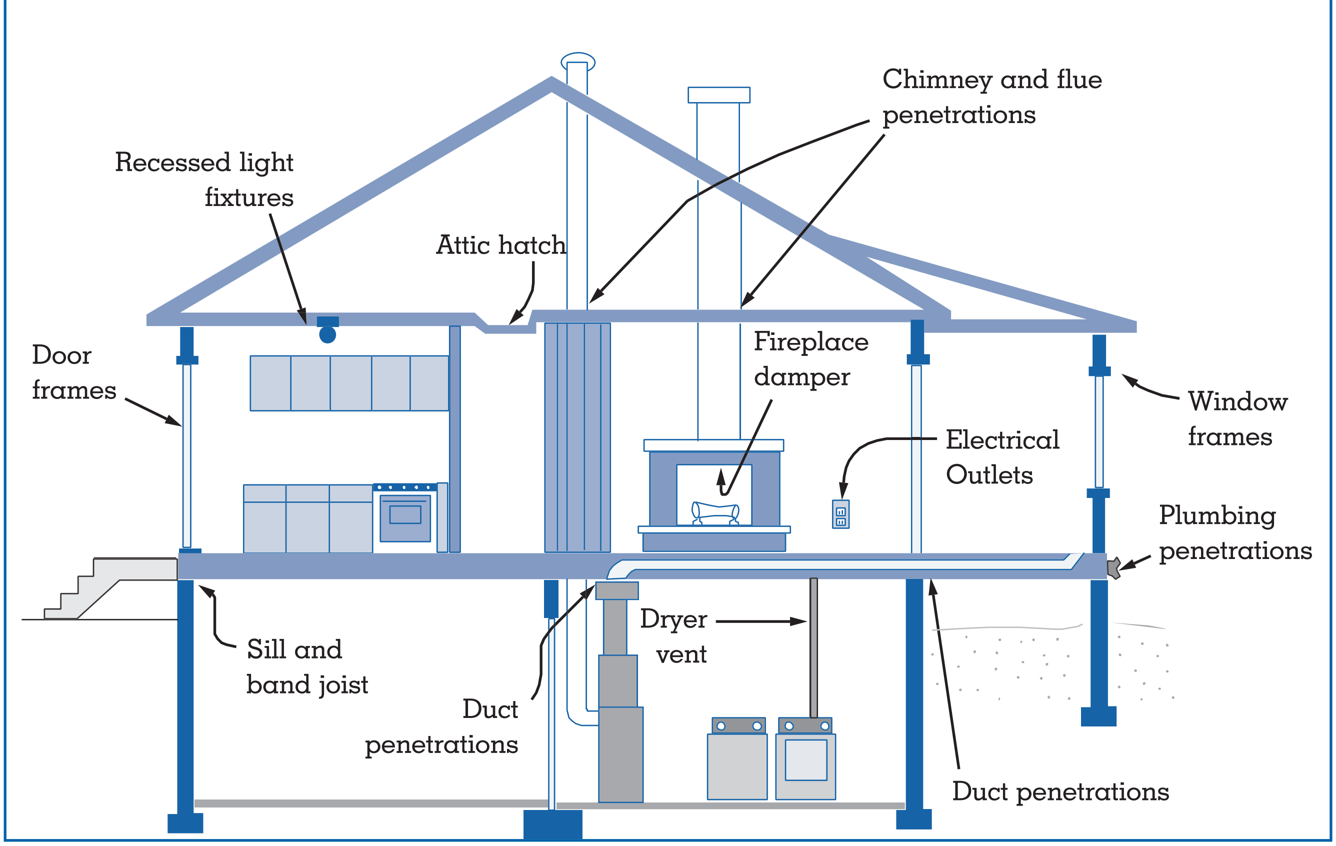


Figure 1.

SOURCE: U.S. Environmental Protection Agency

and gaps in framing along with hundreds of holes for plumbing, mechanical equipment, and electrical wiring (Figure 1). Builders have adopted new job-site management techniques to assure that air sealing is properly completed. One approach is to have a single individual responsible for a comprehensive regimen of airtight construction details. Another approach requires subcontractors to be contractually responsible for sealing all holes they make in the construction process. Diagnostic procedures include blower door testing to measure how well homes are sealed.

Insulation

A home must be correctly insulated to be energy-efficient. Table 1 shows U.S. Department of Energy (DOE) recommended minimum insulation values for different climates throughout the country. The "R-value" is a measure of insulation effectiveness—the higher the R-value, the more insulation is provided. Although there are advantages and disadvantages to each of the wide variety of insulation systems available, the most important criterion for performance is the care with which it is installed. Gaps, holes, voids, and compressions can substantially reduce insulation effectiveness. Again, diagnostic procedures can help builders and contractors maintain quality in the insulation process.

Windows

Window technology has changed dramatically since the mid-1980s. Today, advanced windows are available that approximate the thermal performance of an insulated four-inch wall and use low-emissivity (low-E) coatings, microscopic layers of metallic material applied to interior glass surfaces of double-glazed windows. These coatings prevent undesirable solar heat gain in the summer and excessive heat loss in the winter.

Cost Effective Insulation R-Values

If you live in a climate that is...	*and your heating system is a...*	*then insulate to these levels in the...*			
		ceiling	*wood frame walls*	*floor*	*basement/ crawl space walls*
Warm with cooling and minimal heating requirements (i.e.,FL & HI; coastal CA; southeast TX; southern LA, AR, MS, AL & GA).	gas/oil or heat pump	R-22 to R-38	R-11 to R-15	R-11 to R-13	R-11 to R-19
	electric resistance	R-38 to R-49	R-11 to R-22	R-13 to R-25	R-11 to R-19
Mixed with moderate heating and cooling requirements (i.e., VA, WV, KY, MO, NE, OK, OR, WA, & ID; southern IN, KS, NM, & AZ; northern LA, AR, MS, AL, & GA; inland CA & western NV).	gas/oil or heat pump	R-38	R-11 to R-22	R-13 to R-25	R-11 to R-19
	electric resistance	R-49	R-11 to R-28	R-25	R-11 to R-19
Cold (i.e., PA, NY, New England, northern Midwest, Great Lakes area, mountainous areas (e.g., CO, WY, UT, etc.)).	gas/oil	R-38 to R-49	R-11 to R-22	R-25	R-11 to R-19
	heat pump or electric	R-49	R-11 to R-28	R-25	R-11 to R-19

Table 1.

SOURCE: U.S. Environmental Protection Agency

Note: (a) Adopted from the U.S. Department of Energy 1997 Insulation Fact sheet. (b) Insulation is also effective at reducing cooling bills. These levels assume that you have electric air-conditioning. (c) R-Values are for insulation only (not whole wall) and may be achieved through a combination of cavity (batt, loose fill, or spray) and rigid board materials. (d) Do not insulate crawl space walls if crawl space is wet or ventilated with outdoor air.

Duct Sealing

American homes often utilize duct systems to distribute heated and cooled air to each conditioned room. These ducts can leak as much as 25 to 35 percent of the air they are supposed to deliver to the living space. Figure 2 shows some of the common areas of duct leakage. In addition to wasting energy and money, leaky ducts can lead to durability and indoor air quality problems.

Duct mastic has replaced duct tape as the most effective sealing material. Mastic is a fibrous, elastomeric compound that permanently seals duct connections and seams. In addition, the “boot” connections between ducts and floors and between walls and ceilings should be fully caulked and/or foamed airtight. Lastly, air handlers that house the heating and cooling coils and circulation fans are often extremely leaky and should be fully sealed while still allowing access for service and filter replacement. Use of new airtight airhandlers provides even better performance.

Heating and Cooling

Bigger is not better when it comes to heating and cooling equipment. Oversized equipment is not only more expensive to purchase but it also can reduce comfort. The most prevalent examples are the rapid temperature fluctuations and humidity control problems caused by frequent on/off cycling of oversized equipment. Poor dehumidification occurs because the cooling coils do not sustain cold temperatures long enough to effectively condense moisture out of the air.

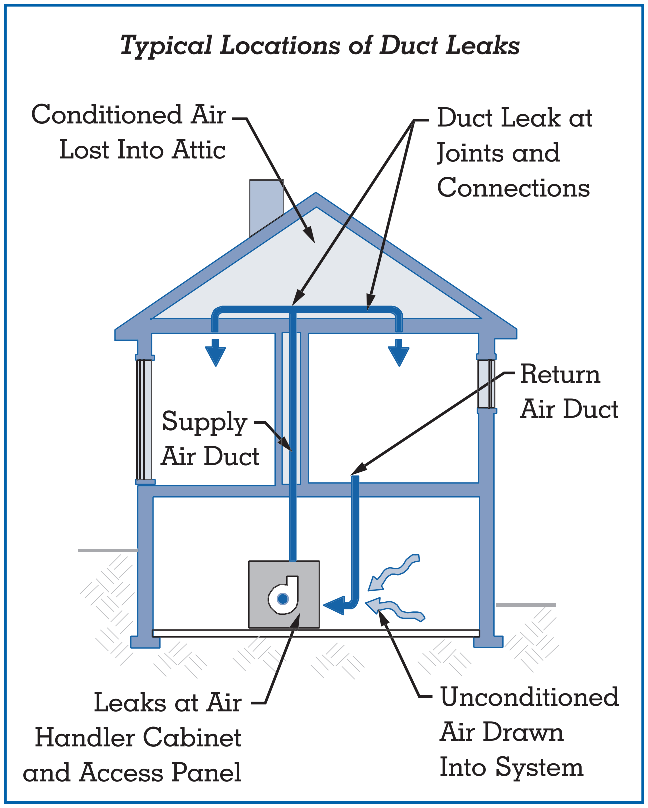


Figure 2.

SOURCE: U.S. Environmental Protection Agency

The impact of incorrect equipment sizing practices is magnified in an energy-efficient home that has dramatically reduced heating and cooling loads.

Once right sizing is addressed, new high-efficiency equipment, including furnaces, boilers, heat pumps, and air conditioners provide further energy savings and other quality improvement benefits. Geothermal heat pump equipment offers further efficiency gains by taking advantage of much more stable year-round ground temperatures (warmer than ambient air in the winter and cooler than ambient air in the summer) with buried heat exchanger loops.

Lighting and Appliances

High-efficiency lighting and appliances also can provide significant energy savings. Whole-house lighting designs typically include hard-wired compact fluorescent lighting (e.g., recessed, sconce, surface-mounted and exterior-wall-mounted fixtures), Thinner diameter (1 inch is 1/4 inch) high-efficiency fluorescent lighting with electronic ballasts, and motion sensors in rooms where lights might normally be left on during long periods of non use (e.g., laundry rooms and children's bathrooms and bedrooms). These efficient fixtures and controls provide high-quality lighting at lower cost than incandescent bulb fixtures with conventional controls. High-efficiency refrigerators and horizontal-axis clothes washers, clothes dryers, and dishwashers also offer significant energy savings. The U.S. Environmental Protection Agency (EPA) and Department of Energy work with lighting and appliance manufacturers to offer the ENERGY STAR label on high-efficiency models so that consumers can easily identify them.

Mechanical Ventilation

Ventilation is required to maintain indoor air quality and comfortable moisture levels. American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) guidelines require that a properly ventilated home replace its full volume of air approximately every three hours. Without proper ventilation, moisture generated indoors from human respiration, bathing, cooking, indoor plants, and cleaning can condense on wall and window surfaces, eventually leading to mold and mildew as well as structural damage to the house. Poorly balanced air flows due to duct leakage and inadequate air sealing can exacerbate the problems. In addition, there are internal sources of air pollution, such as off-gassing of construction materials and furnishings, solvents, combustion appliances, candles, dust mites, and pets. In most homes, particularly energy-inefficient ones, ventilation is provided accidentally by leakage of air through cracks and holes in the building envelope. Depending on wind and weather conditions, this may or may not provide adequate ventilation. Moreover, air drawn in from attics, crawl spaces, garages, or even outdoors can bring a wide range of undesired pollutants indoors (e.g., dust, pollen, radon, mold, mildew, moisture, carbon monoxide, and solvent fumes).

To provide a constant, controlled supply of fresh air to maintain proper indoor air quality and prevent moisture damage, there are three basic ventilation strategies: constant supply, constant exhaust, and balanced systems. Constant supply systems provide a continuous flow of fresh air to living areas and force stale air out through air leaks and kitchen and bathroom fan vents. Constant exhaust systems continuously remove stale air while allowing fresh air to enter either through air leaks or special vents in walls or windows. Balanced systems provide the greatest control of indoor air quality by combining both supply and exhaust functions. Some systems also incorporate a heat exchanger to pre heat or pre cool the incoming air, and dehumidifiers to remove moisture from humid outdoor air. There are new developments in lower-cost, balanced systems that utilize heating and cooling system ducts to deliver fresh air while stale air is removed by exhaust fans.

Diagnostic Testing

To build a home with maximum speed and quality, builders often hire subcontractors to install plumbing, insulation, heating and cooling systems, gypsum board, and other building components. Each trade brings specialized experience to a particular job but often fails to realize the full impact of its work on other parts of the home building process. Since the builder has the ultimate responsibility for the performance of the home, whole house diagnostics such as blower door and duct blaster tests or infrared imaging can catch problems before the buyer moves in. A blower door depressurizes the house to measure whole house infiltration and locate leaks in the building envelope. Duct blaster tests measure duct leakage and identify potential sources of comfort complaints due to reduced air flow. Infrared cameras take a thermal picture of a house to identify insulation voids and gaps, thermal bridging, and hot or cold spots. By implementing diagnostic procedures as part of a quality assurance program, the builder can avoid costly customer callbacks. Diagnostic testing services as well as advice on efficient design are available from a wide variety of experts, including local utility staff, energy consultants, home energy raters, weatherization specialists, and some mechanical, insulation, or air sealing subcontractors.

SYSTEMS SOLUTIONS

A house is a complex system of interacting parts that contribute to overall performance including comfort, energy use, health, maintenance, and longevity. For example, a common air distribution system utilizes supply ducts running through the attic and return ducts tied directly to the air handler inside the home. If the ductwork is not properly sealed and there are combustion appliances in the home, this configuration can lead to health and fire hazards because the return ducts in this case draw more air than the leaky supply ducts are able to provide to the home. The resulting negative pressure can draw combustion appliance exhaust back into the home and even create flame rollout at the equipment combustion chambers. A systems approach properly integrates tight construction, sealed ductwork, and properly functioning ventilation systems to save energy and eliminate indoor air pollution dangers.

Systems thinking also takes advantage of cost reduction opportunities. For instance, energy-efficient homes can be heated and cooled with smaller, lower-cost equipment. In some cases, energy-efficient homes have such small heating requirements that a single water heater can be used for space and water heating. In addition, a home with air-sealed construction, a well-insulated envelope, and high-performance windows no longer needs duct systems extended to outside walls to maintain comfort. A much more compact duct system will cost less, be quieter, and increase efficiency while delivering superior performance. All of these savings can be used to offset the extra cost for other energy features. As a result, an energy-efficient home based on systems solutions delivers improved weather protection, affordability, comfort, indoor air quality, and durability.

WEATHER PROTECTION

To varying degrees, code-compliant homes offer reasonable protection from the rigors of harsh weather conditions. However, features that improve the energy efficiency of homes also offer better weather protection. For example, increased protection from severe weather such as tornadoes and hurricanes can require building systems with greater structural integrity, such as six-inch framed walls, insulated concrete form (ICF) walls, and structural insulated panels (SIP). These systems are also much more energy-efficient because they allow for a better insulated building envelope and tighter construction. Protection from another, more pervasive weather condition, moisture, requires energy-efficient building practices. Building an air-sealed and properly insulated home provides obvious advantages by blocking water vapor from entering the home. Mechanical ventilation systems prevent excessive accumulation of moisture inside the home. In addition, right-sized air conditioners operate efficiently, and more effectively remove moisture from the air.

	Standard Code Home	***Energy Efficient Home***
Initial Cost	$130,000	$132,500
Mo. Mortgage:	$1,075	$1,095
Mo. Utility Cost:	$110	$80
Total Mo. Cost:	**$1,185**	**$1,175**
Mo. Cost Advantage:	**$0**	**$17**

Table 2.
Cost Advantage of Energy Efficient Homes
Cash flow example for an "average" home.

Affordability

Homes are typically purchased with 15- or 30-year mortgages. This long-term financing allows most new homes to be built much more efficiently than required by code while costing less. The reason is that monthly utility bill savings can easily exceed small increases in the monthly mortgage for the additional energy features. Table 2 shows an example where $2,500 of energy efficiency improvements cost only $20 more on the monthly mortgage but reduce monthly utility bills by $30 for a positive cash flow every month of $10, beginning the day the buyer moves in. This cost advantage can increase substantially as builders apply systems solutions to construct homes that are 30 percent to 50 percent more efficient than code at little or no additional cost.

Comfort

Drafts, condensation on windows, ice damming, excessive noise from outdoors or equipment operation, and rooms that are cold in winter and hot in summer will diminish comfort in a home. Air-sealed construction, improved insulation, high-performance windows, right-sized, efficient heating/cooling distribution systems, and mechanical ventilation commonly found in energy-efficient homes all work together to effectively eliminate these problems.

Health

Good indoor air quality, critical for a healthy home, requires effective control of pollutants and moisture. To some extent this can be accomplished by thoughtfully selecting materials that contain low levels of volatile organic compounds and formaldehyde. In addition, hard surfaces that can be easily maintained can help prevent dust mite activity and accumulation of organic matter. Beyond these basic material choices, the path to improving indoor air quality is through energy-efficient building practices. Air-sealed construction and duct systems block a wide array of pollutants from attics, garages, crawl spaces, and basements from penetrating indoors. In addition, because energy-efficient homes avoid moisture problems, they also avoid molds and mildew that can cause serious health problems associated with a number of allergies and asthma. Lastly, mechanical ventilation replaces stale indoor air with fresh outdoor air.

Durability

The features that contribute to the energy efficiency of the home also yield maintenance and durability benefits. High-performance windows block out ultraviolet (UV) radiation that can accelerate wear on carpets, interiors, and furnishings. Correctly sized heating and cooling equipment will operate near its design conditions and will likely last longer. More importantly, proper systems solutions protect against moisture damage arising from poor duct sealing, ventilation, or air sealing. A house designed as a system will likely last longer and be easier to maintain.

PROJECTIONS FOR THE FUTURE

Systems Applications

Future builders will continue the current trend of integrating systems solutions into the construction process to reduce costs, conserve resources, increase energy efficiency, provide greater control of the indoor environment, and reduce the cost of home ownership. Mechanical systems such as heating, ventilation, and hot-water heaters will be combined to improve comfort and efficiency. Homes will become healthier and "greener" as awareness of and concerns about indoor and outdoor environments grow. Builders will adopt diagnostic procedures to provide quality assurance, reduce liability, and increase customer satisfaction.

There are a surprising number of computer-controlled systems in homes today, all acting independently to control heating, cooling, security, lighting, appliances, entertainment equipment and even coffee makers. Houses of the future will integrate all of these computerized functions into centralized home automation systems that will help optimize energy efficiency, comfort, and function at minimal cost to homeowners.

Construction Applications

New building systems include structural insulated panels, insulated concrete forms, and autoclaved concrete walls. These building systems are inherently better insulated, have better air sealing, and are stronger, less tool-intensive, and less wasteful while speeding and simplifying the construction process. Their superior performance results in a home with increased energy efficiency and durability and greater comfort for the residents. All of their advantages, combined with the rising cost of quality lumber, will position building systems at the forefront of the future housing market.

Factory-Made Housing

The two major types of factory-made housing are manufactured (HUD code) and modular. The major distinction between the two is that a manufactured home has a permanent chassis so it can be moved and complies with the national HUD code, while a modular home is permanently installed on a traditional foundation and adheres to the relevant local codes. As of 2000, manufactured and modular homes represent nearly half of all new homes sold, and this fraction is growing. The advantages of a factory-made housing include greater control of the production process, full protection from adverse weather, less uncertainty about materials and labor availability, reduced construction time, and a less wasteful production process. In addition, factory-made housing can be built to higher efficiency standards because the factory setting allows greater consistency and quality control of key measures such as air infiltration reduction, duct design and sealing, and insulation installation.

Indoor Air Quality

Home owners are increasingly aware of the importance of indoor air quality. This is especially true as connections are being made between rapidly increasing cases of allergies and asthma and indoor environments. High-quality filters on air distribution systems and mechanical ventilation solutions that provide both high-quality fresh air and dehumidification will emerge as standard features. Building materials will be selected with better consideration for air-quality impacts. For example, carpets are being developed that protect against dust mite and mold problems.

Environmental Factors

The average American home is responsible for more annual pollution than the average car. This often comes as a surprise because the pollution attributed to homes is produced miles away at a power plant or out-of-sight from roof exhaust flues. However, every time someone flips a switch, activates the air conditioning, or takes a shower, pollution is being produced. There will be growing appreciation for energy efficient homes that help prevent pollution.

In both the materials used and the construction process, sustainability and efficiency will become standard. Improvements in job site management and building design will reduce waste and cost of construction. Materials such as engineered wood, recycled carpeting, and cellulose insulation, which have lower environmental impact, will become cheaper and more widely available. Not only are buildings constructed with such material friendlier to the environment, they also provide higher-quality, lower-cost solutions.

Information Technology

The home sale process will drastically change in response to the information technology revolution. Internet sites are already being established as alternative methods for selecting neighborhoods and homes, completing purchases, and arranging for financing. Thus consumers will have much greater access to information about the comfort, quiet, durability, indoor air quality and resale benefits of energy-efficient housing. As home buyers learn to distinguish between the asking price and the actual cost of homes, builders will incorporate more energy-saving features to drive down total ownership costs.

Sam Rashkin
Japhet Koteen

See also: Air Conditioning; Air Quality, Indoor; Building Design, Commercial; Building Design, Energy Codes and; Domestic Energy Use; Insulation; Windows.

BIBLIOGRAPHY

Carmody, J.; Selkowitz, S.; and Heschong, L. (1996). *Residential Windows*. New York: W. W. Norton.

Gardstein, C., and Zuckman, W. (1999). "Don't Just Raise the Bar, Redefine It." *Spotlight on SIPA* 9(1).

Heede, R. (1995). *Homemade Money: How to Save Energy and Dollars in Your Home*. Snowmass, CO: Rocky Mountain Institute.

Jump, D. A.; Walker, I. S.; Modera, M. P. (1996). "Field Measurements of Efficiency and Duct Retrofit Effectiveness in Residential Forced Air Distribution Systems." In *Proceedings of the ACEEE 1996 Summer Study on Energy Efficiency in Buildings*, Vol. 1, pp. 147–155. Washington, DC: American Council for an Energy Efficient Economy.

Nevin, R. and Watson, G. (1998). "Evidence of Rational Market Valuations for Home Energy Efficiency." *The Appraisal Journal* (October): 401–409.

U.S. Environmental Protetion Agency. (1993). *U.S. EPA National Vehicle and Fuel Emissions Laboratory Fact Sheet: Annual Emissions and Fuel Consumption for "Average" Passenger Car*. Washington, DC: Author.

U.S. Environmental Protetion Agency. (1998). *U.S. EPA Energy Star Homes Fact Sheets: Duct Insulation*. Washington, DC: Author.

BUTANE

See: Liquefied Petroleum Gas

C

CAPACITORS AND ULTRACAPACITORS

Capacitors store electrical energy in the form of an electric field between two electrically conducting plates. The simplest capacitor is two electrically conducting plates separated spatially. By inserting a dielectric material (a poor conductor of electricity) between the two plates the capacity can be greatly increased (Figure 1). The dielectric material used determines the major characteristics of the capacitor: capacitance, maximum voltage or breakdown voltage, and response time or frequency. The first capacitor, the Leyden jar accidentally discovered in 1745, is a glass jar coated with copper on the inside and outside. The inside and outside copper coatings are electrically connected to a battery. The two spatially separated copper plates are the electrodes, and the glass is the dielectric of the Leyden jar capacitor. The capacity to store electrical energy at certain frequencies and to provide high-power discharges makes a capacitor an essential component in most electrical circuits used in electronics, communication, computers, manufacturing, and electric vehicles.

Capacitance is related to the area of the plates (*A*), the distance between the plates (d), and the dielectric constant (ε) of the material between the plates (Figure 2, equation I). The dielectric constant or permittivity of a material is the increased capacitance observed compared to the condition if a vacuum was present between the plates. Common dielectric materials are polystyrene (ε = 2.5), mylar (ε = 3), mica (ε = 6), aluminum oxide (ε = 7), tantalum oxide (ε = 25), and titania (ε = 100). In the Leyden jar the dielectric is silica.

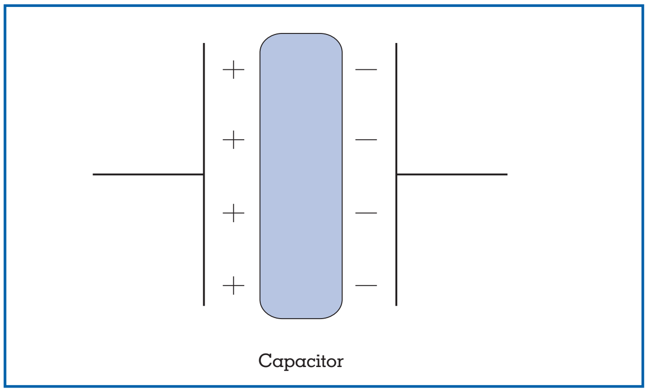


Figure 1.

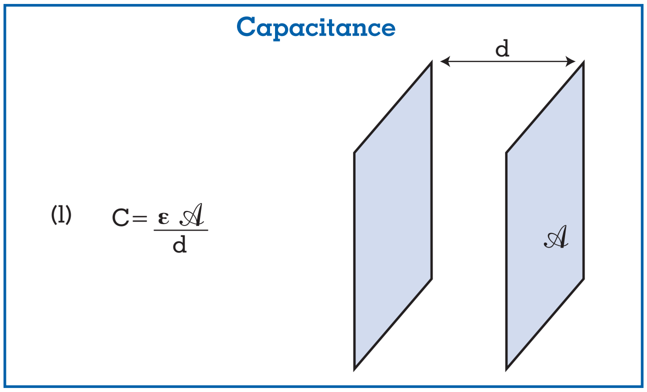
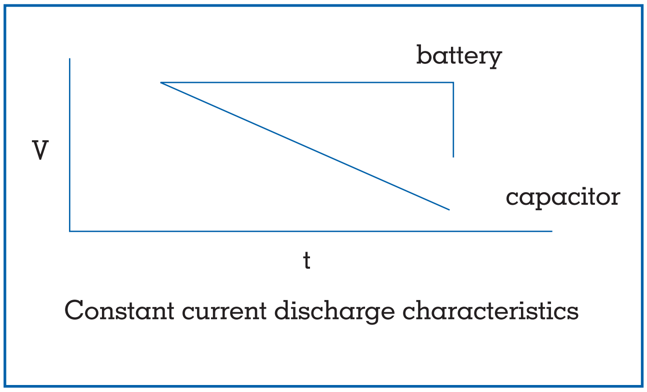


Figure 2.

A capacitor, previously called a condenser, stores electrical energy based on the relationship between voltage (V) and stored charge (Q) in coulombs as shown in the equation C=QU. One farad of capacitance is a coulomb per volt of stored charge. The voltage limit of a capacitor is determined by the breakdown potential of the dielectric material.



Constant current discharge characteristics

Figure 3.

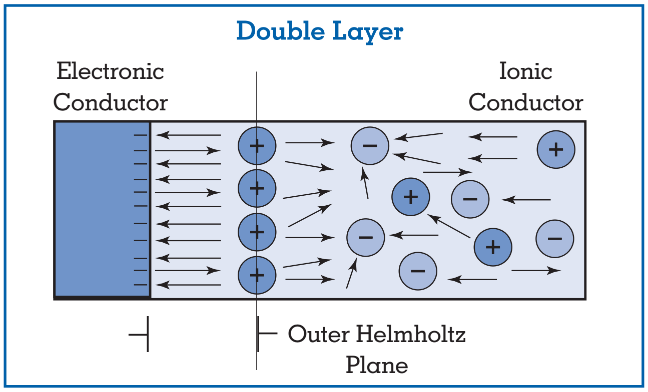


Figure 5.

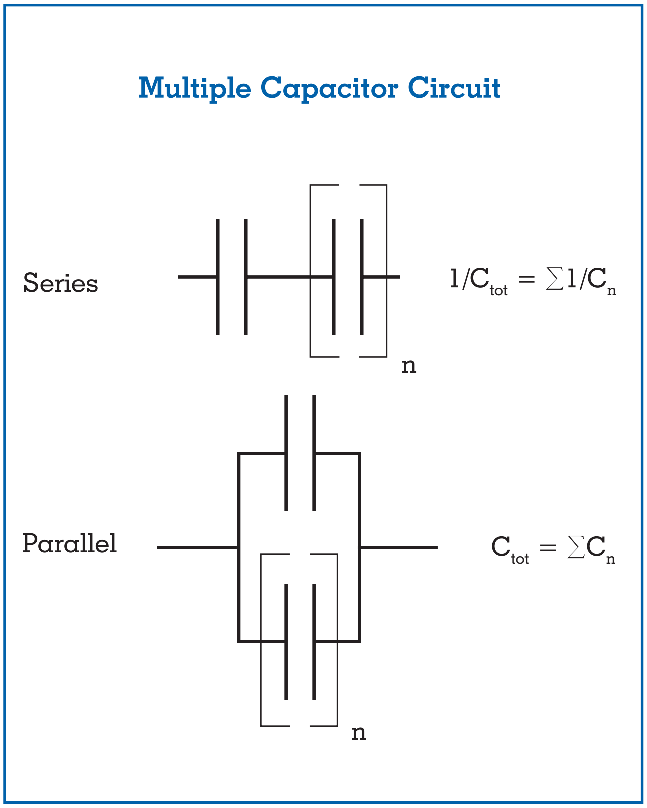


Figure 4.

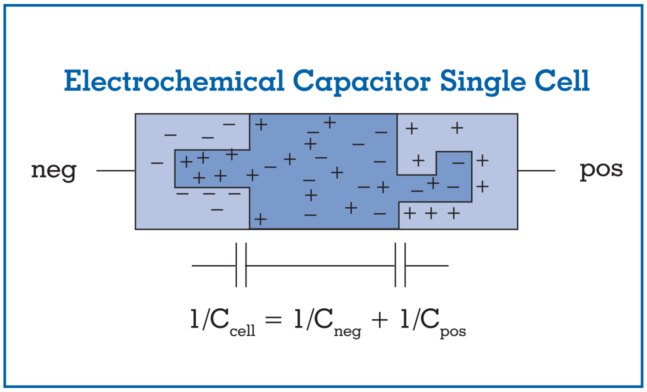


Figure 6.

Like a battery, a capacitor is an electrical energy storage device. There are, however, significant differences in how a battery and a capacitor store and release electrical energy. A battery stores electrical energy as chemical energy and can be viewed as a primary source. Capacitors need to be charged from a primary electrical source. During a constant current discharge, a battery will maintain a relatively constant voltage. In contrast, a capacitor's voltage is dependent on the "state of charge," decreasing linearly during a constant current discharge (Figure 3). The energy of a capacitor in joules is defined in the equation

$$E = \tfrac{1}{2}CV^2.$$

Capacitors are often combined in series or parallel, with the resulting circuit capacitance calculated as depicted in Figure 4. An important relationship is the time constant of a capacitor. The time constant is based on the product of the resistance and capacitance and is known as the RC time constant. A capacitor in a dc circuit will charge or discharge 63.2 percent in one RC time constant. The time dependence of a capacitor is shown in the equations.

$$V(t)=V_i e^{-t/RC}$$

and

$$I=CdV/dt$$

Electrochemical capacitors are also known as double layer capacitors, ultracapacitors, or supercapacitors. These devices are based on either double-layer charge storage or pseudocapacitance. Electrochemical double-layer capacitors, originally developed by Standard Oil Company during the 1960s, store charge at the interface between an electrically conducting electrode such as carbon and an ionically conducting electrolyte such as sulfuric acid. The double layer, first described by Hermann von Helmholtz in 1853, can be considered the equivalent of a parallel plate capacitor wherein the distance of charge separation is given by the ionic radius of the electrolyte, while the solvent continuum is the dielectric (Figure 5). The large charge storage offered by electrochemical capacitors is due to amplifying the double-layer capacitance ($\approx 15 \times 10^{-6}$ F/cm^2) by a large surface area electrode ($\approx 2 \times 10^7$ cm^2/g). Electrochemical capacitors typically have capacitance values of millifarads to tens of farads, in contrast to electrolytic capacitors, which typically have values in the range of picofarads to microfarads.

The single cell of an electrochemical capacitor consists of two electrodes separated by an electrolyte (Figure 6). The cell voltage is limited to the oxidation and reduction limit of the electrolyte, about 1.2V for aqueous and 3-4V for organic electrolytes. To obtain high-voltage electrochemical capacitor devices, single cells are connected in series to achieve the desired voltage. In contrast, electrolytic capacitors can have single-cell voltages of several hundred volts, depending on the dielectric.

Pseudocapacitance is used to describe electrical storage devices that have capacitor-like characteristics but that are based on redox (reduction and oxidation) reactions. Examples of pseudocapacitance are the overlapping redox reactions observed with metal oxides (e.g., RuO_2) and the p- and n-dopings of polymer electrodes that occur at different voltages (e.g. polythiophene). Devices based on these charge storage mechanisms are included in electrochemical capacitors because of their energy and power profiles.

A Ragone plot (Figure 7) compares the power and energy density of electrical energy storage devices. Electrolytic capacitors, based on an oxide dielectric, for example, are associated with high-power densities and low energy densities. Batteries, on the other hand, have high-energy density but limited power. Electrochemical capacitors have good energy and power densities.

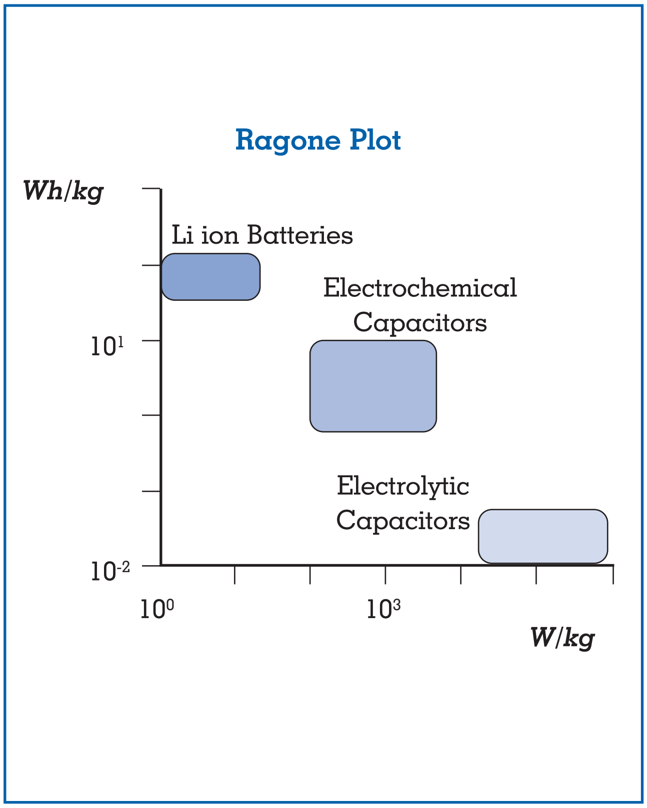


Figure 7.

Capacitors are used in many applications. Every type of capacitor has an optimum performance, depending on the voltage, capacitance, weight and volume, and frequency criteria. Optimization of circuit design requires knowledge of the performance attributes and limitations of each type of capacitor. Typically electrolytic capacitors are high-voltage, low-capacitance devices used as filters in circuits or for fast-time constant (<10 mS) circuits. Ultracapacitors are lower-voltage high-capacitance devices used as standby power for random access memory devices, power sources for actuators, circuit elements in telephone equipment, and for long-time constant (>10 mS) circuits. Ultracapacitors are used in electric vehicles and cellular phones. The rapid charging characteristics and high energy make ultracapacitors useful in smart-card applications.

Alan B. McEwen

See also: Batteries; Electricity; Electric Motor Systems; Electric Powers, Generation of; Helmholtz, Hermann von.

BIBLIOGRAPHY

Horowitz, P., and Hill, W. (1981). *The Art of Electronics.* Cambridge, Eng.: Cambridge University Press.

Conway, B. (1999). *Electrochemical Supercapacitors: Scientific Fundamentals and Technological Applications.* New York: Kluwer Academic/Plenum.

CAPITAL INVESTMENT DECISIONS

Capital investment decisions in equipment, buildings, structures and materials have important economic implications over the life of the investment. Examples of capital investment decisions include selecting the insulation for a building, choosing a heating and air conditioning system, purchasing a fuel-efficient vehicle that satisfies an individual's needs, and selecting an energy-generation facility for an industrial plant.

Capital investment decisions are best made within the context of a life-cycle cost analysis. Life-cycle cost analysis focuses on the costs incurred over the life of the investment, assuming only candidate investments are considered that meet minimally acceptable performance standards in terms of the non-monetary impacts of the investment. Using life-cycle analysis, the capital investment decision takes into account not just the initial acquisition or purchase cost, but maintenance, energy use, the expected life of the investment, and the opportunity cost of capital. When revenue considerations are prominent, an alternative method of analysis such as net benefit or net present value may be preferred.

The economic problem is to identify the best capital investment from a set of possible alternatives. Selection is made on the basis of a systematic analysis of expected costs, and revenues if they differ, over time for each project alternative.

APPROACH

A systematic approach for economic evaluation of projects includes the following major steps:

1. Generate a set of project or purchase alternatives for consideration. Each alternative represents a distinct component or combination of components constituting investment. We shall denote project alternatives by the subscript x, where x = 1, 2, … refers to alternatives 1, 2, and so on.
2. Establish a planning horizon for economic analysis. The planning horizon is the set of future periods used in the economic analysis. The planning horizon may be set by organizational policy (e.g., 5 years for new computers or 50 years for new buildings), by the expected economic life of the alternatives (i.e., the period over which an alternative remains the cost-effective choice), by the period over which reasonable forecasts of operating conditions may be made, or by the period of interest of the investor. The planning horizon is divided into discrete periods, usually years, but sometimes shorter units. We shall denote the planning horizon as a set of t = 0, 1, 2, 3, …, n, where t indicates different points in time, with t = 0 being the present, t = 1 the end of the first period, and t = n the end of the planning horizon. When comparing mutually exclusive alternatives using either life-cycle costing or net-benefits methods, the same planning horizon must be used.
3. Estimate the cash-flow profile for each alternative. The cash-flow profile should include the costs and revenues if they differ, for the alternative being considered during each period in the planning horizon. For public projects, revenues may be replaced by estimates of benefits for the public as a whole. If revenues can be assumed to be constant for all alternatives, only costs in each period are estimated. Cash-flow profiles should be specific to each alternative. We shall denote revenues for an alternative x in period t as B (t,x), and costs as C (t,x). By convention, cash flows are usually assumed to occur at the end of the time period, and initial expenditures to occur at the beginning of the planning horizon, that is, in year 0.
4. Specify the discount rate, or minimum attractive rate of return (MARR) used to discount cash flows to a common time basis. Discounting recognizes that revenues and costs incurred at different times in the future are generally not valued equally to revenues and costs occurring in the

present. Money received in the present can be invested to obtain interest income over time. The MARR represents the trade-off between monetary amounts in different time periods. The MARR is usually expressed as an annual percentage rate of interest. The value of the MARR may be set for an entire organization based upon the opportunity cost of investing funds internally rather than externally in the financial markets, or it may be set for different classes of investment depending on their riskiness. For public projects, the value of the MARR, often called the social rate of discount, is specified in accordance with public policy.The future equivalent value of a dollar one period out is calculated as $(1 + \text{MARR})$, and the equivalent value two periods in the future is $(1 + \text{MARR}) \times (1 + \text{MARR}) = (1 + \text{MARR})^2$. In general, if you have Y dollars in the present (denoted $Y(0)$), then the future value in time t (denoted $Y(t)$) is:

$$Y(t) = Y(0) \times (1+\text{MARR})^t$$

and the present value, $Y(0)$ of a future dollar amount $Y(t)$ is:

$$Y(0) = Y(t)/(1+\text{MARR})^t$$

5. Establish the criterion for accepting or rejecting an alternative and for selecting the best among a group of mutually exclusive alternatives. When both revenues and costs are to be considered, the mutually exclusive alternative with the greatest net heights is selected. For example, the alternative heating system for a building might be selected on the basis of lowest life-cycle cost, and the alternative airport configuration might be selected on the basis of highest net benefits. When all alternatives are assumed to be feasible and have equal benefits, the alternative with the smallest discounted total cost is selected.
6. Perform sensitivity and uncertainty analysis. Calculation of life-cycle costs and net benefits assumes that cash-flow profiles and the value of MARR are reasonably accurate. In most cases, uncertain assumptions and estimates are made in developing cash flow profile forecasts. Sensitivity analysis can be performed by testing how the outcome changes as the assumptions and input values change.

CALCULATIONS

Calculations of life-cycle costs, net benefits, or other measures of economic performance are commonly performed on electronic calculators, commercial spreadsheet software, or by hand. The calculation approach for life-cycle costs is to first compute the net cost amount in each period for each alternative, $C(t,x)$. The life-cycle cost (LCC) of each alternative is then calculated as the sum of the discounted values of $C(t,x)$ over the entire planning horizon:

$$LCC(x) = \Sigma^{n}_{t-0} C(t,x) \,/\, (1 + \text{MARR})^t$$

For a set of mutually exclusive alternatives, the project with the smallest LCC is the most cost-effective.

Other discounting conventions can be used for selecting capital investments. Time-adjusted cash flows can be expressed not only in terms of present value, but also as future, and equivalent annual values.

Capital investments can also be selected on the basis of other measures of performance such as return on investment, internal rate of return, and benefit-cost ratio (or savings-to-investment ratio). However, care must be taken in the application of these methods, as an incremental analysis is required to ensure consistent comparison of mutually exclusive alternatives. Also, rather than requiring a separate value to be calculated for each alternative, as in the case of the life-cycle cost method, these other methods incorporate the difference between two mutually exclusive alternatives within a single measure. For example, the net benefits measure directly pressures the degree to which one alternative is more economically desirable than another.

Special cases of capital investment decisions include lease or buy decisions, when-to-replace decisions, which design to choose, and comparison of alternatives with unequal service lives. These special cases are covered in Park (1997) and in Ruegg and Marshall (1990), as are the other methods for capital investment decisions.

Example 1.

Consider two alternative building designs (Ruegg and Marshall, 1990). One is a conventional building design; the other is an energy conserving design. Table 1 summarizes cost for both designs.

Assume the expected life of the building is 20 years and the discount rate or MARR is 8 percent. The net

	Conventional Design	*Energy Conserving Design*
Construction cost	$9,130,000	$9,880,000
Annual maintenance and operation (nonfuel)	$90,000	$60,000
Major repair (every 10 years)	$100,000	$50,000
Annual energy consumption	$55,000	$12,000

Table 1.
Conventional Design and Energy Conserving Design

present value of the costs over the life of building—that is, the life-cycle cost—for the conventional design is $9,834,635 and for the energy-conserving design, $9,834,068. Assuming both designs have the same functionality, the energy-conserving design is slightly preferable from a cost standpoint. Sensitivity analysis, however, may reveal little or no significant cost difference in the two choices.

Example 2.

Assume that a consumer is interested in purchasing a compact car. The consumer plans to keep the car for 10 years, at which point he or she assumes it will be worth nothing. A manufacturer offers a conventional model averaging 36 mpg, and a gasoline-electric hybrid car offering 70 mpg. The conventional car costs $13,000 and the hybrid car costs $19,000. The typical consumer drives 20,000 miles per year and gasoline costs $1.50 per gallon. The cars provide the same performance, comfort and reliability, and their costs of repair are approximately the same. Assume the consumer has a discount rate of 5 percent. Capital investment analysis can be used to compare the costs of the two cars over their lives.

For the conventional car, $833—(20,000 miles per year / 36 miles per gallon) × ($1.50 per gallon)—will be spent each year on gasoline. Life-cycle costs of the car for the purchase of the car and gas are

$$\begin{aligned} \text{LCC (conventional)} &= \$13{,}000 + \Sigma^{10}_{t-1}\$833 / (1.05)^{t} \\ &= \$13{,}000 + \$6{,}435 \\ &= \$19{,}435 \end{aligned}$$

For the hybrid car, $429—(20,000 miles per year / 70 miles per gallon) × ($1.50 per gallon)—will be spent each year on gasoline. A similar calculation to that for the conventional car reveals that LCC (hybrid) = $22,309. The life-cycle cost analysis indicates that from an economic point of view, the conventional car is the better purchase.

We can also explore the sensitivity of this conclu-

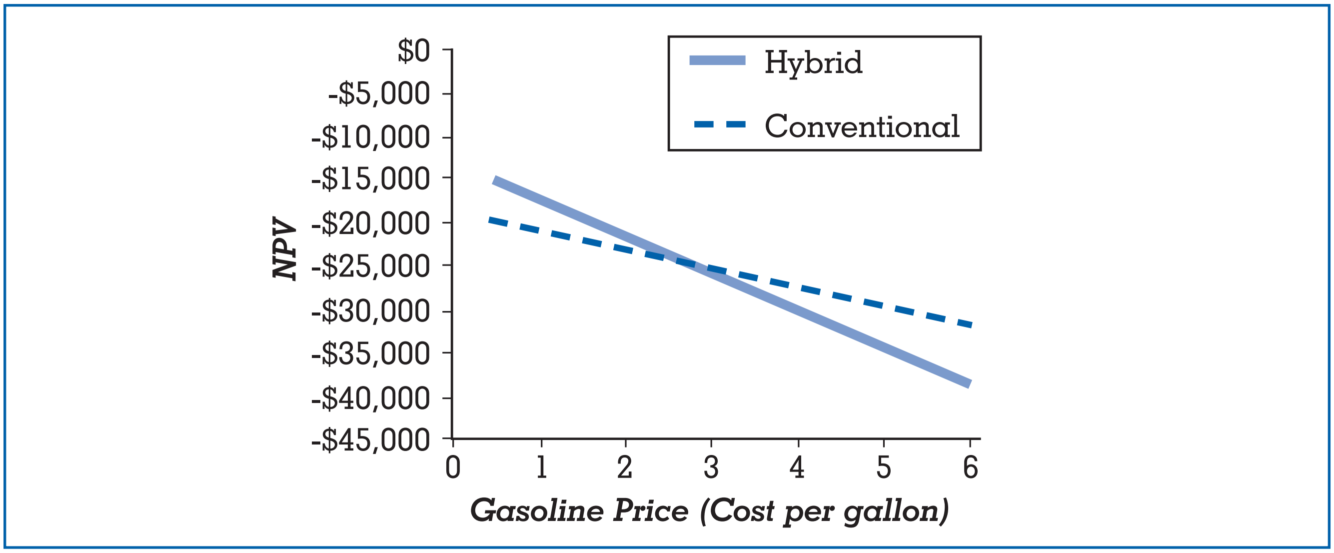


Figure 1.
Sensitivity of life-cycle cost analysis to gasoline prices.

sion to our assumptions about the initial purchase price and the cost of gasoline. Figure 1 shows the LCC of the hybrid and the conventional car over the ten-year period as a function of the cost of gasoline. When gas prices are approximately $3 per gallon, the two cars cost about the same. This value is referred to as the break-even point. If gas prices reach $3.75 per gallon, the approximate cost in Japan, the hybrid car is more economical. Sensitivity analysis can also be conducted for other input variables, such as initial purchase price, miles driven per year and actual fuel economy.

Not all decisions are made on the basis of economics alone. Consumers choose capital investments on the basis of efficiency, aesthetics and perceived benefits. An important environmental benefit of hybrid cars is the reduction in emissions. This is of particular relevance to an environmentally conscious consumer, and may shift the choice even though the benefit to the consumer is difficult to measure in dollars.

Sue McNeil

BIBIOGRAPHY

Au, T., and Au, T. P. (1992). *Engineering Economics for Capital Investment Analysis*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall Inc.

Hendrickson, C., and Au, T. (1989). *Project Management for Construction*. Englewood Cliffs, NJ: Prentice-Hall Inc.

Park, C. S. (1997). *Contemporary Engineering Economics*. Menlo Park, CA: Addison Wesley.

Ruegg, R., and Marshall, H. E. (1990). *Building Economics*. New York: Van Nostrand Reinhold.

CARBON DIOXIDE

See: Air Pollution; Climatic Effects

CARNOT, NICOLAS LÉONARD SADI (1796–1832)

Sadi Carnot, French physicist and engineer, was born in Paris during the French Revolution. He was the eldest son of Lazare Carnot, a military engineer known for writings on mathematics, mechanics, military strategy and politics. Lazare was a major government official under the First Republic that had begun in 1789, and often was described as the "Organizer of Victory" for his training and equipping of the raw recruits called to arms to defend France. Lazare had a great interest in the theory and building of machines of all kinds, an interest he passed on to his son.

Sadi Carnot's early education in mathematics, physics, languages and music was provided at home by his father. In 1812 he entered the prestigious Ecole Polytechnique, a college intended mainly for future military officers, and received a rigorous training in mathematics and science. In 1813 Carnot petitioned Napoleon to allow the Polytechnique students to help defend France against their European attackers. Napoleon granted this request and in March 1814 Carnot and his comrades from the Polytechnique fought in vain to keep the attacking armies out of Paris. Later that year Carnot graduated tenth in his class of sixty-five from the Polytechnique. He then studied military engineering at the Artillery College in Metz, and from 1816 to 1818 served as a second lieutenant in charge of planning fortifications.

Carnot soon realized that he did not have the temperament of a soldier and in 1818 left the army. After leaving the army Carnot took up residence in his father's former Paris apartment, and was presumably supported by his family whiile he attended classes at Sorbonne, the College de France, and the Conservatoire des Arts et Metiers. He also frequently visited factories and workshops, both to see steam engines actually in use, and to learn more about the economics of such industrial use of energy. There were rumors that he did at least on a few occasions receive some consultant's fees for his advise, but there was no clear documentary evidence of this. In 1827 he returned to active military service with the rank of captain, but this lasted only a little more than a year. He resigned in 1828 and died of cholera four years later in Paris.

Sadi Carnot has been called "the founder of the science of thermodynamics." In 1824, when he was twenty-eight, he first became interested in steam engines. All that time, Great Britain led the world in the design and improvement of such engines for industrial purposes. Always the French patriot, Carnot wanted his country to surpass the British, who had spawned the Industrial Revolution. He thought

Nicolas Léonard Sadi Carnot. (Library of Congress)

that a more scientific discussion of steam engines based on sound physical principles might reveal more about such engines than had the highly practical, engineering-type approach of the British pioneers.

Carnot developed the concept of an ideal "Carnot engine" (i.e., one in which all physical processes were completely reversible) to study engine efficiency and to apply his results to practical steam engines. Carnot demonstrated that this idealized engine would be more efficient than any practical steam engine ever built or that could be built. Carnot's engine consisted of a gas in a cylinder fitted with a frictionless piston. He imagined a "cycle," in which the gas absorbed heat from a hot source at temperature T_H, expanded and did work by pushing the piston outward, gave up heat to a colder condenser at temperature T_C, and then contracted, returning to exactly the same state as at the beginning of the cycle. This cycle later came to be called a "Carnot" cycle by researchers in the fields of heat and thermodynamics.

From his study of this cycle, Carnot concluded that the engine efficiency was independent of the working substance (e.g., steam or air). He also found that the maximum possible efficiency of an ideal engine (i.e., the ratio of the work done to the heat delivered from the hot source) was in every case

$$e_{max} = 1 - T_C/T_H .$$

There is a problem with Carnot's analysis, however, since at that time almost all physicists (including Carnot) thought heat consisted of a substance called "caloric," which could not be created or destroyed. As a result, the amount of heat taken from the hot source at temperature T_H would have to be the same as that delivered to the cold reservoir at temperature T_C. Because no heat was converted into work, the efficiency of such an engine would be zero.

It is noteworthy, however, that some of Carnot's notes that were published together with his classic treatise *Réflexions sur la puissance motrice de feu* (Reflections on the Motive Power of Fire), written in 1824 but only formally published in 1878, contain the following sentences: "When a hypothesis no longer suffices to explain phenomena, it should be abandoned. This was the situation with caloric, which physicists regarded as matter, as a subtle fluid." He goes on to say that Count Rumford's experiments (1798) had shown that heat is produced by motion, for example, by the rubbing of two objects together, which increases the molecular motion in the objects. When this concept of heat was introduced into Carnot's 1824 manuscript, there emerged a lucid statement of the conservation of energy principle—what became the first law of thermodynamics.

Carnot's research also made a major contribution to the second law of thermodynamics. Since the maximum efficiency of a Carnot engine is given by $1 - T_{C/T}H$, if the engine is to be 100 percent efficient (i.e., $e_{max} = 1$), T_C must equal zero. This led William Thomson (Lord Kelvin) to propose in 1848 that T_C must be the absolute zero of the temperature scale later known as the "absolute scale" or "Kelvin scale."

Because Carnot's 1824 manuscript remained unpublished at the time of his death in 1832, it was left to Kelvin and Rudolf Clausius to show how the second law of thermodynamics was implicit in Carnot's work. For this reason Kelvin once referred to Carnot as "the profoundest thinker in thermodynamic philosophy in the first thirty years of the nineteenth century."

Joseph F. Mulligan

BIBLIOGRAPHY

Challey, J. F. (1971). “Carnot, Nicolas Léonard Sadi”. In *Dictionary of Scientific Biography, Vol. 3,* ed. Charles Coulston Gillispie. New York: Scribner.

Gillispie, C. C. (1971). *Lazare Carnot Savant.* Princeton: Princeton University Press.

Harismendy, P. (1995). *Sadi Carnot: L’ingénieur de la Républic.* Paris: Perrin.

Kuhn, T. S. (1955). “Carnot’s Version of Carnot’s Cycle.” *American Journal of Physics* 23:91–95.

Kuhn, T. S. (1977). “Energy Conservation as an Example of Simultaneous Discovery.” In *Kuhn: The Essential Tension,* pp. 66–104. Chicago: University of Chicago Press.

Mendoza, E., ed. (1960). *Reflections on the Motive Power of Fire by Sadi Carnot and Other Papers on the Second Law of Thermodynamics by E. Clapeyron and R. Clausius.* New York: Dover Publications, Inc.

CARS

See: Automobile Performance

CARSON, RACHEL (1907–1964)

Nature writer and marine biologist Rachel Carson set off a storm of controversy in 1962 with the publication of her book *Silent Spring.* In her exposé on the dangerous consequences of the indiscriminate use of pesticides, Carson questioned the benefits of the synthetic chemical DDT, condemned scientific conceit, chastised the chemical industry for pursuing dollars at the expense of nature, and chided agriculturists and government officials for polluting croplands and roadsides. Calling such behavior irresponsible, Carson suggested that if people were not careful, they would eventually destroy the natural world so completely that one day they would awaken in springtime to find no birds chirping, no chicks hatching, and a strange shadow of death everywhere.

Agriculturists and chemical officials scorned Carson’s jeremiad and argued that she misrepresented the evidence, while conservationists such as Supreme Court Justice William O. Douglas praised her for writing what he called “the most important chronicle of this century for the human race.” From the Supreme Court to Ladies’ Home Gardening Clubs, Americans discussed Carson’s polemic, and in 1963, President John F. Kennedy entered the fray by commissioning a study on pesticides. In sum, *Silent Spring* contributed to a broader discussion of the environment and served as one significant catalyst for the emergence of the modern environmental movement.

Carson joined a growing number of voices that expressed concern about wilderness preservation, clean air and water, and nuclear energy. And, in concert with others, her writings and action—lobbying Congress, giving speeches, writing letters-to-the-editor, and working with conservationist organizations—spawned a host of environmental regulations that helped shape the contours of government policy, industrial action, scientific development, and individual lifestyles. In 1969, Congress created the Environmental Protection Agency. In 1970, the first Earth Day was held. Thus, Carson’s work helped revolutionize the ways that many thought about the environment. President Jimmy Carter honored Carson posthumously in 1980 with the Medal of Freedom, saying, “Always concerned, always eloquent, she created a tide of environmental consciousness that has not ebbed.”

Pesticide use was not Carson’s chosen topic. She preferred to author works that simply fostered a deeper appreciation of nature. A shy and soft-spoken woman, Carson wrote with an Albert Schweitzer-like reverence for life. All was sacred to her. Her style was lyrical, vivid, and romantic, falling mostly within the nature-writing tradition. She gave her creatures anthropomorphic characteristics, set them in dramatic situations, hoping, she said, “to make animals in the woods or waters, where they live, as alive to others as they are to me.”

Born in 1907 to Robert and Maria Carson, Rachel developed her admiration for nature in the woods and wetlands of her home in the Allegheny hills of western Pennsylvania. Her mother nurtured this interest with nature-study books. Simultaneously, Rachel cultivated her desire to write, publishing her first piece at eleven in the children’s magazine *St. Nicholas.*

One of the greatest influences in Carson’s life, next to her mother who was her lifelong companion, was her biology teacher at Pennsylvania College for Women. After a required course from Mary Scott Skinker, Carson switched her major from English to biology and, following her mentor's footsteps, pursued her studies with a master's in marine zoology from Johns Hopkins University. After a short stint of teaching at the University of Maryland, Carson landed a job in 1935 with the U.S. Bureau of Fisheries (later the Fish and Wildlife Department). In this position, Carson pushed for the protection of natural resources in twelve pamphlets she wrote for the Department's "Conservation in Action" series. To supplement her income, Carson wrote nature articles for popular magazines and completed the first two books of her sea trilogy.

The publication of her second book in 1951 brought Carson much acclaim and changed her life significantly. *The Sea Around Us* garnered positive reviews and was on the *New York Times* best-seller list for eighty-six weeks. Among other honors, it won the National Book Award and the John Burroughs Medal for excellence in nature writing. In 1951, Carson gained an esteemed Guggenheim Fellowship, and in 1952 she resigned from her government position. Book royalties from *The Sea Around Us* made it possible for her to live by writing. In 1955, she released her third book on the sea, again to widespread praise and more prizes.

While Carson's favorite topic was the sea and coastal shores, she saw the "contamination of man's total environment" as "the central problem of [her] age." Watching her own Allegheny Hills change in the wake of burgeoning industrial activity fed this concern. The atomic bomb and the dumping of nuclear wastes on the ocean's floor increased her anxiety. But it was the spread of synthetic pesticides that disturbed her the most. Although troubled by research on DDT in the 1940s, Carson did not get embroiled in the issue until 1957 when she followed a trial in Long Island, New York between local citizens and the U.S. Department of Agriculture. At issue was the spraying of pesticides over private land. Disturbed by plaintiff complaints of poisoned bird sanctuaries and gardens, Carson spent the next four years researching and writing about the impact of synthetic chemicals on the ecosystem.

Rachel Carson. (Corbis-Bettmann)

Carson's work on pesticides and her writings on the sea are two parts of the same message. In all, she wanted to communicate the wonder she felt for the natural world, a world she saw as harmonious, balanced, and beautiful. And, in each, she challenged her fellows to reverence nature and act responsibly to preserve and protect natural habitats. Disputing the notion that humans live separate from and in dominion over the rest of nature, Carson placed people within a "vast web of life" connected to all parts of the ecosystem. Humans were but one small piece. They ought, she emphasized, to respect that place and not squander the world's resources. In the interests of energy use, the message was clear. Harvesting and employing the earth's natural resources was no longer a matter of human need only. Energy officials and policy-makers must consider the requirements of other constituents in the ecosystem, and must heed possible long-term environmental consequences.

Carson succeeded in conveying this view, in part, because she was not the only one fighting for environmental protection. For close to a century, since at least the writings of George Perkins Marsh, the conservationist movement had been building. In the 1940s, conservationists Aldo Leopold and Paul Sears espoused a similar ecological ethic and, for much of the 1950s, David Brower, head of the Sierra Club, fought for public attention in the hopes of saving natural wilderness areas. In 1956, Brower and others successfully resisted the damming of Echo Park in

Colorado, a moment now marked as a turning point in environmental protection, and one that had immediate consequences for developers interested in harnessing the energy of water. In the same year that Carson released *Silent Spring*, Murray Bookchin offered a parallel warning in his book *Our Synthetic Environment*. It is in this context that the success of Carson's work makes sense. Her writings complemented other actions. Still, it was Carson's lay appeal—her refusal to use technical and scientific jargon—that popularized this ecological vision and catapulted concerns about nature into the mainstream of American life, making the environment a crucial part of the agenda in any future technological decisions.

The modern environmentalist's interdependent understanding of the world that tied humans to nature has had multiple implications for technology and energy issues, both nationally and globally. Beginning in the 1950s and 1960s, technological decisions and government, business, and individual energy needs could not easily be divorced from social or ecological concerns. By the 1970s the common assumption was that human actions almost always changed the environment, often in irretrievable ways. Questions of costs and benefits and issues of short term and long-term effects dominated policymaking. This became most apparent in 1973 when the energy crisis intersected with the environmental movement. With the OPEC (Organization of Petroleum Exporting Countries) embargo, costs of oil soared, forcing the nation to focus on the depletion of natural resources and America's high energy-dependent lifestyles. Energy issues became inextricably tied to environmental and economic factors.

Rachel Carson played a crucial role in the ways Americans interpreted these events. Though a quiet person, more a recluse than an activist, her naturalist concerns won out. Propelled away from her refuge by her beloved sea to write a book on pesticides, Carson's activism grew and she spent her last days in the halls of Congress lobbying for environmental legislation. Her fight was short-lived, however. Carson died of breast cancer in 1964, two years after the publication of *Silent Spring*.

Linda Sargent Wood

BIBLIOGRAPHY

Bookchin, M. (1962). *Our Synthetic Environment*. New York: Knopf.

Carson, R. (1941). *Under the Sea-Wind: A Naturalist's Picture of Ocean Life*. New York: Simon and Schuster.

Carson, R. (1951). *The Sea Around Us*. New York: Oxford University Press.

Carson, R. (1955). *The Edge of the Sea*. Boston: Houghton Mifflin Company.

Carson, R. (1962). *Silent Spring*. Boston: Fawcett Crest.

Lear, L. (1997). *Rachel Carson: Witness for Nature*. New York: Henry Holt and Company.

Lear, L., ed. (1998). *Lost Woods: The Discovered Writing of Rachel Carson*. Boston: Beacon Press.

Lutts, R. H. (1985). "Chemical Fallout: Rachel Carson's *Silent Spring*, Radioactive Fallout and the Environmental Movement." *Environmental History Review* 9:210–225.

Melosi, M. V. (1985). *Coping with Abundance: Energy and Environment in Industrial America*. Philadelphia: Temple University Press.

Worster, D. (1977). *Nature's Economy: A History of Ecological Ideas*. New York: Cambridge University Press.

CATALYSTS

A catalyst is a substance whose presence increases the rate of a chemical reaction. The exercise of using catalysts is called catalysis. Today the vast majority of all commercially important chemical reactions involve catalysts, especially in the fields of energy production, petrochemicals manufacture, pharmaceuticals synthesis, and environmental protection.

HISTORY

Catalysis was practiced long before it was recognized as a scientific discipline. The earliest example of catalytic reactions was in the generation of alcoholic beverages through biocatalysis dating from the Neolithic Age. About 2,500 years ago a base-catalyzed (potash lye) process was used to manufacture soap in the Mediterranean area. Although the details are not known, in the 1500s alchemists claimed to have prepared sulfuric acid (sulfuric ether and oil of vitriol they called it) by a mysterious process that would probably be classified as a heterogeneous catalytic reaction today.

During the first third of the nineteenth century, several systematic observations led researchers to conclude that the mere presence of metals induced chemical transformations in fluids that would otherwise not have occurred. Early on, Thenard had observed that ammonia could be decomposed when passed through a red-hot porcelain tube, but only if the tube contained iron, copper, silver, gold, or platinum. Humphrey Davy observed that a warm Pt wire glowed red when placed into a gas-air mixture, and the gas was combusted. His cousin Edmond Davy was able to combust alcohol when exposed to finely-divided Pt particles even at room temperature. Döbereiner combined these discoveries with a hydrogen generator to produce a commercial lighter. Michael Faraday commented on Döbereiner's work in 1823 and, during three months of experiments in 1835, demonstrated catalyst poisoning. Faraday considered catalysts to be just one manifestation of ordinary chemical reactions. Eilhardt Mitscherlich summarized these and other strange results in 1834 and attributed the phenomena to being in "contact" with the substances. Five years before the word was catalysis was coined, in 1831 Peregrine Phillips obtained the first patent in field for an improved method of making sulfuric acid.

In 1836 Jons Jakob Berzelius considered eight seemingly unrelated experimental results and concluded that there was a common thread among them. The commonality he defined as catalysis. In doing this, Berzelius proposed that a "catalytic force" was responsible for catalytic action. The concept of catalysis is today considered by most researchers to be due to Berzelius, probably because of the popularity of his annual *Handbook of Chemistry* where he published his definition of catalytic action. For the next one hundred years many referred to the phenomenon as "contact catalysis" or "contact action," as proposed by Mitscherlich.

Justus von Liebig was another leader in the training of chemists, and many of his students were placed in influential positions throughout the scientific world during the mid-1800s. Liebig was a forceful personality who defended his "turf" with vigor. His concept of catalysis was strongly influenced by purification, a subject poorly understood at that time. Making an analogy with spoilage, Liebig proposed that catalytic action is based on an induced vibration. Just as one rotten apple will eventually cause all apples in a barrel to rot, so Liebig considered that a substance that vibrates at just the right frequency will induce vibrations in certain other molecules through contact and thereby enhance the rate of their reaction. Liebig used this concept as his basis of catalysis and to explain many other phenomena; he even considered catalysis as being analogous to a perpetual motion machine. Liebig's view was seriously considered for nearly half a century.

German chemist Wilhelm Ostwald. He was awarded a Nobel Prize for his work on chemical equilibrium. (Library of Congress)

Wilhelm Ostwald was also defining physical chemistry during the 1880s. As an editor of a new journal devoted to physical chemistry, he wrote brief critical comments about many papers. Reviewing a paper that used Liebig's vibrational theory to explain results, Ostwald provided a new definition of a catalyst that was widely accepted and led to his being awarded the Nobel Prize in 1909.

Ostwald first came to catalysis through his work on the acceleration of homogeneous reactions by acids. This work was popular at the time although ultimately it would be shown to be incorrect because he believed that the acid, acting as a catalyst, did not enter into the chemical change which it influenced but rather acted by its mere presence (contact catalysis).

Discarding Liebig's theory as worthless because it could not be subjected to experimental verification, Ostwald contended that a catalyst merely sped up a reaction that was already occurring at a very slow rate. He also indicated that a catalyst cannot change the equilibrium composition. By analogy, he considered a catalyst to be like oil to a machine, or a whip to a horse; both caused the rate to increase.

Catalysis was soon divided into two classes: positive ones that accelerate reactions and negative ones that suppress reactions. It is now recognized that what was viewed for many years as negative catalysis was actually a case of catalyst poisoning; that is, some material is so strongly adsorbed that it effectively reduces the number of catalytic sites available to the reactant, thereby decreasing the reaction rate. This led to an understanding that while the catalytic steps themselves regenerate the original catalyst, most catalytic reactions are accompanied by side reactions that irreversibly decrease the catalytic activity. Some of these include sintering of highly dispersed metal particles; chemical poisoning by reactants, products, or impurities in the feed stream; physical blockage of active sites; or mechanical damage to catalytic particles. In spite of these effects, Haensel once calculated that each Pt atom in a petroleum reforming catalyst could convert a staggering 10 million hydrocarbon molecules into higher octane fuels during its lifetime before the catalyst had to be regenerated.

A recent definition of catalysis that is based on thermodynamics was advanced by the Subcommittee on Chemical Kinetics, Physical Chemistry Division, IUPA:

> "A catalyst is a substance that increases the rate of a reaction without modifying the overall standard Gibbs energy change in the reaction; the process is called *catalysis*, and a reaction in which a catalyst is involved is known as a *catalyzed reaction*."

While this definition does not address the question of "how" catalysts effect rate increases, it does ensure that a catalyst cannot cause the equilibrium composition to deviate from that of the uncatalyzed reaction.

In 1947 Sir Hugh S. Taylor summarized the state of catalysis in a "Science in Progress" article as follows:

> "Catalysis has been employed in science to designate a substance which by its mere presence facilitates or enhances the rate of chemical reactions. As such it was a cloak for ignorance. When the states of an over-all catalytic process can be described in terms of a well-defined succession of chemical and physical processes the details of which are well understood or are quite plausible, then the necessity for employing such a word as catalysis to mask our ignorance no longer exists. . ."

HOMOGENEOUS CATALYTIC REACTION MECHANISMS

Compared with uncatalyzed reactions, catalysts introduce alternative pathways that, in nearly all cases, involve two or more consecutive reaction steps. Each of these steps has a lower activation energy than does the uncatalyzed reaction. We can use as an example the gas phase reaction of ozone and oxygen atoms. In the homogeneous uncatalyzed case, the reaction is represented to occur in a single irreversible step that has a high activation energy:

$$O + O_3 \rightarrow 2\,O_2 \qquad \text{(Reaction 1)}$$

When chlorine acts as a catalyst, the reaction can be considered as two steps with the Cl being depleted in Reaction 2 and regenerated in Reaction 3:

$$O_3 + Cl \rightarrow ClO + O_2 \qquad \text{(Reaction 2)}$$

$$O + ClO \rightarrow Cl + O_2 \qquad \text{(Reaction 3)}$$

The activation energies of Reactions 2 and 3 are each much lower than the activation energy of the uncatalyzed case [1]. Thus, the kinetic definition could be stated along the following lines: A catalyst effects the rate increase by altering the homogeneous reaction pathway to a polystep reaction pathway, wherein each catalyzed step has a lower activation energy than the single homogeneous reaction, thereby increasing the rate of reactant conversion above that of the uncatalyzed reaction.

HETEROGENEOUS CATALYTIC REACTION MECHANISMS

Heterogeneous catalytic reactions always involve more than one phase with an interface separating them. The chemical reactions occur at that interface, as shown in Figure 1. A fluid molecule (e.g., gaseous) to be converted must react with a surface (usually solid) to form a surface adsorbed species. That species then reacts either with another adsorbed molecule (or a molecule from the fluid phase, or it may act unimolecularly as in Figure 1) to be transformed into an adsorbed product molecule, which then desorbs into the fluid phase. Each step (dashed lines) must have an activation energy that is lower than the homogeneous barrier height (solid curve). The depth of the potential energy curve indicates the strength with which

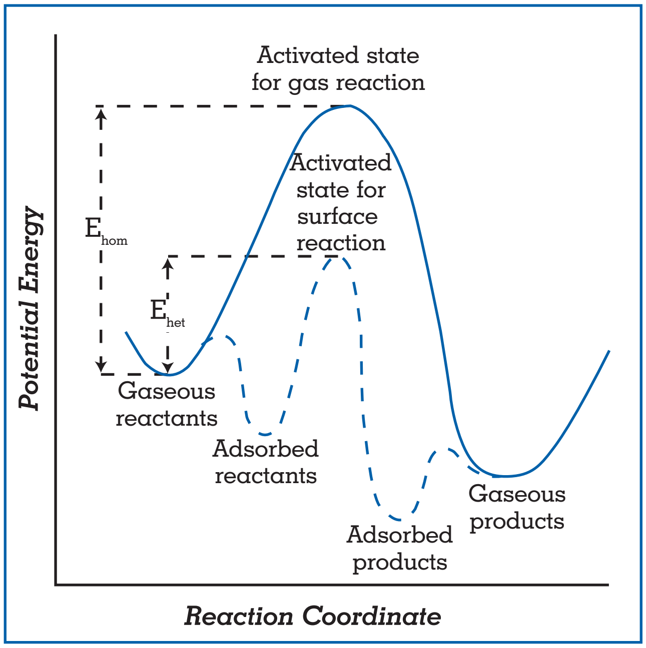


Figure 1.
Potential energy curves for a reaction proceeding homogenously (full curve) or on a surface (dotted line).

each of the species is adsorbed. If the energy decrease is very large, the molecules are strongly adsorbed. When the strongly adsorbed molecule is the reaction product, it may be difficult to remove and can cover the active sites to the point where the reaction rate is actually inhibited by its presence.

Heterogeneous catalytic systems offer the advantage that separation of the products from the catalyst is usually not a problem. The reacting fluid passes through a catalyst-filled reactor in the steady state, and the reaction products can be separated by standard methods. A recent innovation called catalytic distillation combines both the catalytic reaction and the separation process in the same vessel. This combination decreases the number of unit operations involved in a chemical process and has been used to make gasoline additives such as MTBE (methyl tertiary butyl ether).

KINETICS

All catalytic reactions involve chemical combination of reacting species with the catalyst to form some type of intermediate complex, the nature of which is the subject of abundant research in catalysis. The overall reaction rate is often determined by the rate at which these complexes are formed and decomposed. The most widely-used nonlinear kinetic equation that describes homogeneous reactions involving enzyme catalysts was developed by Leonor Michaelis and Maude Menten:

$$\text{Rate} = k\,\theta_A = k\,C_{Eo}\,C_A/(C_M + C_A)$$

where A is the reacting species, E the enzyme catalyst, C_i is the fluid phase concentration of each species i, k is the temperature dependent reaction rate constant, and C_M is the Michaelis constant that has the same dimensions as do the concentration terms. θ_A is the fraction of enzyme molecules tied up in the intermediate complex at any time. Note that when C_A is much smaller than C_M (weak binding) the rate depends linearly on both the enzyme concentration C_E and the reactant concentration C_A. However, if C_A is much larger than C_M (strong binding), then θ_A approaches unity and the reaction rate depends only on the concentration of the enzyme and is independent of the reactant concentration.

A similar nonlinear equation for heterogeneous catalytic systems was developed empirically by Olaf Hougen and Kenneth Watson and derived on a more scientific basis by Irving Langmuir and Cyril Hinshelwood. When applied to fluid reactants and solid catalysts, the nonlinear equation in its simplest form becomes

$$\text{Rate} = k\,\theta_A = k\,K_A\,C_A/(1 + K_A C_A)$$

where again k is the reaction rate constant, θ_A is the fraction of active sites covered with adsorbed A, and K_A is the adsorption equilibrium constant (a large value means A is strongly adsorbed).

If the three-parameter Michaelis-Menten equation is divided by C_M, it becomes the same as the three-parameter Langmuir-Hinshelwood equation where $1/C_M = K_A$. Both these rate equations can become quite complex when more than one species is competing with the reactant(s) for the enzyme or active sites on the solid catalyst.

TRANSPORT EFFECTS

It is not unusual for the full chemical potential of a reaction to be diminished by slower transport processes (i.e., to be transport limited). In fast liquid phase enzyme reactions, mechanical stirring rates can have a strong influence on the observed kinetics that may be limited by the rate of contacting of the reactants and enzymes. Most heterogeneous catalytic reactions take place on catalysts with surface areas of 100 to 1,000 m^2/g. These high surface areas are usually generated by preparing catalysts in the form porous pellets (a few mm in diameter) containing a network of interconnecting pores that may be in the range of a few nanometers in diameter. Diffusion into these small pores by reacting molecules whose size is the same order of magnitude can be extremely slow. Assuming all pores are uniform cylinders, a typical silica-alumina cracking catalyst with surface area of 300 m^2/g and pore volume of 0.25 cm^3/g would contain pores the order of 33 Å diameter and more than 17,000,000 miles/g total length if they were all connected end to end.

For reactions taking place in a flow reactor packed with catalyst particles, each reacting molecule must negotiate a series of seven consecutive steps to accomplish its conversion. It must diffuse across the external boundary layer surrounding the pellet; diffuse inside the pores to an active site; adsorb on, react, and desorb from the active sites; then the liberated product molecules must diffuse back out of the pellet and across the boundary layer before being counted in the product stream. Any one of these sequential steps could be the bottleneck that limits the overall performance of the catalyst. Moreover, heat generated (or absorbed) during the reactions must be accounted for in order to avoid damage to the catalyst and/or hazards to personnel and the environment. This is why reaction engineering plays such an important role in optimizing catalytic processes.

MULTIFUNCTIONAL CATALYSTS

Catalytic processes frequently require more than a single chemical function, and these bifunctional or polyfunctional materials must be prepared in a way to assure effective communication among the various constituents. For example, naphtha reforming requires both an acidic function for isomerization and alkylation and a hydrogenation function for aromatization and saturation. The acidic function is often a promoted porous metal oxide (e.g., alumina) with a noble metal (e.g., platinum) deposited on its surface to provide the hydrogenation sites. To avoid separation problems, it is not unusual to attach homogeneous catalysts and even enzymes to solid surfaces for use in flow reactors. Although this technique works well in some environmental catalytic systems, such attachment sometimes modifies the catalytic specificity of the homogeneous catalyst due to the geometric constraints imposed on the molecules by the solid. With so many factors contributing to the interdisciplinary field of catalysis, it is not surprising that almost all branches of physical science, math, and engineering must be included in the successful development of a catalytic process.

IMPORTANT COMMERCIAL PROCESSES

Industrial catalytic applications comprise four major categories: chemicals manufacturing (25% of money spent on catalytic processes), environmental protection (23%), petroleum processing (26%), and polymers production (26%). In 2003 the total sales of catalysts worldwide is predicted to be $8.9 billion (not including the value of the precious metals and substrates used and includes only manufacturing fees). It has been estimated that about 20 percent of all the world's manufactured products have been touched somewhere along the line by one or more catalytic processes. The field of catalysis has obviously blossomed during the twentieth century, and without any doubt it will be a major factor in the world economy during the foreseeable future.

Burtron H. Davis
Joe W. Hightower

See also: Faraday, Michael; Thermodynamics.

BIBLIOGRAPHY

Boudart, M., and Diega-Mariadassou, G. (1984). *Kinetics of Heterogenous Catalytic Reactions*. Princeton, NJ: Princeton University Press.

Hill, C. G. (1977). *An Introduction to Chemical Engineering Kinetics and Reactor Design*. New York: John Wiley & Sons.

Prettre, M. (1963). *Catalysis and Catalysts*, tr. D. Antin. New York: Dover Publications.

Roberts, W. W. (2000). "Birth of the Catalytic Concept." *Catalysis Letters* 67:1.

Satterfield, C. G. (1970). *Mass Transfer in Heterogeneous Catalysis*. Cambridge, MA: MIT Press.

CATAPULTS

See: Elastic Energy

CHARCOAL

Charcoal is perhaps the oldest known fuel, having been found in archeological sites dating as far back as the Pleistocene era. Charcoal is a relatively smokeless and odorless fuel, and thus ideal for cooking and heating.

HISTORY

As humans entered the Bronze Age, charcoal was the only material that could simultaneously heat and reduce metallic ores. Later, the addition of an air blower made it possible to achieve temperatures high enough to soften or melt iron. During the Industrial Revolution, charcoal was largely displaced in most ironworks by coke derived from coal. However in Brazil, which lacks adequate coking coal resources, most of the charcoal produced is still used to reduce iron ore.

Charcoal was produced in pits, and later in kilns, by burning wood with air insufficient for complete combustion. The heat generated drives off the volatile materials in the wood, leaving a char that contains 60 to 90 percent carbon. In the nineteenth and early twentieth centuries, these volatiles were a major source of raw materials, chiefly acetic acid, acetone, methanol and creosote, for the burgeoning organic chemical industry. However charcoal's utility as a source of starting compounds was short-lived because petroleum-derived feedstocks proved to be cleaner and cheaper sources of these chemicals. By World War II, U.S. charcoal production declined by two-thirds, but since then, the popularity of backyard cooking raised charcoal production to an all-time high of about 1.5 million tons in 1996. As of 1999, worldwide production is roughly 30 million tons, with about half of this in Asia and India, and one-third in South America. Precise data are not available from developing countries, where production is fragmented and manufacturing information is not systematically monitored.

Charcoal typically has the following properties:	
density, as formed	0.3 to 0.5 gm/cubic cm
density, compacted	1.4 gm/cubic cm
fixed carbon	70 - 90 percent
volatile matter	5 - 25 percent by weight
ash	5 percent
heating value	12,000 BTU per pound (30 KJ per gm)

Table 1.
Properties of Charcoal

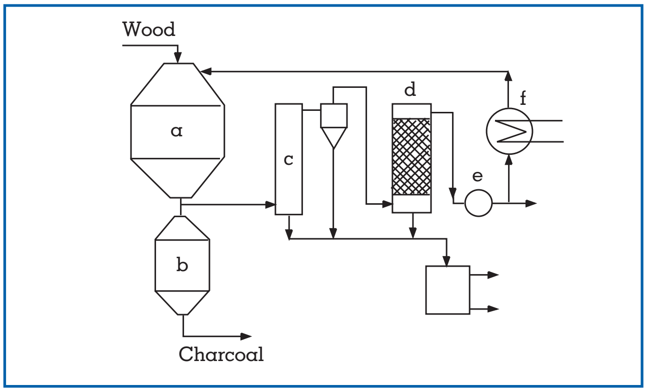


Figure 1.
Typical batch process from a DeGussa patent.

USE

Charcoal is used in electrically heated furnaces to smelt specialty metals such as ferrosilicon. It is a preferred household fuel in developing countries with adequate forest resources. In the United States 95 percent of charcoal use is for barbecuing, while in Japan and Europe charcoal use is split evenly between cooking and industrial needs.

Much has been said about the backyard barbecue as a major air pollutant. However, most cooking smoke comes from food, not from the fuel used. Charcoal contains almost no sulfur, the major pollutant from burning coal. Of the total U.S. 1996 energy consumption of 93.8 quadrillion Btus, only 3.2 percent came from burning biomass. Most of this was derived from direct combustion of fireplace wood and industrial wastes like sawdust, bark and bagasse. In developing countries charcoal may constitute up to 40 percent of energy use. This energy source may not be sustainable because of conversion of forests to farmland. In Brazil, most of the charcoal destined for metallurgical use is made from fast-growing wood species raised on plantations.

Other markets for charcoal are as a filtration medium, a horticultural soil improver, and an adsorbent. Its large surface area of hundreds of square meters per

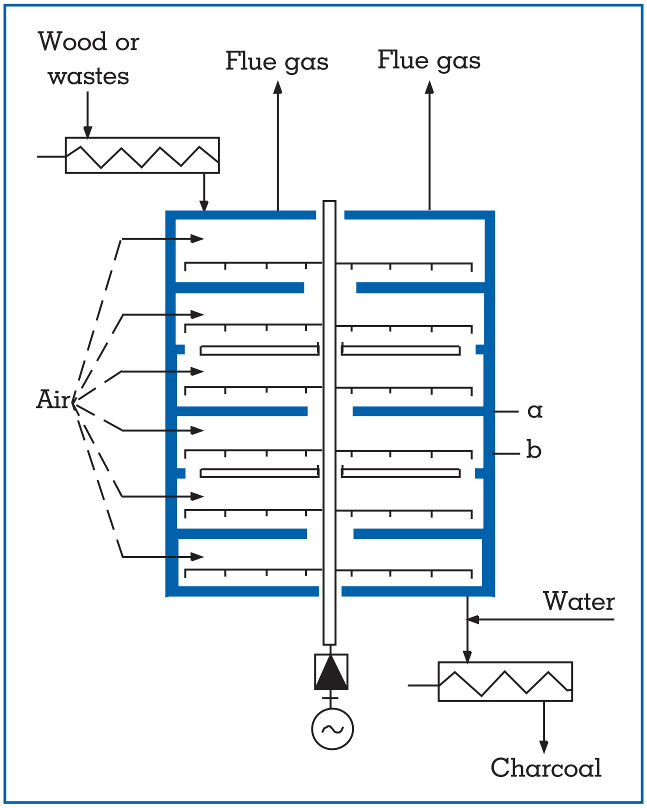


Figure 2.
Typical multiple hearth furnaces of the Hereshoff patent.

gram enables it to adsorb large organic molecules and thus improve color, taste, and/or odor in liquids and gases. Its adsorptivity can be improved by steaming in the presence of certain chemicals. The resulting "activated carbon" is used for a wide variety of applications ranging from water purification to poison-gas masks.

PROCESS

Modern charcoal retorts are charged with wood, biowaste (bark, sawdust, etc.), peat, and sometimes low-rank coals. Yield and properties (hardness, density, surface area, etc.) can vary widely so the desired end use must be considered. Charcoal from coniferous trees is soft and porous, while that from hardwoods is dense and strong. For barbecuing, charcoal is usually compressed into briquettes, with binders and additives chosen to improve handling and ease of ignition.

The manufacturing process usually involves slow heating to about 275°C in a kiln or retort. The reaction is exothermic, and about 10 percent of the heat of combustion of the original woody feed is lost. The gases and volatile liquids generated are usually burned to supply process heat. Then the wood is further carbonized by heating to about 500°C without air. Modern plants generally employ closed retorts where dried wood is heated by external means. This allows for better process control and enhanced pollution abatement. Both batch and continuous methods are used.

There are many process variations. A typical batch process from a DeGussa patent is shown in Figure 1. The reactor is 100 cubic meters and wood pieces are fed by a conveyer belt. Hot, 550°C, wood gas is fed concurrently. A carbonization zone (where the char, mostly carbon, is being formed) travels downward as the wood is pyrolyzed (heated intensely in the absense of air), and the off gases are burned to generate all of the process heat plus some of the energy needed to dry the raw wood. Energy requirements for producing 1 kg. of charcoal, including the drying operation, are 2.5 MJ of heat and 0.25 MJ of electrical energy. The process also requires 0.05 cubic meters of cooling water. A retort of this size and design will produce about 300 tons of charcoal a month.

Continuous production of charcoal is typically performed in multiple hearth furnaces, as illustrated in the Herreshoff patent shown in Figure 2. Raw material is carried by a screw conveyor to the uppermost of a series of hearths. Air is supplied countercurrently and burns some of the wood to supply process heat. As the layers of wood carbonize, they are transported to the lower (hotter) hearths by rakes. The hot charcoal product is discharged onto a conveyor belt and cooled with a water spray.

Fresh charcoal is a strong absorbent for gases, and this is an exothermic process. The heat generated can be enough to cause spontaneous ignition in some cases. Hence it is customary to age charcoal by exposure to air and thus cover the absorption sites with a layer of nitrogen gas. Larger molecules will desorb and replace smaller molecules, so the charcoal will still be effective as a decolorant or deodorizer.

Herman Bieber

See also: Explosives and Propellants.

BIBLIOGRAPHY

Bansal, R. C. (1998). *Active Carbon*. New York: M. Dekker.
DeGussa, German Patent 744,135 (1937).

Johansson, T. B.; Kelly, H.; Reddy, A. K.; and Williams, R. H., eds. (1993). *Renewable Energy: Sources for Fuels and Electricity*. Washington, DC: Island Press.
Nichols, Belgian Patent 309,763 (1968).
Proceedings of the ESCAPE/FAO/UNEP Expert Group Meeting on fuelwood and Charcoal. (1982). Energy Resources Development Series #24. New York: United Nations.
U.S. Bureau of the Census. (1998). *Statistical Abstract of the United States*. Washington, DC: U.S. Government Printing Office.

CHEMICAL ENERGY, HISTORICAL EVOLUTION OF THE USE OF

BEGINNINGS

Electric power, which is produced as a result of chemical reaction, and chemical change that is initiated by the flow of electricity embody the science and technology of electrochemistry. Its history is fairly recent, though electrochemistry might have been used in early historic times for medicinal purposes and electroplating with gold, a suggestion based on vessels containing copper tubes and central iron rods unearthed at various sites in Mesopotamia.

Public awareness of electrochemical processing dates from a meeting of the Royal Society in 1807. English society paid large admission fees to hear Humphry Davy lecture on electrically inducing chemical reactions and to witness him produce sodium and potassium from potash. In a spectacular flourish, Davy dropped the amalgamated product into water, at which moment the alkali metals reacted violently with the water to generate hydrogen, which burst into flame.

The true beginnings of electrochemistry were the experiments of Luigi Galvani and Alessandro Volta in the late eighteenth century. In 1780, Galvani, an Italian anatomist, generated what he believed to be animal electrical energy. Volta, a contemporary Italian physicist, doubted its "animal" nature and, in setting out to disprove it, invented what we now call the "galvanic" or "voltaic" cell (Figure 1), so named to honor these scientific pioneers.

ANIMAL ELECTRICAL ENERGY

Galvani in 1780 noticed that the leg muscles of a freshly killed frog, which was hung from an iron hook, twitched when a nerve was touched with a copper scalpel at the time an electric spark was being produced on a nearby frictional arc. Beginning in 1786, Galvani experimented with this system, inducing muscular contractions by physically contacting copper with iron, with one metal connected to the nerve of a frog, the other to its muscle. Galvani described his frog experiments in 1791 in a publication entitled *De veribus electricitatis in motu musculari commentarius* (Latin: *Commentary on the origins (springs) of electricity in muscular motion*), attributing the phenomenon to "animal electricity." Galvani was influenced by knowledge of electric fish and the effects of electrostatic charges in causing muscles to contract. He likely was also aware of the use of glass rods charged by rubbing with wool to electrify people, first done by Stephan Gray in 1730, with a youth suspended in air to prevent electricity from leaking to the ground.

In 1796, Volta demonstrated that Galvani's electricity was generated by contact between dissimilar metals. Using zinc and silver discs, Volta showed that brine-soaked cardboard or leather could be used in place of the frog. Though Galvani's experiments did not demonstrate "animal electricity," his intuition had validity. Emil du Bois-Reymond, using instruments more sensitive than those available to Galvani, showed in the 1840s that electrical currents do flow in nerves and muscles. Much later, Julius Bernstein in 1902 proposed that nerve cells can become electrically polarized, and David Nachmansohn in 1938 showed that current travels along the nerve to the nerve ending, releasing acetylcholine. The nerve signal, upon reaching a muscle causes a chain of reactions culminating in the breakdown of adenosine triphosphate (ATP) and the release of energy. Today's electrocardiogram (EKG), used in medical examinations, records such electrical impulses generated by the heart.

VOLTA'S PILES

The galvanic cell invented by Volta in 1800 was composed of two dissimilar metals in contact with mois-

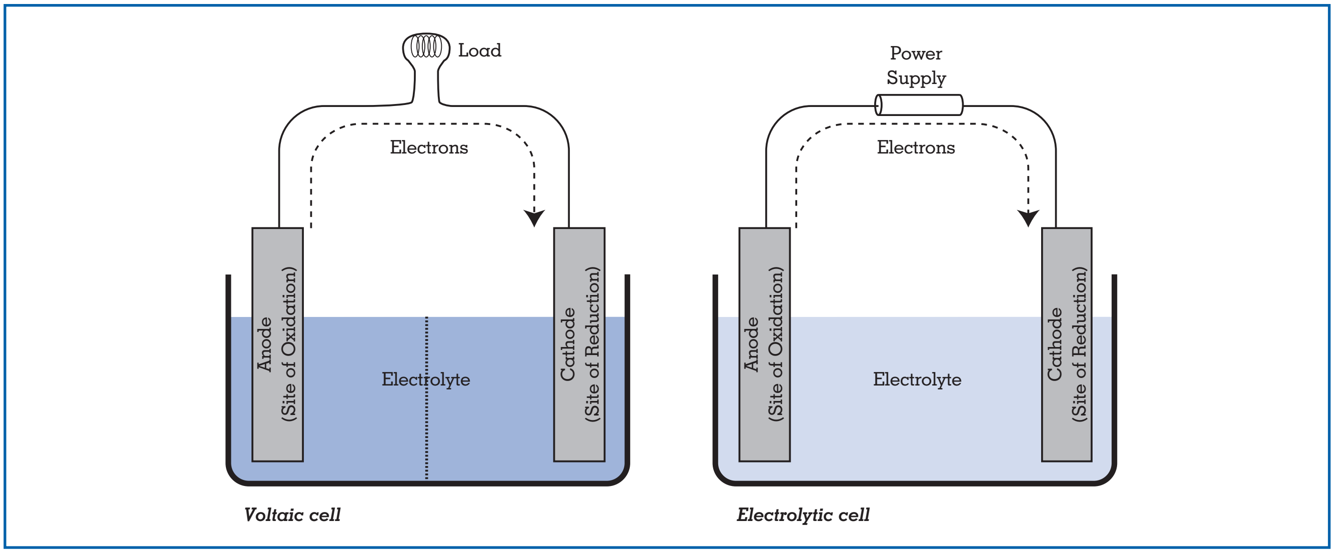


Figure 1.
Comparison of the voltaic/galvanic cell with the electrolytic cell.

tened paper or a salt solution. The first battery, two or more cells in series, consisted of a pile of alternating silver and zinc discs with separators in between soaked in electrolyte, the term now used for a liquid capable of conducting an electric current. Volta soon developed an improved version called "couronne de tasses" (French: crown of cups), which permitted drawing electric current at controlled rates. "Volta's Piles" were the only practical source of continuous electricity until they were displaced near the end of the nineteenth century by dynamos in stationary power applications.

A pile was the first source of energy capable of converting water into its separate elements in significant quantities. Using a seventeen-cell pile, William Nicholson and Anthony Carlisle in 1800 found hydrogen and oxygen evolving in the ratio of two to one. Also in 1800 in other firsts, William Cruickshank generated chlorine from sodium chloride and precipitated metals, notably copper, from solutions of their salts. Davy used the pile to produce an electric arc between two sticks of charcoal, or "poles," spaced three inches apart. Electric furnace poles now can be fifteen feet long and weigh ten tons. Davy deemed the poles positive or negative, depending upon which side of the battery they were connected, and his assistant, Michael Faraday, called the poles electrodes. William Hissinger and Jons Jakob Berzelius in 1803 showed that oxygen and chlorine evolve at the positive poles, and alkalis, metals, and hydrogen at the negative poles. From this came the later development of the chlor-alkali industry.

ELECTROCHEMICAL THEORY

Although theoretical understanding propels further discovery and application, practical application more often precedes understanding. It took most of the nineteenth century to reach an understanding of the underlying electrochemical processes in which (1) an electric current driven by an imposed electromotive force or "voltage" results in storing the potential chemical energy in reaction products (electrolysis), and (2) a generated voltage induces electrical current (as in batteries).

Electrolysis in electrolytic cells involves the migration of ions through the electrolyte and the subsequent reactions of the ions at the two electrodes. Voltaic (or galvanic) batteries do the reverse, using electrochemical processes to convert the chemical energy stored in the reactants to electrolytic energy. In both cases, electrons flow through the external circuit so that there is no net production or loss of charge in the overall process.

The first electrolytic theory was expounded in 1805 by Christian Grothuss who postulated that an electric field rotates the molecules so that their positive and negative components ("ions") face their

Figure 2.
Volta's Electrochemical Series.

opposing poles. These "ions" (not ions in the later sense) separate and move toward their opposing pole until they meet and combine with ions of the neighboring molecules. Soon after, Volta deduced that electric current in a battery would naturally flow from electrodes containing more "electropositive" metals to the less electropositive ones when the electrodes are connected to each other. Using this phenomenon, Volta established the hierarchical Electrochemical Series of Elements (Figure 2). In an analogous fashion, current in an electrolytic cell is forced to flow from the more "electropositive" metal to the less electropositive one. Berzelius in 1811 extended the Electrochemical Series to include non-metals. Volta further differentiated between groups, such as metals, carbon and some sulfides, which conduct electricity without undergoing chemical change, and salt solutions, which are decomposed by the electric current.

The convention that electric current in the external circuit (the wire connecting the electrodes) flows opposite to the direction of electron flow, that is, from negative to positive, proved to be an unfortunate source of confusion. The selection originated in Benjamin Franklin's suggestion in 1747 that rubbing a glass rod with a dry hand causes an electric fluid to pass between hand and rod and results in an excess or "plus" of electric fluid in the rod, and a corresponding deficit or "negative" charge on the hand.

Between 1831 and 1834, using a voltaic pile, Faraday discovered the inductive effects of the electromagnet — the basic principle underlying the operation of electrical machinery — and experimentally quantified the direct proportionality between electrical current and the rates of electrochemical reactions (Faraday's Law). Working with voltaic piles, a great battery of eight Leyden jars charged with static electricity, and using data obtained on other devices by other investigators, Faraday established that electricity was universal and not dependent on its source. During the course of this intense activity, he also invented both the direct and alternating current dynamos.

Seemingly unrelated work by Wilhelm Ostwald in 1884 on a physical phenomenon now called osmotic pressure, provided an important link in understanding electrochemical behavior. Osmotic pressure is generated when a porous but impermeable membrane is interposed between solutions containing differing concentrations of a dissolved substance. The membrane prevents the dissolved substance from diffusing from the more concentrated into the less concentrated zone, thus giving rise to the difference in pressure. Shortly after, as published between 1885 and 1888, Jacobus Henricus van't Hoff noted that osmotic pressure was proportional to concentration, but most important, the pressures were often multiples of predictions. Van't Hoff attributed this observation to the solute dissociating into several ions. Ostwald recognized that this dissociation into ions was the electrolytic process as described by Svante

Arrhenius in 1884. The three (Ostwald in Germany, van't Hoff in Latvia, and Arrhenius in Sweden) were dubbed "Die Ioner" (German: the ionists) by contemporaries, for their close, friendly, and productive collaboration.

It fell to Hermann Nernst, working alongside Arrhenius in Oswald's laboratory, to develop between 1887 and 1889 the thermodynamic relationships between a cell's open circuit potential (the voltage with no electricity flowing) and the amount of heat released or consumed by the net chemical reaction occurring at the two electrodes in a cell. This was conclusively demonstrated by Fritz Haber in 1905. Haber also developed the glass electrode, a premier measuring tool of analytical chemistry. This invention led to its further use by Walter Wright and Lucius Elder in measuring pH in 1928, and to Arnold Beckman commercializing a sturdy and stable pH meter in 1935. Beckman's pH meter is one of the most useful instruments in all of science.

GOLDEN AGE OF ELECTROCHEMISTRY

The nineteenth century has been termed the golden age of electrochemistry. At first, only limited amounts of energy, sufficient for experimental purposes, were available from the cells patterned after Volta's original pile. In 1836, John Daniell introduced a more powerful battery containing zinc and copper electrodes that was capable of maintaining a reliable steady current for long periods of time. In its wake came electroplating, developed by George Richards Elkington (1836), followed by electrotyping by Moritz Hermann Von Jacobi for reproducing articles such as engravings, medals, and printer type setups (1839). Wright's basis, little changed today except in detail, for selecting solutions to plate anything from eating utensils, art objects, tiny computer electrical contacts to expensive automobile bumpers, came in 1840. Wright also developed cadmium plating (1849), rediscovered after World War I when the metal became available in commercial quantities. Robert Wilhelm von Bunsen produced magnesium in 1852 and lithium in 1855 from their salts, but their exploitation did not begin until well into the twentieth century. Friedrich Wohler made calcium carbide in 1862, but thirty years elapsed before Thomas Willson's accidental rediscovery while attempting to produce aluminum. However, Willson went a step further, reacting the carbide with water to produce acetylene, and launching acetylene's use for heat, light and power, and more importantly for metallurgical applications such as producing alloy steels in the 1890s.

Battery System	*Net Electrochemical Reaction*
Rechargeable Cells	
Plante or Lead-Acid Cell	$PbO_2 + Pb + H_2SO_4 = 2PbSO_4$
Nickel Cadmium	$2NiOOH + Cd + 2H_2O = Ni(OH)_2 + Cd(OH)_2$
Nickel Hydrogen	$2NiOOH + H_2 = 2Ni(OH)_2$
Non-Rechargeable Cells	
Leclanche or dry cell	$Zn + 2MnO_2 = ZnO + Mn_2O_3$
Alkaline Cell	$Zn + 2MnO_2 = ZnO + Mn_2O_3$
Silver-Zinc	$Ag_2O_2 + 2Zn + 2H_2O = 2Ag + 2Zn(OH)_2$
Reuben Cell	$HgO + Zn + H_2O = Hg + Zn(OH)_2$
Zinc-Air	$Zn + O_2 + 2H_2O = 2Zn\ (OH)_2$
Fuel Cell	$2H_2 + O_2 = 2H_2O$
Lithium Iodine	$2Li + I_2 = 2LiI$
Lithium-Sulfur Dioxide	$2Li + 2SO_2 = Li_2S_2O_4$
Lithium-Thionyl Chloride	$4Li + 2SOCl_2 = 4LiCl + S + SO_2$
Lithium-Manganese Dioxide	$Li + Mn(IV)O_2 = LiMn(IV)O_2$
Lithium-Carbon Monofluoride	$nLi + (CF)_\alpha = nLiF + nC$

Table 1.
Examples of Commercially Developed Batteries

A great step forward came in 1859 Gaston Plante learned to store electricity in his lead storage cell batteries. Plante had been investigating the effect of current drain on different metals when he noticed the surprising behavior of the lead plates in dilute sulfuric acid. Plante discovered that he could increase the amount of electricity stored in the cell by allowing the cell to rest after it was discharged. In what we now call "formation," the lead peroxide coating on the positive lead electrode is converted to the sulfate during the rest period. Reversing cell polarities from time to time, and repeating these charging and discharging cycles results in a buildup of electrical storage capacity in the cell. Such a rechargeable battery is called a secondary battery, in contrast to a primary battery, one that cannot be recharged. The lead storage cell battery remains the dominant secondary battery, with 40 percent of the world's lead production used in this application.

The lead storage cell battery was ahead of its time. Battery recharging was tedious and difficult, and electrolytic processing was not economical until electricity became commercially available in the late 1800s. Plante had to use a series of Bunsen mercury-zinc primary cells to recharge his battery. The Bunsen cell, invented in 1850, in which a platinum electrode was replaced with carbon actually was a modification of a cell by William Grove. Grove is better known for his invention of the hydrogen-oxygen fuel cell in 1839, which was proven practical and reliable as part of the Gemini and Apollo space programs in the 1960s.

Another major step was the dry cell, first produced by C. Gassner in 1888. The dry cell was an outgrowth of Georges Leclanche's 1868 cell in which a carbon rod was contained in a porous cup filled with a crushed, then compressed, carbon and manganese dioxide mixture. In its later form, a zinc electrode was suspended in the mixture, then molded together into a cylindrical form using a binder. Billions of such batteries, but improved over the years in reliability, leak-proofing, endurance, and shelf life, are used worldwide to power flashlights, toys, and a host of portable electronic devices.

By the end of the nineteenth century, most battery types had been identified, and several hundred had experienced some commercial application. Batteries became so rugged and reliable that Robinson in 1871 used them in train signaling, and Thomas Alva Edison in powering his quadruplex telegraph (simultaneous messages on one wire) in 1874. Alexander Graham Bell's telephone exchanges, starting in New Haven, have operated on batteries since 1878.

Nickel-cadmium and nickel-iron are prime examples of rechargeable (secondary) batteries, invented in 1901 by Waldemar Jungner and Edison, respectively. In the 1920s and 1930s, such batteries powered radios prior to rural electrification. The sealed nickel-cadmium battery is now the most widely used battery in consumer products.

RISE OF THE ELECTRICAL INDUSTRY

The era of large electrical power generation began in 1875, when Cornell University was lit with electric arc lights using Gramme dynamos, continuous high-voltage electric current generators invented in 1869 by Zenobe-Theophile Gramme. This was followed in 1879 by the introduction of the carbon-filament incandescent lamp by Thomas Edison, and the carbon-arc lamp by Charles M. Brush of Cleveland. Neither application was feasible using batteries alone to supply the power. However, electrochemical technology was instrumental in furthering the growth of the electric power industry. Lead storage cell batteries remained a secondary source of electric power, to accommodate sudden surges and provide incremental power during periods of high demand.

Electrochemical processing proved essential to efficient power transmission. J. B. Elkington, building on family electroplating technology, in 1869 used electrowinning (recovery of metals from solutions by electrolysis) to remove impurities from copper and consequently double the electrical conductivity of commercial copper. Other important metals so refined include silver, gold, nickel, antimony and bismuth. In electrowinning, an imposed electric field causes the targeted metal, in this case copper, to be dissolved from impure slabs, then to be selectively deposited on thin sheets. Sludge, periodically removed from the bath, is stripped of its valuable byproducts of gold, silver and tin.

In 1885, Charles Martin Hall invented his aluminum process and Hamilton Young Castner in 1890 developed the mercury-type alkali-chlorine cell, which produced caustic (sodium hydroxide) in its purest form. Edward G. Acheson in 1891, while attempting to make diamonds in an electric furnace, produced silicon carbide, the first synthetic abrasive, second to diamond in hardness. Four years later, Jacobs melted aluminum oxide to make a superior "emery cloth." Within two decades, these two abrasives had displaced most natural cutting materials, including naturally occurring mixtures of aluminum and iron oxides.

Great achievements came with the introduction of cheap hydroelectric power, starting with a 100,000 hp installation on the Niagara River in 1894. The first contracts were not for lighting, but for the electrolysis of alumina with the Pittsburgh Reduction Company (now Alcoa) and the manufacture of silicon carbide by the Carborundum Co. In 1895, Moorehead produced 50 percent ferrosilicon in an electric furnace, leading to ferroalloys with qualities not possible in blast furnaces. In 1898, the Mathieson Alkali Works and Dow Chemical began the manufacture of bleaching powder from electrolytic chlorine, exemplifying the accelerating opportunity for chemical manufacture in the growing textile, paper, fertilizer, soap, and glass industries.

FUEL CELLS

The concept of the fuel cell, that is, a cell in which inert electrodes immersed in an electrolyte could be intimately contacted with a reacting fuel (e.g., hydrogen) and oxidant (e.g., air) and so generate an electric current, was demonstrated in 1839 by Grove and intensively studied by him during the next decade.

Grove recognized that electrodes above the surface of an electrolyte, (e.g., sulfuric acid) would be wetted by capillary action and so allow the platinum electrodes to catalyze the electrochemical reactions of a fuel and oxidant such as hydrogen and oxygen.

The poor efficiencies of coal-fired power plants in 1896 (2.6 percent on average compared with over forty percent one hundred years later) prompted W. W. Jacques to invent the high temperature (500°C to 600°C [900°F to 1100°F]) fuel cell, and then build a 100-cell battery to produce electricity from coal combustion. The battery operated intermittently for six months, but with diminishing performance, the carbon dioxide generated and present in the air reacted with and consumed its molten potassium hydroxide electrolyte. In 1910, E. Bauer substituted molten salts (e.g., carbonates, silicates, and borates) and used molten silver as the oxygen electrode. Numerous molten salt battery systems have since evolved to handle peak loads in electric power plants, and for electric vehicle propulsion. Of particular note is the sodium and nickel chloride couple in a molten chloroaluminate salt electrolyte for electric vehicle propulsion. One special feature is the use of a semi-permeable aluminum oxide ceramic separator to prevent lithium ions from diffusing to the sodium electrode, but still allow the opposing flow of sodium ions.

It 1932, ninety years after Grove, Francis Bacon developed "gas diffusion" electrodes. These electrodes were unique in having the reacting gases fed to one side of each electrode while the other side was in contact with the alkaline electrolyte. A further improvement was E. Justi's "double skeleton catalyst electrodes" in 1952. Justi used finely divided nickel and silver catalysts incorporated onto nickel skeletons as hydrogen and oxygen electrodes, respectively. In 1953 Karl Kordesch's further improvement in oxygen diffusion electrodes involved inclusion of a water-repellent, such as Teflon, in carbon-based electrodes. Leonard W. Niedrach in 1965 also made improvements using an ion exchange membrane fuel cell. Since the 1960s, a number of fuel cell batteries have been developed for use as remote power sources, fueled with methanol, hydrazine or hydrogen obtained subsequent to partially oxidizing a hydrocarbon mixture (e.g., natural gas) or an alcohol (e.g., methanol). Such indirect hydrogen fuel cell batteries are finding application, by General Electric for one, in providing residential power where transmission lines have limited access and are costly.

VEHICULAR POWER SOURCES

The crowning use of electrochemical energy may come with its broad application to vehicular propulsion. Such engines offer the promise of both superior efficiency and exceptionally clean gaseous emissions. Numerous possibilities have received close scrutiny during the past four decades. Zinc-air, zinc-bromine, and lithium-sulfur batteries were subjects of substantial investigation in the 1960s, then sodium-sulfur, zinc-nickel and nickel-iron in the 1970s and 1980s. Contenders added in the 1990s included sodium-nickel chloride, lithium-ion and nickel-metal hydride batteries.

At least eight electric vehicles using lead-acid, nickel-metal hydride, or sodium nickel chloride batteries were commercially available in 1998, but with limited acceptance because power and capacity were insufficient and at too high a price. Consequently, hybrid engines, that is, internal combustion or diesel engines in combination with batteries or fuel cells, may offer greater prospects near-term. The first mass-produced vehicle with a hybrid engine was the Toyota Prius (Latin: pioneering), a subcompact on sale in Japan beginning in 1997. Larger hybrid-powered vehicles are being tested by Toyota and other manufacturers. Twenty hybrid electric buses and trucks were under study in 1998, typically targeted for 25 percent lower diesel fuel consumption and 30 to 50 percent reduction in emissions.

Prius uses an internal combustion engine operating at constant speed to drive a generator that powers a motor connected to the drive train. Excess energy is used to charge a nickel-metal hydride battery, which is intermittently used to handle the power demands of steep climbs and rapid accelerations. When the vehicle slows, the motor behaves as a generator, converting kinetic energy into electricity.

EMERGING OPPORTUNITY

Research, development and entrepreneurial activities are pervasive, fueled by ever-growing demands for lighter, smaller and more powerful batteries with even more reliability and extended shelf life for industrial, automotive, military, business, medical, and home uses. Mobile communications, computing, data storage and access, instrumentation, intelligent sensors, controls for individual devices and systems, power tools, home appliances, systems backup, and remote standby applications provide countless opportunities for improved battery technology.

Materials science is contributing to better fabrication and use of extremely reactive metals. In the 1960s emerged highly reactive and powerful lithium batteries using carbon, titanium sulfide, and transition metal oxide structures to tie up (intercalcate) the reactive lithium. Notable among these is the lithium-ion battery, widely used in cellular phones, video cameras and laptop computer display panels. In this battery, both electrodes are lithium intercalcating electrodes, with lithium intercalcated into carbon and either nickel or cobalt oxide, respectively. An offshoot of the lithium-ion development is the rechargeable lithium-polymer battery in which complexes of lithium salts with certain polymers (e.g., polyethylene) can serve as electrode separators, but still allow transfer of lithium ions between electrodes.

Proliferation of new materials is also important to improving mature batteries. Lead-coated fiberglass is used to reduce the weight of lead acid batteries, and microporous polyethylene is added as separators to contain electrolyte and so prevent its leakage. The nickel-metal-hydride battery is essentially the nickel-cadmium battery, but with the cadmium replaced with a metal alloy. Hydrogen generated during the charging cycle is stored as a hydride and then released and reacted during the discharge cycle. The metal alloys consist largely of nickel and readily available rare earth, modified with cobalt, manganese and aluminum, to improve corrosion resistance. Commercial introduction of new technology comes at a modest pace in this field, and more chapters of electrochemical history are yet to be written.

Barry L. Tarmy

See also: Faraday, Michael; Nernst, Walther Hermann; Volta, Alessandro.

BIBLIOGRAPHY

Bowden, M. E. (1997). *Chemistry Is Electric!* Philadelphia: Chemical Heritage Foundation.

Design News. (10/5/98). “Out of Juice!” Annual Auto Issue: Battery Technology, pp. 98–108. <http://www.design-news.com>.

Graham, R. W., ed. (1981). *Primary Electrochemical Cell Technology: Advances Since 1977.* Park Ridge, NJ: Noyes Data Corporation.

Liebhafsky, H. A., and Cairns, E. J. (1968). *Fuel Cells and Fuel Batteries.* New York: John Wiley and Sons.

Natishan, P. M., ed. (1997). *What is Electrochemistry?*, 4th ed. Pennington, NJ: The Electrochemical Society.

Mantell, C. L. (1950). *Industrial Electrochemistry.* New York: McGraw-Hill.

Schallenberg, R. H. (1982). *Bottled Energy: Electrical Engineering and the Evolution of Chemical Energy Storage.* Philadelphia, PA: American Philosophical Society.

Stock, J. T., and Orna, M. V., eds. (1989). *Electrochemistry, Past and Present.* Washington, DC: American Chemical Society.

Toyota Press Information. (1997). *Toyota Hybrid System.* Tokyo: Toyota Motor Corporation, Planning Group, International Public Affairs Division.

CHERNOBYL

See: Nuclear Energy; Nuclear Energy, Historical Evolution of the Use of

CIRCUIT BREAKERS AND FUSES

See: Electricity

CLAUSIUS, RUDOLF JULIUS EMANUEL

(1822–1888)

Physicist Rudolf Julius Emanuel Clausius was born January 2, 1822, in Koeslin, Pomerania, Prussian province (now Koszalin, Poland). He was the sixth son of eighteen children born to teacher and a Protestant minister. Clausius attended gymnasium (secondary school) in Stettin and from 1840 the University in Berlin, where he studied mathematics and physics mainly as a student of Gustav Magnus. In 1851 he joined The Physics Society, which was formed in 1845 by Magnus's students. In tandem with his university studies, he taught in a Berlin gymnasium to sponsor the education of his younger sisters and brothers. In 1848 he earned a doctorate degree, and in 1850 he became lecturer at Berlin University while simultaneously teaching at the Artillery School of the Prussian army. Clausius left Berlin in 1855 and moved to Zurich, where he became a professor of physics at Polytechnicum, then later at the University of Zurich. Between 1867 and 1869 he resided in Wuerzburg, and finally he settled in Bonn, where he spent the last two decades of his life as a professor at the university. Clausius married in 1859. His wife died tragically in 1875, while giving birth to their sixth child. Again the sense of duty towards his family was a foremost concern in Clausius's life, and he sacrificed his scientific pursuits to supervise the education of his children. A serious knee injury that he had acquired in the Franco-Prussian war further limited his scientific activity, causing him discomfort until he died on August 24, 1888, in Bonn.

From the beginning of his studies Clausius's interest was mathematical physics. He exhibited keen mathematical thinking, although he was not particularly communicative. His textbooks were popular. The works of Julius Mayer, James Joule, and Hermann Helmholtz were the basis of the knowledge of heat nature in the 1840s and the starting point for Clausius's research. In 1850 Clausius gave a lecture in Berlin in which he formulated the basis for heat theory. He stressed that heat is not a substance, but consists of a motion of the least parts of bodies. Furthermore he noted that heat is not conserved. Clausius formulated the equivalence between heat and work and the first law of thermodynamics as follows: mechanical work may be transformed into heat, and conversely heat into work, the magnitude of the one being always proportional to that of the other. During the next fifteen years Clausius concentrated his research on thermodynamics. He gave several formulations of the second law and finally defined and named the concept of entropy. He started his theoretical investigations with the consideration of the idealized process of the so-called Carnot cycle on the basis of the ideal gas. For this purpose, one assumes that the gas is enclosed in an expansible wrap through which no heat exchange is possible. The Carnot cycle consists of four steps:

Rudolf Julius Emanuel Clausius. (Corbis Corporation)

1. The gas of initial temperature T1 in a volume V expands (V rises) isothermally (T1 stays constant). To keep temperature T1, the gas comes in contact with body B1 of constant temperature T1 and receives the quantity of heat Q1.
2. One takes the body B1 away. The gas continues to expand and its temperature decreases to T2.
3. The gas at temperature T2 compresses (V decreases) isothermally (T2 stays constant). The superfluous heat Q2 is transformed to the body B2 of constant temperature T2.
4. One takes the body B2 away. Through compression the gas is brought to the initial state raising the temperature to T1. The cycle is complete. The gas is in the same state as at the beginning of the cycle; it is therefore a reversible process.

During the expansion the gas produces the work, in contrast to the compression where the exterior work is needed. One needs part of the heat to perform work, so Q1 is larger than Q2. Defining the withdrawn heat as negative heat, every reversible cycle system satisfies the relation Q1/T1+Q2/T2=0. In 1854 Clausius further considered processes where the receiving of positive and negative quantities of heat take place at more than two temperatures. Next he decomposed every reversible cyclic process into an infinite number of Carnot processes. In this way, instead of Q/T, d Q/T appears as a differential of a new physical quantity, which depends only on the state parameters of the gas (in general of the body), but it does not depend on the closed path of integration needed to run through the complete cycle. Denoting the new quantity S, one arrives at the relation d Q =T d S.

In 1865, in a talk given in Zurich, Clausius introduced the notion of entropy for the function S, the Greek word for transformation. He meant that the important functions in science ought to be named with the help of the classical languages. After passing through the described reversible cycle process, entropy remains constant. A larger and more common group of thermodynamic processes are irreversible. According to Clausius, the entropy is the measure of the irreversibility of a process and increases constantly to the maximum. In the same lecture he brought the two laws of mechanical heat theory into the neat form: The energy of the universe is constant. The entropy of the universe increases. The definitive physical interpretation of entropy was first made possible in the framework of statistical mechanics by the Austrian physicist Ludwig Boltzmann in 1877.

Clausius was among the first researchers to look for a foundation of thermodynamics in the realm of kinetic theory. In a paper from 1857 he described the qualitative distinction of characteristic features of phases of matter (solid state, liquid, gas) from a microscopic, that is molecular, point of view and derived the formula connecting the pressure of an ideal (i.e. non-interacting) gas to the mean kinetic energy of the molecular constituents of the gas. An 1858 paper was inspired by the critique of Christoph Buys-Ballot, who noted the apparent conflict between the claim of kinetic theory that the molecular constituents of gases move with velocities of the order of some hundred meters per second at room temperature and the everyday experience of the comparatively slow evolution of diffusive processes. (It takes some minutes to smell the effect of overcooked milk outside the kitchen.) Clausius identified the multi-scattering of molecules among each other as the main reason for the slowing down of diffusive processes. He introduced the notion of the mean free path length, which is the average length a molecule flies between two consecutive collisions with some of its neighbours. This notion has become an integral part of modern transport theory.

In an obituary talk given at the Physical Society of Berlin in 1889, Hermann Helmholtz stressed that Clausius's strict formulation of the mechanical heat theory is one of the most surprising and interesting achievements of the old and new physics, because of the absolute generality independent of the nature of the physical body and since it establishes new, unforeseen relations between different branches of physics.

Barbara Flume-Gorczyca

BIBLIOGRAPHY

Clausius, R. (1850). Poggendorff's Annalen 79:368–397, 500–524.
Clausius, R. (1854). Poggendorff's Annalen 93:481–506.
Clausius, R. (1857). Poggendorff's Annalen 100:353–384.
Clausius, R. (1858) Poggendorff's Annalen 105:239–256.
Clausius, R. (1865). Poggendorff's Annalen 125:353–400.
Helmholtz, H. von. (1889). Verhandlungen der physikalischen Gesellschaft zu Berlin: 1–6.

CLEAN FUELS

See: Alternative Fuels

CLIMATIC EFFECTS

Earth's climate has fluctuated over the course of several billion years. During that span, species have arisen, gone extinct, and been supplanted by others. Humankind and our immediate ancestral species have survived the sometimes drastic fluctuations of the past several million years of our evolving climate quite handily. Although human beings and their ancestors have always changed the environment in which they have lived, human actions since the advent of the Industrial Revolution may be producing impacts that are not only local but also global in

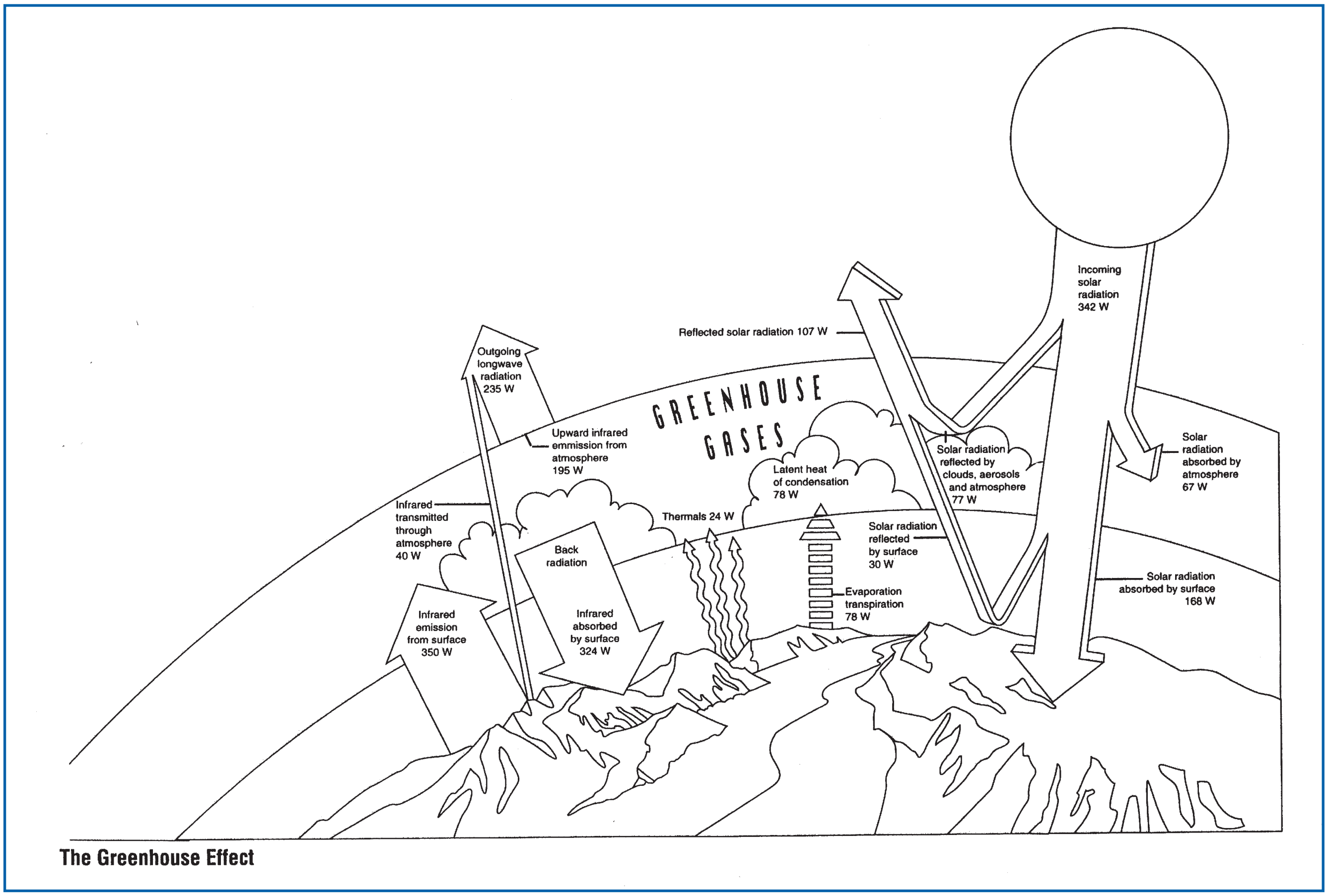


Illustration of the greenhouse effect. (Public domain)

scope. Ever-growing fossil fuel use and other activities associated with industrialization have altered the concentrations of several atmospheric gases, including carbon dioxide. Theoretically, these alterations can cause alterations in the overall heat retention of Earth's atmosphere. Scientists around the world are engaged in studies to determine whether these changes in the atmosphere pose a risk of significant, negative environmental and human impacts.

CLIMATE CHANGE THEORY

Climate change means different things to different people. To some, climate change refers to physical changes in climate and little more. To others, climate change is a theory of how certain gases in the atmosphere influence the climate. Still others focus on the human aspect, and consider climate change only in regard to the way human activity influences the climate. Climate change theory is quite complex. Unlike Albert Einstein's $E = mc^2$, for example, the theory of human-driven climate change is not a singular theory but consists of several interlocking theories, some better defined than others.

At the heart of climate change theory is a more humble theory called the greenhouse effect. This theory, first quantified by mathematician Joseph Fourier in 1824, has been repeatedly validated by laboratory experiments and by millions of greenhouse owners. The greenhouse effect is simple. When sunlight reaches the surface of Earth, some of its energy is absorbed by the surface (or objects on the surface), some is reflected back toward space unchanged, and some is first absorbed by objects or the surface of Earth and then reemitted in the form of heat. Over a bare patch of ground, this dynamic would cause no net increase in the temperature over time because the heat absorbed during the day would be reradiated toward space overnight.

But if there is a greenhouse on that patch of ground, things are different. The sunlight enters as usual, and some of it is reflected back out as usual, but part of the incoming solar energy that was held by the surface and reemitted as heat is prevented from passing back out by the glass, and the greenhouse warms up a bit. If there are water-beating plants or materials inside the greenhouse, water vapor concentration will increase as evaporation increases. Since water vapor also can trap heat, the greenhouse warms still further. Eventually the air reaches maximum humidity, and the system reaches temperature equilibrium.

Scientists have known for a long time that the greenhouse effect applies not only to greenhouses but to Earth as a whole, with certain gases (called greenhouse gases) playing the role of the glass in the example above. The primary greenhouse gases are water vapor (H_2O), carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O), ozone, and chlorofluorocarbons (CFC). When scaled up to the entire planet, the natural greenhouse effect produces a pronounced warming effect. This warming is a natural aspect of Earth's environment, crucial for the maintenance of life on Earth. In fact, without Earth's natural greenhouse effect, and the warming that goes with it, Earth would be a much colder planet, inhospitable to life as we know it. The greenhouse effect has also been seen to maintain warmer planetary atmospheres on Mars and Venus. While the effect is mild on Mars, the high carbon dioxide level on Venus (78,000 times that of Earth) keeps the atmosphere about 500°C (900°F) higher than it would otherwise be.

Against the backdrop of an Earth warmed by its own greenhouse effect, other forces operate that can increase or decrease the retention of heat by the atmosphere. Some of these forces are of human origin, some are produced by nature, and some are produced by mutual feedback reactions.

Which forces have dominated in the recently observed warming of the climate is still an open question. The last landmark report of the United Nations' Intergovernmental Panel on Climate Change (IPPC) concludes:

> Although these global mean results suggest that there is some anthropogenic [of human origin] component in the observed temperature record, they cannot be considered as compelling evidence of a clear cause-and-effect link between anthropogenic forcing and changes in the Earth's surface temperature. It is difficult to achieve attribution of all or part of a climate change to a specific cause or causes using global mean changes only. The difficulties arise due to uncertainties in natural internal variability and in the histories and magnitudes of natural and human-induced climate forcings, so that many possible forcing combinations could yield the same curve of observed global mean temperature change [IPCC, 1995, p. 411].

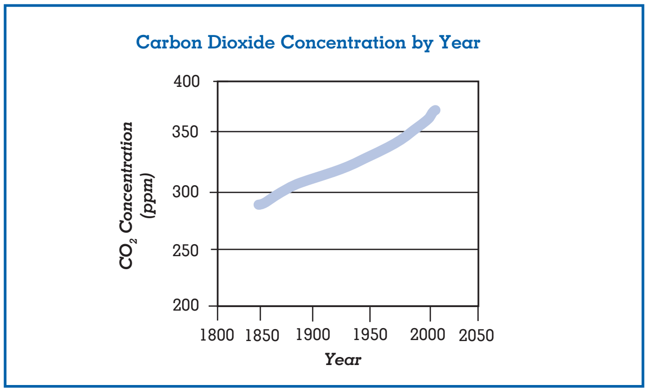


Figure 1.

SOURCE: <http://www.giss.nasa.gov/data/si99/ghgases>.

Nonetheless, the IPCC's chapter on the attribution of climate change concludes:

> The body of statistical evidence. . . when examined in the context of our physical understanding of the climate system, now points towards a discernible human influence on global climate. Our ability to quantify the magnitude of this effect is currently limited by uncertainties in key factors, including the magnitude and patterns of longer-term natural variability and the time-evolving patterns of forcing by (and response to) greenhouse gases and aerosols [IPCC, 1995, p. 439].

WARMING AND COOLING FORCES

Human activities (as well as nonhuman biological, chemical, or geological processes) release a variety of chemicals into the atmosphere, some of which, according to climate change theory, could exert a warming or a cooling effect on Earth's climate. In climate change literature, these are referred to as "climate forcings." Some of these forcings are actually secondhand responses to changes in the climate caused by others. In such a case, a forcing might be referred to as a feedback.

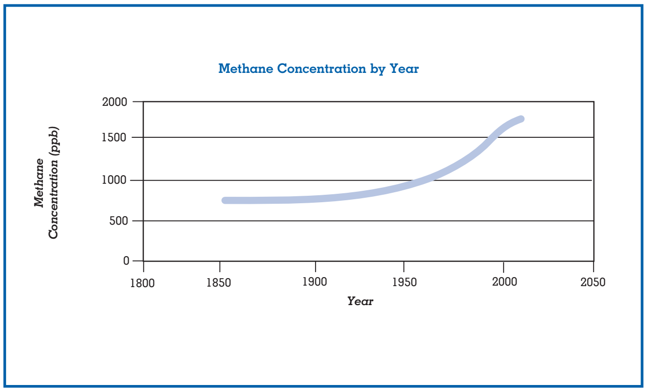


Figure 2.

SOURCE: <http://www.giss.nasa.gov/data/si99/ghgases>.

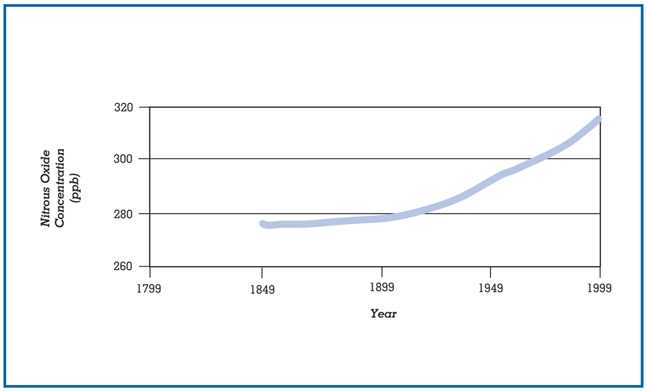


Figure 3.

SOURCE: <http://www.giss.nasa.gov/data/si99/ghgases>.

Carbon Dioxide

Carbon dioxide, considered a warming gas, comprises about 0.036 percent of the atmosphere by volume. As Figure 1 shows, carbon dioxide levels have increased as a component of the atmosphere by nearly 30 percent from the late eighteenth century to the end of the twentieth century, when the level was close to 365 parts per million by volume. Prior to the period of industrialization, carbon dioxide levels were largely stable, at about 280 parts per million, though fluctuations as low as 200 parts per million or as high as 300 parts per million have been observed through analysis of air bubbles trapped in arctic ice cores.

Carbon dioxide is released into the environment by human activities such as fuel burning, cement production, and land use.

Since highly accurate, direct measurement of carbon dioxide levels began only in the late 1950s, most of our understanding of carbon dioxide's historical patterns of fluctuation come from indirect measurements, such as the analysis of gas bubbles trapped in glaciers and polar ice caps. Though indirect measurements carry greater uncertainty than direct measurements of carbon dioxide levels, indirect measurements have contributed to our understanding of Earth's carbon cycle. Still, significant gaps in our understanding remain, specifically involving questions of time lag, the impact of world vegetation on atmospheric carbon dioxide levels, other processes that might lock carbon dioxide away from the atmosphere, and the role of carbon dioxide as a causal agent of climate change.

Methane

Methane is a greenhouse gas up to fifty-six times as powerful a warming agent as carbon dioxide, depending on the time scale one considers. In a twenty year time frame, for example, a given quantity of methane molecules would have fifty-six times the impact of the same quantity of carbon dioxide molecules, but since carbon dioxide has a longer lifespan, this ratio declines over time. As an atmospheric component, methane is considered a trace gas, comprising approximately 0.00017 percent of the atmosphere by volume. As Figure 2 shows, methane levels in the atmosphere increased nearly 120 percent from the middle of the nineteenth century to the end of the twentieth century, when the levels were the highest ever recorded, though the pattern of methane emissions has been highly irregular and actually showed downturns toward the end of the twentieth century for reasons that are not clear. Studies of methane concentrations in the distant past show that methane concentrations have fluctuated significantly, from as few as 400 parts per billion to as many as 700 parts per billion, due to changes in wetlands and other natural sources of methane. Today methane comes from a variety of sources, some of human origin, others of nonhuman origin.

Nitrous Oxide

Nitrous oxide is a long-lived warming gas with a relative warming strength 170 to 310 times that of carbon dioxide, depending on the time scale one considers. Nitrous oxide, like methane, is considered a trace gas in the atmosphere, but at considerably lower

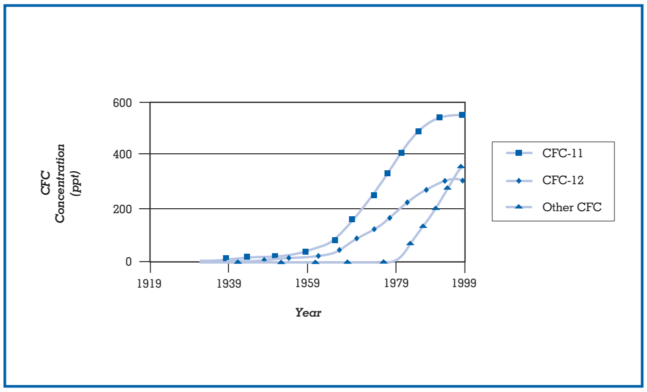


Figure 4.

SOURCE: <http://www.giss.nasa.gov/data/si99/ghgases>.

levels, about 0.00003 percent of the atmosphere by volume. As Figure 3 shows, nitrous oxide concentrations have increased significantly from the middle of the nineteenth century to the end of the twentieth century. Prior to the industrial period, nitrous oxide concentrations fluctuated at an average of 270 parts per billion by volume, though fluctuations as low as 200 parts per billion by volume have been measured from the distant past.

Nitrogen oxides (of which nitrous oxide is the major component) come from a variety of sources, only some of which are of human origin.

Chlorofluorocarbons (CFCs)

Chlorofluorocarbons (CFCs) are man-made compounds used as cooling agents and propellants in a broad range of applications. There are many different species of CFCs, some of which have been banned. CFCs are very powerful warming gases. Some species are more than ten thousand times more capable of trapping heat than is CO_2. Of course, CFCs also are found at much lower concentrations than the other greenhouse gases. Whereas carbon dioxide is measured in parts per million, and methane in parts per billion, CFCs are measured in parts per trillion. Figure 4 shows the concentration of the three major CFCs from 1929 to 1999.

Some CFCs also can break down ozone. Ozone-depleting CFCs can exert either warming or cooling effects, depending on where they are found. In the lower atmosphere, ozone-depleting CFCs exert a warming effect through the absorption of heat reradiated from Earth's surface. In the upper atmosphere, ozone destruction exerts a cooling effect by destroying some of the high-altitude ozone that can either warm or cool the surface in different circumstances. On a net basis, our current understanding is that the ozone-depleting CFCs (banned by the Montreal Protocol in 1987) exerted a cooling effect. Replacement chemicals for the ozone-depleting CFCs are considered pure warming gases, but with a considerably lower warming potential than the chemicals they replaced. Because of the complexities of ozone chemistry in the atmosphere and uncertainties regarding the warming or cooling potential of remaining ozone-depleting CFCs and replacement compounds, the ultimate impact of CFCs on climate change is highly uncertain.

Aerosols

Aerosols are not gases in the strictest sense of the word, but are actually liquid or solid particles small enough to stay suspended in air. Both human-made and natural processes generate aerosols. Some aerosol particles tend to reflect light or cause clouds to brighten, exerting a cooling effect on the atmosphere. Other aerosol particles tend to absorb light and can exert a warming effect. Most human-made aerosols seem to exert a cooling effect on the climate. On a global basis, some have estimated that this cooling effect offsets about 20 percent of the predicted warming from the combined greenhouse warming gases, but that cooling is not uniform: the offsetting impact varies geographically, depending on local aerosol concentrations.

The omission of aerosol considerations in earlier climate models led to considerable overprediction of projected global warming and predicted regional impacts, though newer models have done much to internalize the cooling effect of aerosols. Aerosols act as cooling agents through several mechanisms, however, some of which are only poorly understood. Besides directly scattering incoming sunlight, most particulates also increase the reflectivity, formation, and lifetime of clouds, affecting the reflection of incoming solar radiation back to space.

Water Vapor

Water vapor is the most abundant of the greenhouse gases and is the dominant contributor to the natural greenhouse effect. About 99 percent of all the moisture in the atmosphere is found in the troposphere, which extends about 10 to 16 kilometers above sea level. Only about one-third of the precipitation that falls on Earth's continents drains to the oceans. The rest goes back into the atmosphere as a result of evaporation and transpiration.

In the lower part of the atmosphere, the water vapor content of the atmosphere varies widely. On a volume basis, the normal range is 1 to 3 percent, though it can vary from as little as 0.1 percent to as much as 5 percent.

Water vapor can be a climate warming force when it traps heat, or can cause either climate warming or cooling when it takes the form of clouds, which reflect incoming solar energy away from Earth.

Most climate models predict that a warming of Earth's atmosphere would be accompanied by an increased level of water vapor in the lower atmosphere, but determining whether this has happened in response to recent climate warming is difficult. Data on water vapor concentrations are limited, and the data suffer from a range of limitations, including changes in instrument type, limited geographic coverage, limited time span, and so on. Data from satellites may offer some relief for these problems, but such data have been gathered only for a few years.

Some researchers have observed what appear to be slight increases in water vapor in various layers of the atmosphere, ranging up to 13 percent. Others have analyzed satellite data, and seen what appears to be a drying of the atmosphere, rather than increased moisture levels.

Solar Activity

Rather than burning with a steady output, the sun burns hotter and cooler over time. Several cycles of increased or decreased solar output have been identified, including cycles at intervals of eleven years, twenty-two years, and eighty-eight years.

Though measurements of solar output have been taken only for the past eighteen years, longer trend patterns can be derived from indirect data sources, such as ice cores and tree rings. Cosmic rays, which fluctuate with the sun's activity, also strike constituents of the atmosphere, creating radioactive versions of certain elements. Beryllium, in particular, is ionized to ^{10}Be by cosmic rays. The ^{10}Be then gets incorporated into trees as they grow, and is trapped in bubbles in ice masses, as is carbon dioxide.

A 1995 reconstruction of historical solar output levels from 1600 to 2000 shows that solar irradiance has risen over time, but with many short-term peaks and troughs in the overall curve of increase, increasing the level of solar output that constitutes the main driver for the climate system's temperature.

Studies suggest that increased solar output may have been responsible for half of the 0.55°C (1°F) increase in temperature from 1900 through 1970, and for one-third of the warming seen since 1970.

Ozone

Ozone is a highly reactive molecule composed of three atoms of oxygen. Ozone concentrations vary by geographical location and by altitude. In addition, ozone exerts a different climate-forcing effect, depending upon altitude.

At lower, tropospheric altitudes, ozone exerts a warming force upon the atmosphere. Tropospheric levels of ozone have been increasing in the Northern Hemisphere since 1970, and may have doubled in that time. Ozone concentrations in the Southern Hemisphere are uncertain, while at the poles, tropospheric ozone concentrations seem to have fallen since the mid-1980s. At higher, or stratospheric altitudes, ozone exerts a cooling force upon the atmosphere. Ozone concentrations in the stratosphere have been declining over most of the globe, though no trend is apparent in the tropics. Much of the decline in stratospheric ozone concentrations has been attributed to the destructive action of the chlorofluorocarbons discussed previously.

Section Summary

It is clear that human action can affect seven of eight of the major greenhouse "forcings": carbon dioxide, methane, nitrous oxide, ozone, CFCs, aerosols, and water vapor. As studies of solar variation have shown, it is also clear that human action is not the only factor involved in determining the impact of these forcings. There is still substantial uncertainty regarding the actual climate impact of the climate forcings.

OBSERVED CLIMATE CHANGES

Part of the concern about global climate change stems from the human tendency to seek meaning in events that may or may not be more than simply a random event. A particularly cold winter, a particularly hot summer, an especially rainy season, or an especially severe drought will all send people off on a search for the greater meaning of the phenomenon. Is it a pattern, or a one-time event? Must we build a dike, or has the danger passed? Since the summer of

1988, virtually all unusual weather events seem to have triggered questions about global climate change.

Our ability to really know what the climate is doing is limited by a short observational record, and by the uncertainties involved in trying to figure out what climate was like in the past, or might be like in the future, for comparison with recent climate changes. While Earth's climate has been evolving and changing for more than four billion years, recordings of the temperature cover only about 150 years, less than 0.000004 percent of the entire pattern of evolving climate. In fact, temperature records are spotty before about 1960 and cover only a tiny portion of the globe, mostly over land. In addition to that 150-year conventional surface temperature record, temperature readings taken from weather balloons cover the years since 1970, and satellite temperature readings cover the years since 1982. Modern, reliable measurements of greenhouse gases are also very recent sources of data, beginning with carbon dioxide measurements at the South Pole in 1957, at Mauna Loa in 1958, and later for methane, nitrogen oxides, and chlorofluorocarbons.

Aside from temperature readings, other climate trends proposed as secondary effects of global warming carry information about the state of the climate. Changes in absolute humidity, rainfall levels, snowfall levels, the extent of snowfall, the depth of snowfall, changes in ice caps, ice sheets, sea ice, and the intensity or variability of storms have all been proposed as secondary effects of global warming. But because the history of recording climate trends is extremely short, most evidence regarding nontemperature-related changes in Earth's climate and atmospheric composition prior to the recent history of direct measurements is gathered from indirect sources such as air bubbles trapped in polar ice, or the study of fossils. This evidence, while interesting as a potential "reality check" for global human-made climate change models, is considered far less reliable than direct observational data.

These limitations in our evidence make it difficult to draw hard-and-fast conclusions regarding what changes have actually occurred recently in comparison to past climate conditions. More importantly, these limitations make it difficult to determine whether those changes are beyond the range of previous climate trends, happening at a faster rate than previous climate trends, or are being sustained for longer than previous climate trends. These are all critical questions when evaluating whether humanity is causing changes to Earth's normal climate patterns.

Nevertheless, scientists have evidence at hand regarding recent changes in both atmospheric composition and global climate trends that suggest that humanity has at least changed Earth's atmospheric composition in regard to greenhouse gases and other pollutants. These changes may or may not be contributing to recently observed changes in global warmth. A quick review of the climate changes suggested by the available evidence follows.

Temperature Trends

Besides readings of Earth's surface temperatures taken with standard glass thermometers, direct readings of atmospheric temperatures have been taken with satellites and weather balloons. In addition to direct measurements of Earth's recent temperatures, proxy measurements of temperatures from farther in the past can be derived from borehole temperature measurements, from historical and physical evidence regarding the extent and mass of land and sea ice, and from the bleaching of coral reefs.

This information is in relatively good agreement regarding what seems to be happening to global temperatures, at least in the recent periods of change spanning the past few hundred years, though there are discrepancies among some of the data sets. Temperatures recorded at ground-based measuring stations reveal a mean warming trend ranging from 0.3°C to 0.6°C (0.5°F to 1.1°F) since about 1850, with 0.2°C to 0.3°C (0.4°F to 0.5°F) of this warming occurring since 1960. The warming is not uniform, either in chronology or in distribution. More of the change occurs over land than over water. More of the warming happens at night, resulting in warmer nighttime temperatures rather than hotter daytime temperatures. More of the warming is noticeable as a moderation of wintertime low temperatures rather than as an increase in summertime high temperatures. Temperatures taken from weather balloons (also called radiosondes) and from satellites span a much shorter period of time (though, arguably, a more rigorously standardized measuring technique), and there is controversy over what they indicate, and how much weight should be given to such a short data set. Some analysts contend that the satellite and balloon recordings show a slight cooling trend in the tropics (about 0.1°C [0.2°F] per decade) since 1982. Others contend that the discrepancy is only an artifact caused by a limited data set, and the recent, unrelated increase in the strength of the El Niño southern oscillation.

And even here, taking the simplest of physical measurements, temperature, uncertainties are present. Temperature readings (satellite or ground station) were not taken specifically for the sake of evaluating the climate patterns of the entire Earth. Consequently the readings were taken from a variety of locations, cover only selected parts of the atmosphere, and are not necessarily well placed to be most informative about the climate as a whole. Further, measurement techniques and stations varied over the course of the temperature record, with data adjustments of a full degree occasionally needed to make the different sets of data compatible with each other. Satellites and balloons measure a different part of the atmosphere than ground stations do, making the comparability of such records questionable. In addition, the shortness of the satellite data record, punctuated as it has been by impacts of volcanic eruptions and the El Niño southern oscillation, further complicate the evaluation of temperature data.

Finally, perspective is important. While the past ten thousand years have been abnormally placid as far as climate fluctuations go, evidence of prior climate changes show an Earth that is anything but placid climatically. Some 11,500 years ago, for example, there is evidence that temperatures rose sharply over short periods of time. In Greenland, temperatures increased by as much as 7°C (12.6°F) over only a few decades, while sea surface temperatures in the Norwegian Sea warmed by as much as 5°C (9°F) in fewer than forty years. There is also evidence of about twenty rapid temperature fluctuations during the last glaciation period in the central Greenland records. Rapid warmings of between 5°C and 7°C (9°F to 12.6°F) were followed by slow returns to glacial conditions over the course of 500 to 2,000 years.

Rainfall Trends

Changes in precipitation trends are, potentially, a form of indirect evidence reflecting whether Earth is currently experiencing man-made climate change. Climate change models suggest that an enhanced greenhouse effect could cause changes in the hydrologic cycle such as increased evaporation, drought, and precipitation. But the IPCC warns that "our ability to determine the current state of the global hydrologic cycle, let alone changes in it, is hampered by inadequate spatial coverage, incomplete records, poor data quality, and short record lengths."

The global trend in rainfall showed a slight increase (about 1%) during the twentieth century, though the distribution of this change was not uniform either geographically or over time. Rainfall has increased over land in high latitudes of the Northern Hemisphere, most notably in the fall. Rainfall has decreased since the 1960s over the subtropics and tropics from Africa to Indonesia. In addition, some evidence suggests increased rainfall over the Pacific Ocean (near the equator and the international dateline) in recent decades, while rainfall farther from the equator has declined slightly.

Sea Level Trends

Changes in sea level and the extent of ice sheets, sea ice, and polar ice caps are still another form of indirect evidence reflecting whether Earth is currently undergoing anthropogenic climate change. Climate change theory would suggest that rising global temperatures would cause sea levels to rise due to a combination of the thermal expansion of water and melting of glaciers, ice sheets, ice caps, and sea ice.

Studies of sea leavels considered to reflect our best understanding of sea-level rise, as summarized in the 1995 reports of the United Nations Intergovernmental Panel on Climate Change, indicate a rise of 18 cm during the twentieth century, with a range of uncertainty of 10 to 25 cm. There is little evidence that the rate of sea-level rise has increased during that time period, though in theory the rate of warming has been accelerating. But thermal expansion of water is only one contributor to sea-level changes. Glaciers, ice sheets, and land water storage all play a role—a highly uncertain role.

Surface waters

Global warming would also be expected to influence surface waters such as lakes and streams, through changes induced in the hydrologic cycle. However, the last published report of the IPCC states no clear evidence of widespread change in annual streamflows and peak discharges of rivers in the world (IPCC, 1995, p. 158). While lake and inland sea levels have fluctuated, the IPCC also points out that local effects make it difficult to use lake levels to monitor climate variations.

Snow and Ice Trends

Global warming would also be expected to influence things such as snowfall, snow depth, and snow coverage (or extent), but studies examining changes in these aspects of the climate are quite mixed. Consistent with the indications of slight warming of the global climate, snow cover has declined in recent years, with a higher percentage of precipitation in cold areas coming down as rain rather than snow. But while the annual mean extent of snow cover over the Northern Hemisphere has declined by about 10 percent since 1979, snowfall levels have actually increased by about 20 percent over northern Canada and by about 11 percent over Alaska. Between 1950 and 1990, snowfall over China decreased during the 1950s but increased during the 1960s and the 1970s. Snowfall over the 45–55-degree-latitude belt has declined slightly. Snow depth levels, which respond both to atmospheric temperature and to the ratio of rainfall to snowfall, show equally mixed changes. Snow-depth measurements of the former Soviet Union over the twentieth century show decreased snow depth of about 14 percent during the Soviet winter, mostly in the European portion, while snow depth in the Asian sectors has increased since the 1960s.

Glaciers, Ice Caps, and Ice Sheets

With regard to glaciers and ice caps, the state of knowledge is even more limited. Glaciers and ice caps may have accounted for 2 to 5 centimeters of the observed sea-level rise discussed above, but the range of uncertainty is high. With regard to ice sheets, data are contradictory: There is not enough evidence to know whether the Greenland and Antarctic ice sheets are shrinking, hence contributing to sea-level rise, or growing, and hence retarding sea-level rise. They may even be doing both—growing on top and shrinking at the margins.

Weather Intensity and Variability Trends

Finally, increases in the intensity or variability of weather are considered another form of indirect evidence reflecting whether Earth is currently undergoing human-driven climate change. Predictions of increased incidence of extreme temperatures, tornadoes, thunderstorms, dust storms and fire-promoting weather have been drawn from basic global climate change theory. However, evidence has not so far borne out these predictions on a global scale. The IPCC concludes:

> [O]verall, there is no evidence that extreme weather events, or climate variability, has increased, in a global sense, through the 20th century, although data and analyses are poor and not comprehensive. On regional scales, there is clear evidence of changes in some extremes and climate variability indicators. Some of these changes have been toward greater variability; some have been toward lower variability [IPCC, 1995, p. 173].

Section Summary

Evidence regarding changes in Earth's climate in the twentieth century is mixed, and encompasses a range of uncertainties. While the most recently published IPCC report holds that there is a discernible human influence on climate, this conclusion is not dependent on the evidence of actual changes in Earth's climate, as shown in this figure. On that note, the IPCC (1995, p. 411) says, "Despite this consistency [in the pattern of change], it should be clear from the earlier parts of this chapter that current data and systems are inadequate for the complete description of climate change." Rather, this conclusion is based on mathematical modeling exercises and "reality checked" with what hard evidence we have.

UNCERTAINTY AND FUTURE RESEARCH NEEDS

While recent studies of climate have contributed a great deal to our understanding of climate dynamics, there is still much to learn. The process of searching for evidence of man-made climate change, in fact, is both a search for new discoveries about how climate works, and continuing refinement of our understanding of the underlying theories we already have.

While greenhouse effect theory is a relatively uncontroversial issue in the scientific sense, the theory of global, human-driven climate change is at a much younger stage of development. Although there are very few articles in science journals that contradict either the overall theory or details of the core greenhouse effect, the same cannot be said for the theory of human-driven climate change and the consequences of that change. Indeed, nearly every month on the pages of leading science journals, studies jockey back and forth about key elements of human-made climate change.

Current climate change models have acknowledged weaknesses in their handling of changes in the sun's output, volcanic aerosols, oceanic processes, and land processes that can influence climate change. Some of those uncertainties are large enough, by themselves, to become the tail that wags the dog of climate change. Three of the major remaining uncertainties are discussed below.

The Natural Variability of Climate

Despite the extensive discussion of climate modeling and knowledge of past climate cycles, only the past thousand years of climate variation are included in the two state-of-the-art climate models referred to by the IPCC. As discussed earlier, however, the framework in which we view climate variability makes a significant difference in the conclusions we draw regarding either the comparative magnitude or rate of climate changes, or the interpretation of those changes as being either inside or outside of the envelope of normal climate change variations. The IPCC report summarizes the situation succinctly:

> Large and rapid climatic changes occurred during the last ice age and during the transition towards the present Holocene period. Some of these changes may have occurred on time-scales of a few decades, at least in the North Atlantic where they are best documented. They affected atmospheric and oceanic circulation and temperature, and the hydrologic cycle. There are suggestions that similar rapid changes may have also occurred during the last interglacial period (the Eemian), but this requires confirmation. The recent (20th century) warming needs to be considered in the light of evidence that rapid climatic changes can occur naturally in the climate. However, temperatures have been far less variable during the last 10,000 years (i.e., during the Holocene) [IPCC, 1995, p. 416].

Until we know which perspective is more reflective of Earth's climate as a whole—the last ten thousand years, or a longer period of time—it will be difficult to put recent warming trends in perspective, or to relate those trends to potential impacts on the climate and on Earth's flora and fauna.

The Role of Solar Activity

At the front end of the climate cycle is the single largest source of energy put into the system: the sun. And while great attention has been paid to most other aspects of climate, little attention has been paid to the sun's role in the heating or cooling of Earth. Several studies in the late 1990s have highlighted this uncertainty, showing that solar variability may play a far larger role in Earth's climate than it was previously given credit for by the IPCC. If the sun has been heating up in recent times, researchers observe, the increased solar radiation could be responsible for up to half of the observed climate warming of the past century. But as with satellite measurements of Earth's temperature, the short timeline of satellite measurements of solar irradiance introduces significant uncertainty into the picture. Most researchers believe that at least another decade of solar radiation measurement will be needed to clearly define the influence of solar input on the global climate.

Clouds and Water Vapor

Between the emission of greenhouse gases and change in the climate are a range of climate and biological cycles that can influence the end result. Such outcome-modifier effects are called "feedbacks" or "indirect effects" in the climate change literature.

One such feedback is the influence of clouds and water vapor. As the climate warms, more water vapor enters the atmosphere. But how much? And which parts of the atmosphere, high or low? And how does the increased humidity affect cloud formation? While the relationships among clouds, water vapor, and global climate are complicated in and of themselves, the situation is further complicated by the fact that aerosols exert a poorly understood influence on clouds.

Earlier computer models, which omitted the recently validated cooling effect of aerosols, overestimated the global warming that we would have expected to see by now, based only on the levels of greenhouse gases that have been emitted. As discussed earlier, aerosols themselves may offset 20 percent of the expected impact of warming gases. In addition, though direct cooling impacts of aerosols are now being taken into account by climate models, aerosol impact on clouds remains a poorly defined effect with broad implications, given a range of additional cooling potential of up to 61 percent of the expected warming impact from the warming greenhouse gases.

The last published report of the IPCC acknowledges that "the single largest uncertainty in determining the climate sensitivity to either natural or anthropogenic changes are clouds and their effects on radiation and their role in the hydrological cycle . . . At the present time, weaknesses in the parameterization of cloud formation and dissipation are probably the main impediment to improvements in the simulation of cloud effects on climate" (IPCC, 1995, p. 346).

THE IMPACTS OF CLIMATE CHANGE

Global warming, and the potential climate changes that might accompany such warming, are estimated using of complex computer models that simulate, with greater or lesser complexity and success, the way Earth's climate would change in response to the level of greenhouse gases in the air. It is widely acknowledged that the potential temperature changes predicted by global warming theory do not pose a direct threat to human life. In fact, since more people die from extremes of cold rather than heat, the actual warming of the atmosphere, on net, could save more lives through warmer winters than it takes through hotter summers.

The major concerns about climate change focus on the second- and thirdhand impacts that would theoretically accompany global warming. Climate change theory suggests that warming of the overall environment could lead to a variety of changes in the patterns of Earth's climate as the natural cycles of air currents, ocean currents, evaporation, plant growth, and so on change in response to the increased energy levels in the total system. The most commonly predicted primary impacts of global warming are increased activity in the hydrologic, or water cycle of Earth, and the possible rise of oceans due to thermal expansion and some melting of sea ice, ice sheets, or polar ice caps. More dynamic activity in the water cycle could lead to increased rainfall in some areas, or, through increased evaporation rates, could cause more severe droughts in other areas. Rising sea levels could inundate some coastal areas (or low-lying islands), and through saltwater intrusion, could cause harm to various freshwater estuaries, deltas, or groundwater supplies.

Some have also predicted a series of thirdhand impacts that might occur if the climate warms and becomes more dynamic. Wildlife populations would be affected (positively and negatively), as would some vegetative growth patterns. The "home range" of various animal and insect populations might shift, exposing people to diseases that were previously uncommon to their area, and so on.

But one need not wade far into the most recently published IPCC report on the potential impacts of climate change before encountering an admission that uncertainty dominates any discussion of such potential impacts:

> Impacts are difficult to quantify, and existing studies are limited in scope. While our knowledge has increased significantly during the last decade and qualitative estimates can be developed, quantitative projections of the impacts of climate change on any particular system at any particular location are difficult because regional scale climate change projections are uncertain; our current understanding of many critical processes is limited; and systems are subject to multiple climatic and non-climatic stresses, the interactions of which are not always linear or additive. Most impact studies have assessed how systems would respond to climate changes resulting from an arbitrary doubling of equivalent atmospheric carbon dioxide concentrations. Furthermore, very few studies have considered greenhouse gas concentrations; fewer still have examined the consequences of increases beyond a doubling of equivalent atmospheric carbon dioxide concentrations, or assessed the implications of multiple stress factors [IPCC, 1995, p. 24].

The IPCC report goes on to point out that this extreme uncertainty is likely to persist for some time, since unambiguous detection of human-made climate change hinges on resolving many difficult problems:

> Detection will be difficult and unexpected changes cannot be ruled out. Unambiguous detection of climate-induced changes in most ecological and social systems will prove extremely difficult in the coming decades. This is because of the complexity of these systems, their many non-linear feedbacks, and their sensitivity to a large number of climatic and non-climatic factors, all of which are expected to continue to change simultaneously. The development of a base-line projecting future conditions without climate change is crucial, for it is this baseline against which all projected impacts are measured. The more that future climate extends beyond the boundaries of empirical knowledge (i.e., the documented impacts of climate variation in the past), the more likely that actual outcomes will include surprises and unanticipated rapid changes [IPCC, 1995, p. 24].

Uncertainties of this scale do not imply, as some analysts have asserted, that there is no reason to fear negative change, nor does it imply that we must fear drastic impacts. Rather, uncertainties of this scale indicate the need for a sustained research program aimed at clarifying our understanding of Earth's climate and how human activities might or might not translate into negative environmental impacts.

Risk-Reduction Benefits - Near Term	
Reduced risk of harm from changing weather patterns:	NONE
Reduced risk of harm from extreme weather events:	NONE
Reduced risk of harm through famine avoidance:	NONE
Reduced risk of harm through disease prevention:	NONE
Reduction in other proposed climate change hazards:	NONE
Reduced risk of harm through avoided economic impacts of climate change:	NONE
Risk-Reduction Benefits - Long Term	
Reduced risk of harm from changing weather patterns:	NONE-HIGH
Reduced risk of harm from extreme weather events:	NONE-HIGH
Reduced risk of harm through famine avoidance:	NONE-HIGH
Reduced risk of harm through disease prevention:	NONE-HIGH
Reduction in other proposed climate change hazards:	NONE-HIGH
Reduced risk of harm through avoided economic impacts of climate change:	NONE-HIGH
Risk Reduction Liabilities - Near Term	
Induced fatalities from reduced disposable income:	Approx. 15,500/yr.
Lives not saved through other available risk-reduction investments:	Approx. 800,000/yr.
NET RISK-REDUCTION BENEFIT - NEAR TERM	***NONE***
NET RISK-REDUCTION BENEFIT - LONG TERM	***NONE-HIGH***

Table 1.
Risk-Alteration Ledger: The Kyoto Protocol

INTERNATIONAL AGREEMENTS ON RELEASE OF GREENHOUSE GASES

In 1997, a treaty was developed to limit the amount of greenhouse gases released into the atmosphere. Under the auspices of the United Nations Secretariat, the Third Conference of the Parties, held in Kyoto, Japan, produced the Kyoto Protocol (Table 1). This protocol calls for reductions in greenhouse gas emissions by various countries, though developing countries, predicted to become the dominant producers of greenhouse gases in the twenty-first century, were not bound to greenhouse gas reductions. Several countries, such as Australia, Iceland, and Norway, were allowed to increase their levels of greenhouse gas emissions under the treaty. The major reductions in emissions were to come from Europe, where Latvia, Estonia, and Lithuania agreed to an 8 percent reduction in emissions relative to 1990 levels. Japan and Canada agreed to 6 percent reductions from 1990 levels, and the United States agreed to reduce greenhouse gas emissions 7 percent below 1990 levels.

The Kyoto Protocol covers reductions in carbon dioxide, methane, nitrous oxide, and three fluorocarbons: hydrofluorocarbons, perfluorocarbons, and sulfur hexafluoride. The protocol also included mechanisms for considering greenhouse gas reductions stemming from changes in land use, and enshrined the principle of international emissions trading, though not the mechanism or specifics, which were left for later Conferences of the Parties to resolve. Finally, the Kyoto Protocol created a “clean development mechanism” by which developing countries could develop advance credits for taking actions that would limit the release of greenhouse gases in the future.

Several obstacles stand in the way of the Kyoto

Protocol. In 1997, prior to President Clinton's acceptance of the Kyoto Protocol, U.S. Senate Resolution 98, the Byrd-Hegel resolution, which was passed by a vote of ninety-five to zero, imposes specific requirements that must be met before the Kyoto Protocol can be ratified. The resolution calls for a specific timeline and commitments by developing countries to reduce greenhouse gas emissions, and evidence that adoption of the Kyoto Protocol would not result in serious harm to the U.S. economy. In addition, the Fifth Conference of the Parties (1999) failed to resolve numerous outstanding issues held over from the previous conference, and put off critical decision making until the Sixth Conference of the Parties in The Hague, Netherlands, in November 2000.

COST OF REDUCING GREENHOUSE EMISSIONS AND IMPACT ON ENERGY SYSTEMS

Reducing emissions of carbon dioxide and other greenhouse gases is not a trivial problem. Fossil fuels provide the overwhelming majority of energy production globally, and are predicted to do so through 2010. In the United States in 2000, fossil fuels were used to produce 70 percent of all energy generated. Alternative technologies such as nuclear power, solar power, wind power, hydropower, geothermal power, and hydrogen power have promise, but also have significant limitations and are considerably more costly than fossil fuel use. Consider that nearly 25 percent of total U.S. energy consumption in 1996 was for transportation, which is nearly all powered by fossil fuels.

Estimates of the cost of reducing greenhouse gas emissions, and the impact that such reductions would have on energy systems, vary widely. Estimates of greenhouse gas reduction costs are critically dependent on the assumptions used in economic models. Models assuming that greenhouse gas reduction targets will be met using international, multiemission trading systems suggest lower costs than those models that assume less international trading, single-gas approaches, carbon taxes, and so on.

In a comparison of nine economic models, estimated costs to the United States as of the late 1990s ranged from a loss in gross domestic product from $40 billion to $180 billion, with assumptions of no emission trading; from $20 billion to $90 billion with trading only among developed countries; and from $5 billion to $20 billion with assumptions of global trading systems. Studies showing values at the higher end of the ranges outnumbered those showing costs at the lower end of the spectrum. The largest costs in such models stem from accelerated fuel substitution and the adoption of more expensive nonfossil fuel forms of fuel generation.

Kenneth Green

See also: Acid Rain; Air Pollution; Atmosphere; Environmental Problems and Energy Use; Ocean Energy Systems; Pollution and Energy Efficiency.

BIBLIOGRAPHY

Broad, W. J. (1997). "Another Possible Climate Culprit: The Sun." *New York Times*, September.

Elliot, W. P. (1995). "On Detecting Long-Term Changes in Atmospheric Moisture." *Climatic Change* 31:349–367.

Etheridge, D. M., et al. (1998). "Atmospheric Methane between 1000 A.D. and Present: Evidence of Anthropogenic Emissions and Climate Variability." *Journal of Geophysical Research* 103(DI 3):15979–15993.

Fluckinger, J., et al. (1999). "Variations in Atmospheric N_2O Concentration During Abrupt Climate Changes." *Science* 285:227–230.

Fourier, J. (1824). "Remarques générales sur la temperature du globe terrestre et des espaces planétaires." *Annals de chimie et de physique* 27:136–167.

Green, K. (1997). *A Plain English Guide to the Science of Climate Change*. Los Angeles: Reason Public Policy Institute.

Green, K., et al. (1999). *Climate Change Policy Options and Impacts: Perspectives on Risk Reduction, Emissions Trading, and Carbon Tax*. Los Angeles: Reason Public Policy Institute.

Hansen, J., et al. (1998). "Climate Forcings in the Industrial Era." *Proceedings of the National Academy of Science* 95:12753–12758.

Intergovernmental Panel on Climate Change (IPCC). (1995). *Climate Change 1995, Impacts, Adaptations and Mitigation of Climate Change: Scientific-Technical Analyses*. Cambridge, Eng.: Cambridge University Press.

Intergovernmental Panel on Climate Change (IPCC). (1995). *Climate Change 1995, The Science of Climate Change*. Cambridge, Eng.: Cambridge University Press.

Kerr, R. A. (1995). "Studies Say—Tentatively—That Greenhouse Warming is Here." *Science* 268:1567.

Kerr, R. A. (1996). "A New Dawn for Sun-Climate Links?" *Science* 271:1360.

Kerr, R. A. (1997). "Did Satellites Spot a Brightening Sun?" *Science* 277:1923.

Lean, J., and Rind, D. (1998). "Climate Forcing by Changing Solar Radiation." *Journal of Climate* 11:3069–3094.

Manahan, S. E. (1984). *Environmental Chemistry*, 4th ed. Monterey, CA: Brooks/Cole Publishing Company.
Rind, D. (1998). "Just Add Water Vapor." *Science* 281:1152.
Schroeder, S. R., and McGuirk, J. P. (1998). "Widespread Tropical Atmosphere Drying from 1979 to 1995." *Geophysical Research Letters* 25(9):1301–1304.
Spencer, R. W., and Christy, J. R. (1992). "Precision and Radiosonde Validation of Satellite Gridpoint Temperature Anomalies, Part 11: A Tropospheric Retrieval and Trends during 1979–90." *Journal of Climate* 5:858–866.
Taubes, G. (1997). "Apocalypse Not." *Science* 278:1004–1006.
Watson, R. T.; Meira Filho, L. G.; Sanhueza, E.; and Janetos, A. (1992). "Greenhouse Gases: Sources and Sinks." In *Climate Change 1992*, ed. J. T. Houghton, B. A. Callander, and S. K. Varney. Cambridge, Eng.: Cambridge University Press.
Weyant, J. P., and Hill, J. N. (1999). "Introduction and Overview." In "The Costs of the Kyoto Protocol: A Multi-Model Evaluation." *The Energy Journal* (special issue).
Willson, R. C. (1997). "Total Solar Irradiance Trend During Solar Cycles 21 and 22." *Science* 277:1963–1965.

COAL, CONSUMPTION OF

INTRODUCTION AND BACKGROUND

Coal, essentially fossilized plant material, has been used as an energy source for centuries. As early plants decomposed in the absence of oxygen (for example, at the bottom of deep lakes), oxygen and hydrogen atoms broke off the long organic molecules in the plant material, leaving mostly atoms of carbon, with some other impurities. Formed as long as ago as 300 million years and more, during the Carboniferous Period (named for obvious reasons), coal's main useful component is carbon.

Coal comes in various forms. Older coal is harder, higher in energy content and has a higher proportion of carbon. Anthracite is the hardest, purest version, with a carbon content of about 90 percent. It is also rarer, found in the United States almost exclusively in eastern Pennsylvania. Its energy content is about 25 million Btus per short ton, 67 percent higher than lignite, the softest form of coal, which sometimes contains less than 40 percent of carbon (most of the balance being water and ash—the latter composed mostly of sodium carbonate and potassium carbonate). Lignite is much younger than anthracite. Some lignite deposits are less than 100 million years old (a third the age of most anthracite), being formed from plants that lived in the Cretaceous and Tertiary Eras. Not all of its impurities, in the form of water, carbonates, or independent atoms of oxygen and hydrogen have had the time to be removed by geological and chemical processes.

Bituminous coal—an intermediate form—is and by far the most widely found and used. Its energy content averages around 21.5 million Btus per short ton. Some forms of coal have a substantial sulfur content and contain other impurities as well, chiefly carbonates. Peat, also used as an energy source through burning, can be viewed as an earlier stage of coal: pressure in the absence of oxygen over a period of tens of millions of years will change peat into coal.

The oxidation of carbon through burning produces a significant amount of heat through an exothermic reaction, a chemical reaction that releases energy. However, since most of the energy gained from burning coal results in the production of carbon dioxide (CO_2), this principal greenhouse emission is higher for coal than for other fossil fuels, namely oil and natural gas, in which other elements are oxidized in addition to carbon. Because the burning of coal produces so much carbon dioxide, its use will be severely affected by compliance with the Kyoto Protocol, discussed in a following section.

The other fuels, when burned, also produce water vapor and—in the case of oil—other hydrocarbons, as well as carbon dioxide. For equal amounts of energy, oil produces about 80 percent of the CO_2 that coal does; natural gas only produces 55 percent of coal's CO_2 level.

EARLY HISTORY OF COAL USE

According to the U.S. Energy Information Administration, although scattered use of coal may have occurred as early as 1100 B.C.E., it was substantially later before coal became widely used. Nevertheless, coal has been utilized for millenia. In China, as early as the fourth century C.E., coal sometimes substituted for charcoal in iron smelting. By the eleventh century, coal had become the most important fuel in China, according to *The Columbia History of the World* (Garrity and Gay, 1972). In the Middle Ages, in various parts of Europe, especially England, coal began to be used as an energy source

Older railroad engines, like the Scottish Railway 4-6-0 engine here, were major consumers of coal. (Corbis-Bettmann)

for smelting, forges, and other limited applications. But it was not until the fifteenth century that it began to be used for residential heating.

England's good fortune in having large deposits of coal was particularly important for that nation, as the English had destroyed most of their forests between the twelfth and sixteenth centuries, in order to produce heat (chiefly residential), as well as charcoal for industrial purposes. By 1840, Britain's coal production was ten times that of Prussia and significantly higher than that of France and other European nations. The availability of coal, along with various inventions, such as the Watt steam engine, which was usually coal-driven, helped the Industrial Revolution to begin in Britain in the last quarter of the eighteenth century. The rest of Europe began catching up a few decades later. But still, around 1870, Britain alone produced over 30 percent of the manufactured goods in the world.

Coal is used to produce coke, manufactured by heating coal in the absence of air. Coke, when heated to high temperatures, yields carbon monoxide, which reduces the iron oxides in ore to iron. Steel contains a small quantity of carbon in an alloy with iron, and coke refining adds carbon to the iron, in addition to refining the ore. As a result of their supply of coal, the British had the capacity to smelt considerably more iron and later, steel, than other countries, resulting in military and economic advantages.

With the development of improvements in the steam engine and other inventions useful for manufacturing—often in the area of textiles—the British were also able to expand new, energy-intensive industries more rapidly than their competitors. At the beginning of the nineteenth century, switching from renewable (animal, wind, hydroelectricity, and mainly wood) energy resources to coal became the hallmark of strong, expanding economies.

HEALTH EFFECTS OF COAL COMBUSTION

Some effects of massive coal burning are quite pernicious, particularly in the absence of strong efforts to reduce the sulfates and other pollutant byproducts. The London "pea soup" fog, conjuring up mystery, intrigue, and respiratory disease, was largely due to the intensive use of coal in conjunction with stagnant and humid meteorological conditions. This local weather effect has virtually vanished in London since the advent of centrally generated electricity as a substitute for coal fires in each individual living area. Much of this electricity is now fueled by nuclear energy among other sources. Another mitigating factor has been the introduction of anti-pollution measures at coal-burning plants, including limestone scrubbers in smokestacks and fluidized bed reactors for coal combustion.

While it was prevalent, the persistent presence of coal-generated smog was found to have serious health effects on the public. This became apparent on days when air pollution was especially bad, and hospital admissions for respiratory ailments increased considerably. The classic example is the "Great London Smog" of December 4, 1952, to which was attributed a net increase of about 4,000 deaths in the following several days. More recently, in Czechoslovakia and East Germany, following the close of the Cold War, researchers discovered that in several severely impacted locations, the rate of upper respiratory disorders—especially among children—was extremely high.

The power plant in the background burned coal as fuel. (Corbis-Bettmann)

REDUCED COAL EMISSIONS THROUGH TECHNOLOGICAL ADVANCES

Over the past decades, advances have been made that reduce environmental impacts of coal burning in large plants. Some are standard and others experimental. Limestone (mainly calcium carbonate) scrubber smokestacks react with the emitted sulfates from the combustion and contain the chemical products, thereby reducing the release of SO_x into the atmosphere by a large factor (of ten or more). Pulverization of coal can also allow for the mechanical separation of some sulfur impurities, notably those in the form of pyrites, prior to combustion. Currently deployed—with more advanced versions in the development stage—are various types of fluidized bed reactors, which use coal fuel in a pulverized form, mixed with pulverized limestone or dolomite in a high temperature furnace. This technique reduces sulfate release considerably. There are pressurized and atmospheric pressure versions of the fluidized bed.

Each technique has its own advantages, but the pressurized version, operating at high temperature, also makes use of jets of the combustion gases to keep the pulverized fuel/limestone mixture in a suspension, later employing both waste heat and a turbine, driven by a jet of combustion gases, to extract more of the energy generated by the reaction. Design engineers hope to increase efficiencies from the standard level of about 33 percent to 40 percent and even, possibly, approaching 50 percent. If successful, this increase in efficiency could have a major positive impact on carbon dioxide emissions, since up to 50 percent more energy could be extracted from the same amount of oxidized coal. Thus, these technologies, while greatly reducing sulfate emissions, can contribute to reducing carbon dioxide emissions by increases in efficiency.

Nevertheless, in the minds of much of the public and of decision–makers, there is a preference for substituting coal with natural gas (which is primarily methane, CH_4) to the degree that is easily feasible. Natural gas produces far less of most types of pollutants, including SO_x, NO_x, and CO_2. The tendency away from coal is mitigated by the fact that coal is much cheaper as a fuel (although the capital cost of building a coal plant is higher). For electricity plants that are already built and whose capital costs are sunk, coal's advantage is clear: it costs $0.85 per million Btu, compared with $2.18 per million Btu for natural gas and $2.97 per million Btu for crude oil.

U.S. COAL CONSUMPTION: QUANTITATIVE EVOLUTION

Throughout the industrialized world over the past two centuries, coal became relied upon as an energy source for industrial processes and for residential heat. In the United States, all the coal consumed before the year 1800—much of it imported from Britain—amounted to only 108,000 tons, which is one ten-thousandth of current annual U.S. production. Until 1840, wood exceeded coal as an energy source. However, coal then began a slow, steady expansion in usage, and, for over a century, until 1951, it was the chief energy source in the United States, contributing in the area of transportation (railroads) as well as the earlier, familiar sectors of industrial processes and residential heat.

With the discovery of oil in Pennsylvania in 1859, the seeds of a new energy era were sown. After an initial period of growth from a zero base, oil began to contribute significantly to the energy budget around the turn of the century. In the years following, oil slowly substituted for coal in transportation, starting with automobiles (where coal-fired steam engines had never been very successful) and continuing in trains. Later, it also displaced coal in residential heat and many industrial processes as well as, to a degree, in electricity generation. The "cleaner" energy sources that smelled better than coal began to be substituted for it, particularly in the post–World War II period. From 1952 to 1983, coal lost its primary position, and oil and natural gas vied for the title of main energy source in the United States (oil always remained ahead). But since 1984, coal has regained the lead in domestic production, although not consumption, due mainly to its renewed leading role in electricity generation. The development of cleaner combustion methods removed some of the political objections to obvious coal pollution, in the form of sulfates and particulates. These methods could be introduced economically, due to the advantages of scale inherent in the centralized nature of electricity production.

Year	Coal	Gas	Oil	Hydro	Nuclear
1950	12.35	5.97	13.31	1.44	0.00
1955	11.17	9.00	17.25	1.41	0.00
1960	9.84	12.39	19.92	1.61	0.01
1965	11.58	15.77	23.25	2.06	0.04
1970	12.26	21.79	29.52	2.65	0.08
1975	12.66	19.95	32.73	3.22	1.90
1980	15.42	20.39	34.20	3.12	2.74
1985	17.48	17.83	30.92	3.36	4.15
1990	19.11	19.30	33.55	3.09	6.16
1995	20.11	22.16	34.66	3.44	7.18
1996	20.99	22.59	35.72	3.88	7.17

Table 1.
U.S. Energy Consumption by Source, 1950–1996
NOTE: Quads = Btu × 10^{12}
SOURCE: Energy Information Administration, U.S. Department of Energy. (1988). *Annual Energy Review.* DOE/EIA-0384(87); Energy Information Administration, U.S. Department of Energy. (1988). *Monthly Energy Review, July 1988.* DOE/EIA-0035(88/07);. Energy Information Administration, U.S. Department of Energy. (1997). *Annual Review of Energy 1997.*

Since coal consumption bottomed out in the 1960s, there has been a steady increase in the fraction of energy produced by coal in the United States. About 90 percent of coal use in 1996 was for electricity generation, and most of the rest of its usage was accounted for in industrial coking processes. Nuclear power had a chance to substitute for coal in electricity generation and, to an extent, made serious inroads until encountering significant political and economic problems related to perceptions of public safety and to the increasing cost of nuclear facilities. Following the Three Mile Island nuclear accident in 1979, the use of nuclear power as an energy source was strongly opposed by many public interest groups. No new plants in the United States that were ordered after 1973 have been completed. Thus, it is unlikely that the United States will see widespread replacement of coal–fired electric plants by the nuclear option for the foreseeable future.

The Organization of Petroleum Exporting Countries (OPEC) oil embargo in 1973 resulted in temporary oil shortages in the United States. This crisis focused attention on the need to minimize oil's use except for transportation, for which no other energy source would provide a very effective substitute. As a result, coal use has increased in both fractional and absolute terms since the 1970s and now is the single largest domestically produced energy source in the United States. In 1996, some 21 quads of energy were produced by coal, and in 1997, coal energy consumption rose to 23 quads. About one-third (31%) of all domestic energy production in the United States now comes from coal, which also produced a majority (52%) of the nation's electricity in 1997. Coal accounted for over one–quarter of energy consumption, since a large amount of petroleum is imported from other countries. In addition, nearly 10 percent of U.S. coal production is exported. Table 1 shows the trends in energy consumption by source in the United States.

WORLD COAL CONSUMPTION

The recent history of the world use of coal roughly follows that of the United States for two reasons. First, the United States and the industrial nations have had, in the aggregate, similar energy behavior in terms of energy sources. Second, the United States itself accounts for about one quarter of world energy use. Thus, world energy use patterns reflect, to a considerable degree, those of the United States.

World coal usage, inclusive of the three major types of coal—anthracite, bituminous (by far the most prevalent form) and lignite—reached a plateau in the first decade of the twentieth century and climbed only very slowly in the half century that followed. By 1880, coal use had equaled wood use on a worldwide basis. The usage around the turn of the century was on the order of 2.2 gigatons per year (around 55 quads), of which about 600 million tons were in the United States. World oil production progressively supplemented the use of coal between 1900 and 1950, increasing by more than an order of magnitude in that period of time, from a little over a quad to some 20 quads. Coal's increase over those years was fractionally much less.

After 1930, four other energy sources began to contribute significantly, as wood use continued its slow decline and coal production was relatively flat. These four were oil, natural gas, nuclear power (beginning in the 1950s), and hydroelectricity. The latter two are relatively small players, but oil and natural gas are major sources of energy, with oil energy production actually exceeding coal in the 1960s.

The first oil crisis in 1973 was a politically–driven event. It resulted from production cutbacks by oil–producing nations following the Yom Kippur War between Egypt and Israel and marked a watershed in patterns of energy use in the industrialized world—especially, as previously noted, in the United States. Oil was saved for transportation to the degree possible, when it became evident that there would be an increased reliance on potentially unreliable foreign sources as domestic sources were depleted. Energy conservation achieved a strong boost from this traumatic event, in which oil prices rose sharply and the uncertainty of oil sources became clear to First World nations.

Since the early 1990s the United States has imported more oil than it has produced for its own use. And, as the nuclear option became frozen, coal has become the chief source for generating electricity, which itself accounts for about 35 percent of the energy sector. In 1997, 52 percent of electricity produced in the United States was generated from coal and in other recent years the fraction has approached 56 percent. Since the United States accounts for one–quarter of total world energy usage, the increase in coal use in the United States alone has a significant impact on worldwide statistics.

Energy Source	*1970*	*1995*	*2010*	*2020*
Oil	97.8	142.5	195.5	237.3
Natural Gas	36.1	78.1	133.3	174.2
Coal	59.7	91.6	123.6	156.4
Nuclear	0.9	23.3	24.9	21.3
Renewables	12.2	30.1	42.4	50.2
Total	**206.7**	**365.6**	**519.6**	**639.4**

Table 2.
Summary and Projections of World Energy Consumption (1970–2020)
NOTE: Quads = Btu × 10^{12}
SOURCE: Energy Information Administration, U.S. Department of Energy. (1998). *International Energy Outlook 1998.* DOE/EIA-0484(98). Washington, DC: U.S. Government Printing Office. (<http://www.eia.doe.gov/oiaf/ieo98>)

EFFECTS OF THE KYOTO PROTOCOL

The U.S. Department of Energy (DOE) has analyzed and projected energy use by sector and energy source for a number of years. One recent forecast (see Table 2) analyzed the implications of the international agreement in the Kyoto Protocol to the United Nations Framework Convention on Climate Change. The United States committed, under the Protocol (Annex B), to reduce greenhouse emissions by 7 percent from 1990 levels by the period between 2008 and 2012. This translated to a 31 percent decline in the production of greenhouse gases (chiefly carbon dioxide) relative to the DOE's assessment of the most likely baseline number, for U.S. energy use and related carbon dioxide emissions predicted for that time.

Other industrial states have committed to reductions nearly as large. However, the world's chief user of coal, China, as a developing nation, has not yet made solid commitments to reduction. Neither has India, another major coal burner, although it uses less than one-quarter as much as China. Limitations on emissions by these emerging world powers are still the subject of discussion. It is clear that, without some limitations on CO_2 emissions by major Third World industrializing nations, the goal of rolling back world greenhouse emissions to the 1990 level will be very difficult to achieve.

In 1995, about 92 quads of coal-fired energy were consumed in the world. This constituted close to one-quarter of the 366 quads estimated to be the world's total energy production. The level of coal use will be a major determining factor in whether greenhouse emission goals for 2010 will be met. Assessing the likely state of affairs for coal use at that time requires predicting the state of the world's economies by region as well as estimating probable technological advances by then.

By the year 2010, one DOE model predicts a world energy level of 520 quads per year. Countries already using relatively large quantities of energy will contribute less to the increase in the total energy use than will large, rapidly developing countries, such as China. The projected U.S. excess of carbon dioxide release over that permitted by the Protocol will be on the order of 550 million tons of carbon per year (1803 million metric tons, rather than 1252). Of course, this projected excess would arise from all fossil fuel use, not just that of coal. However, for comparison, in coal equivalent energy, this amounts to some 700 million tons, or about 20 quads, which amounts to 18 percent of the total projected energy use for the United States in that year. It would come to nearly 90 percent of current coal consumption.

Looking at the allocation of energy production among sources, the Kyoto Protocol is meant to increase preference for those energy sources that do not produce carbon dioxide and, secondarily, for those that produce much less than others. There will be a particular disincentive to use coal, since this source produces the most CO_2 per energy produced.

The most likely substitute for coal is natural gas, which, as noted earlier, releases about 55 percent of the amount of carbon dioxide that coal, on the average, does. In addition, it produces far fewer other pollutants, such as sulfates and polycyclic hydrocarbons, than coal and oil yield on combustion.

Current U.S. coal consumption is just under 1 billion short tons per year—second highest in the world—after China, which produces some 50 percent more. By 2010, the projected U.S. baseline energy case (in the absence of any attempt to meet Kyoto Protocol limits) would raise this level to 1.25 billion tons, in rough numbers. Coal currently accounts for about one-third of all United States carbon dioxide emissions. If the same patterns of energy source use were to hold in 2010—and if one wished to reduce the carbon dioxide emissions by 30 percent, while making no reductions in usage of other fossil fuels—this would mean reducing coal emissions by 90 percent. Such a scenario is clearly highly unlikely, even if one were to take much longer than 2010 to accomplish this goal.

Moreover, substituting 90 percent of coal with natural gas would reduce the level of emissions only by $0.90 \times 0.45 = 0.38$, which is still less than half way to the goal of reducing emissions by an amount equal to that produced by 90 percent of U.S. coal use.

Therefore, merely reducing coal use will not be sufficient to satisfy the Protocol. Any plan to comply with the Protocol needs to assume substitution, first by non-combustion energy sources—that is by renewables or nuclear energy—and second by natural gas. This would have to be accompanied by achievement of far greater efficiencies in energy production (for example by introduction of far more fuel-efficient steam gas turbines, driven by natural gas) and by more efficient use of energy.

The remaining possibility for the United States, under the Kyoto Protocol, would be to compensate for the excess of carbon emissions over the committed goal by either planting more trees, in the United States or elsewhere, or by purchasing carbon "pollution rights," as envisioned by the protocol, from other countries. How either of these schemes will work out is in some question. Both would require bilateral agreements with other countries, probably many other countries. The template of the trade in acid rain "pollution rights" to help all parties meet agreed–upon goals may not be a good analogy for carbon emissions, since acid rain, although international, is generally a regional, not a global problem. Further, carbon sources and sinks are not as well understood as are the sulfate and nitrate sources (chiefly coal) that are responsible for acid rain. This uncertainty will make it more difficult to achieve the international agreements necessary to make the "pollution trade" work as a widely-accepted convention, necessary due to the global nature of the problem.

Anthony Fainberg

See also: Coal, Production of; Coal, Transportation and Storage of; Environmental Problems and Energy Use; Fossil Fuels.

BIBLIOGRAPHY

Bethe, H. A., and Bodansky, D. (1989). "Energy Supply." In *A Physicist's Desk Reference,* ed. Herbert L. Anderson. New York: American Institute of Physics.

Borowitz, S. (1999). *Farewell Fossil Fuels.* New York: Plenum Trade.

Durant, W., and Durant, A. (1976). *The Age of Napoleon.* New York: Simon and Schuster.

Energy Information Administration, U.S. Department of Energy. (1996). *Coal Industry Annual 1996.* Washington, DC: U.S. Government Printing Office.

Energy Information Administration, U.S. Department of Energy. (1997). *Annual Review of Energy 1997.* Washington, DC: U.S. Government Printing Office.

Energy Information Administration, U.S. Department of Energy. (1997). *International Energy Database, December 1997.* Washington, DC: U.S. Government Printing Office.

Energy Information Administration, U.S. Department of Energy. (1998). *International Energy Outlook 1998.* DOE/EIA-0484(98). Washington, DC: U.S. Government Printing Office.

Garrity, J. A., and Gay, P., eds. (1972). *The Columbia History of the World.* New York: Harper & Row.

Howes, R. and Fainberg, A. eds. (1991). *The Energy Sourcebook.* New York: American Institute of Physics.

United Nations Framework Convention on Climate Change. (1997). *Kyoto Protocol.* New York: United Nations.

COAL, PRODUCTION OF

GEOLOGY OF COAL

Coal is a fossil fuel—an energy source whose beginnings can be traced to once-living organic materials. It is a combustible mineral, formed from the remains of trees, ferns, and other plants that existed and died in the tropical forests 400 million to 1 billion years ago. Over vast spans of time, heat and pressure from Earth's geological processes compressed and altered the many layers of trees and plants, slowly transforming these ancient vegetal materials into what we know as coal today. The several kinds of coal now mined are the result of different degrees of alteration of the original material.

It is estimated that approximately 0.9 to 2.1 m of reasonably compacted plant material was required to form 0.3 m of bituminous coal. Different ranks of coal require different amounts of time. It has been estimated that the time required for deposition of peat sufficient to provide 0.3 m of the various ranks of coal was: lignite, 160 years; bituminous coal, 260 years; and anthracite, 490 years. Another estimate indicates that a 2.4 m bed of Pittsburgh Seam (bituminous) coal required about 2,100 years for the deposition of necessary peat, while an anthracite bed with a thickness of 9.1 m required about 15,000 years.

Depending on the environment in which it was originally deposited, coal will have higher sulfur content when it was formed in swamps covered by seawater; generally, low-sulfur coal was formed under freshwater conditions. Although coal is primarily carbon, it's complex chemical structure contains other elements as well—hydrogen, oxygen, nitrogen, and variable trace quantities of aluminum, zirconium, and other minerals.

COAL RANK

Coal is a very complex and diverse energy resource that can vary greatly, even within the same deposit. The word "rank" is used to designate differences in coal that are due to progressive change from lignite to anthracite. Generally, a change is accomplished by increase in carbon, sulfur, and probably in ash. However, when one coal is distinguished from another by quantity of ash or sulfur, the difference is

A longwall shearer cuts swatches of coal 750 feet long and 28 inches thick, 800 feet underground at Consolidation Coal Company's Blacksville #2 mine. (Corbis-Bettmann)

said to be of grade. Thus a higher-grade coal is one that is relatively pure, whereas a higher-rank coal is one that is relatively high on the scales of coals, or one that has undergone devolatization and contains less volatile matter, oxygen, and moisture than it did before the change occurred.

In general, there are four ranks of coal: lignite, subbituminous, bituminous, and anthracite. Lignite, a brownish-black coal with generally high moisture and ash content, as well as lower heating value, is the lowest-rank coal. It is an important form of energy for electricity generation. Some lignite, under still more pressure, will change into subbituminous, the next higher rank of coal. It is a dull black coal with a higher heating value than lignite and is used primarily for generating electricity and space heating. Even greater pressure will result in the creation of bituminous, or "soft" coal, which has higher heating value than subbituminous coal. Bituminous coals are primarily used for generating electricity. Anthracite is formed from bituminous coal when great pressure developed during the geological process, which occurred only in limited geographic areas. Sometimes referred to as "hard" coal, anthracite has the highest energy content of all coals and is used for space heating and generating electricity.

In addition to carbon, all coals contains many noncombustible mineral impurities. The residue from these minerals after coal has been burned is called ash. Average ash content of the entire thickness of a coal seam typically ranges from 2 to 3 percent, even for very pure bituminous coals, and 10 percent or more for many commercial mines. These materials, which vary widely in coal seams with respect to kind, abundance, and distribution, among from shale, kaolin, sulfide, and chloride groups.

WORLD COAL RESERVES AND PRODUCTION

Coal is the most abundant and most economical fossil fuel resource in the world. Proven coal reserves exceed 1 trillion tons, and indicated reserves are estimated at 24 trillion tons. Coal is found in every continent of the world, including Antarctica, although the largest quantities of coal are in the Northern Hemisphere. Coal is mined in some sixty countries in nineteen coal basins around the world, but more than 57 percent of the world's total recoverable reserves are estimated to be in the United States, and China, which together account for more than two-thirds of the world's coal production.

COAL MINING METHODS

Depending on the depth and location of the coalbed and the geology of the surrounding area, coal can be mined using either surface or underground methods. In the United States, coal is usually mined underground if the depth of the deposit exceeds 200 ft.

In surface mining, the covering layers of rock and soil (called "overburden") are first removed using either a power shovel, a dragline (for large surface mines), or bulldozers and front-end loaders (for small mines). Front-end loaders also can be used to load coal. In large mines, coal usually is loaded using power shovels and hydraulic shovels. Depending on the size of the mine, shovels and draglines ranging from 4 cu m to 50 cu m are usually used for loading and excavating. Large-capacity haul trucks, usually in the range of 170 to 240 mt but possibly as big as 320 mt, are then used to transport coal to loading stations for shipping and sold as raw coal, or to a preparation plant for further processing. For post mining reclamation, draglines are used.

Depending on geologic conditions and surrounding terrain, there could be several types of surface mining. If the coal seam is of the same depth in flat or gently rolling land, area mining is developed where the overburden from one cut is used to fill the mined-out area of the preceding cut. Contour mining and mountaintop removal are methods that follow a coalbed along the hillsides. The overburden is cast (spoiled) down-

A dragline at Atlantic Richfield's Black Thunder strip mine in Gillette, Wyoming, loads coal into a dump truck. The coal is transported to cities around the country where it is burned to generate electricity. (Corbis-Bettmann)

hill from this first pit, exposing the coal for loading by trucks. The second pit could then be excavated by placing the overburden from it into the first pit. Digging starts where the coal and surface elevations are the same and precedes toward the center of a hill or mountain until the overburden becomes too thick to remove economically. An open pit combines the techniques of contour and area mining and is used where thick coalbeds are steeply inclined.

After coal is extracted, the pit is backfilled with earth and subsequently reclaimed or restored to its approximately original contour, vegetation, and appearance.

The use of underground mining methods requires integration of transportation, ventilation, ground control, and mining methods to form a system that provides the highest possible degree of safety, the lowest cost per ton of product, the most suitable quality of final product, the maximum possible recovery of coal, and the minimum disturbance of environment. Depending on the location of coal deposits, there can be three different types of underground mines: a drift mine is one in which a horizontal (or nearly horizontal) coal seam crops to the surface in the side of a mountain, and the opening of a mine can be made into the coal seam. Transportation of coal to the outside can be by track haulage, belt conveyor, or rubber-tired equipment.

A slope mine is one in which the coal is of moderate depth and where access is made through an inclined slope (maximum, 16°). This type of mining also may follow the coalbed if the coal seam itself is included and outcrops, or the slope may be driven in rock strata overlying the coal to reach the coal seam. Either a belt conveyor (no more than 30% grade), coal trucks (maximum grade, 18%), or electrical hoist if the slope is steep, can be used to transport coal out of the mine.

When the coal seam is deep, a shaft mine is used because the other two types of access are cost-pro-

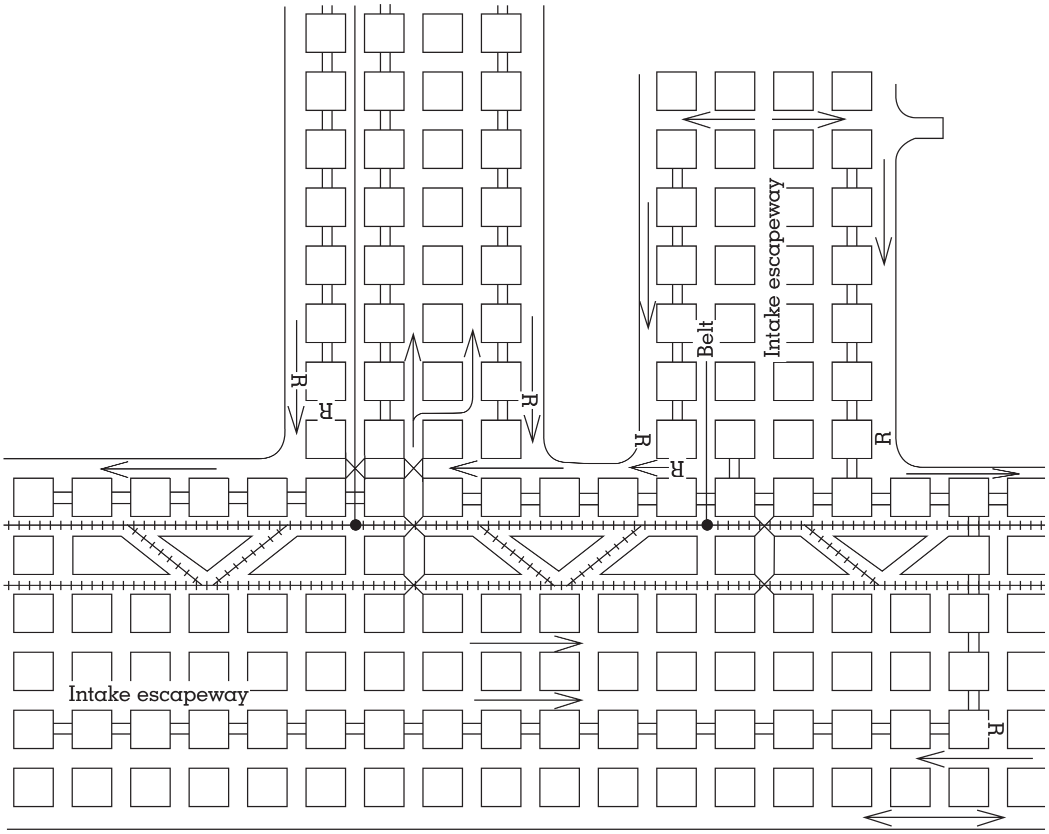


Figure 1.
A room-and-pillar mining system. Pillars are left behind in the rooms to support the roof.

hibitive. Vertical shafts are drilled for both production and ventilation.

Production methods underground are generally classified according to the types of mining equipment used (conventional, continuous mining, or longwall) or by the method in which coal is being extracted (longwall or longwall caving). Both conventional and continuous mining are room-and-pillar systems; even the longwall method uses room-and-pillar during development.

In the conventional mining system, the coal face is first undercut, center cut or top cut using a cutting machine that most nearly resembles a large chain saw on wheels. The outlined coal blocks are drilled in a predetermined drill pattern using a mobile powered drill, with holes charged with explosives, and the coal is dislodged. The broken coal is gathered by a loading machine onto a shuttle car and dumped onto a nearby belt, to be transported out of the mine.

In the continuous mining method, a continuous mining machine (also referred to as a continuous miner) is employed in the extraction process. This machine combines several extracting functions into one continuous process: cutting, loading, and tramming, thereby tearing the coal from a seam and automatically removing it from the area by a machine-mounted conveyor onto a shuttle car, which is used to transport the mined coal to a dumping station, then transported out of the mine using a conveyor belt. Remote-controlled continuous miners allow an operator to control the machine from a distance, increasing safety. The mine roof is further secured using wooden timbers; steel crossbars on posts; or, most commonly, roof bolts.

Both conventional and continuous mining methods use a room-and-pillar system in which the coal is mined by extracting a series of “rooms” into the coalbed, and leaving “pillars,” or columns, of coal to help support the mine roof (Figure 1). Depending on the location, the rooms are generally 20 to 30 ft. wide

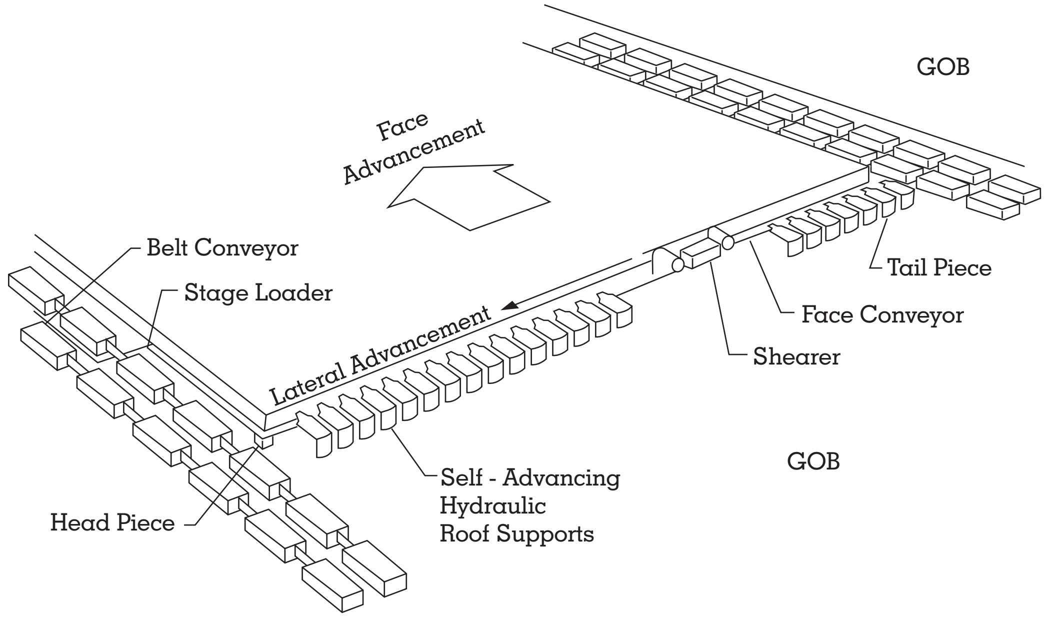


Figure 2.
A typical setup. The panel is 8,000–10,000 feet long. The face is 800–900 feet wide.

and the pillars 20 to 90 ft. wide, with the height determined by the thickness of the coal seam. In the not-too-distant future, robotic versions of these machines, now under development, will allow for enhanced automatic operations and even greater efficiencies than now possible. Although still utilized in stand-alone production operations, continuous miners also are employed for main entry and longwall panel developments.

As a rule of thumb, 50 to 55 percent of coal can be extracted using continuous mining. To improve this extraction ratio, a pillar-recovery process usually is applied when mining reaches the end of the panel and the direction of the mining is reversed. The continuous miner mines into the pillars, recovering as much coal as possible, as the roof is allowed to systematically collapse. Usually this can increase the extraction ratio by up to 5 percent.

Although the development of the continuous mining system in the 1950s consolidated several operations in one machine and have greatly improved coal production, it is still not fully "continuous," as the face haulage and roof support operations remain as major impediments to truly continuous production.

The introduction of the longwall system has provided not only continuous cutting and loading but also continuous haulage and roof support. In the longwall mining system, large blocks of coal, outlined in the development process, are completely extracted in a single, continuous operation.

The longwall consists of a panel of coal, usually 8,000 to 10,000 ft in length and 800 to 900 ft in width (Figure 2). In the face area, a rotating drum (or a plow) is dragged mechanically back and forth across a wide coal seam. The loosened coal falls onto a conveyor for removal from the mine. The system has its own hydraulic roof supports, which advance with the machine as mining proceeds. The supports provide not only high levels of production but also increased miner safety. Newer versions of the longwall system employ sensors to detect the amount of coal remaining in the seam being mined, as well as robotic controls to enhance efficiency. As the face advances after the coal is mined, the roof is systematically allowed to cave behind to form a gob. In general, longwall systems can provide an extraction ratio of up to 80 percent.

Longwall mining has helped revolutionize underground coal mine operations in the past two decades, with its share of total U.S. underground production increasing from 10 percent to 48 percent, surpassing continuous mining tonnage in 1994, and the trend has held true since then.

Several modified versions of longwall methods also are practiced in areas where the coal seam is either thick or steeply inclined. Since the maximum height a shearer can reach is about 14 ft, thicker coal seams have to be mined using either a multiple pass method, where the top seam is mined followed by a lower pass, or a longwall caving method, where the lowest seam is mined using the traditional longwall method and the upper portion of the seam is allowed to cave under gravity; coal is then collected behind the shield support and shipped out of the mine.

COAL MINING AND THE ENVIRONMENT

It has been said that mining is a temporary use of the land. While mining does disturb the land, modern technologies and increased application of environmentally safe mining methods in the United States and other major mining countries have enabled today's coal mining industry to provide the valuable energy resources modern society requires without destroying the environment in the process. In the United States, stringent environmental regulations mandate specific standards for reclamation, quality of water discharge, and other mining practices that may disturb the land.

While some problems still exist, there is no question that coal mining operations are more efficient and safer for workers and leave less of an environmental footprint than operations several generations ago. As society's demand for energy from coal continues to increase and as coal's price declines (between 1978 and 1996 U.S. mine mouth prices fell from $47.08 to $18.50 per ton in constant 1996 dollars), there is certain to be even greater efforts to limit the environmental impact of mining operations.

Jerry C. Tien

See also: Coal, Consumption of; Coal, Transportation and Storage of.

BIBLIOGRAPHY

Coleman, L. L. (1999). *International Coal*, 1998 ed. Washington, DC: National Mining Association.

Fiscor, S. (1998). "U.S. Longwall Thrive" *Coal Age*, February, pp. 22–27.

Hower, J. C., and Parekh, B. K. (1991). "Chemical/Physical Properties and Marketing." In *Coal Preparation*, 5th ed., ed. J. W. Leonard and B. C. Hardinge. Littleton, CO: SME.

Katen, K. P. (1982). "Modern Mining Methods—Longwall, Shortwall," In *Elements of Practical Coal Mining*, 2nd ed., ed. D. F. Crickmer and D. A. Zegeer. New York: Society of. Mining Engineers/American Institute of Mining and Metallurgy.

National Mining Association. (1998). *Facts about Coal: 1997–1998*. Washington, DC: Author.

Reid, B. (1998). "Longwall Production at Record Pace." *Coal Leader* 32(9):1.

Schroder, J. L., Jr. (1982). "Modern Mining Methods—Underground." In *Elements of Practical Coal Mining*, 2nd ed., ed. D. F. Crickmer and D. A. Zegeer. New York: Society of Mining Engineers/American Institute of Mining and Metallurgy.

Simon, J. A., and Hopkins, M. E. (1982). "Geology of Coal," In *Elements of Practical Coal Mining*, 2nd ed., D. F. Crickmer and D. A. Zegeer. New York: Society of Mining Engineers/Institute of Mining and Metallurgy.

Stefanko, R. (1983) "Geology of Coal." *In Coal Mining Technology: Theory and Practice*, ed. C. Bise. New York: Society of Mining Engineers/American Institute of Mining and Metallurgy.

Thakur, P. (1997). "Methane Drainage from Gassy Mines." In *Proceedings, Sixth International Mine Ventilation Congress*, ed. R. V. Ramani. Littleton, CO: SME.

Tien, J. C. (1998). "Longwall Caving in Thick Seams". *Coal Age*, April, pp. 52–54.

COAL, TRANSPORTATION AND STORAGE OF

Coal competes primarily in the market for low-cost boiler fuels. Coal is also characterized by a relatively low energy content per unit of weight (at best two-thirds that of residual oil). Consequently, low-cost and efficient transportation is essential to the competitiveness of coal.

OCEAN TRANSPORTATION

World trade in coal totaled 576 million tons (524 million tonnes) in 1998, of which 523 million tons (476 million tonnes) shipped in oceangoing vessels. Coal shipments use the same dry bulk vessels that transport other bulk commodities, such as iron ore and bauxite, so vessel rates for coal shipments are hostage to wider market forces. However, the cyclic pattern observable in vessel rates disguises the long-term trend in which rates have varied little in nominal terms. For example, spot vessel coal rates in the 1998–1999 time period were about the same as in the mid-1980s, varying between $5 and $10 per ton.

Coal is generally shipped either in vessels capable of transversing the Panama Canal (Panamax vessels of 60,000 dwt) or Capesize carriers of 200,000 dwt and greater. Vessels may be designed for self-unloading or be "gearless" carriers that require onshore bulk-handling equipment.

BARGE TRANSPORTATION

Coal-carrying barges move in tows of fifteen to forty barges, pulled by a single towboat of 2,000 to 10,000 hp. A "jumbo"-size barge carries 1,800 tons (1,633 tonnes) of coal, so a large tow can move 72,000 tons (65,304 tonnes) of coal, as much as five unit trains. These large volumes result in significant economies of scale. Barge rates can run (on a cost-per-mile or cost-per-kilometer basis) a quarter or less of rail rates.

The primary cost variable in barge shipments is fuel; a midsize towboat can consume 5000 gal (18.9 kl) of diesel fuel daily. Barge shipments are also dependent on weather conditions; low water or frozen rivers and canals can halt shipments. As with ocean vessels, the barges that move coal also ship other bulk commodities, making the rates and availability of barges for coal shipments dependent on conditions in other markets. Backhauls (i.e., shipment of one commodity to a terminal and return with a different product) can substantially reduce coal rates.

Barges receive coal at a dock to which the coal is initially transported by rail or truck. These transloading facilities can play an important role in the coal supply system as intermediate storage sites and by providing facilities where different coals can be blended into a custom product.

A cargo barge travels up the Rhine. Though weather conditions can affect timeliness, the economies of scale usually make water transport cheaper than rail. (Corbis Corporation)

RAIL TRANSPORTATION

Rail-transported coal is typically moved in unit trains that operate in dedicated shuttle service between a mine and a destination. Unit trains operating in the western United States and Canada consist of 100 to 120 lightweight aluminum railcars carrying upward of 121 tons (110 tonnes) of coal apiece, or more than 14,000 tons (12,700 tonnes) per train. In the 1990s, distributed power (DP) came into widespread use in the western United States. In this system a remotely controlled engine is put into the middle of a train, allowing greater traction and control of train motion. DP trains can consist of 135 cars and are the most efficient method of rail transportation of coal.

Railroad productivity has increased dramatically since the mid-1970s. In part this reflects reform of outdated labor practices, but the technical sophistication of the rail industry is rarely appreciated. Modern systems use microwave Centralized Traffic Control (CTC) systems to move trains safely with minimal between-train clearance, allowing substantial increases in system capacity. Modern diesel electric locomotives rely on microprocessor and alternating-current motor technology to provide enhanced power (6,000-hp class) and greater traction, allowing two or three engines to do the work of five earlier models. All aspects of the rail system, from operations to invoicing, are heavily dependent on computer processing.

Because rail systems exhibit economies of scale, there is a tendency toward consolidation manifested either in state ownership or merger of privately owned systems into a handful of competitors. High barriers of entry into the rail business allow the exercise of monopoly power over rates to customers with single-rail access. This has been a persistent issue in nations with deregulated rail industries, such as the United States and Canada. Rate complaints by coal mines and consumers have been common in these countries. On the other hand, railroads have had difficulty earning adequate returns on investment, due to intermodal competition, heavy capital investment requirements, and other factors. The tension between shipper demands for low rates and high-quality service, and railroad efforts to improve profitability, was a political controversy in the nineteenth century and continues to be an unresolved public policy issue in the early twenty-first century.

TRUCK TRANSPORTATION

Truck transportation is used to move coal to a transloader for placement onto a water or rail carrier, or for direct shipment to the customer. Trucks have the advantage of routing flexibility and modest capital requirements, but coal can be economically transported for at most about 100 miles (160 km) one-way or less, due to the high unit cost of moving a low-value product in relatively small batches.

Coal-carrying vehicles are typically end-dump trucks with a carrying capacity of roughly 25 to 50 tons (22.7 to 45.4 tonnes), depending on local road conditions and safety regulations. In the 1990s, strides were made toward increasing the productivity of truck operations, such as higher-capacity vehicles. But while these improvements have enhanced the ability of trucks to compete with railroads for short hauls, they have not significantly increased the maximum radius within which truck shipments are economical.

COAL SLURRY PIPELINES

Coal slurry pipelines have been widely discussed, but few slurry pipelines have been built. In addition to the Black Mesa operation in Arizona, a 38-mile (61-km) pipeline was built by the Soviet Union, and a 108-mile (173-km) pipeline in Ohio was mothballed in 1963 after six years of operation. It is arguable to what extent the limited use of slurry pipelines is due to economics or to political opposition from rail carriers and interests concerned with water rights.

The most successful slurry operation is the dedicated pipeline that serves the 1,580-MW Mohave Generating Station in southern Nevada. The plant receives all of its coal via a 273-mile (437-km) pipeline built in 1970 that originates at the Black Mesa mine in Arizona. Coarsely ground coal is mixed with water (the slurry is about 47% solids by weight) and pumped through an 18-inch (46-cm) pipe. At the plant the coal is dewatered using centrifuges. The pipeline has a capacity of about 5 million tons (4.5 million tonnes) annually.

STORAGE AND OXIDATION

Storage is necessary at several points in the coal supply chain. Because coal is transported in batches (e.g., a unit train or a vessel), rather than moved continuously through a network, like natural gas, the supply chain must accommodate surges and lulls in demand at the mine; at the origin and receipt dock or port for water shipment; and at the end user, such as a power plant. The global wave of privatization and deregulation, particularly in the electric sector, has increased pressure on logistics managers to make the coal supply chain as seamless as possible to minimize the amount of coal in storage at any time. Stored coal ties up working capital and, as discussed below, can deteriorate and create safety hazards

About 2.5 million tons (2.3 million tonnes) of coal are burned daily in U.S. power plants. This is equivalent to roughly 21,000 railcars in transit, so it is apparent that coordinating production and consumption is no easy task. Accidents, rail strikes, natural disasters (e.g., floods that take out bridges and rail lines) and severe weather (e.g., deep river freezes that halt barge traffic) can all severely disrupt deliveries for utility customers dependent on a reliable coal supply for base load plants. Nonetheless, to reduce costs U.S. utilities have significantly reduced typical inventory levels over time. Whereas a coal inventory of ninety days of supply was once typical, inventories now frequently run in the range of thirty to forty-five days.

Another reason to keep inventories low is the potential for storage problems. Coal in storage must be carefully handled. Improperly stored coal can oxidize (weather), causing a loss of heat content. And if heat is allowed to build up in a stagnant coal pile (or in a vessel, barge, or railcar), the coal can self-ignite.

Self-ignition is particularly a risk with lower-grade subbituminous coals and lignite. To avoid oxidation, coal piles should be turned frequently so that heat can vent, and piles should be packed and shaped to minimize surface exposure.

Stan M. Kaplan

See also: Coal, Consumption of; Coal, Production of; Locomotive Technology; Transportation, Evolution of Energy Use and.

BIBLIOGRAPHY

Fieldston Company, Inc. (1996). *Fieldston Coal Transportation Manual: 1996–97.* Washington, DC: Author.

Schmidt, R. D. (1979). *Coal in America: An Encyclopedia of Reserves, Production, and Use.* New York: McGraw-Hill.

Singer, J. G., ed. (1981). *Combustion: Fossil Power Systems.* Windsor, Conn.: Combustion Engineering Company.

COGENERATION

Cogeneration is the production of two useful forms of energy in a single energy conversion process. For example, a gas turbine may produce both rotational energy for an electric generator and heat for a building.

During the energy conversion process, an energy converter converts some form of energy to a form having a more suitable use. A light bulb and a gasoline engine are two familiar converters. People invest in electric energy to operate a light bulb because light is useful; likewise, people invest in gasoline for energy to run the automobile internal combustion engine because automobiles are useful. The laws of nature require that there be no loss of energy in the conversion. If 100 joules of energy are converted, then 100 joules remain after the conversion. However, the laws of nature neither require the converted energy to be in the form we desire, nor do they require that the other forms be useful. If the converter of 100 joules were a light bulb, only about 10 joules would emerge as light. The other 90 joules would be heat. Touching an ordinary light bulb when lit attests to the heat that is produced.

Efficiency is a practical measure of the performance of a converter: efficiency is equal to the desired form of energy divided by the total energy converted. If the light converted 100 joules of energy into 10 joules of light energy, we would say its efficiency is $10 \div 100 = 0.1$ or 10 percent.

Heat is always produced to some extent in energy conversion. In fact, when energy has gone through all possible conversions, it ends up as thermal energy in the environment. The efficiency of a steam turbine at a large electric power plant is about 50 percent. This means that 50 percent of the energy converted is rejected as heat to the environment by massive cooling towers that are prominent at power plant sites. Heat is a useful energy commodity, so one must wonder why rejected heat is not put to some use. The idea of cogeneration is to do just that.

Evaluating the practical worth of thermal energy in a substance such as water requires consideration of both temperature and the amount of the substance. To understand this we say the thermal energy of a substance is equal to the number of molecules times the energy per molecule. The thermal energy per molecule (i.e., the second factor) increases with increasing temperature. So, even if the temperature is high, making energy per molecule larger, the total will still be small if there are only a few molecules. Similarly, if the temperature is low and a large number of molecules is involved, the total thermal energy can be large. The temperature of the water removing heat from a steam turbine is relatively low—only 10 to 15°C above the temperature of the environment—but a huge amount of water is needed to remove the heat from the turbine, so the thermal energy transferred to the water must be quite large. The thermal energy, although low-grade (about 80°F or 30°C), is appropriate for heating buildings. In a scheme of relatively small scale called district heating, buildings are heated in some towns and cities. But usually a power plant, especially a nuclear power plant, is well removed from the city and the economics of piping the heat to where it is needed is very unfavorable—requiring not only longer runs of piping, but resulting in greater heat loss from those longer runs. Consequently, for remotely sited plants, the thermal energy is rejected to the environment and goes unused.

Industry needs both electricity and heat. It is possible for an industry to produce its electricity from gas-fired turbogenerators and use the rejected heat for industrial purposes. The rejected heat can be at relatively high temperature, making it more useful if some sacrifice is made in the efficiency of the turbogenerator. It is in areas like this that cogeneration has its greatest potential and one sees commercial cogeneration enterprises evolving to provide a growing share of energy production.

Joseph Priest

BIBLIOGRAPHY

Horlock, J. H. (1997). *Cogeneration—Combined Heat and Power: Thermodynamics and Economics*. Malabar, FL: Krieger Publishing Company.

Spiewak, S. A., and Weiss, L. (1994). *Cogeneration and Small Power Production Manual*, 4th ed. Liburn, GA: Fairmont Press.

COGENERATION TECHNOLOGIES

Cogeneration or combined heat and power is the simultaneous production of heat and power in a single thermodynamic process that has a history going back several centuries. Originally employed to save labor, its inherent fuel economy took it to the forefront of the industrial revolution in the nineteenth century. More recently the environmental benefits derived from reduced fuel consumption have made cogeneration a significant factor in global environmental strategies, while current trends towards utility deregulation and distributed power generation continue to bolster the market for this technology.

Cogeneration encompasses several distinct thermodynamic processes of simultaneous heat and power production. One utilizes air as a medium, another steam, a third employs heat rejected from a separate combustion process, such as an internal-combustion engine, and a fourth utilizes a thermochemical process such as found in a fuel cell. Although each process is distinct, they are often combined together to maximize the energy production in a single thermodynamic system.

The oldest form of combined heat and power is the smokejack, developed in Tibet to turn prayer wheels during religious ceremonies. Captured Tartar slaves introduced this device into Europe by the early fourteenth century and Leonardo da Vinci sketched one around 1480. Commentators as diverse as Montaigne (1580), John Evelyn (1675), and Benjamin Franklin (1758) mention smokejacks, which were small windmills installed inside a chimney and powered by the hot air rising from fires. The rotary motion of the fan was used to power a spit or lathe. The amount of power produced would be dependent on the velocity and mass flow of the heated air and the efficiency of the blades, but in general use the smokejack delivered approximately one dog-power. Turnspit dogs were specifically bred to turn spits and other apparatus requiring rotary motion, although children, slaves, and servants were also pressed into this labor, which was basically a larger version of a hamster in a wheel. Prior to the widespread electrification of farms in the mid–to–late twentieth century, American farms often had similar devices allowing all members of the farm community to contribute to the domestic workload, reminding us that the current leisurely life of our canine friends is a relatively recent phenomenon.

Franklin also noted that the natural draft of a chimney was also able to turn a smokejack, an idea recently promoted on a generative power using a large natural draft chimney with an air turbine. In 1832, Charles Busby used a smokejack to power a pump to circulate water through pipes for “warming and cooling the interior of buildings.” It is uncertain if Busby’s “Patent Circulation” achieved wide success, although it would have worked well since the flow of exhaust air through the chimney would be directly related to the amount of circulated needed.

By the end of the nineteenth century, smokejacks had evolved into the hot air turbine, which found application as aircraft turbosuperchargers before evolving into gas turbines. Some engines such as the General Electric LM2500 have a separate hot air power turbine that converts hot exhaust air into mechanical power.

An interesting use of air involved the use of compressed air for power distribution in urban areas. Still widely used as a power source within factories, several cities in the mid- to late-nineteenth century had compressed air public utility systems, with Paris being perhaps the largest example. Simple air motors could be installed on a wide variety of equipment, and at least one clothing factory utilized the air exhausted from sewing machine motors to provide ventilation for the sewing machine operators, which also provided cooling since the air expanded passing through the motor.

Despite the widespread use of smokejacks, the industrial revolution could not be sustained by power measured in dog units. Although water and wind had

Electric power cogeneration site. (Greenpeace)

been useful sources of energy for many centuries, they were geographically limited and offered limited, usually seasonable, availability. The solution was to harness the power of steam, of which the earliest example is Heron's rotating steam engine in ancient Greece. Steam power grew from these small-scale applications to larger uses during the eighteenth century, when Savery and Newcomen introduced larger steam engines to pump water out of mines. Savery's engines operated with low pressure (<2 psig) steam and a thermal efficiency of roughly one percent. Large steam engines required large boilers, which in turn created more steam, which Desaugliers in 1720 adapted to manufacturing applications.

James Watt doubled the efficiency of the steam engine by introducing the separate condenser in the 1760s and created a new unit, the horse power, to measure its output. In 1784, Oxford brewer Sutton Thomas Wood obtained a patent for using waste steam from an industrial process to drive a steam engine and also to use the exhaust steam or hot water from a steam engine for heating or manufacturing, marking the first cogeneration patent.

Despite Watt's contribution to the advancement of the steam engine, his reliance on low pressure steam resulted in safe, but large and inefficient engines. Only after his patent rights expired were other inventors, such as Richard Trevithick in England and Oliver Evans in Philadephia, able to build and market high pressure engines that were more efficient, much smaller, and indeed more dangerous than their low pressure ancestors. Higher pressure steam made it possible to operate these engines economically without a condenser, but the exhaust steam was rapidly put to useful purposes.

The idea of cogeneration, fortunately, slowly found its way into many factories. In the mid 1820s a religious community led by George Rapp built a utopian community, Old Economy, on the Ohio River outside Pittsburgh. The Evans engine that had powered their steamboat was reinstalled in a mill and the engine exhaust steam was distributed through pipes to warm the community's buildings in an early example of district heating. By mid-century exhaust steam was widely used in industrial settings both in Britain and America, and English sanitary reformer Edwin Chadwick proposed using the waste heat from factory engines to heat public baths and worker housing. Private entrepreneurs built several such baths, charging a penny for admission with discounts for frequent bathers.

Interestingly, one of the first and greatest works on thermodynamics, Sadi Carnot's 1824 *Reflexions sur la Puissance Motrice du Feu (Reflections on the Motive Power of Heat*) has been one of the foremost impediments to cogeneration practice. Carnot stated flatly that the efficiency of a heat engine was strictly a function of the temperature difference across the engine, which is entirely correct insofar as the engine itself goes but ignores the common situation where the engine is not acting alone in an isolated process. Take, for instance, two Carnot cycle engines, each with the same high temperature condition. One has a low temperature reservoir such as that produced by a large body of water. The reservoir in the other is at a higher temperature, but the heat in the reservoir is then used for a useful purpose, such as space heating. While the first engine has the higher engine efficiency as Carnot postulated, the second might have an overall system efficiency four or five times higher than the first engine. Many students of thermodynamics even today continue to be misled by Carnot's useful, but limited, theories.

As factories grew in size and complexity, managers began to more closely analyze the various cost elements of their businesses. Many enterprises requiring large amounts of power, such as cotton mills, located their factories based on the availability of water power, but a lively public discussion on the relative costs of steam and water power took place in the early 1840s. Several mill owners reported their surprise at discovering that steam power was often no more expensive than water power and often less so, particularly where the exhaust heat of the engine could be utilized for process or space heating. Steam was also much more dependable, and did not rely on the vagaries of weather and excessive demand on available water supplies.

Post–Civil War America witnessed a tremendous growth in steam power, including the 1882 introduction of the central electric station by Thomas Edison. These stations were initially thought of as curious novelties rather than solid investments, a view undoubtedly shared by many early electric pioneers trying to make a profit. Ever mindful of the direct correlation between reducing waste and increasing profits, many electric light plants sold steam and hot water to residential and commercial customers, in many cases making the difference between a profit and loss. In Pennsylvania, state lawmakers regarded cogeneration as unfair competition to traditional utilities and passed a bill outlawing it, although it was quickly repealed when they discovered how widespread it was. By World War I more than four hundred central stations in the United States utilized cogeneration, most small enough to measure their electric power output in kilowatts.

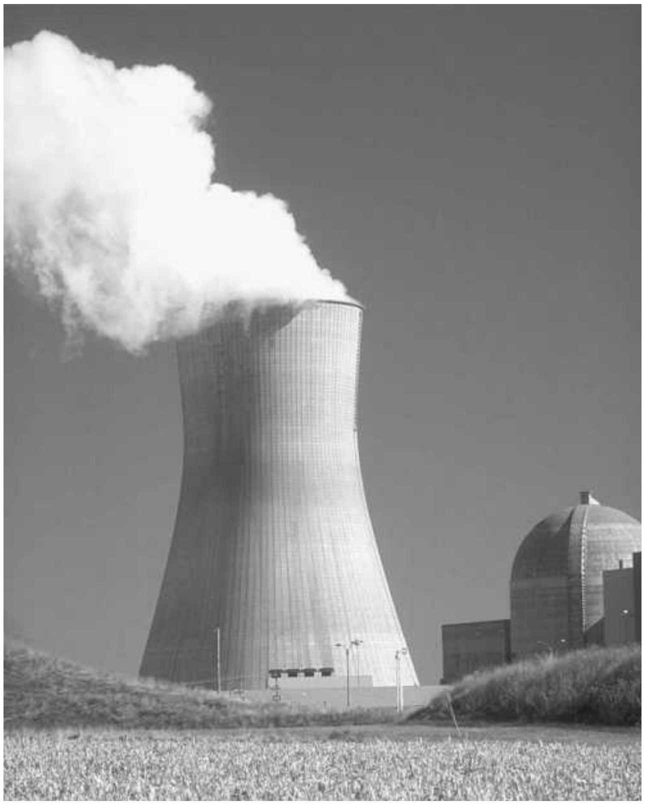
The introduction of large steam turbines, higher boiler pressures and temperatures, and the effects of a war–time energy crisis in America led most utility managers to prefer large power plants located close to coal mines, obviously reducing opportunities to utilize the plant's heat output. Several utilities, however, built plants that could also supply heat in dozens of cities, including New York, Philadelphia, Boston, and Denver, and many utilities supplied heat in order to keep their customers from installing their own small cogeneration plants.

In Europe, industrial use of cogeneration greatly expanded between the wars, while in the new Soviet Union large-scale combined heat and power plants were incorporated into the formal national planning process as the preferred power plant technology. Beginning with the 1930 World Power Conference in Berlin, a lively debate took place for a decade between Soviet engineers promoting large-scale heat and power systems and German engineers arguing for smaller distributed cogeneration schemes. Steam turbines remain the largest source of thermal electric power today and their fuel flexibility and easy adaptability allows them to be commonly employed in cogeneration schemes of all sizes.

The introduction of the internal combustion engine in the late nineteenth century opened up an entirely new approach to combined heat and power. Rather than using the same fluid for the heat and power process as was the case with hot air and steam processes, the tremendous waste heat generated by the internal combustion process can easily be transformed into useful heat. Cogeneration applications using stationary engines were common in Europe prior to World War I and remain quite popular because the heat is relatively easy to capture and transport. The most widespread application of internal combustion cogeneration is most certainly the automobile heater, which was widely introduced in the mid-1920s and is today virtually ubiquitous. This almost always utilizes hot water as a heat transfer medium, although many owners of the Volkswagen Beetle enjoyed the simplicity of hot air heat from its air-cooled engine. Surprisingly, experts continue to predict the imminent demise of the internal combustion engine due to its thermal inefficiency, failing to recognize that cold weather instantly turns it into a very efficient cogeneration plant and that its touted replacement, the electric car, is essentially useless in cold climates. The hot exhaust from internal combustion engines can also be utilized for useful heat and power purposes using a turbosupercharger or heat recovery apparatus.

During the early twentieth century the hot air turbine was reborn as the gas turbine for stationary power applications but was adapted to aircraft propulsion on the eve of World War II. Heat was being recovered from operating stationary gas turbines before the 1950s and both aeroderivative and industrial gas have undergone substantial development in the last half of the twentieth century. Initially used primarily in industrial applications, small gas turbines (along with their internal combustion cousins) enjoyed a large market for “total energy” applications in the 1960s and 1970s. A total energy plant was designed to operate without a connection to the electricity utility grid, either because it was inaccessible or uneconomical. Many American schools, hospitals, and shopping malls had such plants, most of which were eventually connected into utility grids.

The 1978 Public Utility Regulatory Policy Act (PURPA) introduced the word cogeneration, and required utilities to interconnect with and buy power from cogeneration systems at their avoided incremental production cost. This led to a rapid growth in cogeneration capacity in the United States from about 10,000 Mwe in 1980 to almost 44,000 Mwe by 1993. Most new capacity was installed at large industrial plants such as petroleum refineries, petrochemical plants, and paper mills. But PURPA also had some negative impacts that have given cogeneration a bad reputation. A few states required utilities to buy excess power from cogenerators at very high rates, which later proved to be in excess of incremental market costs. Also, a few systems were optimized for electric production rather that overall efficiency, with minimal use of the “waste heat” in order to qualify for the high power buyback rates established under PURPA. But there are only a few documented cases of such practices.



Steam pours from the cooling tower at a nuclear power plant in Stedman, Missouri. (Corbis-Bettmann)

Another cogeneration technology with roots in the nineteenth century is the fuel cell, which was first described by Sir William Grove in 1839 but only in the past few years become a viable competitor in the marketplace, largely due to development funded first by the American space program and more recently by the U.S. Department of Energy and Electric Power Research Institute. Although still relatively expensive compared to other generation technologies, fuel cells have been widely demonstrated as reliable and clean energy producers. A few companies plan to introduce and market fuel cell cogeneration systems in the next several years. Their principal competitor appears to be small microturbines and internal combustion engines.

The United States obtained about 9 percent of its electricity from combined heat and power (cogeneration) systems as of 1997. Cogeneration is more prevalent in some European nations than in the

United States, with Germany, the Netherlands, and the Czech Republic obtaining 15 percent or more of their electricity from cogeneration facilities. Denmark leads the world and gets 40 percent of its electricity from combined heat and power systems

Cogeneration, by whatever name it is known, has survived and prospered for centuries because of its adaptability and inherent fuel efficiency. Many combined heat and power plants consistently achieve an overall thermal efficiency of 80 to 90 percent, more than three times more efficient than average utility plants and 50 percent higher than the newest and most efficient combined cycle gas turbine power plant. Cogeneration can be done with any fuel that can be burned, or even without burning a fuel as in Iceland where steam cogeneration is employed in geothermal fields to extract power from the steam before it is transported several miles as hot water to heat cities. Several nuclear cogeneration plants have been built with varying degrees of success, including an air-transportable model designed for remote military radar stations.

The current movement to deregulate electric utilities has created significant new opportunities for all three combined heat and power markets: industrial, district heating, and individual buildings. In the United States nearly 10,000 Mwe of new cogeneration capacity installed during 1994 to 1996, primarily involving large gas turbine plants located at industrial facilities. A number of European countries are actively cogeneration as a means to reduce global warming and to reduce fuel poverty in lower income households. In the United States, many organizations are working to reduce the institutional and regulatory obstacles to cogeneration, so that adoption of cogeneration can continue to expand. The U.S. Department of Energy has set a goal of doubling cogeneration capacity in the United States by 2010.

Morris A. Pierce

See also: Cogeneration; District Heating and Cooling; Fuel Cells; Industry and Business, Energy as a Factor of Production in; Turbines, Gas; Turbines, Steam.

BIBLIOGRAPHY

Hirsh, R. F. (1989). *Technology and Transformation in the American Electric Utility Industry*. Cambridge, Eng.: Cambridge University Press.

Hirsh, R. F. (1999). *Power Loss: The Origins of Deregulation and Restructuring in the American Electric Utility System*. Cambridge, MA: MIT Press.

COKING

See: Coal, Production of

COMBUSTION

"Combustion" is a term often used synonymously with "burning." However, a distinction can be made that explains why combustion is more than just burning. To burn something is to set it on fire. To combust something is to subject the material (or fuel) to the process of rapid oxidation that leads to the consumption of both the material (or fuel) and the oxidizer (usually the oxygen in air) with the release of heat and light. (Usually the oxidizer is oxygen but there can be nonoxygen species, that under certain circumstances fit the definition of an oxidizer being a substance that can accept electrons in a chemical reaction.) Fires and burning involve combustion, but not all combustion involves fire in the form of visible, hot flames. There are flames that are invisible but release heat, and there are flames that emit light but have so little evolution of heat that they are called "cool flames." By making this distinction between burning and combustion, many features of combustion such as ignition, extinction, and flames can each be discussed separately from a scientific perspective.

Combustion is the entire process by which something is oxidized. It is part of the use of gasoline or diesel fuel in automobiles and trucks, as well as part of propulsion in aircraft either in jet engines or propeller engines. This latter association is so often made that the propulsive devices in aircraft are called combustors. Similarly, furnaces and boilers, that often involve flames for the production of heat, are combustion devices involving many of the elements of the complete process. Incinerators, too, are commonly associated with combustion of fuel in the form of waste materials. Other common manifestations of combustion are house, forest, and chemical fires;

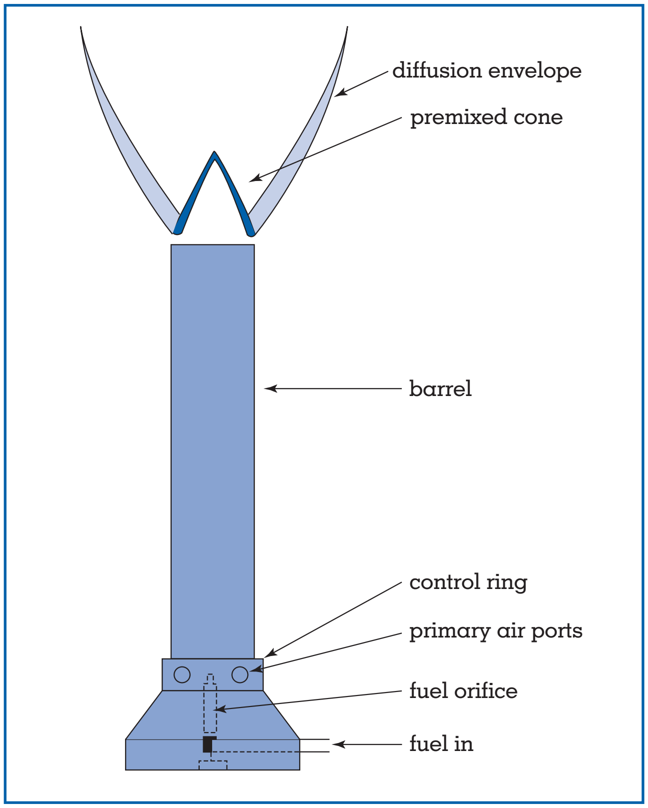


Figure 1.
A Bunsen burner.

explosions of flammables; and air pollution from cars and incinerators.

Because of the beneficial and adverse aspects of combustion, it is necessary to better understand it (i.e., all the components in the process of oxidizing a fuel with attendant heat and light).

COMBUSTION SCIENCE

Flames

A flame is a thin region of rapid, self sustaining oxidation of fuel that is often accompanied by the release of large amounts of heat and light. Flames are what we most commonly associate with combustion. One part of combustion science focuses on the different ways flames can be formed and the scientific and practical consequences of each.

Premixed Flame. For this type of flame, the fuel and oxidizer—both gases—are mixed together before flowing to the flame zone (the thin region of the flame). A typical example is the inner core of a Bunsen burner (Figure 1), or combustion in an automobile engine cylinder. (In general, a burner is the part of the combustion device that supplies fuel and sometimes air and where the flame is produced and stabilized.)

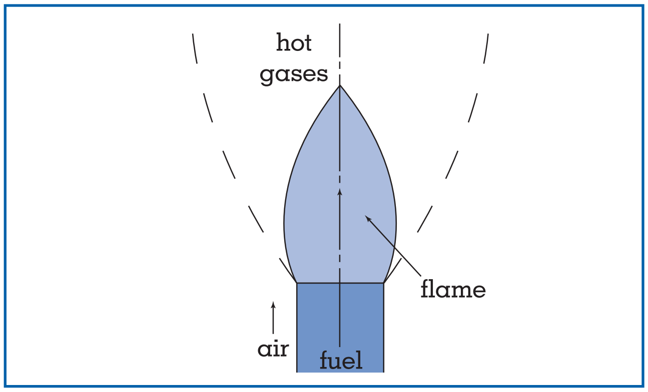


Figure 2.
Laminar diffusion flame jet.

Diffusion Flame. When the fuel and oxidizer are initially unmixed and then mix in a thin region where the flame is located, the flame is called a diffusion flame (Figure 2). The word diffusion is used to describe the flame because the fuel and oxidizer are mixed on the molecular level by the random thermal motion of the molecules. An example of a diffusion flame is a candle flame or flares at an oil refinery.

Laminar Versus Turbulent Flames. Premixed and diffusion flames can be either laminar or turbulent gaseous flames. Laminar flames are those in which the gas flow is well behaved in the sense that the flow is unchanging in time at a given point (steady) and smooth without sudden disturbances. Laminar flow is often associated with slow flow from small diameter tubular burners. Turbulent flames are associated with highly time dependent flow patterns, often random, and are often associated with high velocity flows from large diameter tubular burners. Either type of flow—laminar or turbulent—can occur with both premixed and diffusion flames.

Droplets/Sprays. Flames can also be established with fuels that are initially liquids. A typical example is the flame around a droplet of hydrocarbon fuel such as diesel fuel. Droplets can burn individually, with a gaseous diffusion flame surrounding the evaporating liquid fuel center (Figure 3). When many droplets are combined into an array, a spray is formed. Burning of the droplets in the spray may

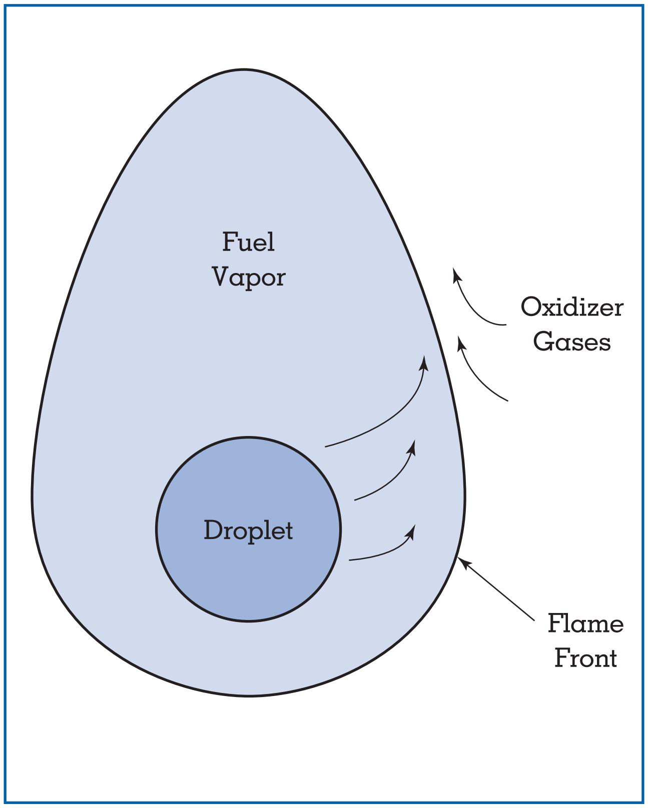


Figure 3.
Shapes of diffusion flames surrounding a burning spherical fuel droplet.

consist of a continuous flame if the droplets completely evaporate or of isolated flames around each droplet if fuel evaporation is slow. An intermediate situation occurs when incomplete evaporation takes place. The types of flames formed depend on the fuel droplet sizes and spacing.

Liquid Pool Flames. Liquid fuel or flammable spills often lead to fires involving a flame at the surface of the liquid. This type of diffusion flame moves across the surface of the liquid driven by evaporation of the fuel through heat transfer ahead of the flame. If the liquid pool or spill is formed at ambient conditions sufficient to vaporize enough fuel to form a flammable air/fuel mixture, then a flame can propagate through the mixture above the spill as a premixed flame.

Solid Fuel Flames. The flames from the combustion of solids such as coal and wood are the result of a combination of processes including the burning of gases that have been released from the heated solid (devolatization) that burn in the gas phase as diffusion flames. The remaining nonvolatile material, the char, then is oxidized on its surface and in its pores as oxygen diffuses into the interior. If the particles are large, devolatization and char burning occur simultaneously.

Very small solid fuel particles such as sawdust, agricultural grains, or coal dust can sustain flames when they are suspended in air. In fact, very serious fires have occurred in grain storage towers and coal mines because of the flammability of suspended dusts. The combustion of the individual particles follows the usual pattern of solid particle burning—devolatization and char burning. The combustion of the whole cloud of particles is similar to spray combustion and its characteristics depend on the nature of the fuel, size of the particles, and the number of particles in a given volume.

Other seemingly solid fuel flames such as those from the burning of plastics are actually more like liquid pool flames because the plastic melts and volatizes ahead of the advancing flame front.

Ignition

Even if a fuel and oxidizer are present in proportions that could sustain combustion, nothing may occur unless the combustible mixture is brought to the right conditions by an ignition source. Typical ignition sources are spark plugs in car engines, pilot flames in gas stoves, and matches for lighting barbecues. In order for an ignition source to be effective, it has to raise the temperature of the combustible mixture enough so that combustion can continue after the ignition source is removed. This means that the amount of heat added during the ignition process must be adequate to overcome any heat loss, to the engine walls for example, and still raise the temperature of the gas region to a value high enough to cause the flame to propagate.

Extinction

Once propagating, flames will continue to propagate unless they are extinguished or quenched. An obvious cause of extinction is the depletion or cessation of fuel flow. Flames can be extinguished by heat loss (e.g., by passage through very small passageways that accentuate heat loss), through smothering by water or chemical fire extinguishers that slow the combustion process, or blowing the flame away with high velocity flows. Flames can also be extinguished by removing one of the reactants, such as air.

Detonations

These are types of combustion waves (actually shock waves—extremely thin regions in which flow properties such as pressure and temperature change enormously—sustained by combustion) that consume fuel at supersonic speeds and create very large pressure and temperature increases. Detonations are formed only under special conditions that convert an ordinary flame propagating through a combustible mixture to a detonation. When these special conditions exist, then the detonation can have devastating consequences (e.g., explosions in buildings containing natural gas from leaks or in mines filled with natural gas as a result of the mining operation).

FUELS FOR COMBUSTION

Fuels for combustion are initially gases, liquids, or solids. A fuel initially in one phase may be transformed into another during the burning process (i.e., liquids vaporized to gases). The factors involved in the selection of the fuel phase or its physical and chemical characteristics for an application such as burning in an automobile or jet aircraft involve many different considerations such as price, availability, and source.

Among the various selection considerations are specific combustion characteristics of different fuels. One of the combustion characteristics of gaseous fuels is their flammability limit. The flammability limit refers to the mixture proportions of fuel and air that will sustain a premixed flame when there is either limited or excess air available. If there is a large amount of fuel mixed with a small amount of air, then there is a limiting ratio of fuel to air at which the mixture will no longer sustain a flame. This limit is called the rich flammability limit. If there is a small amount of fuel mixed with excess air, then there is a limiting ratio of the two at which the flame will not propagate.This limit is called the lean flammability limit. Different fuels have different flammability limits and these must be identified for each fuel.

The combustion characteristics of liquid fuels are similarly determined by measures of their ability to sustain a flame. Two measures of the combustion characteristics of liquid fuels especially related to safety are flash point and autoignition temperature. The flash point is the maximum temperature at which a liquid fuel can be maintained in an open vessel exposed to air before which it will sustain a flame in the presence of a pilot flame. The autoignition temperature is a similar concept except no pilot flame is present. The autoignition temperature is the maximum temperature at which a liquid fuel can be maintained in an open vessel exposed to air before which the fuel bursts into flame without the presence of an external ignition source.

Solid fuels, unlike gases and liquids, are entirely characterized by their composition. For example, coal can be characterized by its carbon, hydrogen, oxygen, sulfur, and nitrogen content. The water and mineral content of coal are also important means of differentiating coals from various sources.

THE CHEMISTRY OF COMBUSTION AND ITS EFFECT ON THE ENVIRONMENT

The most commonly used fuels for combustion are hydrocarbons, materials that are compounds of only hydrogen and carbon. Occasionally, fuels such as alcohols, that contain oxygen, are burned. When hydrocarbon fuels with or without oxygen are burned in air (combusted) to completion, the products are water, from the hydrogen part of the fuel, and carbon dioxide, from the complete conversion of the carbon part. If oxygen is present in the fuel, it shows up in the final product as part of either the water or carbon dioxide.

Carbon dioxide has been implicated as a contributing factor in global warming. Increased global warming has been associated with increased release of carbon dioxide into the atmosphere attributed in part to an increase in the combustion of hydrocarbon fuels. Carbon dioxide is an inevitable consequence of the complete combustion of hydrocarbons in air. If combustion devices are made more efficient, less fuel is required and less carbon dioxide is released into the atmosphere.

Unlike carbon dioxide and water that are the inevitable by products of complete combustion of hydrocarbons, species such as carbon monoxide, ethene, toluene, and formaldehyde can be emitted because combustion has been interrupted before completion. Many factors lead to emissions from incomplete combustion. Emitted unburned hydrocarbons and carbon monoxide are regulated pollutants that must be eliminated. In automobiles with spark ignited engines, these emissions are almost entirely removed by the catalytic converter.

Soot particles, that are comprised primarily of carbon and hydrogen in an 8 to 1 ratio and are about 20-50 nm in diameter when first formed (coagulation and surface growth ultimately leads to a chain of soot spheres much larger than 50 nm), are the result of incomplete combustion of hydrocarbons. Soot, too, is a regulated combustion pollutant and is a particular problem with diesel engines. The black clouds emitted from the vertical exhaust stacks of trucks are laden with soot particles. The nature of diffusion flames precludes a practical way to reduce soot emissions in the diffusion flame processes that occur in diesel engines.

In contrast to carbon monoxide, small hydrocarbon molecules and soot that result from incomplete conversion of the hydrocarbon fuels, nitric oxide and nitrogen dioxide, are noxious emissions that result from the oxidizer—air. However, fuel components that contain nitrogen may also contribute, in a lesser way, to the formation of the oxides of nitrogen.

The nitrogen component of air, normally inert and unreactive, reacts at the very high temperatures of combustion. It reacts in a series of simple steps with atomic and molecular oxygen to yield NO, nitric oxide, that is subsequently converted to NO_2, nitrogen dioxide, in the atmosphere. Nitrogen dioxide, is a regulated, undesirable emission because it forms a brownish haze, leads to acid rain, and is a component of photochemical smog. Both nitric oxide and nitrogen dioxide can be considered inevitable byproducts of high-temperature combustion in air. The concentration of NO emitted into the atmosphere can be reduced either by lowering the temperature of combustion by various engineering techniques (with negative effects on performance) or through catalytic conversion to molecular nitrogen in postcombustion cleanup as is done in automobiles.

RESEARCH ACTIVITIES

The following synopsis of current research activities in the field of combustion is organized around the list of papers presented at the 27th International Symposium on Combustion (1998) under the auspices of The Combustion Institute.

Elementary Reaction Kinetics, Kinetic Mechanisms, Models, and Experiments

Examining the details involved in the oxidation and pyrolysis (thermal decomposition) of fuel molecules is very important. The results of these research activities will permit predictions about the chemicals emitted during incomplete combustion because reaction rate constants and chemical pathways will be evaluated and determined.

Laminar Premixed Flames

Research in this area focuses on understanding the chemical, thermal, and fluid-mechanical (behavior of fluids) structure of these types of flames. Recent advances in computer based modeled flames requires the knowledge developed in this type of research for calibration, validation, and prediction.

Laminar Diffusion Flames

Here, too, computer based predictions about the nature of these flames require information about the chemicals and science of diffusion flames for the predictions to be accurate. The predictions are made accurate by comparison with measured chemical species concentrations, measured temperatures, and flow characteristics.

Premixed Turbulent Combustion and Nonpremixed Turbulent Combustion

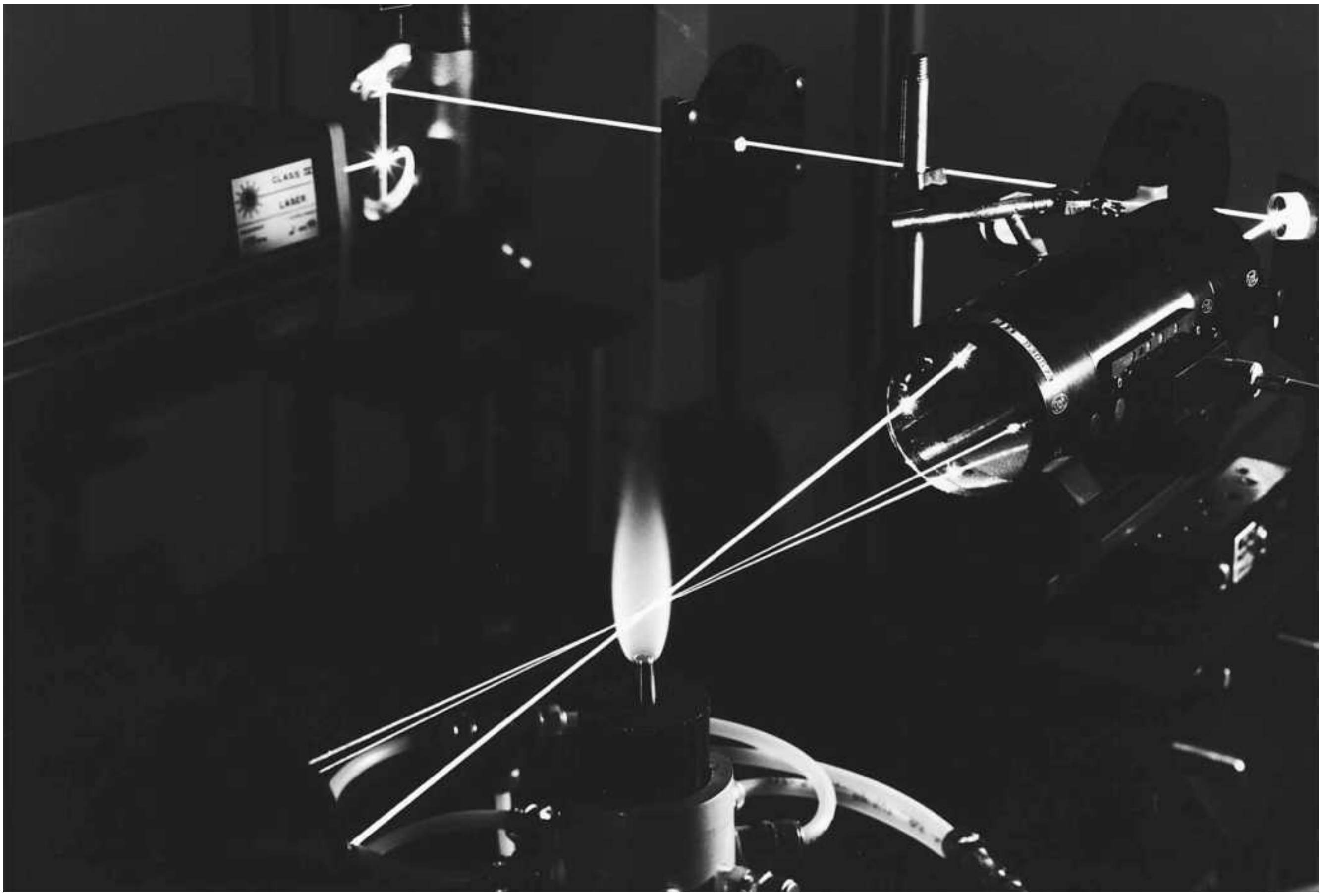
Because many practical flames are turbulent (spark ignited engine flames, oil field flares), an understanding of the interaction between the complex fluid dynamics of turbulence and the combustion processes is necessary to develop predictive computer models. Once these predictive models are developed, they are repeatedly compared with measurements of species, temperatures, and flow in actual flames for iterative refinement. If the model is deficient, it is changed and again compared with experiment. The process is repeated until a satisfactory predictive model is obtained.

Incineration, NO_x (NO and NO_2) Formation and Control, Soot Formation and Destruction

Environmental consequences of combustion are still a high priority requiring investigation of the chemistry and process effects on the emissions. Effective means of eliminating the pollutants is also a subject of further research.

Gas Turbines, Diesel Combustion, and Spark Ignition Engines

Application of combustion science to practical power source devices is one of the ultimate aims of developing a fundamental understanding of combustion.



Advanced lasers and computers at the Combustion Research Facility in Livermore, California, are used to study exactly how and why fuels burn. (U.S. Department of Energy)

Using combustion science to improve performance through design changes and engineering techniques is an ongoing research subject.

Droplet and Spray Combustion and Pool Fires

The combustion of liquids is a fertile area for further study. Kowledge of the combustion science of individual droplets as well as groups of droplets helps improve performance of devices that rely on spray burning, particularly diesel engines. Understanding of the science of liquid pool fires potentially effects safety during spills.

High-Speed Combustion, Metals Combustion, and Propellants

Research in these specialized areas is aimed at developing improved and new methods for advanced propulsion. For example, an understanding of high-speed combustion is used in the development of supersonic ram jet engines, which are simple alternatives to conventional turbojet engines. Knowledge of metals combustion is relevant to improving the use of metal additives in solid propellants to increase impulse and stability. In general, the study of propellant combustion aids in the development of more stable and longer range rocket engine performance.

Catalytic and Materials Synthesis

These two research areas share the common characteristic of involving inorganic solids in the combustion process. Catalytic combustion research focuses on using the solid to facilitate the oxidation of well-known fuels such as hydrogen and methane. Materials synthesis research focuses on using combustion as a means to react the solids either with each other or a gas, such as nitrogen (which in this case acts as an oxidizer), to make new solid materials.

Microgravity Combustion

Microgravity refers to the environment of extremely low gravity commonly known as a weightless environment. Under microgravity conditions, combustion phenomena that are affected by gravity, such as flames, behave differently than at Earth gravity conditions. Research in this area focuses on using the special microgravity conditions to understand, by contrast, basic combustion processes on Earth.

Fire Safety Research

Research is conducted to increase understanding of the science of combustion specifically as it relates to fires involving homes and plastics, wood, and large-scale spills. This is helpful in the development of fire prevention and extinction techniques.

Detonations

Examining the conditions for the formation and propagation of detonation waves is relevant to special applications of detonations to propulsion as well as safety.

Coal and Char Combustion

The importance of coal as an energy source motivates further research into the combustion characteristics and chemical kinetics of both coal and the material that remains after devolatization, char. Further research will aid in making coal a cleaner and more efficient energy source.

Fluidized Beds, Porous Media Fixed Bed Combustion, and Furnaces.

Specialized practical configurations for combustion have a number of practical applications such as coal burning for energy production. The study of these specialized combustion setups is necessary for better application.

Kenneth Brezinsky

See also: Catalysts; Coal, Production of; Conservation of Energy; Explosives and Propellants; Heat Transfer.

BIBLIOGRAPHY

Borman, G. L., and Ragland, K. W. (1998). *Combustion Engineering*. Boston: McGraw-Hill.

Chomiak, J. (1990). *Combustion: A Study in Theory, Fact, and Application*. New York: Abacus Press.

Combustion and Flame (series). New York: Elsevier.

Combustion Science and Technology (series). New York: Gordon and Breach.

Glassman, I. (1996). *Combustion*, 3rd ed. San Diego, CA: Academic Press.

International Symposium on Combustion (series). Pittsburgh: The Combustion Institute.

Progress in Energy Combustion Science (series). New York: Pergamon.

Turns, S. R. (2000). *An Introduction to Combustion*, 2nd ed. Boston: McGraw-Hill.

Williams, F. A. (1985). *Combustion Theory*, 2nd ed. Menlo Park, CA: Benjamin/Cummings.

COMMUNICATIONS AND ENERGY

To propel or move anything requires energy. Light, electrical waves, or sound waves used in communication are no exception. What is different about communication is the energy used to transmit data.

BASIC SIGNALING

The history of communications and energy shows a relationship that can be expressed by the popular adage, “What goes around, comes around.” In ancient times humans communicated through the use of torches, fire, and smoke signals. All three methods required a great deal of energy for the amount of information generated and the short transmission distance. While fire was more than likely first used by prehistoric peoples to cook food and as a source of warmth, it also provided a mechanism for performing a basic signaling method. This method evolved into the foundation for modern communications systems. If we fast-forward to the new millennium, the opposite is true: Communication over great distances with little energy is accomplished through the use of microprocessors and microelectronics. Microprocessors and microelectronics provide global terrestrial and satellite communications, whose use also facilitates the exploration, recovery, and distribution of different types of energy.

Although unknown at the time, some of the earliest uses of fire represented a binary signaling system that was even used during the American Revolution. On the evening of April 18, 1775, Joseph Warren sent for Paul Revere and instructed him to ride from

Charlestown to Lexington, Massachusetts, to warn Samuel Adams and John Hancock that British troops were coming to arrest them. The prearranged signal, which began his historic ride, was based on the placement of lanterns in the bell tower of Christ Church in Boston. Two lanterns were hung to indicate that the redcoats would come by sea rather than by land; the latter would have been signaled by one lantern.

Another popular use of fire for communications during human evolution involved the use of torches. In many locations throughout the world underground deposits of oil seeped to the surface, providing fuel for torches. They could only be used when it was dark and required a line of sight between originator and recipient. Torches were supplemented by smoke signals during daylight. It wasn't until the invention of electricity that communications over relatively long distances were only marginally affected by the elements.

Although many people credit American Indians with first using smoke signals to communicate information, African tribes and ancient Greeks and Romans also used smoke signals. Burning brush or wood and placing cloth or animal hides over the fire enabled people to generate puffs of smoke.

With the discovery of electricity, the evolution of communications occurred at a rapid rate. Smoke signals that were used for hundreds of years to convey information at a word or two per minute were first replaced by the use of copper-based conductors, such as the telegraph, that transmitted twenty to forty words per minute. By the end of the twentieth century, lasers with fiber optic wires transmitted an entire book around the world in under a second.

THE TELEGRAPH

The first notable development in the field of data communications occurred in 1832 when an American, Samuel F. Morse, invented the electric telegraph. When an operator at one location depressed a key, an electrical path was established that allowed current to flow through the path connecting the two locations. At the distant location, electricity flowed through a wire coil, forming a magnet that caused a metal plate (key) to click as it was attracted to the coil. When the key was released it opened the circuit; the distant key was released and it struck a stop bar with a slightly different sounding click. This resulted in two distinct clicks that defined the duration of the operator's key depression and enabled Morse to develop his well-known code. That code used the time between clicks to represent a dot (short time) or dash (long time). Morse demonstrated the practicality of the telegraph in 1844 when he transmitted the now famous message, "What hath God wrought!," over a wire routed from Washington, D.C., to Baltimore. Similar to torches, lanterns, and smoke signals, the telegraph represents a two-state system, with current either flowing or not. As communications progressed from the telegraph to radio, microwave, satellite, light on fiber, and other systems, the two-state communications system continued to be used. It is the simplest to construct and eventually was developed into the binary system used by all modern communications methods.

In addition to the telegraph, the invention of electricity resulted in the development of other communications systems that paved the way for the modern communications infrastructure. The telegraph was followed by the laying of the transatlantic cable (1858), the invention of the telephone (1876), Marconi's wireless telegraph (1897), television (1927), and satellite communications (1957). Although early implementation of each technology was restricted to a minimum level of practical communication, each technology eventually evolved into a mass market for use by businesses and residential customers.

THE COMPUTER REVOLUTION

The development of the computer fostered the growth in communications. The first electronic computer, developed at the University of Pennsylvania's Moore School of Engineering during World War II, quickly moved from high-energy consumption vacuum tube technology to low-energy consumption semiconductor-based technology by the 1960s. As computers became smaller and used less power, they became more practical for use in telephone company offices. Computerized switches gave automated operator-dependent functions, which enhanced the ability of consumers to rapidly and easily communicate with people around the globe. The invention of the laser and microprocessor during the 1960s and 1970s eventually resulted in communications systems applications. Microprocessors enabled low-cost voice-digitization methods to be implemented that enabled telephone companies to digitize analog conversations at their central offices. Use of lasers made it possible to transmit digitized data over fiber cables using light energy instead of electrical energy, elimi-

A large microwave antenna transmits communications for the Kennedy Space Center in Florida. (Corbis-Bettmann)

nating the adverse effect of electromagnetic interference generated by lightning and machinery on transmission via copper circuits and microwave.

The use of digital technology and the evolution of the communications infrastructure of carriers to lasers and fiber optic media reduced the energy required to transmit information and increased transmission capacity. Digital transmission is based on the use of microprocessors and digital signal processors (DSPs) that consume less energy than filters and modulators used by analog transmission systems. The use of lasers to transmit information through fiber cables was also more energy efficient because it eliminated the 5 to 10 percent energy loss of copper circuits.

Beginning in the late 1970s most telephone companies in the world began to convert trunk circuits that interconnected central offices from microwave and copper-based transmission to fiber-optic-based transmission. By the mid-1990s almost all of Western European and North American long-distance telephone transmissions resulted in analog voice conversations being digitized and transmitted over fiber optic cable. The use of fiber optic cable not only provided immunity to electromagnetic interference but, in addition, provided a bandwidth several orders of magnitude greater than that obtainable on conventional copper-based media. This enabled a single strand of glass or plastic to transport tens of thousands of simultaneous voice conversations.

As transmission systems were developed to take advantage of the evolution of technology it became possible to communicate further using less energy. While communications carriers were initially the prime beneficiary of the evolution of products that expanded their capability to transport voice, data, and video, the resulting efficiencies were eventually passed along to the consumer in the form of new offerings as well as lower communications costs. By 2000 it was rare to pay more than a dime a minute to make a coast-to-coast long-distance telephone call that cost more than $2 per minute during the 1950s. In addition, the modern telephone call is highly automated and only requires operator intervention for special services, such as setting up an international conference call or reversing charges. In comparison, many long-distance telephone calls made during the 1950s required operator intervention.

Today the majority of energy in the telephone system occurs on the legacy local loop routed from the telephone company central office to the subscriber. The red colored wire in the home or office carried negative voltage (48 Vdc) relative to ground. When a phone is on-hook, it presents a dc resistance near infinity. When a phone is off-hook, it presents approximately 600 ohms to the central office and the current flow rises to 25 mA. As a result of going off-hook, the voltage will drop to between 12 and 15 Vdc. While 25 mA represents a low current, when you multiply this by hundreds of millions of subscribers the energy used by the phone system becomes considerable. Home computers and phone lines that surf the Internet account for approximately 5 percent of electricity consumed in the United States.

THE EVOLUTION OF TELEVISION

Although television dates back to 1927 and was highly publicized during the 1936 Olympics, it was not until post–World War II prosperity that vacuum-tube-based television products reached the mass consumer market. The initial series of televisions manufactured during the 1950s had screen displays that made the viewer squint and the vacuum tubes used so much electricity that they dimmed the lights in the home. Semiconductors also had a profound effect on the television industry because they made televisions smaller and significantly reduced their power consumption. Television became popular home entertainment and it also allowed networks to provide viewers with news broadcasts, presidential press conferences, and special reports. Unfortunately, hills, mountains, and the canyons formed by buildings within cities caused broadcast reflections, echos, and periodically the inability to receive signals. Such problems resulted in the development of the cable television (CATV) industry that evolved into several national multimedia conglomerates. The initial development of the CATV industry was focused on one-way communications. However, by the late 1990s many CATV operators began to replace their one-way amplifier infrastructure with a bidirectional amplifier infrastructure. This upgrade enabled the support of cable modems that allow subscribers to access the Internet at data rates up to 10 Mbps compared to the maximum transmission rate of 56 kbps obtainable over the public switched telephone network (PSTN). CATV modems, as well as those on the PSTN, represent considerable advances in the use of solid state technology over modems developed during the 1960s through 1980s. Packing circuitry built with microprocessors more powerful than mainframe computers that reached the market during the 1980s, today's PSTN and CATV modems consume only a fraction of the electricity of the prior generation of modems, yet have a signal processing capability as great as those of mainframe computers manufactured just a decade earlier.

SATELLITE-BASED SYSTEMS

No discussion of the history of energy and communications would be complete without covering the role of satellites. Although Sputnik communications during 1957 was limited to a series of beeps, today there are hundreds of communications satellites in geostationary and low Earth orbit. Geostationary communications satellites support communications over a predefined arc, while low Earth orbit satellites commonly operate in groups that circle the earth and provide communications coverage over a much wider area. Applications from satellite television broadcasts to Internet access, and global positioning communications to marine transmission to ships at sea are now a reality via the use of satellites. The evolution of satellite communications is closely linked to advances in semiconductor technology that enable the design of miniaturized circuitry requiring less power. This in turn enables communications satellites to be designed employing more transponders (transmitters/receivers) while using the same or a similar solar panel used on earlier orbited satellites.

Through the use of satellite communications, the adage "What goes around, comes around" is a reality today with respect to the relationship of energy and communications. As the search for energy to include petroleum reserves expanded to less populated, more rugged terrain, the need to communicate with field workers increased. Satellite communications is now commonly used as a standard way to facilitate communications with energy exploration workers in the foothills of Montana, at the Arctic Circle, and drilling for gas and oil at various offshore locations throughout the world. A second related area between energy and communications involves the distribution of production. To facilitate communications to mobile locations such as ships delivering oil, an international cooperative known as Inmarsat began service in 1979. Today Inmarsat operates a series of geostationary satellites that provide worldwide communications coverage for telephone, fax, email, and data services. Operating forty land stations in thirty-one countries that transmit and receive communications through Inmarsat satellites, maritime customers are no longer dependent on bouncing radio frequency communications off the ionosphere in an attempt to connect to land.

The role of energy in providing a communications capability is indispensable to maintaining our living standards. Whether it is cell phones powered by batteries, checking our bank account via a toll free telephone number, or surfing the Internet, each of these activities depends on the use of energy. In the history of communications and energy, advances in technology made possible the ability to transmit further and faster using less energy, which has improved productivity and living standards. Within a few years, low power and low cost semiconductor charged coupled devices (CCDs) will more than likely enable the Picturephone to replace the telephone, while other advances in semiconductors may make the Dick Tracy watch telephone a reality. As communications technology progresses, we can continue to expect to do more using less energy.

Gilbert Held

See also: Electricity, History of; Energy Economics; Transportation, Evolution of Energy Use and.

BIBLIOGRAPHY

Held, G. (1997). *The Complete Modem Reference*, 3rd ed. New York: John Wiley & Sons.

Held, G. (1999). *Cisco Security Architectures.* New York: McGraw-Hill.

Held, G. (1999). *Data Communications Networking Devices*, 4th ed. New York: John Wiley & Sons.

Held, G. (1999). *Enhancing LAN Performance: Issues and Answers*, 3rd ed. New York: John Wiley & Sons.

Held, G. (1999). *Understanding Data Communications*, 6th ed. Indianapolis: New Riders Publication.

Held, G. (1999). *Voice & Data Internetworking.* New York: McGraw-Hill.

COMPRESSED AIR STORAGE

See: Storage Technology

CONDUCTION

See: Heat Transfer

CONSERVATION OF ENERGY

Conservation of energy has two very different meanings. In the popular sense, "conserve" means to "save" or "preserve." Electric energy for lights likely had its origin in burning coal, so turning off lights tends to preserve coal, a valuable natural resource. In the scientific sense, conservation alludes to constancy. Succinctly stated, the energy of the universe is constant. Energy can be converted from one form to another, but ultimately there are as many joules of energy after the conversion as before. Every second, an operating 100-watt lightbulb converts 100 joules of energy. If 10 joules are in the form of light, then 90 joules are in some other form, notably heat to the room. Even though the total amount of energy following the conversion is unchanged, the energy may not be available for some desired purpose. For example, the heat produced by a lightbulb is in the surroundings and is no longer available for other uses. In fact, when energy runs through all possible conversions, the energy ends up as thermal energy in the environment. Addition of thermal energy to the environment can produce local increases in temperature, leading to what is called thermal pollution. This effect is not to be confused with a possible global increase in temperature due to accumulation of carbon dioxide and other gases in the atmosphere.

CONSERVATIVE FORCES

The illustration in Figure 1 depicts a person pushing a box up a ramp. In the process, the person works against the gravitational force on the box and a frictional force between the box and the ramp. The person, the gravitational force, and the frictional force all do work on the box. The same would be said if the ramp were made longer. But interestingly, from the way work is defined, the work done by the gravitational force depends only on the vertical height through which it moves. The work is the same no matter how long or how short the ramp, as long as the vertical height is the same. If the work done by a force depends only on where it started and where it ended up, the force is said to be conservative. Unlike the gravitational force, the frictional force is nonconservative because the work done by it does depend on the path of the movement, that is, the length of the ramp. Potential energy is associated with the work done by a conservative force. For a mass (m) a height (h) above its lowest level the potential energy is $U = mgh$ where g is the acceleration due to the gravitational force ($9.80\ m/s^2$).

MECHANICAL ENERGY

Kinetic energy (the energy of motion) and potential energy (the energy based on position) added together are called mechanical energy. Mechanical energy equals kinetic energy plus potential energy.

$$E = K + U, \text{ with } K = \tfrac{1}{2}mv^2$$

In an isolated system—one devoid of friction—the mechanical energy does not change. Although friction can never be totally eliminated, there are situations where it is small enough to be ignored. For example, when you hold a book in your outstretched hands, it has potential energy but no kinetic energy because its speed is zero. When dropped, the book acquires speed and kinetic energy. As its height above the floor decreases, its potential energy decreases. The gain in kinetic energy is balanced by a loss in potential energy, and the sum of kinetic energy and potential energy does not change. This is the idea of the conservation principle. Of course, this assumes that there is no friction (a nonconservative force), which is rarely the case in the real world. Frictional forces will always extract energy from a system and produce heat that ends up in the environment. The falling book will come to rest on the floor and have neither kinetic energy nor potential energy. All the energy it had before being dropped will have been converted to heat (and a very small amount of sound energy from its impact with the floor).

CONSERVATION OF ENERGY WITH A SIMPLE PENDULUM

A simple pendulum with ignorable friction illustrates the conservation of mechanical energy. Pulling the bob (the mass) from its lowest position and holding it, the pendulum has only potential energy; the mechanical energy is all potential. When released, the pendulum bob gains kinetic energy and loses potential energy, but at any instant the sum never differs from the sum at the beginning. At the lowest point of the movement the potential energy is zero and the mechanical energy is all kinetic. As the pendulum bob moves to higher levels, the potential energy increases, and the kinetic energy decreases. Throughout the motion, kinetic energy and potential energy change continually. But at any moment the sum, the mechanical energy, stays constant.

A simple pendulum isolated from nonconservative forces would oscillate forever. Complete isolation can never be achieved, and the pendulum will eventually stop because nonconservative forces such as air resistance and surface friction always remove mechanical energy from a system. Unless there is a mechanism for putting the energy back, the mechanical energy eventually drains and the motion stops. A child's swing is a pendulum of sorts. If you release a swing from some elevated position it will oscillate for a while but eventually will stop. You can push the swing regularly and keep it going, but in doing so you do work and put energy back into the system.

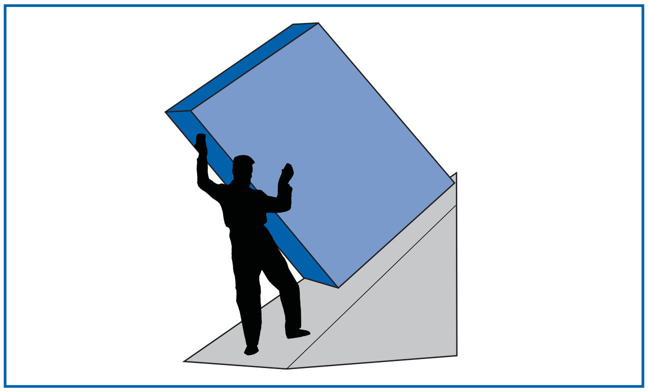


Figure 1.
A person does work on a box when pushing it up a ramp and the potential energy of the box increases.

CONSERVATION OF ENERGY IN A SPRING-MASS SYSTEM

When a spring is stretched by pulling on one end, the spring pulls back on whatever is pulling it. Like the gravitational force, the spring force is conservative. Accordingly, there is potential energy associated with the spring that is given by $U = kx^2$, where k, the spring constant, reflects the strength of the spring and x is the amount of stretching from the relaxed position.

A horizontal spring with a mass attached to one end is a form of oscillator—that is, something that periodically returns to the starting position. To the extent that the spring-mass system can be isolated, the mechanical energy is conserved. When stretched but not yet in motion, the system has only potential energy. When released, the mass gains kinetic energy, the spring loses potential energy, but the sum does not change; it is conserved. Much like the oscillating pendulum, both kinetic energy and potential energy change continually, but the sum is constant. Most people do not think of atoms in a molecule as being connected by springs, yet the forces that bind them together behave like springs, and the atoms vibrate. The mechanical energy of the spring-like atomic system is an important energy attribute. A better understanding of the spring-like characteristics of atoms in materials has made possible many advances in sporting goods equipment, from graphite composite vaulting poles to titanium drivers for golfers.

Even though mechanical energy is rigorously conserved only when a system is isolated, the principle is elegant and useful. Water flowing over the top of a dam has both kinetic energy and potential energy. As it plummets toward the bottom of the dam, it loses potential energy and gains kinetic energy. Impinging on the blades of a paddle wheel, the water loses kinetic energy, which is transferred to rotational kinetic energy of the wheel. The principle provides an accounting procedure for energy.

THE FIRST LAW OF THERMODYNAMICS

The first law of thermodynamics is a statement of the principle of conservation of energy involving work, thermal energy, and heat. Engines converting heat to useful work are widely employed in our society, and the first law is vital for understanding their operation. The first law of thermodynamics accounts for joules of energy in somewhat the way a person accounts for money. A person uses a bank to receive and disperse money. If money comes into a bank and nothing is dispersed, the bank account increases. If money is dispersed and nothing comes in, the bank account decreases. If we call U the money in the bank, Q the money coming in or out, and W the money dispersed then the change in the account (ΔU) can be summarized as

$$\Delta U = Q - W$$

Algebraically, Q is positive for money coming in, negative for money going out. W is positive for money going out. If a person put $100 into the bank and took $200 out, the change in the account would be

$$\Delta U = \$100 - \$200 = -\$100$$

The account has decreased $100.

The energy content of a gas is called internal energy, symbol U. The gas can be put in contact with something at a higher temperature, and heat will flow in. If that something is at a lower temperature, heat will flow out. If the gas is contained in a cylinder containing a movable piston, then the gas can expand and push against the piston doing work. In principle, the work is equivalent to a person pushing and moving a box on a floor. Some agent can push on the piston doing work on the gas. If heat enters the gas and no work is done, the internal energy increases. If no heat enters or leaves the gas and if the gas expands doing work, the internal energy decreases. All of this is summarized in the equation

$$\Delta U = Q - W$$

If 100 J of energy (heat) entered the gas and the gas expanded and did 200 J of work, then

$$\Delta U = 100\text{ J} - 200\text{ J} = -100\text{ J}$$

The internal energy decreased 100 J.

The internal energy of all gases depends on the temperature of the gas. For an ideal gas, the internal energy depends only on the temperature. The temperature is most appropriately measured on the Kelvin scale. The contribution to the internal energy from the random kinetic energy of the molecules in the gas is called thermal energy.

In a monetary bank, money is added and withdrawn, but if in the end the bank is in exactly the same state, the net amount of money in the bank has not changed. Similarly, a gas may expand or be compressed and its temperature may undergo changes, but if at the end it is in exactly the same thermodynamic condition, the internal energy has not changed.

CONSERVATION OF ENERGY AND HEAT ENGINES

Heat engines work in cycles. During each cycle, heat is absorbed, work is done, and heat is rejected. At the beginning of the next cycle the gas is in exactly the same state as at the beginning of the previous cycle; the change in internal energy is zero. To illustrate, consider a conventional automobile engine. The cycle (Figure 2) starts with the piston moving down and pulling a mixture of gasoline vapor and air through the open intake valve and into the cylinder. The intake valve closes at the bottom of the downward motion, the piston moves up, and the gaseous mixture is compressed. At the top of the stroke the mixture is ignited; the gas expands doing work. At the end of the stroke the exhaust valve opens, the cooler gas is forced out the automobile's exhaust, and a new cycle begins.

Because the change in internal energy in a cycle is

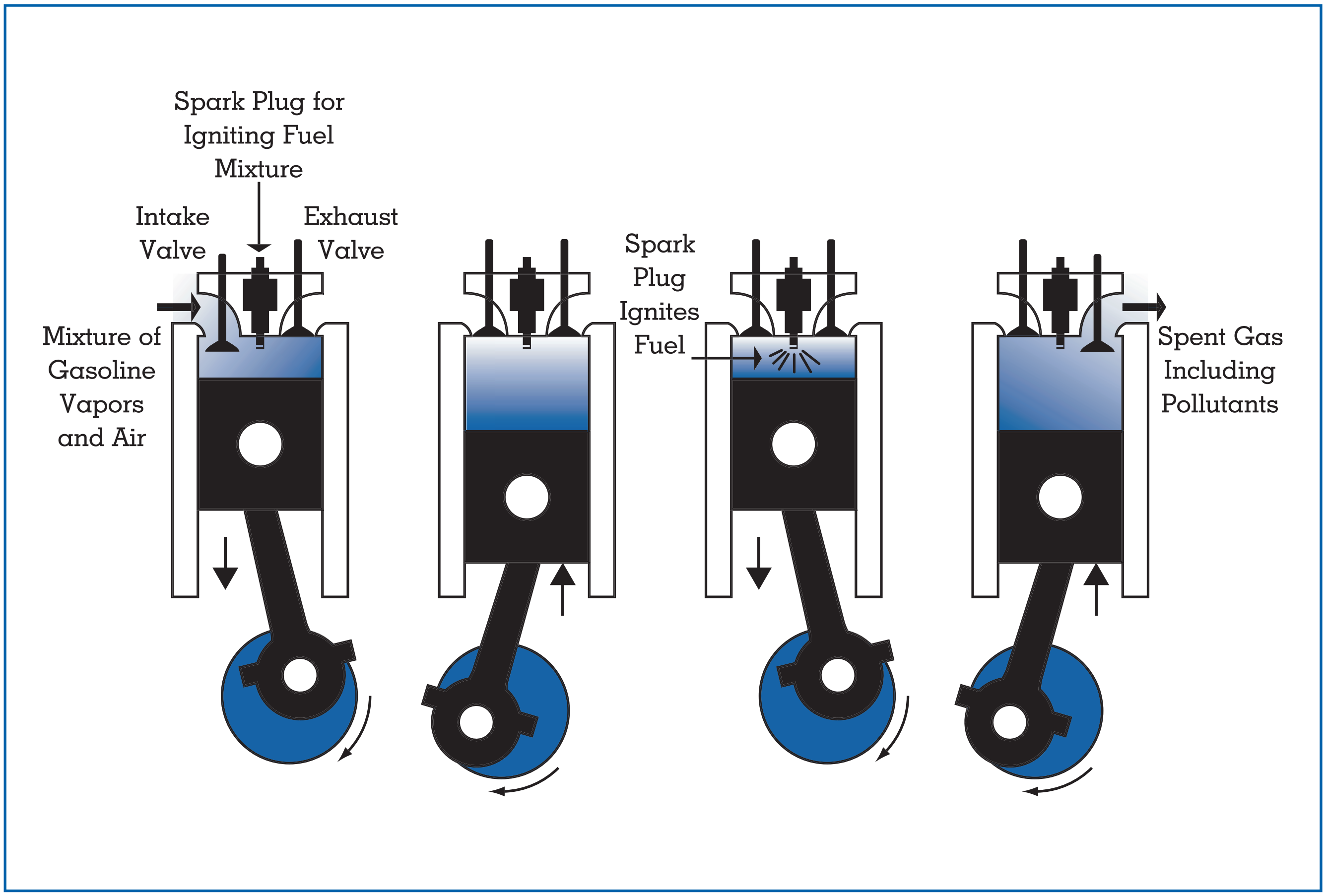


Figure 2.
The cycle of a conventional internal combustion engine.

zero, the first law of thermodynamics requires Q=W. In words the net heat exchanged equals the work done in the cycle. The difference between the heat absorbed and the heat rejected has gone into useful work. The energy absorbed by a gas always takes place at a temperature higher than the temperature at which energy is exhausted. An automobile engine exhausts energy into the environment at a temperature of about 300 K, which is much lower than the temperature of the gasoline vapor-air mixture at the moment of ignition, about 1,000 K.

Engines are used to do work, and a quantitative measure of the performance of an engine is efficiency:

$$\text{efficiency} = \text{work done} \div \text{energy absorbed}$$

For a given amount of energy absorbed at a high temperature, the more work obtained in a cycle, the more efficient the engine. In symbols,

$$e = W \div Q_{high}.$$

The first law of thermodynamics requires $W = Q_{high} - Q_{low}$ so the efficiency may be written

$$\begin{aligned} e &= (Q_{high} - Q_{low}) \div Q_{high} \\ &= 1 - (Q_{low} \div Q_{high}) \end{aligned}$$

If all the heat absorbed were converted into work, the efficiency would be 1, or 100 percent. If none of the heat absorbed was converted into work, the efficiency would be 0. The first law of thermodynamics limits the efficiency of any heat engine to 1 but does not prevent an efficiency of 1. The efficiency of practical heat engines is always less than 1. For example, the efficiency of a large steam turbine in an electric power plant is about 0.5, which is considerably more efficient than the typical 0.35 efficiency of an auto engine.

When two objects at different temperatures are in contact, heat always flows from the hotter one to the cooler one. There is nothing in the first law of thermodynamics that prevents the opposite. The first law only requires that energy be conserved. A heat engine relies on heat flowing from some reservoir to a reservoir at a lower temperature. It is somewhat like a hydroelectric system relying on water flowing from a higher level to a lower level. The efficiency of a hydroelectric system increases as the difference in heights of the water levels increases. The efficiency of a heat engine increases as the difference in temperatures of the two reservoirs increases. In 1824, at age twenty-eight, the French engineer Sadi Carnot reasoned that the maximum efficiency of a heat engine depends *only* on the Kelvin temperatures of the two reservoirs. Formally,

$$\text{maximum efficiency} = e_{\text{max}} = 1 - (T_{\text{low}} \div T_{\text{high}})$$

For a heat engine like a steam turbine in an electric power plant the low temperature is determined by the outdoor environment. This temperature is about 300 K. Engineering considerations limit the high temperature to about 800 K. The maximum efficiency according to Carnot is 0.63 or 63 percent. No matter how skilled the builders of a steam turbine, if the temperatures are 300 K and 800 K, the efficiency will never exceed 63 percent. When you realize that the efficiency can never be larger than about 63 percent, a realizable efficiency of 50 percent looks quite good.

An important message in this discussion is that all the thermal energy extracted from a reservoir is not available to do work. Some will always be lost, never to be recoverable. The principle of conservation of energy guarantees that thermal energy exhausted to the environment is not lost. But the principle does not say that the energy can be recovered. To recover the energy there must be a reservoir at a lower temperature for the heat to flow into. When everything in the environment comes to the same temperature, there is no reservoir at a lower temperature for heat to flow into.

CONSERVATION OF ENERGY AND REFRIGERATORS

Water never flows spontaneously from the bottom of a dam to the top. Water can be forced to flow to a higher level, but it requires a pump doing work. Similarly, heat never flows spontaneously from a lower temperature to a higher temperature. Heat can be forced to flow from a lower temperature to a higher temperature, but it requires work. A household refrigerator is a good example. The noise emanating from a refrigerator is due to an electric motor doing work, resulting in heat flowing from the cool interior to the warmer surroundings. In principle, a refrigerator is a heat engine running backward. The refrigerator operates in cycles and subscribes to the first law of thermodynamics. Work must be done on the working substance in order for heat (Q_{low}) to flow from the lower temperature. At the end of each cycle this heat as well as the work done (W) is rejected at the higher temperature. Conserving energy the first law requires

$$Q_{\text{high}} = Q_{\text{low}} + W$$

Work is invested to force heat to flow from the interior of the refrigerator. A measure of the performance, called coefficient of performance (COP), is

$$\text{COP} = Q_{\text{low}} \div W$$

If for every joule of work done, 2 joules of heat flowed out of the refrigerator then the performance would be 2. Using $Q_{\text{high}} = Q_{\text{low}} + W$, the performance equation can be written

$$\text{COP} = Q_{\text{low}} \div (Q_{\text{high}} - Q_{\text{low}})$$

It is quite appropriate to think of a refrigerator as a heat pump. It pumps energy from one region and dumps it into another region at a higher temperature. A commercial heat pump does just this to warm a building during the heating season. There is a lot of energy outside of a building, even though the temperature may be 30°F (16.7°C) lower. This is because there are a lot of molecules in the outdoors, and each molecule contributes to the total energy. The performance of a heat pump decreases as the outside temperature decreases, but if the temperature remains above the freezing point of water (0°C or 32°F) commercial heat pumps can achieve performances between 2 and 4. If it were 2, this means that 2 joules of energy are deposited in the building for every joule of work. This is a significant gain. The temperature below the surface of the ground is several degrees higher than at the surface. Heat pumps drawing energy from the subsurface tend to be more efficient than an above-ground heat pump. This is because it is easier to extract heat from the solid subsurface than it is from air, and the subsurface temperature is higher.

CONSERVATION OF ENERGY IN CHEMICAL REACTIONS

Conservation of energy is mandatory in chemical and nuclear reactions. To illustrate, consider the combustion of gasoline. One type of gasoline, isooctane, has the molecular form C_8H_{18}. Thermal energy is released when isooctane combines with oxygen according to

$$2C_8H_{18} + 25O_2 \rightarrow 16XO_2 + 18H_2O + \text{ενεργψ}$$

From an energy standpoint the arrow is an equals sign. The total energy on the left side of the equation must equal the total energy on the right side. The sixteen carbon dioxide (CO_2) molecules and eighteen water (H_2O) molecules have lower total energy than the two isooctane (C_8H_{18}) molecules and twenty-five oxygen (O_2) molecules from which they were formed. The difference in energy is liberated as heat. It is possible to have chemical reactions in which the total energy of the molecules on the left side of the equation is less than the total energy of the molecules on the right. But for the reaction to proceed, energy must be added on the left side of the equation.

CONSERVATION OF ENERGY IN NUCLEAR REACTIONS

No name is more universally known in science than that of Albert Einstein, and no equation of physics is more recognizable than $E = mc^2$. In this equation E, m, and c stand for energy, mass, and the speed of light (300,000,000 m/s). Taken literally, the equation suggests that anything having mass has energy. It does not mean the energy of a mass moving with the speed of light. The energy, mc^2, is intrinsic to any mass whether or not it is moving. Because the speed of light is such a large number, the energy of anything is huge. For example, the mass-energy of a penny having a mass of 0.003 kg (3 grams) is 270 trillion joules. This is roughly 10,000 times the energy liberated from burning a ton of coal. As strange as it may seem, mass and energy, are equivalent and it is demonstrated routinely in nuclear reactions that liberate energy in a nuclear power plant.

The nucleus of an atom consists of protons and neutrons that are bound together by a nuclear force. Neutrons and protons are rearranged in a nuclear reaction in a manner somewhat akin to rearranging atoms in a chemical reaction. The nuclear reaction liberating energy in a nuclear power plant is called nuclear fission. The word "fission" is derived from "fissure," which means a crack or a separation. A nucleus is separated (fissioned) into two major parts by bombardment with a neutron.

Uranium in the fuel of a nuclear power plant is designated ^{235}U. The 92 protons and 143 neutrons in a ^{235}U nucleus sum to 235, the number in the ^{235}U notation. Through interaction with a neutron the 92 protons and 144 neutrons involved are rearranged into other nuclei. Typically, this rearrangement is depicted as

$$^{235}U + n \rightarrow {}^{143}B\alpha + {}^{90}K\rho + 3n + \text{ενεργψ}$$

A barium (Ba) nucleus has 56 protons and 87 neutrons; a krypton (Kr) nucleus has 36 protons and 54 neutrons. The 92 protons and 144 neutrons being rearranged are accounted for after the rearrangement. But the mass of a ^{235}U nucleus, for example, differs from the sum of the masses of 92 free protons and 143 neutrons. When you account for the actual masses involved in the reaction, the total mass on the left side of the arrow is less than the total mass on the right. The energy released on the right side of the equation is about 32 trillionths of a joule. The energy equivalent of the mass difference using $E = mc^2$ accounts precisely for the energy released. In the reaction described by the equation above, one-thousandth of the mass of the ^{235}U has been converted into energy. It is energy from reactions like this that ultimately is converted into electric energy in a nuclear power plant. The illustration of a nuclear fission reaction using an arrow becomes an equality for energy when the equivalence of mass and energy is taken into account.

The energy liberated in nuclear reactions is of such magnitude that mass differences are relatively easy to detect. The same mass-energy considerations pertain to chemical reactions. However, the energies involved are millions of times smaller, and mass differences are virtually impossible to detect. Discussions of energies involved in chemical reactions do not include mass energy. Nevertheless, there is every reason to believe that mass energy is involved.

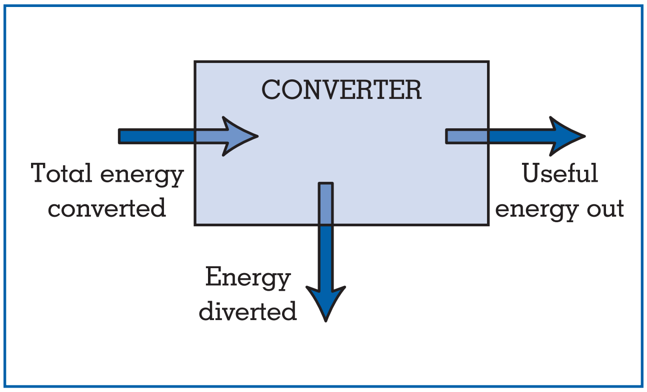


Figure 3.
A descriptive model for an energy converter.

A MODEL FOR CONSERVATION OF ENERGY

Big or small, simple or complex, energy converters must all subscribe to the principle of conservation of energy. Each one converts energy into some form regarded as useful, and each one diverts energy that is not immediately useful and may never be useful. Because energy is diverted, the efficiency defined as

$$\text{efficiency} = \text{useful energy} \div \text{total energy converted}$$

can never be 100 percent. Generally, several energy conversions are involved in producing the desired form. The food we convert to energy in our bodies involves several energy conversions prior to the one a person performs. Energy conversions are involved in our sun to produce light. Photosynthesis producing carbohydrates for the food entails energy conversions. Even the carbon dioxide and water require energy conversions for their formation. Tracking energy conversions is facilitated with the descriptive model shown in Figure 3.

An energy analysis of the production of electric energy in a coal-burning power plant provides an opportunity to illustrate the model. Generating electricity requires burning coal for heat to vaporize water to steam, a turbine driven by steam to drive an electric generator, and an electric generator to produce electric energy.

The diagram in Figure 4, with typical efficiencies for the three converters, describes the fate of 1 J of energy extracted from burning coal. It is important to note that only 0.39 J of electric energy was derived from the 1 J. More energy in the form of heat (0.61 J) was rejected to the environment than was delivered as electric energy (0.39 J). Thinking of the three converters as one unit converting the 1 J into 0.39 J of electric energy the efficiency of the converter is 0.39. This is just the product of the efficiencies of the three converters:

$$\text{efficiency} = 0.88 \times 0.45 \times 0.99 = 0.39$$

from burning coal into electric energy.

The overall efficiency is smaller than the lowest efficiency in the chain. No matter how efficient all converters in a chain are, the efficiency will always be smaller than the lowest efficiency. As long as the steam turbine is used in the commercial production of electricity, the overall efficiency will be relatively low. Electric power plants using nuclear reactors also use steam turbines, and their efficiency is essentially the same as for a coal-burning plant. Energy in the form of heat is lost in the transmission of electricity to consumers, which reduces the overall efficiency to about 0.33. This means that every joule of electric energy paid for by a consumer requires 3 joules of energy at the input of the chain of energy converters. It also means that every joule of energy saved by turning off lights when not needed saves 3 joules at the input.

The scientific principle of conservation of energy is imbedded in the model for evaluating the performance of a chain of energy converters. Applying the model, we find that the overall efficiency of a chain of energy converters is always less than the smallest efficiency in the chain. Realizing this, strong arguments can be made for conserving energy in the sense of saving. If the overall efficiency of conversion is 10 percent, as it is for an incandescent lightbulb, then saving one unit of energy at the end of the chain saves ten units of energy at the beginning of the chain.

Joseph Priest

See also: Carnot, Nicolas Leonard Sadi; Climatic Effects; Engines; Matter and Energy; Nuclear Energy; Nuclear Fission; Refrigerators and Freezers; Thermal Energy.

BIBLIOGRAPHY

Aubrecht, J. G. (1995). *Energy.* Upper Saddle River, NJ: Prentice-Hall.

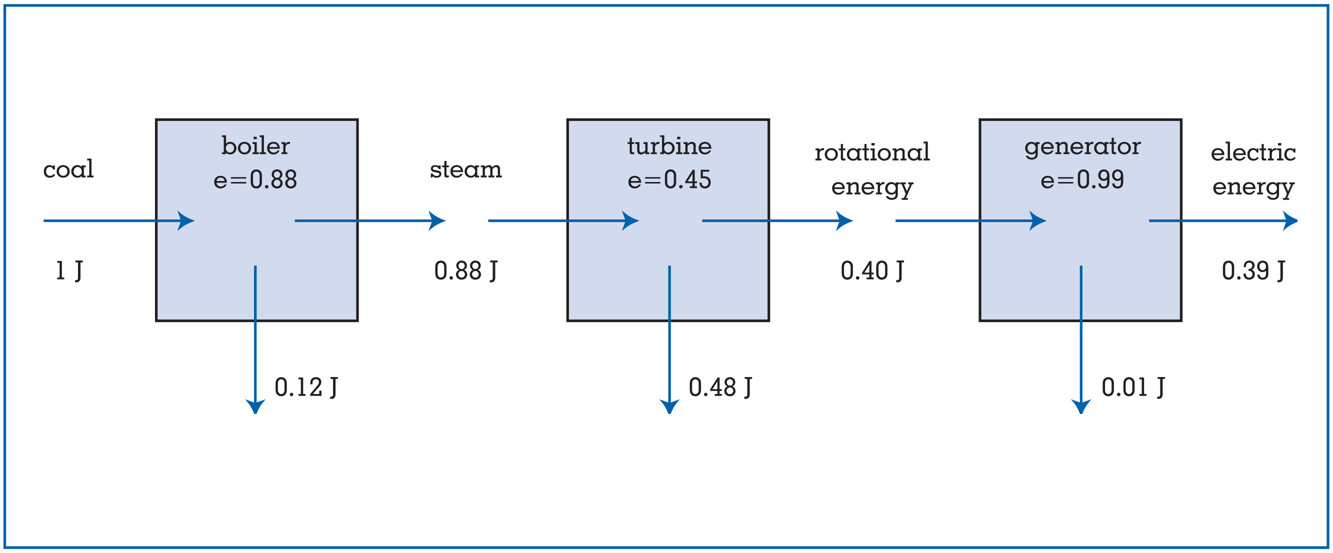


Figure 4.
A model for evaluating the conversion of energy.

Bodansky, D. (1996). *Nuclear Energy: Principles, Practices, and Prospects*. Woodbury, NY: American Institute of Physics.

Brescia, F.; Arents, J.; Meislich, H.; and Turk, A. (1998). *General Chemistry*, 5th ed. San Diego: Harcourt Brace Jovanovich, Inc.

Ebbing, D. D., and Wrighton, M. S. (1987). *General Chemistry*. Boston: Houghton Mifflin Company.

Energy. (1978). San Francisco: W. H. Freeman.

Hecht, E. (1980). *Physics in Perspective*. Reading, MA: Addison-Wesley Publishing Company.

Hecht, E. (1996). *Physics*. Pacific Grove, CA: Brooks/Cole Publishing Company.

Hill, J. W. (1999). *General Chemistry*. Upper Saddle River, NJ: Prentice-Hall.

Hobson, A. (1998). *Physics: Concepts and Connections*. Englewood Cliffs, NJ: Prentice-Hall.

Kondepudi, D. K., and Prigogine, I. (1998). *Modern Thermodynamics: From Heat Engines to Dissipative Structures*. Chichester, NY: John Wiley and Sons, Ltd.

Krame, K. S. (1995). *Modern Physics*. New York: John Wiley and Sons, Inc.

Krauskopf, K. B., and Beiser, A. (2000). *The Physical Universe*. New York: McGraw Hill Higher Education.

Priest, J. (2000). *Energy: Principles, Problems, Alternatives*, 5th ed. Dubuque, IA: Kendall/Hunt Publishing Company.

Ristinen, R. A., and Kraushaar, J. J. (1999). *Energy and the Environment*. New York: John Wiley and Sons, Inc.

Sandfort, J. F. (1979). *Heat Engines*. Westport, CT: Greenwood Publishing Group.

Serway, R. A. (1998). *Principles of Physics*, 2nd ed. Fort Worth, TX: Saunders College Publishing.

CONSERVATION SUPPLY CURVES

The cost of conserved energy (CCE) and its extension, supply curves of conserved energy, are useful tools for investigating the technical potential and economics of energy conservation measures. The CCE is an investment metric that is well suited for analysis of energy conservation investments, and the supply curve approach provides a bookkeeping framework that is ideal for diverse conservation investments. Several people, including Amory Lovins, John Sawhill, and Arthur Rosenfeld, independently developed the general approach in the late 1970s. However, Alan Meier, along with Janice Wright and Arthur Rosenfeld, systematized the concepts and procedures in the early 1980s. This article introduces the cost of conserved energy and supply curves of conserved energy, and explains their application to energy-efficiency issues.

THE COST OF CONSERVED ENERGY

Energy conservation typically involves making an investment that results in lower energy running costs. An investor (or policymaker) is often confronted with a list of possible conservation measures. The investor needs a way to rank the measures and then decide which are worth undertaking. He or she ranks the measures with the help of an investment metric, such as the simple payback time, the benefit-cost ratio, or return on investment. The investment metric provides a means of ranking the opportunities, and then separates the attractive investments from those in which the money would be better invested elsewhere.

Each investment metric has strengths and limitations. For example, the simple payback time indicates the time required to recover the investment, but it ignores any benefits that may occur after the payback time, so measures offering many years of benefits appear no better than short-lived ones. A common drawback of these investment metrics is that the price of energy must be assumed. If the energy price changes, then the payback time must be recalculated.

The CCE spreads the investment over the lifetime of the measure into equal annual payments with the familiar capital recovery factor. The annual payment is then divided by the annual energy savings to yield a cost of saving a unit of energy. It is calculated using the following formula:

$$\text{CCE} = (I/\Delta E) \times [d/(1 - (1 + d)^{-n})]$$

where I is the investment or cost of the measure; ΔE is the energy savings (per year); d is the real discount rate; n is lifetime of measure.

The CCE is expressed in the same units as the cost and the energy savings. For example, if the investment is entered in dollars and the savings are in gigajoules (GJ), then the CCE will have the units $/GJ.

For example a consumer wishes to buy a new refrigerator. The high-efficiency model (offering services identical to the standard model) costs $60 more but uses 400 kWh/year less electricity. The consumer expects to keep the refrigerator for ten years and has a discount rate of 5 percent. The cost of conserved energy in this case is calculated as follows:

$$\begin{aligned}\text{CCE} &= (\$60/400\text{kWh/y}) \times [(0.05/\text{y}) \div (1 - (1 + 0.05)^{-10})] \\ &= \$0.02/\text{kWh}\end{aligned}$$

Here of the cost of conserving a kilowatt hour is much less than the typical residential electricity price $0.08/kWh.

A collection of conservation measures can be ranked by increasing CCE. The measures with the lowest CCE are the most economically attractive. A measure is cost-effective if its cost of conserved energy is less than the price of the energy it displaces. For example, if a lighting retrofit has a CCE of 3 cents/kWh, then it will be worth doing wherever the electricity tariffs are above 3 cents/kWh. Note that the price of energy does not enter into the CCE calculation, only the decision about economic worthiness.

SUPPLY CURVES OF CONSERVED ENERGY

A supply curve of conserved energy is a devices for displaying the cumulative impact of a sequence of conservation measures. It shows the potential energy savings and CCE of each measure. Figure 1 is an example of a supply curve of conserved electricity for a commercial refrigerator. Each step represents a conservation measure. The step's width is its energy savings and the height is its cost of conserved energy.

The supply curve is useful because it shows which measures should be selected first—the ones on the left—and the cumulative energy savings. Measures with CCEs less than the price of the saved energy are cost-effective. In the example, an energy price line has been drawn to show the cut off point; those measures below the energy price line are cost-effective.

Behind the supply curve approach is a consistent bookkeeping framework. The same data for each conservation measure must be collected and the same CCE calculation performed. This encourages comparison among measures and is important when trying to assess the overall impact of many small measures. Consistent treatment also permits generalizations about the impact of alternative sequences of measures and errors in estimates of energy savings, and minimizes double-counting of energy savings. For example, if a measure is implemented before its position in the sequence shown on the curve, then the energy savings will equal or exceed those indicated, and the CCE will be lower than in the original calculation. These features make the overall approach and results more robust even when some numbers are not accurately known.

The supply curve of conserved energy is useful when trading off the benefits of additional supply against reduced demand through energy conservation, such as homes operating on photovoltaic power systems. There, the costs of supplying additional electricity can easily be compared to the costs of reducing electricity demand because both are

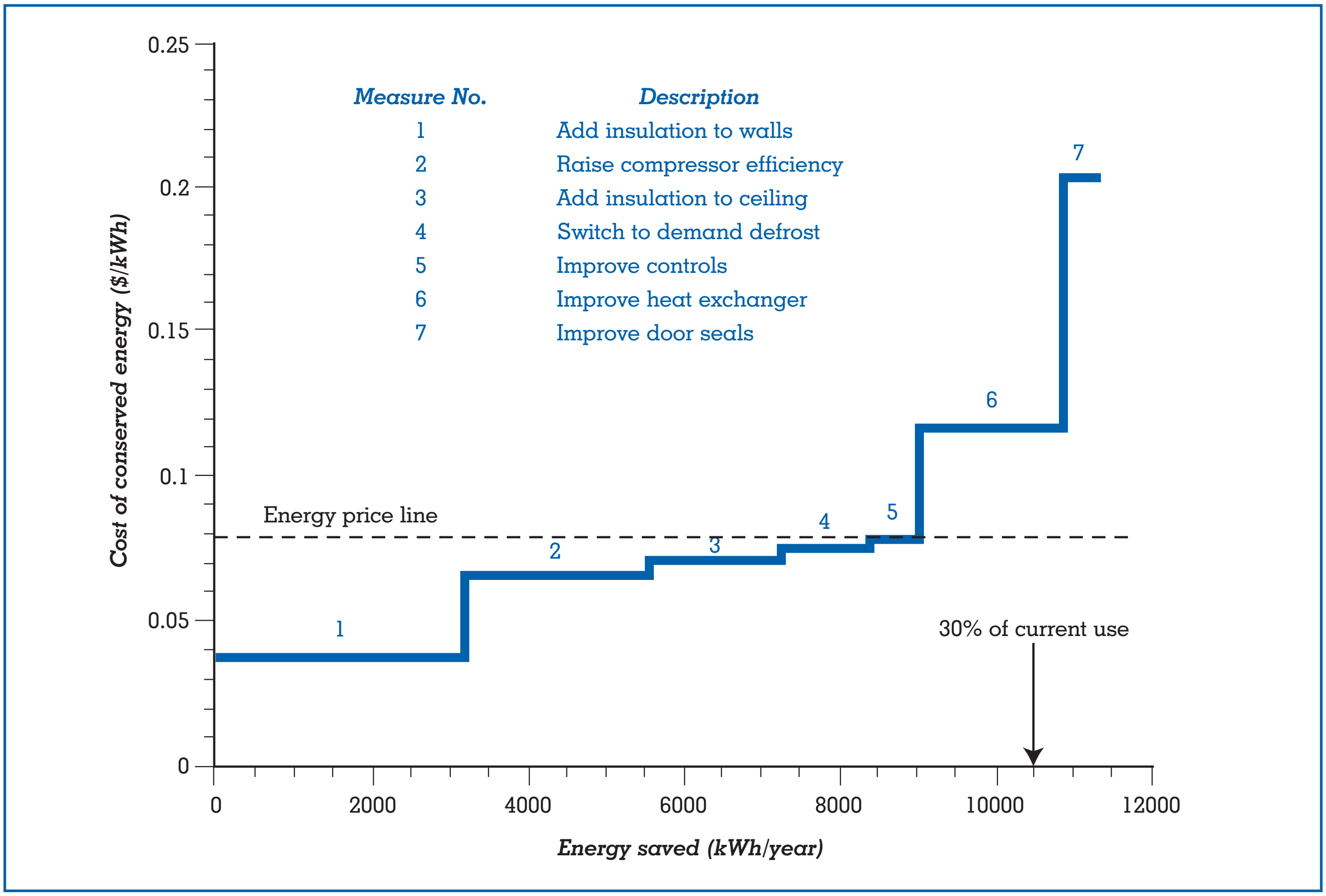


Figure 1.
A micro supply curve of conserved energy for a large commercial refrigerator. Each step represents a conservation measure. The numbers above the steps are keyed to the measure descriptions in the legend. Note that measures 1–4 are cost-effective because the cost of conserved energy is less than the energy price. If energy prices rise, then measures 6 and 7 may also become cost-effective.

expressed in same units, that is, cents/kWh. The economically optimal system will occur when the costs of supply and conservation are equal.

MACRO SUPPLY CURVES OF CONSERVED ENERGY

Figure 1 depicts a micro supply curve of conserved electricity for a single device; however, it is also useful to make macro supply curves of conserved energy, showing the potential cost and energy savings from widespread installation of conservation measures. Figure 2 shows an example of a supply curve of conserved electricity for the U.S. residential sector. Again, each step represents a conservation measure, but here the savings apply to the entire stock rather than to a single unit. This figure shows the technical potential and should not be confused with a forecast.

These aggregated, or macro, curves are especially useful for policymakers because they show the potential tradeoffs between energy conservation policies and investments in new supplies. This is otherwise difficult because most energy conservation measures are small, highly dispersed, and cannot be instantly undertaken, while energy supplies (such as power plants) typically appear in a few large units.

Consistent bookkeeping is also an important feature of the macro supply curve of conserved energy. Each measure requires, in addition to the data used to calculate the CCE data on the stocks of equipment, turnover rates, etc. The consistent inputs encourage

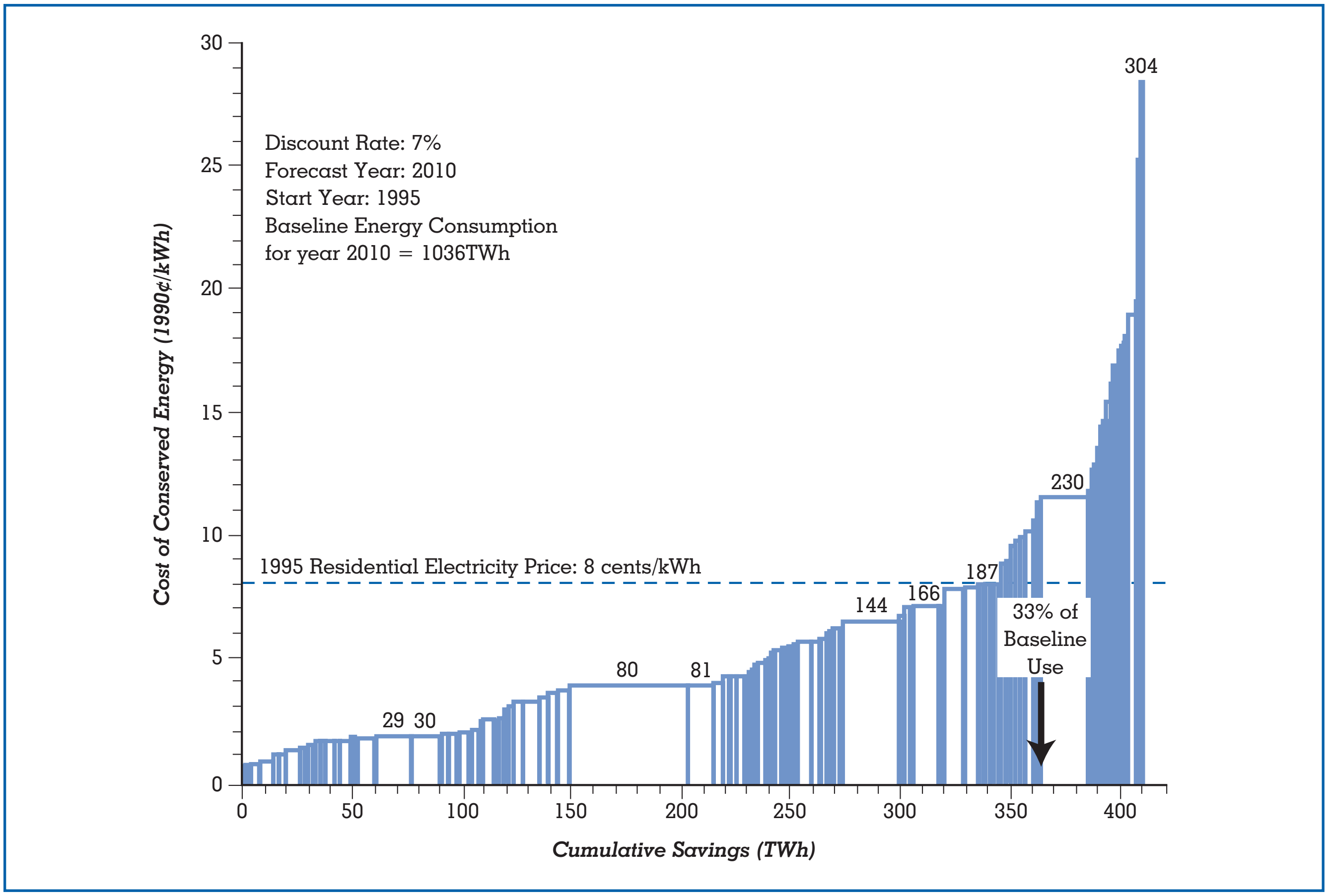


Figure 2.
A macro supply curve of conserved electricity for the U.S. residential sector. Key assumptions are given inside the chart. This supply curve shows estimated savings potentials from 304 different measures. The associated table describing the measures is too long to present here, but certain measures with numbers on top of them are noteworthy from a policy perspective. For example, measure 80 is conversion from conventional water heaters to heat pump water heaters.

confidence in comparisons among measures and in their cumulative impacts.

LIMITATIONS OF THE CCE AND SUPPLY CURVES

Some conservation measures do not easily fit into the form of an initial investment followed by a stream of energy savings because there will be other costs and benefits occurring during the measure's operating life. Furthermore, it is difficult to incorporate peak power benefits into the CCE approach. In these situations, a more precise analysis will be necessary. However, the CCE and supply curve approaches provide first-order identifications of cost-effective conservation, those measures that should be implemented first, and the overall size of the conservation resource.

Alan K. Meier
Arthur H. Rosenfeld

BIBLIOGRAPHY

American Council for an Energy-Efficient Economy. (1986). *Residential Conservation Power Plant Study.* Washington, DC: ACEEE.

Hunn, B. D.; Baughman. M. L.; Silver, S. C.; et al. (1986). *Technical Potential for Electrical Energy Conservation and Peak Demand Reduction in Texas Buildings.* Austin, TX: Center for Energy Studies, University of Texas.

Interlaboratory Working Group. (1997). *Scenarios of U.S. Carbon Reductions: Potential Impacts of Energy-Efficient* *and Low-Carbon Technologies by 2010 and Beyond*. ORNL-444. Oak Ridge, TN: Oak Ridge National Laboratory.

Meier, A. K. (1982). "Supply Curves of Conserved Energy." Ph.D. diss. University of California, Berkeley.

Meier, A. K. (1983). "What is the Cost to You of Conserving Energy?" *Harvard Business Review* 61(1):36–38.

Meier, A. K.; Wright, J.; and Rosenfeld, A. H. (1983). *Supplying Energy Through Greater Efficiency*. Berkeley, CA: University of California Press.

Panel on Policy Implications of Greenhouse Warming of the National Academy of Sciences. (1992). *Policy Implications of Greenhouse Warming: Mitigation, Adaptation, and the Science Base*. Washington, DC: National Academy Press.

Ruegg, R. T., and Marshall, H. E. (1990). *Building Economics: Theory and Practice*. New York: Van Nostrand Reinhold.

CONSUMPTION

Energy is a commodity, and like any commodity, its level of consumption is largely a reflection of its price. However, unlike most other commodities, energy is not valued as a good unto itself, but as a means to achieve an end—the power behind the technology that makes it possible to do more and improve standards of living. For this reason, huge sums are spent by governments, utilities, and businesses to gather the statistics to forecast consumption patterns. They want to know not only how much energy can be found, extracted, transported, and converted to useful forms, but also how that energy powers today's technology and the emerging technologies of tomorrow.

CONSUMPTION IN THE UNITED STATES

From 1967 to 1973, U.S. energy consumption dramatically increased, from 57.57 quadrillion Btus to 74.28 quadrillion Btus (see Figure 1). In these years leading up to the 1973 Arab oil embargo, energy was inexpensive and growth in consumption was closely linked to population and economic growth. Forecasters could look at population and economic growth trends and accurately project a similar growth in energy consumption. The oil price shock changed this dynamic by spurring energy-efficiency improvements. Overall energy consumption declined to 70.52 quadrillion Btus by 1983 thanks, in large part, to dramatic improvements in energy efficiency. People in the United States, as well as much of the developed world, learned how to do more with less. More fuel-efficient cars, new energy-efficient homes and office buildings, the retrofitting of existing structures, and more energy-stingy appliances and lighting all played a part in lowering energy consumption from the mid-1970s to the mid-1980s.

The years 1983 through 2000 were marked by stable or dropping energy prices. Adjusting the price of energy for inflation, it was about the same in 1998 as it was in 1967. In the case of gasoline, the inflation adjusted price was actually lower in 1999 than it was before the oil embargo. As a result, people put less effort into conserving energy and became more open to incorporating new energy-hungry devices into their lives. Products that were virtually nonexistent in the 1960—dishwashers, microwaves, telephone answering machines, and personal computers to name a fe—are now considered necessities by many. Americans felt an urge to splurge as consumption gradually reached 94.21 quadrillion Btus by 1997.

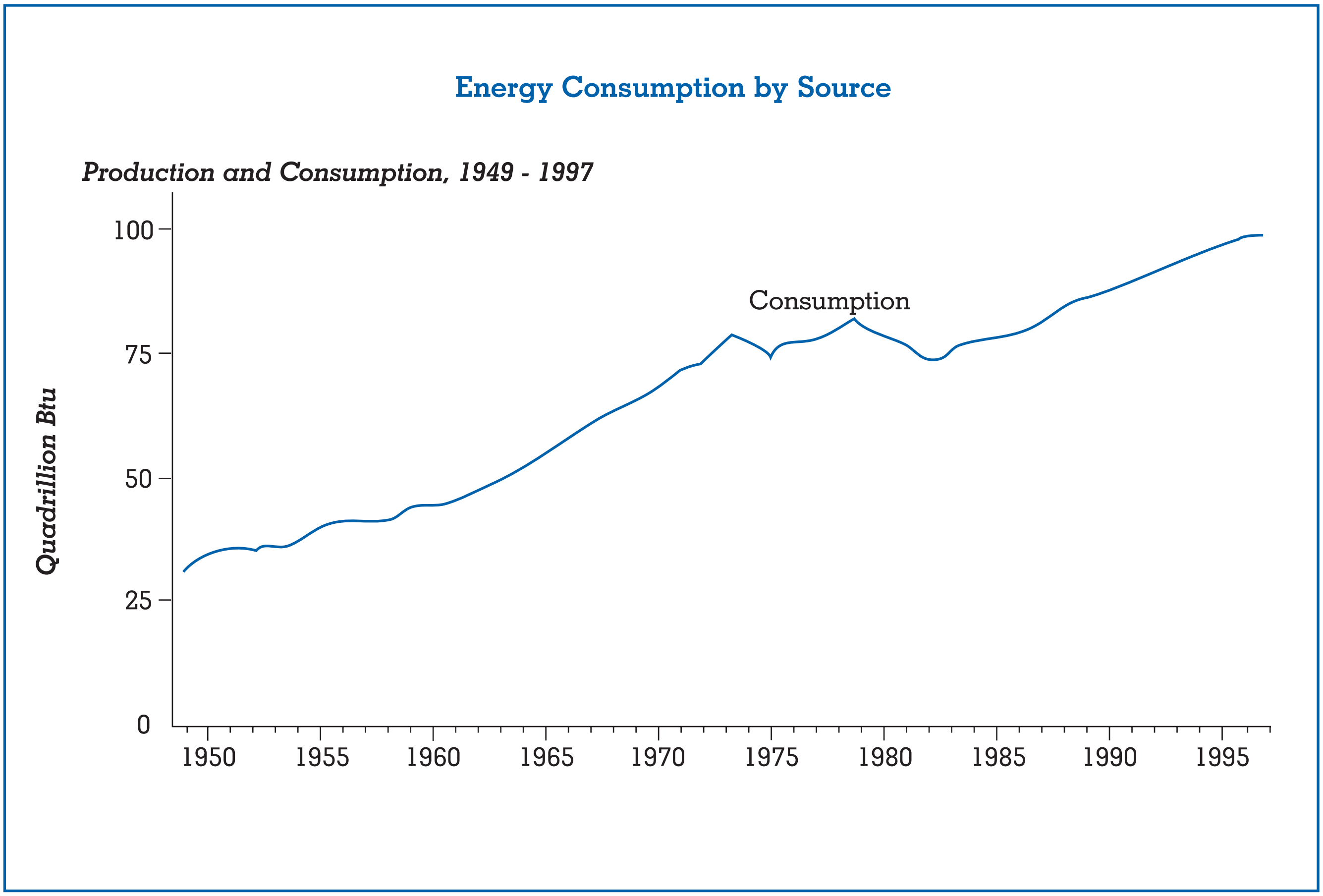
Energy-efficiency improvements have offset much of the increased demand from these new appliances. But demand has increased, and will continue to increase, as the combination of continually increasing disposable income and cheap energy accelerates ever more novel inventions and innovations requiring energy.

The U.S. Department of Energy (DOE) projects that the prices of oil and natural gas will rise modestly until the year 2020, and the price of coal and electricity will fall approximately 20 percent by then, which will help push energy consumption to 119.4 quadrillion Btus by 2020, more than double the level of consumption in 1967.

THE DUAL ROLE OF EFFICIENCY

Efficiency has a dual role. While improvements were being made on the use side, major efficiency improvements were happening on the conversion side—the converting of heat energy to electricity. From 1970 to 2000, the efficiency of combined-cycle natural gas plants jumped from about 38 percent to 58 percent. The average efficiency of coal-burning steam turbine plants also improved, with the average for better units attaining 34 to 38 percent efficiency, and a few new units reaching the 45 to 48 percent range.

Efficiency improvements benefit a nation because


Energy Consumption by Source

Production and Consumption, 1949 - 1997

Figure 1.

NOTE: 1×10^{15} British thermal units = One quadrillion Btus

SOURCE: Department of Energy, *Annual Energy Review 1997.*

it takes half as much coal to produce a given amount of electricity with a 40 percent efficient coal-fired steam plant in 2000 as it did with a 20 percent efficient plant in 1970. Even though electricity consumption is expected to increase by more than 20 percent from 2000 to 2020, inefficient plants will continue to be retired, retrofitted, or replaced by efficient plants, so that it will take far less than 20 percent more fossil fuel to meet this greater consumption.

A secondary benefit is that efficiency gains in fossil fuel generation also reduce all types of harmful emissions, even carbon dioxide—the greenhouse gas suspected by many as a major culprit of climate change. A 45 percent efficient plant releases approximately 40 percent less CO_2 per megawatt-hours of electricity produced than a 25 percent efficient plant that it might be replacing.

ENERGY CONSUMPTION BY SOURCE

The U.S. government has tracked consumption by energy source since 1949, the year when petroleum overtook coal as the major source of energy (see Figure 2). Petroleum consumption continued to increase throughout the 1950s and 1960s due to increases in transportation and industrial demand, and as a coal replacement for heating and electric power generation. Natural gas consumption also increased during this period as it became the fuel of choice for home heating.

Following the Arab oil embargo of 1973, natural gas consumption declined until the mid-1980s primarily due to the assumption that the nation was running out of natural gas and because of legislation outlawing the use of natural gas for "low priority" uses. Energy conservation efforts in the industrial, commercial, and residential sectors, primarily the improved energy efficiency of new furnaces and boilers, also were instrumental in this decline. Petroleum consumption peaked later, in 1978, and then began to fall as older vehicles were replaced by more fuel-efficient models, and because of the effort of utilities to switch from petroleum as a fuel for generating electricity.

In the early 1970s, coal consumption once again equaled its earlier peak in the early 1950s and continued to grab a larger share of the electricity-generation market due to the price and supply problems of petroleum and natural gas.

Beginning in 1986 and through the 1990s, natural gas consumption rose again as the Federal Energy Regulatory Commission began deregulating natural gas, and natural gas electricity generation became the choice due to innovations improving the efficiency of generating technology. These new plants were not only more efficient than coal-fired plants, but also less expensive and time-consuming to construct. By 1998, natural gas consumption equaled its 1972 peak of 22.6 quadrillion Btus.

Many nuclear power plants were ordered in the 1960s and early 1970s, but construction slowed in the mid-1970s and halted in the early 1980s because of the high cost of construction, problems with radioactive waste disposal, and political obstacles. Despite no new power plants being built, better management and technology resulted in more energy generation during the late 1980s and 1990s. In 1996, nuclear facility efficiency (the amount of power generated divided by the maximum possible generation) reached an all-time high of 76.4 percent. Although the amount of nuclear-generated electricity more than doubled between 1980 and 1996 (2.74 to 7.17 quadrillion Btus), the future contribution is certain to fall through 2020 for three important reasons: limited potential for further gains in efficiency, many nuclear facilities are scheduled for retirement, and no new facilities are planned.

Out of the 7 quadrillion Btus contributed by renewable energy, more than 95 percent comes from hydroelectric power and biofuels (waste energy, wood energy, and alcohol). Geothermal, solar, and wind are all very minor contributors. Renewable energy's share is unlikely to grow because hydroelectric power faces political and environmental concerns about dams, no new geothermal sites are planned, and biofuel potential is limited. Another major factor hindering growth in renewables was the much lower than expected electric generating cost for coal and natural gas. The projections made in about 1980 for the year 2000 were way off target. In terms of dollars per million Btus of energy, coal was widely projected to reach $3 to $5, not $1, and natural gas was projected to reach $4 to $8, not $2.

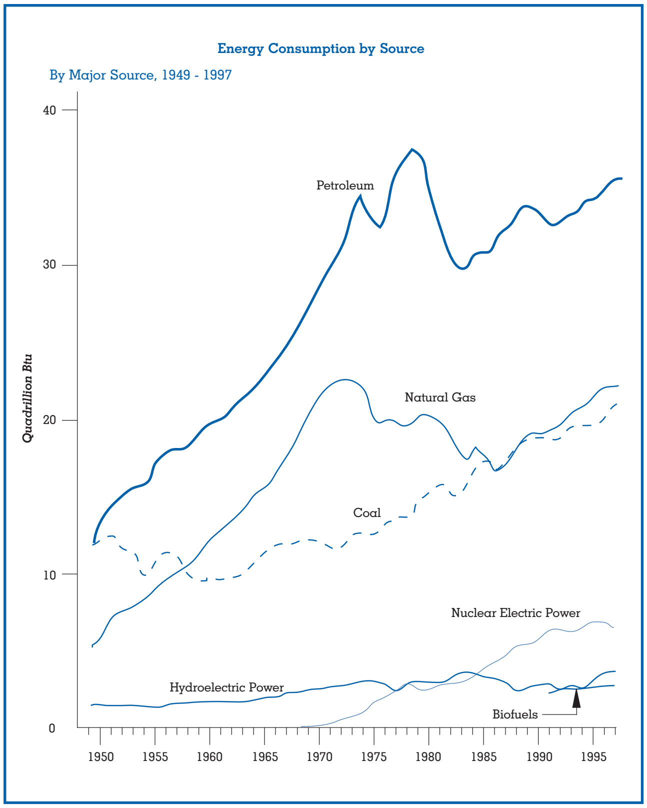


Figure 2.

Note: Because vertical scales differ, graphs should not be compared.

SOURCE: Department of Energy, *Annual Energy Review 1997.*

In the future, if the criterion for selecting new generating capacity was solely fuel cost, coal will be the number one choice. But the much greater costs of coal-fired plants (primarily to meet local and federal emission standards), as well as the potential of tighter standards, will make gas more attractive in many cases. And although natural gas prices may rise, the fuel costs per kilowatt-hour for gas-fired power plants should remain unchanged as efficiency gains offset the rise in fuel prices.

CONSUMPTION BY SECTOR

Among the four major sectors of the economy—residential, commercial, industrial, and transportation—

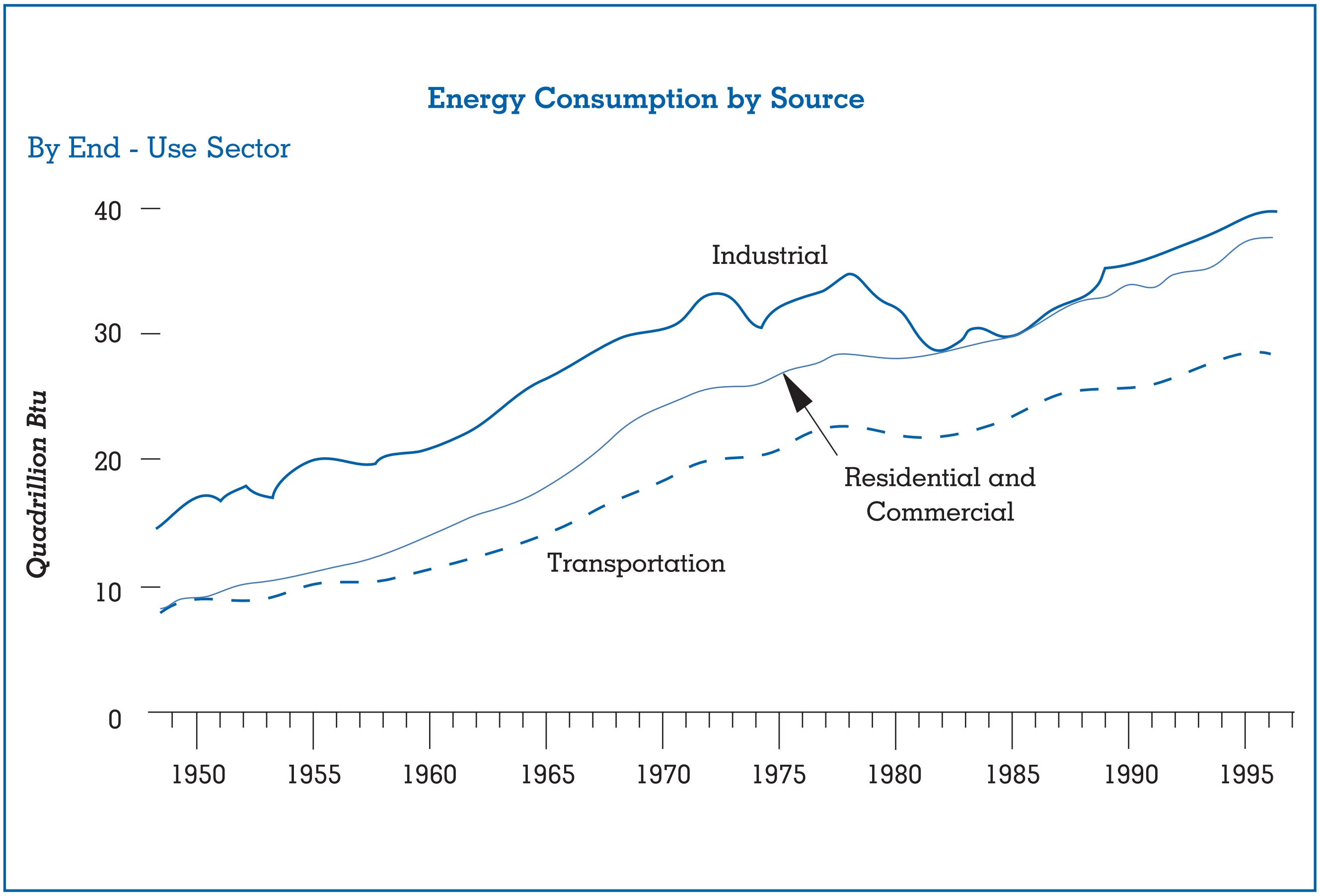


Figure 3.

SOURCE: Department of Energy, *Annual Energy Review 1997.*

the industrial sector has historically been the greatest user of energy (see Figure 3). The industrial sector includes energy-intensive industries such as those for paper, metals, chemicals, and petroleum products as well as light-energy industries such as textiles and furniture manufacturers.

In 1951, industrial consumption was 17.13 quadrillion Btus, nearly as much as that for residential and commercial and transportation combined. From the 1950s through the early 1970s all three sectors continued to grow, with the residential and commercial sector showing the greatest increase, a reflection of the growth in office and housing stock.

The period 1973 to 1983 was one in which industrial consumption declined significantly, falling from 31.53 quadrillion Btus to 25.75 quadrillion Btus. There were a number of factors responsible for this steep drop. Foremost among them were the increase in the price of energy, a drop in output in many energy-intensive industries, and vast improvements in efficiency. Per thousand dollars of manufacturing sales, energy use declined by more than 25 percent between 1977 and 1997. The concerted energy-efficiency efforts of the industrial sector resulted in consumption in the residential and commercial sector briefly surpassing the industrial sector by 1986.

From 1983 to 1998, all three sectors of the economy experienced an increase in consumption largely because the price of all energy sources dropped. In the residential and commercial and the industrial sectors, the greatest growth was in natural gas and electricity consumption. Although new homes were more energy-efficient, they were on average much larger and filled with more energy-hungry housing features that either did not exist or were considered luxuries—air conditioning, Jacuzzis, microwaves, dishwashers, and security systems, to name a few—for home buyers in the 1970s. Some of the consumption increase can be attributed to the need to space-condition more area, as the average home size expanded from 1,600 square feet in 1970 to 2,100 square feet in 1998. The amount of commercial sector space also expanded, which required more energy for air conditioning, heating, and lighting.

Consumption in the transportation sector historically has been less than either the industrial sector or the residential and commercial sector, but it is of greater concern because consumption is almost entirely petroleum as opposed to a mix of petroleum, natural gas, and electricity. And because electricity can be generated from a number of different sources, and because there is greater opportunity to substitute one source of energy for another in generating electricity, the price and supply security that exists in other sectors does not apply to the transportation sector.

Another problem with the dependence of transportation on petroleum is the enormous military, strategic, economic, and political costs of securing the petroleum supply. Estimates of the military costs alone associated with the Mideast region for the period from 1980 to 1993 ranged from $100 billion to more that $1 trillion.

Transportation is also the emissions leader. About 75 percent of carbon dioxide emissions and 45 percent of nitrogen oxide emissions come from the transportation sector. If rising levels of CO_2 are found to be responsible for global warming, and measures are put in place to severely curtail CO_2 emissions, the measures will have the greatest impact on the transportation sector.

As the other sectors have shifted from petroleum beginning in the mid-1970s, the transportation sector has not because of the formidable petroleum-consuming internal-combustion engine. Although competing technologies (electric, fuel cells, and hybrid vehicles) create less emissions, internal-combustion-engine technology continues to make dramatic steps in limiting emissions. It is likely to remain the dominant means of transportation propulsion through 2020, which is why petroleum consumption in the transportation sector will remain high. In 1997, a total of 24.04 quadrillion Btus out of the 36.31 quadrillion Btus of petroleum consumption found use in the transportation sector, consuming nearly all the 24.96 quadrillion Btus that had to be imported.

Close to 65 percent of the petroleum consumed for transportation is gasoline, followed by 18 percent for diesel fuel and 13 percent for jet fuel. The DOE expects that the rise of "cleaner burning" alternative fuels should lower the share of gasoline and diesel, and expects greater air travel to increase the share of jet fuel.

Vehicle technology exists to improve fuel economy to 45 miles per gallon (if 45 miles per gallon was the average, the United States could be a net exporter of petroleum), yet the average consumer of the 1990s values performance and other features far more than fuel economy. This shifting priority away from fuel economy is why the steady improvements in gas mileage through the late 1970s and 1980s halted in 1990s. The trends of more people, more vehicles per person, and more miles driven per vehicle all point toward transportation petroleum consumption increasing significantly in the future.

THE WORLD PICTURE

Despite the Asian economic crisis of the late 1990s, the world is projected to consume more than twice as much energy in 2000 as in 1970, and more than three times that amount by 2020 (see Figure 4). In the world at large, energy consumption is greatest in North America, followed by Europe and the Pacific Rim countries. In the early 1970s, approximately 60 percent of all energy consumption occurred in North America. Consumption growth of the Asian Pacific economies led the world in the 1980s and 1990s, which brought the North American share of energy consumption down, to about 50 percent. Still, looked at from a consumption per capita basis, the 4 percent of the world population who reside in the United States consume 50 percent of all energy. Many feel that this fact could have dire consequences. If the developing world follows the energy-intensive U.S. model, the world will have to deal with the environmental problems associated with much greater energy use.

The United States is by far the most energy-intensive user. Americans consume more than twice as much energy per capita as in Europe or Japan, and are the largest emitter of CO_2. Of the 6 billion metric tons of carbon emissions worldwide in 1996, the United States emitted 1.466 billion metric tons, followed by China at 805 million, Russia at 401 million, and Japan at 291 million. Although many countries consume more energy than China, China consumes

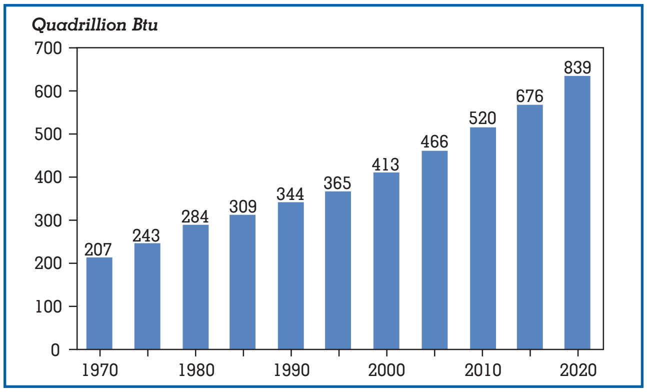


Figure 4.
World Energy Consumption 1970-2020.
SOURCE: U.S. Energy Information Administration.

vast amounts of coal—the worst fossil fuel for CO_2 emissions.

As long as the source for consumption is mainly fossil fuels, who is doing the consuming and how much is being consumed promise to be two of the most contentious aspects of international relations for the foreseeable future. Energy consumption presents a troubling paradox: Greater energy consumption is the result of a higher standard of living, but this greater consumption also portends greater, and often unacceptable, environmental costs.

John Zumerchik

See also: Biofuels; Coal, Consumption of; Economically Efficient Energy Choices; Economic Growth and Energy Consumption; Emission Control, Vehicle; Fossil Fuels; Gasoline Engines; Hydroelectric Energy; Natural Gas, Consumption of; Petroleum Consumption; Population Growth and Energy Consumption; Supply and Demand and Energy Prices; Transportation, Evolution of Energy Use and.

BIBLIOGRAPHY

BP Amoco. (1998). *BP Statistical Review of World Energy*. London: Author.

Green, D. L. (1996). *Transportation and Energy*. Washington, DC: Eno Transportation Foundation.

International Energy Agency. (1998). *Energy Statistics of OECD Countries*. Paris: Author.

Nye, D. E. (1997). *Consuming Power: A Social History of American Energies*. Cambridge, MA: MIT Press.

U.S. Energy Information Administration. (1998). *Annual Energy Review 1997*. Washington, DC: Author.

U.S. Energy Information Administration. (1998). *Annual Energy Outlook 1999*. Washington, DC: Author.

U.S. Energy Information Administration. (1998). *International Energy Outlook 1998, with Projections through 2000*. Washington, DC: Author.

CONTROL SYSTEMS, HISTORY OF

INTRODUCTION

The control of energy in its various forms was always a necessity that became more relevant with the increasing performance requirements of the twentieth century. The control of energy conversion contributes to the optimization in performance and energy efficiency for all processes, machines, and devices.

Technology developments from 1960 to 2000 in the areas of microelectronics and power electronics made possible the development of more complex, efficient and reliable energy controls. In this century there were significant technology mutations, with controls going from mechanical to electromechanical devices, evolving gradually to full electronic controls without moving parts. Since the 1970s electronic controls have been implemented more and more with programmable systems through the use of microcomputers.

The complexity and reduced time constants of modern processes imply the adoption of high performance programmable controllers. This requires not only higher processing speed but also more advanced control algorithms that can optimize the process operation in real time.

Today, with the pressing need to achieve sustainable development, the reduction of the energy losses and the optimization of all processes has promoted the continuous development and implementation of advanced energy control systems in all sectors.

CONTROL SYSTEMS

Basic Concept of the Control

There is a need to have some measurements or observations made on the relevant variables of the controlled system. The data is compared to a reference, and that will cause some feedback on the process to be controlled, in order that value of the controlled variables approaches the desired reference value.

Types of Controls

Manual Controls. The first methods used in energy control involved human intervention. The operator was the sensor (i.e., using his eyes, ears, and hands or using additional devices to quantify the values of the controlled variables), and he was also the actuator controller. The control of the processes was slow and very ineffective. For example, in an old steam engine control the human operator sees the instantaneous pressure and then manually regulates the power of the device (e.g., by adding fuel to a boiler). But in today's industrial reality, this control is not only ineffective but in most cases is not possible.

Another example is the electric generators' excitation control. Early systems were manually controlled (i.e., an operator manually adjusted the excitation system current with a rheostat to obtain a desired voltage). Research and development in the 1930s and 1940s showed that applying a continuously acting proportional control in the voltage regulator significantly increased generator performance. Beginning in the 1950s most of the new generating units were equipped with continuously acting electronic automatic voltage regulators (AVRs).

Mechanical Controls. Mechanical controls have been widely used in steam and internal combustion engines. Although they are low-cost, they can only implement simple control strategies.

One of the oldest energy control systems is the steam engine speed control device, developed by James Watt, consisting in the regulation of motor speed through input steam flow. This device is purely mechanical, and its physical principle is shown in Figure 1.

The angular velocity (ω_{engine}) control of the steam engine load shaft is controlled by a steam valve. As the angular velocity rises, the centrifugal force ($F_{centrifugal}$) increases in the spheres, becoming higher than spring force. The result is a displacement (D_{linear}) of the shaft that partially closes the steam valve. If, due to a load perturbation, the velocity is higher than desired, the increasing speed will force the valve to partially close, decreasing the steam flow. This is an example of a simple negative feedback closed loop mechanical control. The human operator defines an input (desired engine angular velocity), with the regulation of the shaft length $l_{regulation}$ and the vertical position of the spheres is the angular speed sensor.

Electromechanical Controls. Electro-mechanical control devices are typically used for load control (lighting, ventilation, and heating) in buildings with no feedback signal. The most common device is the electromechanical timer, in which a small motor coupled to a gearbox is able to switch electrical contacts according to a predefined time schedule. They are still in use today, applied to loads with simple scheduling requirements.

Mechanical switches provide a simple manual interface to operate all sorts of loads, but they suffer from all the drawbacks of manual control, namely in terms of speed of response, and also requiring permanent operator awareness.

The first power control devices, with electric insulation between the low power input stage and high power output stage, were electromechanical relays, introduced in the late nineteenth century. They are still widely used today. Applying voltage to the magnetic field winding, an attraction force will be generated between the mobile and fixed iron core that will switch the mobile terminal COM (common terminal) between the normally closed (NC) and normally open (NO) position. Their main advantages are high electric input/output insulation, a high input/output power ratio, high efficiency, and low cost. The disadvantages are the limitations in terms of commutation speed and the limited number of operations (10^5–10^7 cycles). With the invention of the transistor, this type of device evolved into the solid-state relay, which has the advantage of smaller size, higher reliability, and lower input power. The operation principle is based on a light emitting diode (LED), fed by a control signal, which with the emitted light will excite the fototransistor. The zero-cross detector module will then control the firing of the solid-state switches (thyristors), closing the load power circuit. In the late 1990s, the integrated silicon electrostatic relay was introduced, available in an integrated circuit package that has the advantages of small size and very low power consumption. However, the output power is very limited, when compared to electromagnetic and solid-state relays. Basically, the input voltage between a conductive fixed plate and a

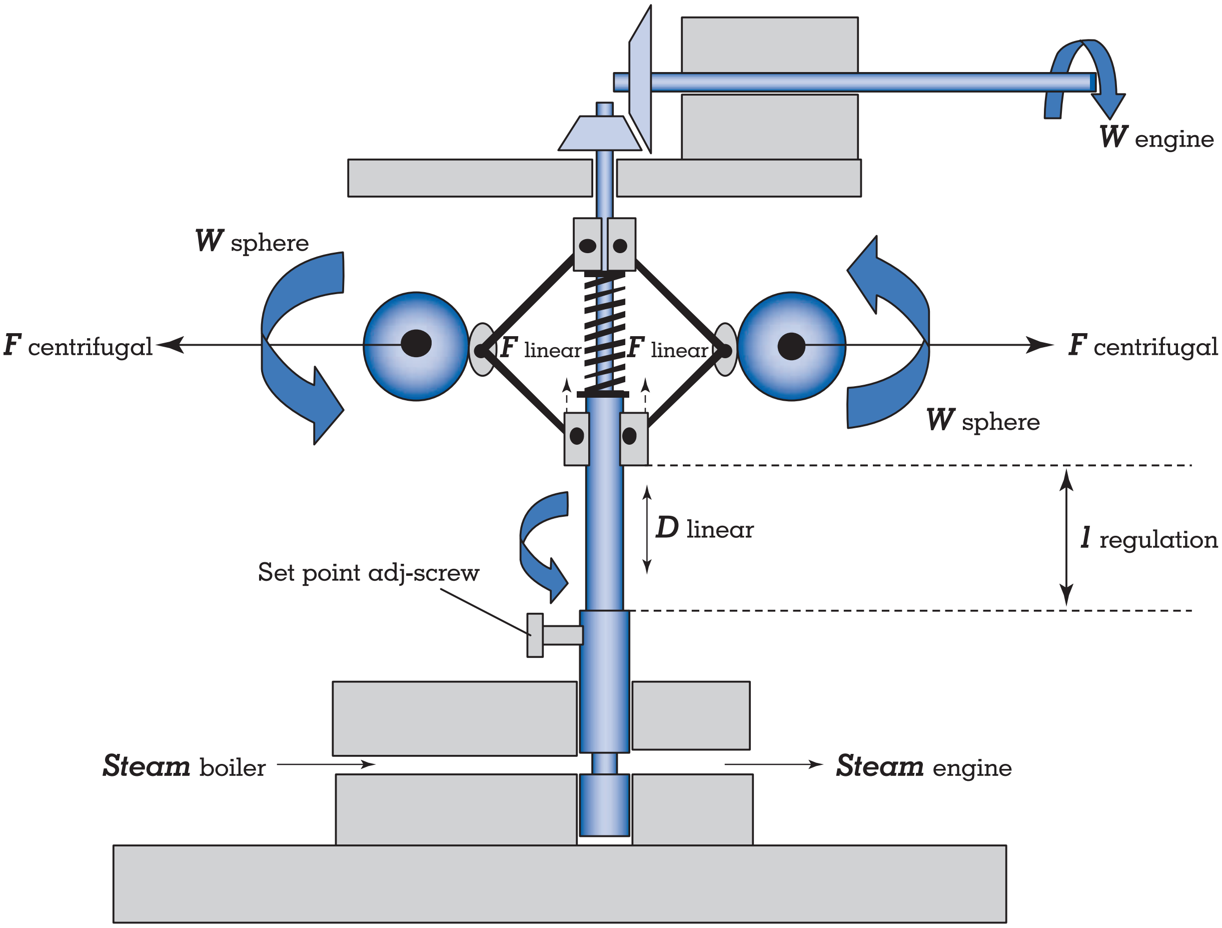


Figure 1.
Example of a pressure control device for a steam engine.

flexible plate creates an electric field between them, which will generate an attraction force. The flexible plate will be deflected and its conductive terminal will switch the contacts.

Electronic Closed Loop Controls. After its development, the concept of closed loop control has become one of the most common tools for systems control. Initially, automatic closed loop controls were widely implemented with electronic analog circuits. Electric power systems were one of the many applications that successfully used these types of controls in power plants.

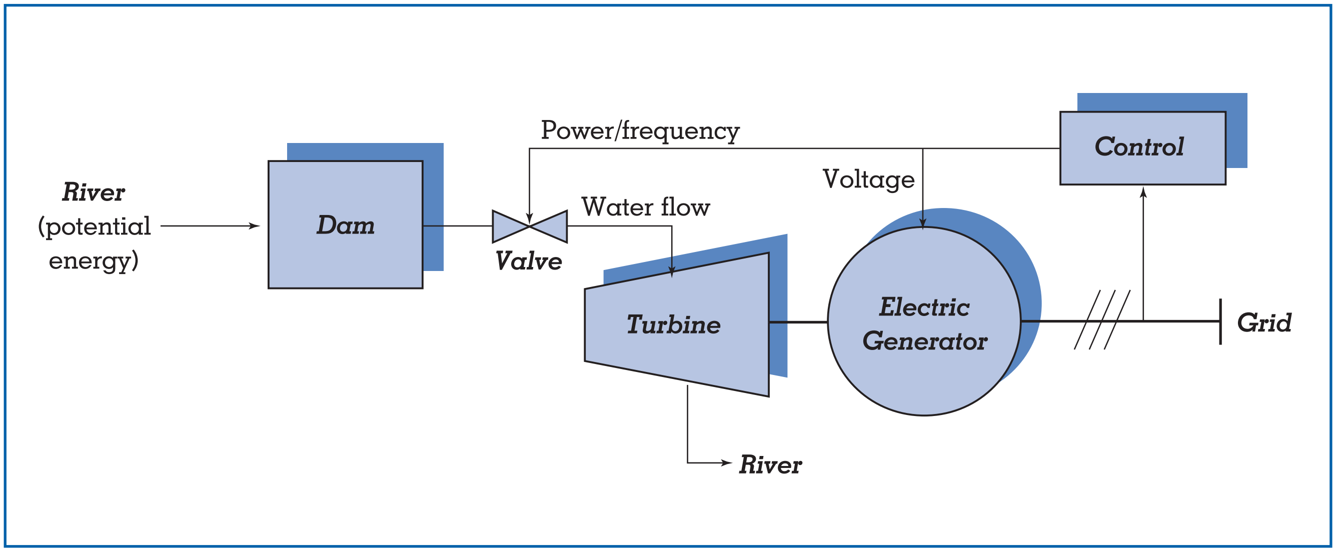


Figure 2.
Simplification of a hydroelectric power plant control system.

In electric power systems, it is essential to have permanent control of the power in electricity production, transportation, and consumption. Because of speed and reliability requirements, electric power systems were the first large systems to use a variety of automatic control devices for the protection of different parts of the system.

A variety of electromechanical relays was used for the detection of abnormal operating conditions (e.g., overload, short-circuit, etc.), leading to the isolation of the faulty components. In the second half of the twentieth century electromechanical relays were progressively replaced by electronic controls that are capable of faster response and higher reliability.

Power generation also makes use of electronic energy controls. A typical hydroelectric power plant using river water flow as its energy source is organized into three main subsystems, corresponding to the three basic energy conversions taking place in the process (Figure 2): the dam, the turbine, and the electric generator (or alternator). The dam stores the water and, using a valve, controls the output water flow. The turbine transforms the kinetic energy of water flow into turbine mechanical power and makes it available at the shaft for final conversion into electrical power in the electric generator. In the generators the electric power and frequency variables are related to the flow of water (or steam in the case of thermal power plants) in the turbines, and the basic closed loop control involves the reading of the two variables: power and frequency. The controller regulates the function of the valves and thus the turbine valve opening. The first power plants used a speed governor in which the valves were controlled by a mechanism similar to the one described in Figure 1. To guarantee constant voltage in the terminals, an automatic voltage regulator (AVR) is necessary, controlling the generator excitation current as a function of the generator output voltage.

A typical excitation system includes a voltage regulator, or exciter, protection circuits, and measurements transducers. If the terminal voltage decreases below rated value, the control system increases the rotor current and thus the magnetic field; as shown in Faraday's law, the generated voltage is thus forced up to the desired value.

Microcomputer Based Electronic Systems. The introduction of microelectronics and computers in energy control systems enabled the implementation of complex closed-loop control algorithms. A typical example of this type of control is an oven temperature control. Figure 3 depicts a diagram of temperature control in an electric oven. The oven temperature is measured by a sensor that gives an analog output. The sensor output signal is converted to a digital signal using an analog to digital converter and then transmitted to a computer that has a control program. The input temperature is compared with a

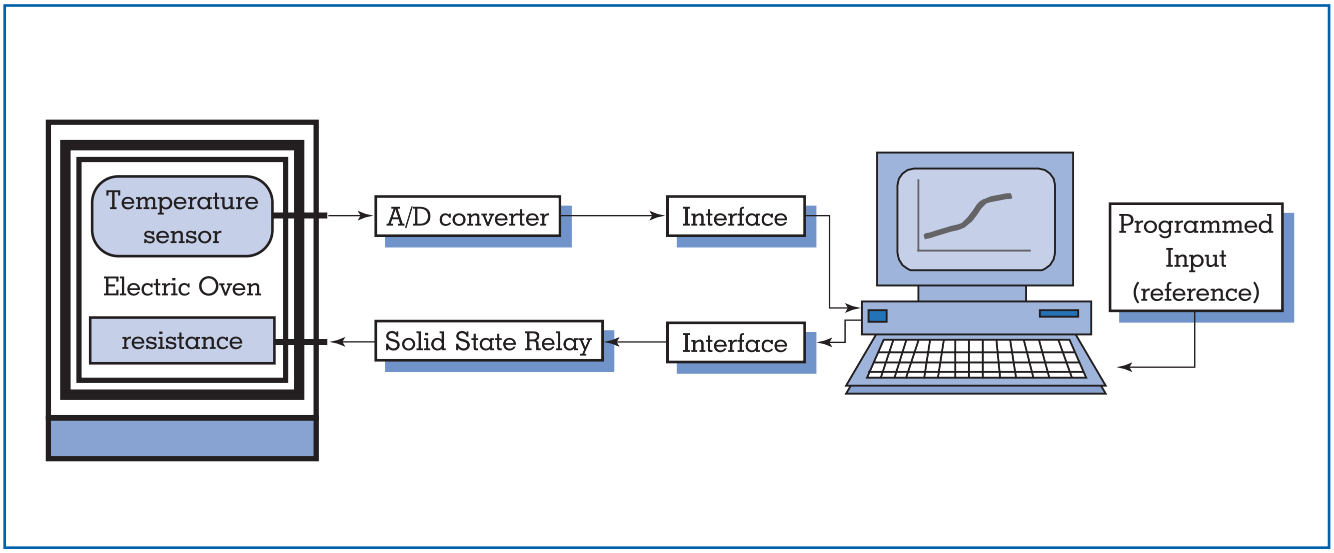


Figure 3.
Temperature control system for an electric oven.

programmed reference temperature, and the computer generates a correcting output to the heater using an interface and a solid-state relay to force the oven temperature to the desired value. A common control algorithm uses a correction signal which is a combination of proportional plus integral plus differential (PID) terms based on the error (difference between desired and actual temperatures).

One of the most recent trends in energy control is the use of fuzzy logic. Fuzzy logic is a multivalued logic that allows intermediate values to be defined between conventional evaluations such as yes/no, true/false, and black/white. Notions such as “rather warm” or “pretty cold” can be formulated mathematically and processed by computers. In this way an attempt is made to apply a more human-like way of thinking in the programming of computers. The employment of fuzzy control is commendable for very complex processes, when there is no simple mathematical model, in highly nonlinear processes, and if the processing of expert knowledge (linguistically formulated) is to be performed. Some applications for fuzzy logic control in the energy field are: automated control of dam gates for hydroelectric power plants, prevention of unwanted temperature fluctuations in air-conditioning systems, improved efficiency and optimized function of industrial control applications, control of machinery speed and temperature for steel works, improved fuel consumption for automobiles, and improved sensitivity and efficiency for elevator control. Fuzzy logic (control) is a way of interfacing inherently analog processes that move through a continuous range to a digital computer that likes to see variables as well-defined numeric values.

Let us consider for example the system to control the temperature in a building is directed by a microcontroller that has to make decisions based on indoor temperature, outdoor temperature, and other variables in the system. The variable temperature in this system can be divided into a range of “states”: “cold,” “cool,” “nominal,” “warm,” and “hot.” However, the transition from one state to the next is hard to pin down. An arbitrary threshold might be set to divide “warm” from “hot,” but this would result in a discontinuous change when the input value passed over that threshold. The microcontroller should be able to do better than that. The way around this is to make the states “fuzzy,” that is, allowing them to change gradually from one state to the next. The input temperature states can be represented using “membership functions.”

The input variables’ state now no longer jumps abruptly from one state to the next, but loses value in one membership function while gaining value in the next. At any one time, the “truth value” of the indoor or outdoor temperature will almost always be in some degree part of two membership functions:

0.6 nominal and 0.4 warm, or 0.7 nominal and 0.3 cool, and so on. Given "mappings" of input variables into membership functions and truth values, the microcontroller then makes decisions as to which actions to take based on a set of "rules" that take the form:

IF *indoor temperature* IS *warm*
AND *outdoor temperature* IS *nominal*
THEN *heater power* IS *slightly decreased.*

Traditional control systems are in general based on mathematical models that describe the control system using one or more differential equations that define the system response to its inputs. In many cases, the mathematical model of the control process may not exist or may be too "expensive" in terms of computer processing power and memory. In these cases a system based on empirical rules may be more effective. In many cases, fuzzy control can be used to improve existing controller systems by adding an extra layer of intelligence to the current control method.

Analog Displays

Analog displays are simple devices like mercury thermometers and pressure gauges in which the variation of a physical variable causes a visible change in the device display. This information can be used for monitoring purposes and for manual process control. Although simple and low-cost, most analog displays suffer from poor accuracy. Analog displays have been progressively replaced by electronic sensors.

Sensors

Both electronic and microcomputer-based controls require information about the state of the controlled system. Sensors convert different physical variables into an electric signal that is conditioned and typically converted to a digital signal to be used in microcontrollers. The trend in the construction techniques of modern sensors is the use of silicon microstructures because of the good performance and the low cost of this type of device. In the energy control scope the main quantities to be measured are the temperature, pressure, flow, light intensity, humidity (RH), and the electric quantities of voltage and current.

Temperature. The simplest temperature sensor/control systems typically use a bimetallic thermostat that integrates two superimposed metal plates with different expansion coefficients. Thus, the deflection of the beam caused by a temperature increase will cause the opening of the contacts. This device is typically used to control electrical heater equipment operation. The most common used devices in temperature measurement are thermocouples, thermistors, semiconductor devices, platinum resistance thermometers, and infrared radiometers.

One widely used temperature sensor is the integrated circuit AD590 introduced by Analog Devices. It generates a current whose value in μA (microamperes) is equal to the temperature in degrees Kelvin (K).

Humidity. Any instrument capable of measuring the humidity or psychrometric state of air is a hygrometer, and the most common types used are psychrometers, dew-point hygrometers, mechanical hygrometers, electric impedance and capacitance hygrometers, electrolytic hygrometers, piezoelectric sorption, spectroscopic (radiation absorption) hygrometers, gravimetric hygrometers, and calibration. The sensor response is related to factors such as wet-bulb temperature, relative humidity, humidity (mixing) ratio, dew point, and frost point.

Pressure. Pressure so defined is sometimes called absolute pressure. The differential pressure is the difference between two absolute pressures. The most common types of pressure-measuring sensors are silicon pressure sensors, mechanical strain gauges, and electromechanical transducers.

Fluid Velocity. The flow of air is usually measured at or near atmospheric pressure. The typical instruments to measure fluid velocity are airborne tracer techniques, anemometers, pilot-static tubes, measuring flow in ducts.

Flow Rate. The values for volumetric or mass flow rate measurement are often determined by measuring pressure difference across an orifice, nozzle, or venturi tube. Other flow measurement techniques include positive displacement meters, turbine flowmeters, and airflow-measuring hoods.

Light. Light level, or illuminance, is usually measured with a photocell made from a semiconductor such as silicon. Such photocells produce an output current proportional to the illumination on the sensor incident luminous density.

Occupancy. In the scope of energy management in buildings, occupancy is an important parameter.

This type of sensor typically includes infrared (IR) and ultrasound sensors. IRs detect the heat released from humans, and ultrasound sensors detect the movement of the human occupant (i.e., the device compares the reflection in different instants; if they are different, something is moving in the sensor distance range).

EXAMPLES OF ENERGY AND LOAD CONTROL

Lighting Controls

Generally in all sectors, light energy consumption is very significant, and typically, control has been done manually or by electromechanical timers. The most sophisticated systems integrate sensors and programmed microcomputer controls. In less important places or with irregular human presence, it is common to use occupancy sensors that activate the lights. Photosensors packaged in various configurations allow the control ambient lighting levels using building automation strategies for energy conservation. Some examples include ceiling-mounted indoor light sensors that are used to continuously dim the available lights in order to produce the desired light level.

Motor Controls

Most induction ac motors are fixed-speed. However, a large number of motor applications would benefit if the motor speed could be adjusted to match process requirements. Motor speed controls are the devices which, when properly applied, can tap most of the potential energy savings in motor systems. Motor speed controls are particularly attractive in applications where there is variable fluid flow. In many centrifugal pump, fan, and compressor applications mechanical power grows roughly with the cube of the fluid flow. To move 80 percent of the nominal flow only half of the power is required. Centrifugal loads are therefore excellent candidates for motor speed control. Other loads that may benefit from the use of motor speed controls include conveyers, traction drives, winders, machine tools and robotics.

Conventional methods of flow control used inefficient throttling devices such as valves, dampers, and vanes. These devices, although they have a low initial cost, introduce unacceptable running costs due to their inefficiency. Several speed control technologies can be used to improve motor system operation. Electronic variable speed drives are the dominant motor speed control technology. Figure 4 shows the power consumed by a motor driving a fan, using different flow control methods. The main benefits of adjusting motor speed through the use of VSDs include better process operation, less wear in mechanical equipment, less noise, and significant energy savings (50% or more for some type of applications).

With the development of power electronics, the introduction of variable speed drives (VSDs) to control induction motor speed has become widespread. VSDs produce a variable frequency and voltage output that will regulate the motor speed and torque. In the case of closed loop control, the use of speed sensors (encoders) allows more precise control of the speed. The most common VSD type is the inverter-based VSD, in which the 3-phase supply is converted from ac to dc using a solid-state rectifier. Afterwards the inverter uses this dc supply to produce a 3-phase adjustable frequency, adjustable-voltage output which is applied to the stator windings of the motor. The speed of the motor will then change in proportion to the frequency of the power supply. Usually output voltage waveforms can be synthesised over the frequency range of 0–100 Hz.

HVAC Controls

Heating, ventilation, and air conditioning (HVAC) system controls are the link between varying energy demands on a building's primary and secondary systems and the approximately uniform demands for indoor environmental conditions. Without a properly functioning control system, the most expensive, most thoroughly designed HVAC system will be a failure. It simply will not control indoor conditions to provide comfort. The main controlled variable in buildings is zone temperature. The control of zone temperature involves many other types of control within the primary and secondary HVAC systems, including boiler and chiller control, pump and fan control, liquid and air flow control, humidity control, and auxiliary system control subsystems. There are two fundamentally different control approaches—pneumatic and electronic. Various kinds of sensors, controllers, and actuators are used for principal HVAC applications.

Pneumatic Control Systems. The first widely adopted automatic control systems used compressed air as the operating medium. A transition to electronic controls has been occurring since the 1970s. Pneumatic controls use compressed air for the operation of sensors and actuators. The most common pneumatic sensor is the temperature sensor. A typical method of sensing temperature and producing a control signal is a bellows (or diaphragm): if the temperature rises, it swells and moves the flapper to open or close the flapper/nozzle gap. This flapper could also be a bimetallic strip. Humidity sensors in pneumatic systems are made from materials that change size significantly (1–2%) with humidity (typically synthetic hygroscopic fibers). Because the dimensional change is relatively small, mechanical amplification of the displacement is used. An actuator converts pneumatic energy to motion—either linear or rotary. It creates a change in the controlled variable by operating control devices such as dampers or valves. The opening of a pneumatically operated control is controlled by the pressure in the diaphragm acting against the spring. Pneumatic controllers produce a branch line pressure that is appropriated to produce the needed control action for reaching the set point.

Electronic Control Systems. Electronic controls are being increasingly used in HVAC systems in commercial buildings. The main advantages are precise control, flexibility, compatibility with microcomputers, and reliability. With the continuous decrease in microprocessor cost and associated increase in capabilities, its cost penalty is virtually disappearing, particularly when calculated on a per function basis. Modern electronic systems for buildings are based in direct digital controllers (DDC) that enhance the previous analog-only electronic system with digital features. Modern DDC systems use analog sensors (converted to digital signals within a computer) along with digital computer programs to control HVAC systems. The output of this microprocessor-based system can be used to control electronic, electrical, or pneumatic actuators or a combination of these. DDC systems have the advantage of reliability and flexibility (it is easier to accurately set control constants in computer software than by making adjustments at a controller panel with a screwdriver). DDC systems also offer the option of operating energy management systems.

Anibal T. De Almeida
Hashem Akbari
Fernando J. T. E. Ferreira

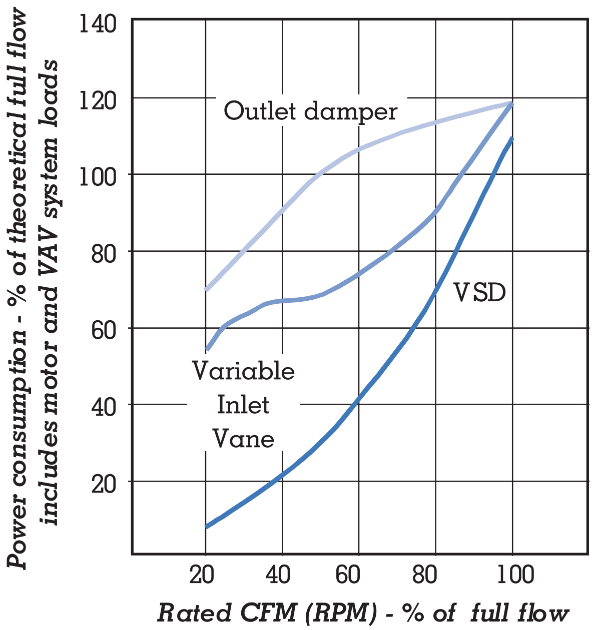


Figure 4.
Power consumption with different flow control methods in a fan system.

See also: Auditing of Energy Use; Behavior; Building Design, Commercial; Building Design, Energy Codes and; Building Design, Residential; Conservation of Energy; District Heating and Cooling; Electric Motor Systems; Heat Transfer.

BIBLIOGRAPHY

de Almeida, A; Bertoldi, P.; and Werner L. (1997). *Energy Efficiency Improvements in Electric Motors and Drives.* Heidelberg, Germany: Springer-Verlag.

Doebelin, E. O. (1990). *Measurement Systems-Application and Design.* New York: McGraw-Hill, Inc.

Eley Associates, Lighting. (1992). *Fundamentals Handbook.* Palo Alto, CA: Author.

Howell, R.; Saner, H.; and Coad, W. (1998). "Principles of Heating, Ventilating and Air Conditioning." Atlanta: ASHRAE, Inc.

Ogata, K. (1990). *Modern Control Engineering.* Englewood Cliffs, NJ: Prentice-Hall.

Weedy, B. M. (1998). *Electric Power Systems.* London: John Wiley & Sons.

CONVECTION

See: Heat Transfer

COOL COMMUNITIES

As modern urban areas have grown, there has been a corresponding growth in darker surfaces and a decline in vegetation, affecting urban climate, energy use, and habitability. Dark roofs heat up more and thus raise the summertime cooling demands of buildings, collectively with reduced vegetation, warming the air over urban areas and creating "heat islands." On a clear summer afternoon, the air temperature in a typical city can be as much as 2.5°C (4.5°F) higher than in surrounding rural areas. (The peak heat island effect occurs during cold winter evenings and is caused primarily by the rapid cooling of the rural areas—the thermal storage of pavements and dark-roofed buildings is much greater than greenery.) Peak urban electric demand rises by 2 to 4 percent for each 1°C (1.8°F) rise in daily maximum temperature above 15 to 20°C (59 to 68°F), so the additional air-conditioning use caused by higher urban air temperature is responsible for 5 to 10 percent of urban peak electric demand, costing U.S. ratepayers several billion dollars annually.

Temperatures in cities are increasing. Figure 1 depicts summertime monthly maximum and minimum temperatures between 1877 and 1997 in downtown Los Angeles, clearly indicating that these maximum temperatures are now about 2.5°C (4.5°F) higher than in 1920. Minimum temperatures are about 4°C (7.2°F) higher than in 1880. In California from 1965 to 1989, the average urban-rural temperature differences, measured at thirty-one pairs of urban and rural stations, have increased by about 1°C (1.8°F). In Washington, D.C., temperatures rose 2°C (3.6°F) between 1871 and 1987. This recent warming trend is typical of most U.S. metropolitan areas and exacerbates demand for energy. In Los Angeles, we estimate a heat-island induced increase in power consumption of 1 to 1.5 GW, costing ratepayers $100 million per year.

Besides increasing systemwide cooling loads, summer heat islands increase smog production. Smog production is a highly temperature-sensitive process. At maximum daily temperatures below 22°C (71.6°F), maximum ozone concentration in Los Angeles is below the California standard of 90 parts per billion. At 35°C (95°F), practically all days in Los Angeles are smoggy (see Figure 2).

HEAT ISLAND MITIGATION

When sunlight hits an opaque surface, some energy is reflected (this fraction is known at the albedo or reflectivity); the rest is absorbed. Use of high-albedo urban surfaces and planting urban trees are inexpensive measures that reduce summertime temperatures. The effects of planting trees and increasing albedo are both direct and indirect. Planting trees around a building or using reflective materials on roofs or walls has a direct effect: altering the energy balance/cooling requirements of that building. Planting trees and modifying albedo throughout the city has an indirect effect: citywide climate modification. By reducing air temperature, the air conditioning requirements of all buildings are reduced. Planting trees also sequesters atmospheric carbon through photosynthesis. Figure 3 depicts the overall process that impacts energy use and air quality within an urban area.

COOL ROOFS

When dark roofs are heated by the sun, they directly raise summertime building cooling demand. For highly absorptive (low-albedo) roofs, the surface/ambient air temperature difference may be 50°C (90°F), while for less absorptive (high-albedo) surfaces with similar insulative properties (e.g., white-coated roofs), the difference is only about 10°C (18°F), which means that "cool" surfaces can effectively reduce cooling-energy use.

For example, a high-albedo roof coating on a house in Sacramento, California, yielded seasonal savings of 2.2 kWh/day (80% of base-case use), and peak-demand reductions of 0.6 kW (about 25% of base-case demand). Field studies of nine homes in Florida before and after applying high-albedo coatings to their roofs yielded an air-conditioning energy-use reduction of 10 to 43 percent, saving on average 7.4 kWh/day (19%). The peak-demand reduction at 5:00 P.M. was 0.2 to 1.0 kW, an average of 0.4 kW (22%).

Researchers have simulated the impact of the urbanwide application of reflective roofs on cooling-energy use and smog in the Los Angeles Basin. They

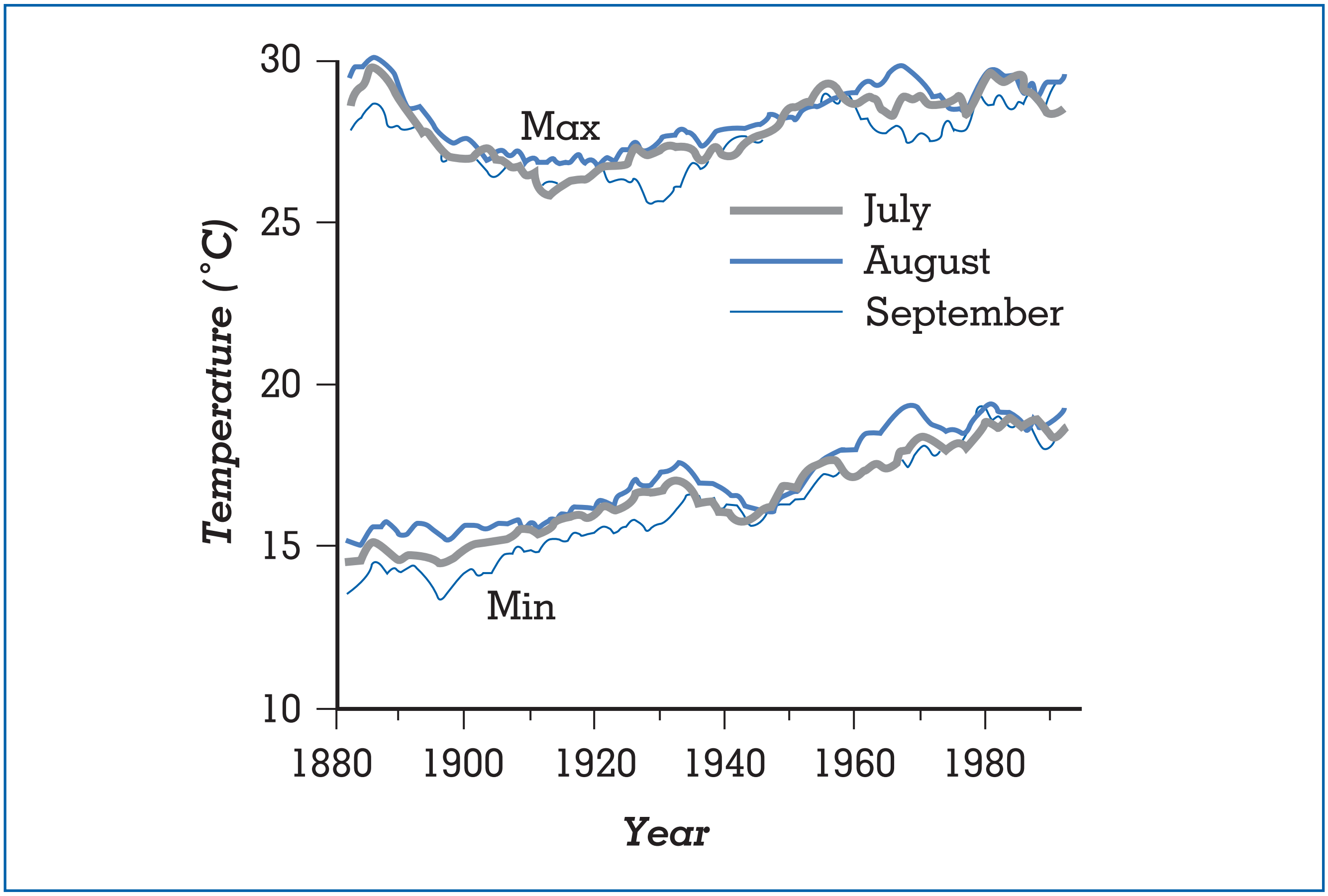


Figure 1.
Ten-year running average summertime monthly maximum and minimum temperatures in Los Angeles, California (1877–1997). The average is calculated as the average temperature of the previous four years, the current year, and the next five years. Note that the maximum temperatures have increased about 2.5°C (4.5°F) since 1920.

estimate that roof albedos can realistically be raised by 0.30 on average, resulting in a 2°C (3.6°F) cooling at 3:00 P.M. on a sunny August day. This temperature reduction significantly reduces building cooling-energy use further. The annual electricity savings in Los Angeles are worth an estimated $21 million. Cooling the air also results in a 10 to 20 percent reduction in population-weighted exposure to smog.

Other benefits of light-colored roofs include a potential increase in their useful life. The daily temperature fluctuation and concomitant expansion/contraction of a light-colored roof is less than that of a dark one. Also, materials degradation because of absorption of ultra-violet light is temperature-dependent. Thus, cooler roofs may last longer than hot roofs of the same material. Cool roofs incur no additional cost if color changes are incorporated into routine reroofing and resurfacing schedules.

URBAN TREES

The beneficial effects of trees are also both direct and indirect: shading of buildings and ambient cooling. Their shade intercepts sunlight before it warms buildings, and their evapotranspiration cools the air. In winter, trees shield buildings from cold winds. Urban shade trees offer significant benefits by reducing building air conditioning, lowering air temperature, and thus improving urban air quality (reducing smog). Over a tree's life, savings associated with these benefits vary by climate region and can be up to $200

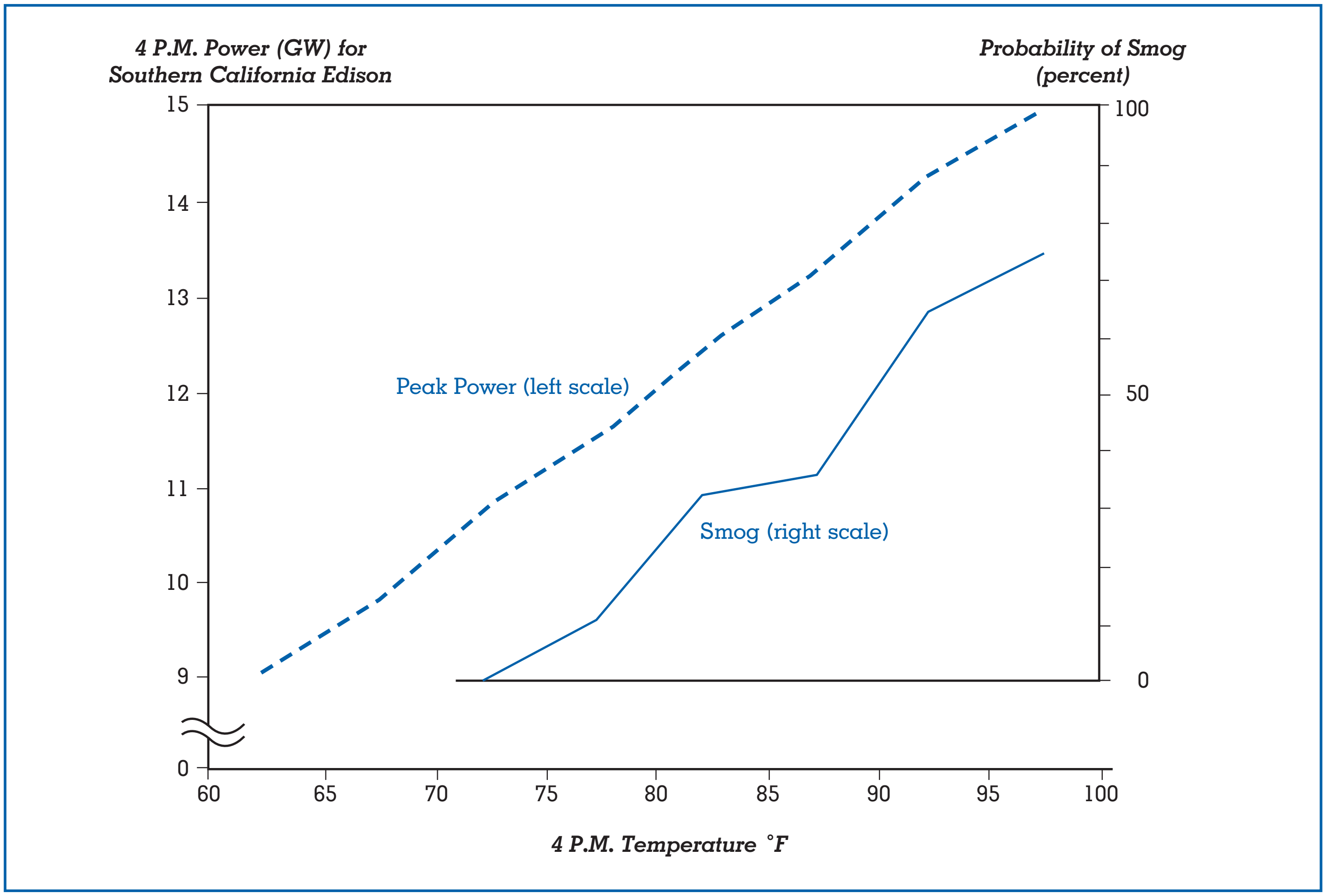


Figure 2.
Ozone levels and peak power for Southern California Edison versus 4 P.M. temperature in Los Angeles as a predictor of smog.

per tree. The cost of planting and maintaining trees can vary from $10 to $500 per tree. Tree planting programs can be low-cost, offering a good return on investments for communities.

Data on energy savings from urban trees are rare but impressive. In one experiment, the cooling-energy consumption of a temporary building in Florida was cut by up to 50 percent after adding trees and shrubs. Cooling-energy savings from shade trees in two houses in Sacramento were about 30 percent, corresponding to average savings of 3.6 to 4.8 kWh/day.

Simulations of the meteorological and energy impact of large-scale tree-planting programs in ten U.S. metropolitan areas show that on average trees can cool cities by about 0.3°C to 1.0°C (0.5°F to 1.8°F) at 2:00 P.M. The corresponding annual air-conditioning savings from ambient cooling by trees in hot climates range from $5 to $10 per 100m^2 of roof area of residential and commercial buildings. Indirect effects are smaller than direct shading, and, moreover, require that the entire city be planted.

There are other benefits associated with urban trees. These include improvement in life quality; increased property values; and decreased rain run-off and hence flood protection. Trees also directly sequester atmospheric CO_2, but the direct sequestration of CO_2 is less than one-fourth of the emission reduction from savings in cooling-energy use.

COOL PAVEMENTS

Urban pavements are made predominantly of asphalt concrete. The advantages of this smooth and all-

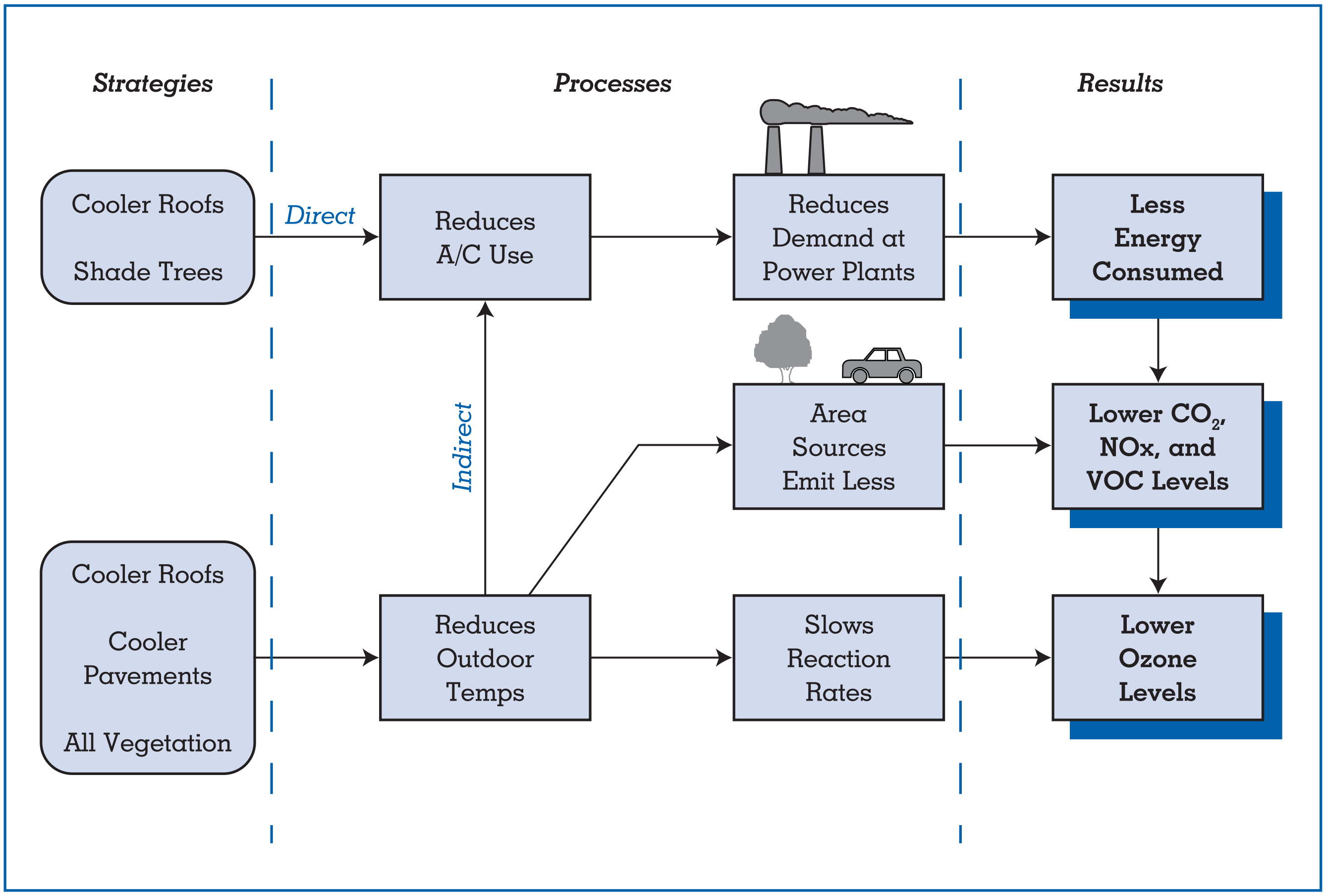


Figure 3.
Methodology to analyze the impact of shade trees, cool roofs, and cool pavements on energy use and air quality (smog).

weather surface for vehicles are obvious, but some associated problems are perhaps not so well appreciated. Dark asphalt surfaces produce increased heating by sunlight. Experimentally, the albedo of a fresh asphalt concrete pavement is only about 0.05. The relatively small amount of black asphalt coats the lighter-colored aggregate. As an asphalt concrete pavement is worn down and the aggregate is revealed, albedo increases to about 0.15 for ordinary aggregate. If a reflective aggregate is used, the longterm albedo can approach that of the aggregate.

The benefits of cool pavements can be estimated by first finding the temperature decrease resulting from resurfacing a city with more reflective paving materials. Cool pavements provide only indirect effects through lowered ambient temperatures. Lower temperature has two effects: (1) reduced demand for electricity for air conditioning and (2) decreased smog production.

Furthermore, the temperature of a pavement affects its performance: cooler pavements last longer. Reflectivity of pavements is also a safety factor in visibility at night and in wet weather, reducing electric street-lighting demand. Street lighting is more effective if pavements are more reflective, increasing safety. In reply to concerns that in time dirt will darken light-colored pavements, experience with cement concrete roads suggests that the light color of the pavement persists after long usage.

SUMMARY

Cool surfaces (cool roofs and cool pavements) and urban trees can reduce urban air temperature and hence can reduce cooling-energy use and smog. A

	Measures			
Benefits	***Cooler roofs***	***Trees***	***Cooler pvmnts***	***Totals***
1. Direct				
a A/C energy savings (M$/yr)	**46**	**58**	**0**	**104**
b Δ *Peak power (GW)*	*0.4*	*0.6*	*0*	*1.0*
c Present value ($)	153	64	0	
2. Indirect				
a A/C energy savings of 3°C cooler air (M$/yr)	**21**	**35**	**15**	**71**
b Δ *Peak power (GW)*	*0.2*	*0.3*	*0.1*	*0.6*
c Present value ($)	25	24	18	
3. Smog				
a 12% ozone reduction (M$/yr)	**104**	**180**	**76**	**360**
b Present value ($)	125	123	91	
4. Total				
a All above benefits (M$/yr)	**171**	**273**	**91**	**535**
b *Total* Δ *peak power (GW)*	*0.6*	*0.9*	*0.1*	*1.6*
c Total present value ($)	303	211	109	

Table 1.
Energy, Ozone Benefits, and Avoided Peak Power of Cooler Roofs, Pavements, and Trees in Los Angeles Basin
NOTE: The present value and surcost data for surfaces are calculated for 100 m^2 of roof or pavement area, and for one tree.
SOURCE: Rosenfeld et al., 1998.

thorough analysis for Los Angeles (see Table 1) showed that a full implementation of heat island mitigation measures can achieve savings of more than $500 million per year. Extrapolating those results, we estimate that national cooling demand can be decreased by 20 percent. This equals 40 TWh/year savings, worth over $4 billion a year by 2015 in cooling-electricity savings alone. If smog reduction benefits are included, savings could total to over $10 billion a year.

Hashem Akbari
Arthur H. Rosenfeld

See also: Air Conditioning; Air Pollution; Atmosphere; Climatic Effects; District Heating and Cooling; Domestic Energy Use; Efficiency of Energy Use; Energy Economics; Environmental Economics; Environmental Problems and Energy Use; Geography and Energy Use.

BIBLIOGRAPHY

Akbari, H., ed. (1998). "Energy Savings of Reflective Roofs." *ASHRAE Technical Data Bulletin* 14(2).

Akbari, H.; Bretz, S.; Taha, H.; Kurn, D.; and Hanford, J. (1997). "Peak Power and Cooling Energy Savings of High-Albedo Roofs." *Energy and Buildings* 25(2):117–126.

Akbari, H.; Davis, S.; Dorsano, S.; Huang, J.; and Winnett, S., eds. (1992). *Cooling Our Communities: A Guidebook on Tree Planting and Light-Colored Surfacing.* Washington, DC: U.S. Environmental Protection Agency.

Akbari, H.; Kurn, D.; Taha, H.; Bretz, S.; and Hanford, J. (1997). "Peak Power and Cooling Energy Savings of Shade Trees." *Energy and Buildings* 25(2):139–148.

Hall, J. V.; Winer, A. M.; Kleinman, M. T.; Lurmann, F. M.; Brajer, V.; and Colome, S. D. (1992). "Valuing the Health Benefits of Clean Air." *Science* 255:812–817.

Heat Island Group. (1999). <http://eetd.lbl.gov/HeatIsland>. Berkeley, California: Lawrence Berkeley National Laboratory.

Pomerantz, M.; Akbari, H.; Chen, A.; Taha, H.; and Rosenfeld, A. H. (1997). "Paving Materials for Heat Island Mitigation." Report LBNL-38074. Berkeley, CA: Lawrence Berkeley National Laboratory.

Rosenfeld, A.; Romm, J.; Akbari, H.; and Pomerantz, M. (1998). "Cool Communities: Strategies for Heat Islands Mitigation and Smog Reduction." *Energy and Buildings* 28(1):51–62.

Rosenfeld, A.; Romm, J.; Lloyd, A.; and Akbari, H. (1997). "White and Green." *MIT's Technology Review* February/March:54-59.

Taha, H. (1996). "Modeling the Impacts of Increased Urban Vegetation on the Ozone Air Quality in the South Coast Air Basin." *Atmospheric Environment* 30(20):3423–3430.

Taha, H. (1997). "Modeling the Impacts of Large-Scale Albedo Changes on Ozone Air Quality in the South Coast Air Basin." *Atmospheric Environment* 31(11):1667–1676.

Taha, H.; Konopacki, S.; and Gabersek, S. (1999). "Impacts of Large-Scale Surface Modifications on Meteorological Conditions and Energy Use: A 10-Region Modeling Study." *Theoretical and Applied Climatology* 62(3–4):175–185.

COOLING TOWERS

See: Heat and Heating

CRACKING

See: Refineries

CRUDE OIL

See: Petroleum

CULTURE AND ENERGY USAGE

INTRODUCTION

The earliest energy theorists were largely physical scientists, some of whom held that the growth and increasing complexity of society were largely synonymous with "progress," construed as a movement toward the higher, the better, and the more desirable. Nobel laureate chemist Wilhelm Ostwald (1907) was of this camp, although other eminent scholars such as the Nobel laureate physicist Fredrich Soddy (1912) and Alfred Lotka (1925), a founder of mathematical biology, also ventured ideas on the relation of energy and evolution yet made no explicit connection.

Energy theorists of cultural evolution are concerned with the whole sweep of cultural evolution, from prehistoric hunters and gatherers to modern industrial societies. This global, secular perspective is useful in assessing the relevance of ideas advanced to account for short periods of time in the history of particular societies. Those who propose an energy theory of cultural evolution emphasize the problem of causality-whether or not the amount of energy a society uses can be manipulated, and if so, to what extent, by what means, and to what effect (Nader and Beckerman, 1978).

In the social sciences, steps toward an energy theory of cultural evolution were made by British archeologists V. Gordon Childe (1936) and Graham Clarke (1946). However, the first figure in the social sciences who fully developed an energy theory of cultural evolution is anthropologist Leslie White. White (1959) held that culture advances as a consequence of the ability to harness more energy, although we are not to conclude that people, either individually or collectively, can choose to vary their energy harnessing technology and thus vary the rest of their culture. Causality in White's view, runs from materialistic forces like environmental change, population pressure, culture contact, and the like, to "superorganic" technological systems, and thence to superorganic social and ideological systems. Technological systems may determine the rest of the culture, but specific technology in turn can come about and continue in use through forces completely outside the conscious command of the participants in culture.

Sociologist Fred Cottrell's thesis (1955, p. 2) was that "the energy available to man limits what he *can* do and influences what he *will* do." He later added that both material phenomena and choice are involved in any human situation. However, human choice for Cottrell is not directed. To varying degrees choices can be predicted, given information on individual values, the costs to the individuals of making various choices, and the power of the individuals in question to achieve their choices. One assumes that some element of chance is involved, but in a given situation a particular choice may be predicted with a high level of confidence. Although Cottrell is far from White in his rejection of radical determinism, he is equally far from many political philosophers prominent in the history of the West, who assume that society simply represents the ongoing result of innumerable unconstrained individual choices.

Howard T. Odum (1971), an ecologist, is the most diligent in his attempts to reduce all-or nearly all-cultural phenomena to the currency of energy. On first examination his approach to causality is strikingly reminiscent of White. On closer examination, it seems that Odum holds a "possibilistic" position on causality, similar to that of Cottrell. In a section on theories of history, he comments apropos of the rise and fall of civilizations (1971, p. 229) and refers to the expansion of fossil fuel energy use as the "basic cause" of the population explosion. Energy use may cause other social phenomena, as asserted by White, but energy use itself is not seen as the *primum mobile* of all human affairs, else there would be no point in suggesting that humans alter their energy use. Thus, the gross amount of energy harnessed by a society is only a starting point. What matters is the presence and sensitivity of feedback loops by which energy is channeled into "useful work," and not discharged in destructive "short circuits."

As an example, Odum (1971, p. 291) mentions the relations in some states among hunters, the Fish and Game Department, and the game animal populations. Hunters pay a significant amount of money for their hunting licenses. This money is spent by the Fish and Game Department on preserving and augmenting animal populations that are hunted. When the animal populations grow too large to be supported in their natural habitat, more hunting is permitted; when the animal populations fall, hunting is curtailed. He specifically (1971, p. 300) calls for this sort of loop in regard to energy indicating the right to inject fuel into the overheated world economy must be regulated.

Lastly, Richard N. Adams (1975) claims intellectual descent from White. He also ranks the cultural evolution of societies by the amount of energy harnessed, and sees the drive toward the harnessing of increasing amounts of energy by the whole fabric of human cultures as inevitable, as long as energy is available to be harnessed. However, this statement does not amount to a prediction of the course of any particular society. Adams has modified White's determinism so that only global processes are held to be deterministic while local events may manifest a high degree of indeterminism (Lovins, 1976). Just how much room this reexpression allows for the manipulation of a particular society's energetic parameters by national policy decisions is an open question.

All four theorists agree that the amount of energy available constrains possibilities for social change and social action. They also agree on a relationship between energy use and the increase of what is socially desirable. White decouples increasing amounts of harnessed energy from ideas of what is more desirable. Cottrell is also concerned with cultural evolution on a macroscopic scale, focusing on the contrast and transition between "low energy" (unindustrialized, or "third world") societies and "high-energy" (industrialized) societies. Cottrell also recognizes that there is no necessary coupling, especially in the short term, between cultural evolution (defined again as the harnessing of increasing amounts of energy) and the increase of what is socially desirable. Odum's emphasis on feedback loops is that intricate feedback linkages and a high degree of role specialization are necessary for the realization of individual worth. He places a positive value on systemic stability. This stability is achieved through the use of more energy than that employed by tribal peoples, whom Odum considers to exist at the whims of a fickle natural environment, but less energy than currently employed by industrial countries, which Odum considers to be running, as cancer does, out of control. For him, the position seems to be that quality of life relates not to gross magnitudes of energy but to the complexity and stability of the system of energy production, distribution, and use. There is explicit coupling of growth of energy use without feedback controls with a deterioration of something like the quality of life. On the other hand, Adams argues that a deterioration in the quality of life for some members of a society is an inevitable correlate of increased energy flow.

Energy policy debates having to do with the immediate future in the United States take a more restricted range of time and space than do the global schemes of energy theorists. Further, some of the primitive assumptions of these policy arguments run directly counter to assumptions of global theorists. Policy statements almost always assume, for instance, that energy use is the dependent variable which can be manipulated at will by political decisions (Nader and Milleron, 1979).

Despite general acceptance of some measure of harnessed energy as an index of cultural evolution, it has been more than fifty years since anyone seriously argued that cultural evolutionary "advance" was in itself a movement toward the higher, the better, and the more desirable. Societies may indeed grow larger, more centralized, more internally differentiated, and more powerful as they consume more energy, but as time unrolls these changes have resulted in both improvement and degeneration in the lives of members of those societies, some members experiencing both (Daly, 1974; Duncan, 1975).

The conclusions of the cultural theorists are pertinent because of the clear dangers of putting too much stock in local and short term experience, since technology rarely exists free of people. For example, it is popularly held that energy use per capita versus quality of life in the United States in the period between 1955 and 1975 shows a negative correlation: energy use generally increased while quality of life generally declined. Conversely, the period 1935 to 1955 shows a positive correlation: both energy use and quality of life increased. Neither correlation shows the true relationship, because per capita energy use and quality of life are related to each other in complex ways. While this situation can hardly be illuminated by a simplistic, unidirectional attribution of causality, it does make the point that a mere increase in energy use cannot force a rise in the quality of life (Nader and Beckerman, 1978).

CONTEMPORAY PERSPECTIVE

Many issues in energy policy have been and will continue to be argued on the basis of presumed effects of energy use and energy policy on the quality of life. Some international comparisons of energy use and well-being are useful in their focus on efficiency of energy use (Goen and White, 1975). Others, such as the comparison between the United States and Sweden (Schipper and Lichtenberg, 1976), conclude that far from suggesting an inevitable coupling between level of economic activity and energy use, comparisons actually suggest ways in which more well-being can be wrought from every BTU of fuel and kilowatt-hour of electricity consumed in a given place. Furthermore, the data are persuasive that there is no direct relationship between per capita energy use and standard of living as measured by Gross National Product (GNP) (Weizdäcker, Lovins, and Lovins, 1997).

If the gross per capita energy use in a society cannot be held to correlate with the quality of life therein, what energy measures can? It seems likely that rapid change in the amount of energy used, either up or down (unless driven by zeal), will have a deleterious effect on the quality of life. If we examine American history over the last century or so, and if we compare our experiences with those of other nation-states, it appears that evidence for sharp increases or decreases in energy production and consumption correlates with social downturn. Our own experience of the great depression, as well as the fears that industrial nations share of an oil boycott, lead to the conclusion that a rapid increase in energy use will bring unemployment, shortages, and declining economic indicators of all kinds. On the other hand, experience from the last part of the nineteenth and first part of the twentieth centuries, as well as comparable periods of energy growth from the history of England, Germany, and Japan suggest that rapid increases in energy use are accompanied by fluctuations in the roles of social institutions, unmanageable inequalities in the distribution of power, class, regional, and ethnic conflicts of a serious nature, and a pervasive sense of anomie.

Many non-Western cultures, including the Chinese, tend to use more flexible forms of transportation than automobiles. (Corbis-Bettmann)

In the social science literature life style has had an accepted meaning-value preference as expressed in consuming behavior. Sociologists have developed methods whereby projections may be made of personal consumption patterns such as a shift in the proportion of consumer dollars spent on services as compared to dollars spent on durable goods. Such research attends to such questions as, “Can consumption patterns change so that less energy is used without altering social preferences?” Other researchers might examine the official statistics on personal consumption expenditures over an extended period of time in order to gain a historical understanding of the nature and scope of changing life styles, up or down. One can observe that Ireland and New Zealand have a very similar consumption of fossil fuel per capita (Cook, 1976). The Republic of Ireland is frequently used as an example of social and economic stagnation and even misery, while New Zealand is typically seen as a society in which everything runs so smoothly and progressively and equitably that its only fault is dullness. Per capita energy use and technologies are parallel. Fads and fashions play a role, as does enthusiasm, zeal, loyalty to tradition, style setting, and being the first on the block.

In Denmark, an intensely ecologically–minded country, moving to wind power made sense because they are located in one of the windiest comers of the world. They moved to wind power development after choosing not to have nuclear power. In 1999, wind energy covers 7 percent of Danish electricity with the goal being 50 percent over the next three decades. Eighty percent of the country’s windmills are owned by individuals and cooperatives. Denmark’s project is part of a larger voluntary venture for sustainability. Self-sufficiency and the environment are values strongly supportive of conservation and individual initiatives in Denmark, whereas in Japan and France the use of nuclear power has increased from 1987–1997. While long-term results of such choices cannot be easily forecast, energy catastrophies in one country could easily change the picture.

The ethnographic literature is replete with examples that illustrate that within a wide range, the total amount of energy that a society uses per capita is less important than the specific uses made of such energy. For example, the use made of energy and the perception of energy surplus among three precolonial African agricultural groups varied. All three groups increased the energy at their disposal by obtaining slaves, but they used these for different purposes: the Bemba principally as a trade item, the Tonga as additional labor about the homestead, and the Chokwa as both porters and trade items. All three groups had an energy surplus over and above what they expended. To have enough energy to cope with occasions that require a high expenditure of energy, any society has to live with most of its members operating below full work capacity most of the time. People mobilize their energy resources for short spurts; there are additional food and other resources known to them that could be harvested if needed but which usually are ignored (Cline-Cole et al., 1990). Without long-term storage, the inefficiencies are a necessary condition for survival over time, unlike the assumption that any energy available ought to be used because it is available and not because there is some human purpose to be obtained as a consequence of its deployment.

ENERGY AND SOCIAL ORGANIZATION

There is reason to believe that social organization, the framework within which we operate on a daily basis, may be strongly influenced by energy capability to produce, consume, store, or distribute. In one well-documented example (White, 1962), the pattern of energy control led to drastic change in social organization. The significant technological innovation responsible for the rise of feudalism in Europe was the introduction of the stirrup. The effect of the stirrup was to permit the concentration of greater force on the tip of a spear or the edge of a sword than could be achieved by a rider without stirrups, and to permit the rider to withstand greater force without being unhorsed. The horse had been in common use in Europe for centuries. What changed was not the gross amount of energy available, but the proportion of it that could be concentrated, and the speed and precision with which that concentrated energy could be released. The single consequence of this innovation was that, as arms and armor evolved to the full potential of a cavalry, and the support and forage of horses became requisite for military success, warfare became too expensive for serfs, who could afford neither the equipment nor the land necessary for mounted war. The result was the “flowering” of knighthood, the code of chivalry, the tying of ownership of land to the vassalage of its occupants, the tying of ownership of wealth to public responsibility, and the rest of the distinctive characteristics of feudalism. The control of a particularly important means of locally distributing energy led to a vast change in society itself.

It is anthropological commonplace that those who control scarce but necessary resources control, in large measure, the society that depends on those resources. Around the turn of the nineteenth century and somewhat later, railroads dominated large scale transportation in this country. The owners of railroad companies also dominated political life to an extraordinary degree (Boyer and Morais, 1955). At the grassroots level there was distrust and outright hatred of the railroad companies and their owners which brought us perhaps closer than we have ever been to class warfare in the United States. Nevertheless, the building of railroads led inexorably to capital intensive, highly centralized control.

The power of the railroads was diminished, not only by federal regulation, as history books sometimes argue, but also by the rise of motor vehicles and public roads as viable alternatives. With private trucks and cars the monopoly of a small segment of the population on the scarce "resource" of long range transportation was broken. There is some parallel with nuclear reactor technology since nuclear power is centralized, heavily regulated, subsidized by government, and crucially dependent upon selling prowess quoting scientific and engineering expertise.

Discussions of technologies have not always related to organization and values. Certainly since 1945 the bulk of the discussion has concentrated upon technical issues, such as the adequacy of the emergency core cooling system in light water reactors or the question of nuclear waste. However, the breakthroughs on safety and vulnerability resulted from studies of the culture of nuclear power. Charles Perrow (1999) argues that accidents are "normal" because they are built into the system. Perrow concludes that some complex systems can never be made accident-free because of "interactive complexity" (technological components are too varied for human operators to predict), and "tight-coupling" (small errors escalate too quickly for operators to figure out what is happening). For him, failure is built into a hard-wired system that does not allow for resilient possibilities of recovery.

In addition, discussions have focused on the effect of technology upon individual civil rights and civil liberties, terrorism, financial issues, and decaying technocarcasses. Such discussions have not utilized in-depth comparisons of strategies as with solar, nuclear, coal, and conservation; rather, each strategy has been considered separately, often independently of end users and consequences of failures (Kuletz, 1998), as shut downs, health and environmental catastrophies, or wars.

Americans tend to commute singly in automobiles, which can lead to traffic jams, such as this one near Longfellow Bridge in Boston, Massachusetts. (Corbis-Bettmann)

American discussions of nuclear energy first explored civil rights and civil liberties. Dangers of theft or sabotage of a plutonium facility require drastic incursions on individual freedom (Ayers, 1975).

At issue are risks of increased civilian surveillance, the extension of the military clearance system to include civilian workers, and increasing steps towards infringement on privacy such as covert airport services and security of government officials. In sum, safety considerations, so vital in dealing with vulnerable technologies, result in restrictive laws affecting all aspects of behavior, such that there are large increases in the numbers of police (Zonabend, 1993). Safety considerations then justify extraordinary investigation, arrest, and regulatory measures.

The sociopolitical consequences of increased commitment to nuclear technologies which represent only 5 percent of world energy, raises questions of democratic decision-making to safeguard the environment and health and safety of the general public (Holdren, 1976). Some ask if it is worth the price. Research on the social and political implications identifies the crucial contrast between vulnerable and nonvulnerable technologies, and between technological waste and social waste.

Areas of consensus and dissent appear, suggesting that the way energy is used and the purpose to which it is put are important to acceptance if not satisfaction. For example, in Europe, North America, and elsewhere a consensus is forming against wasteful engineering design. Few people would express themselves against improved miles per gallon or improved efficiency of refrigerators. There is more likely to be dissent on social waste; people would be more likely to object to carpooling or trading autos for mass transit. With regard to solar strategies there might be consensus on the democratizing effect of direct solar technology-after all, the sun falls on the rich and the poor, the weak and the powerful, the famous and the anonymous. Particularly the issue of decentralized solar power is symbolic of a greater issue: the preservation of liberty and equity through maintaining some independence from the "big system" (Stanford Research Institute, 1976). Centralized solar energy systems would have few of the dangers associated with highly vulnerable supply technologies, but there is expressed dissent at least among experts. Whatever the disagreements, it is clear that at issue is the value placed on freedom.

A SHIFT IN VALUES

Movement to redirect technological progress has brought about an "efficiency revolution" and the notion of a new industrial revolution incorporating, for example, the production of hypercars, compact fluorescent lamps, water drip systems, desk top computers that give more for less. Such movement is fueled by the realization that if Northern lifestyles spread globally it would take several globes to accommodate such life styles. It is well documented that the world's well–to–do minority uses the most energy, produces the greatest amount of pollution, and contributes greatly to the greenhouse effect. The efficiency revolution is in direct contradiction to supply side wizardry or high tech fantasies of fast breeder reactors, mega-fertilizer factories, gigantic water projects, preferring instead direct solar, hydro power, wind power, biomass, fuel-cell cars and the like (Nader, 1995). In the United States and Europe a coalition of business executives, consumers, environmentalists, labor leaders and legislators are the new energy entrepreneurs in a world where oil and coal fuels are increasingly being viewed as sunset industries (Hawken, Lovins, and Lovins, 1999).

Energy is becoming a multi–disciplinary concern. Aspects which were of interest to physicists, chemists and engineers are now a fixed growing concern for a wide variety of people. All energy research is inextricably interwoven with values, such as those relating to scale, complexity, organization, scientific challenge, and cost. The bulk of energy research that deals with scientific and technological questions is often embedded in deeply held beliefs about the human condition and direction. Technical specialists operating within the limits of their competence produce a clouding of the basic human factors that apply to broad understanding of the human dimensions of energy issues.

Energy can be used to foster different values in society: economic values based on efficiency, political values of democratic decision making, aesthetic values of architectural and environmental beauty, or their opposites. The question arises as to whose values will predominate in relation to the amount of energy produced, the purposes for which energy is used, and the forms and consequences of energy production. The perspective from which these relationships are addressed is important (Nader, 1981). The mode of energy use in industrialized society has been largely determined by producers rather than users. Automobile companies produce cars to sell cars, and Los Angeles's city rail systems were eliminated with this in mind (Mokhiber, 1988, pp. 221–228). Oil companies prefer inefficient cars because they use more gasoline (Sampson, 1975). Central power systems allow utilities to retain control over all productive facilities. Illuminating engineers may be apt to illuminate in a way that adds to profit and convenience of the illuminating engineer.

When quality of life becomes central, the dialogue on energy changes. In such a dialogue the individual user must have equal time. Certain choices might never be made if public welfare was the yardstick. Production and per capita expenditures of energy may then be secondary to the purposes for which energy is used and to the form of energy production. While certain broad correlation of energy use with other social parameters have been examined by social scientists for over a generation, findings are often ignored or misinterpreted. For example, technological progress and increased energy is said to have eliminated the drudgery of women's work while there is ample evidence to the contrary (Bendocci, 1993). Yet energy technologies are still being sold as panaceas for a woman's work life.

Our understanding of the relationship between human energy usage and culture has evolved from the simple paradigm wherein increased gross energy expenditure equaled cultural advancement to complex, non-lineal theories which account for interactions between myriad technological and social forces. The diversity of practical technologies world-wide suggests an enormous variety of solutions on hand and in the making for meeting human energy needs in both developed and developing countries.

Laura Nader

See also: Energy Economics; Ethical and Moral Aspects of Energy Use; Geography and Energy Use.

BIBLIOGRAPHY

Adams, R. N. (1975). *Energy and Structure*. Austin: University of Texas Press.

Ayres, R. W. (1975). "Policing plutonium: the civil liberties fallout." *Civil Liberties Law Review* 10(2):369–443.

Bendocci, C. G. (1993). *Women and Technology: An Annotated Bibliography*. New York: Gareaud Publishing.

Boyer, R. O., and Morais, H. M. (1955). *Labor's Untold Story*. New York: Cameron Assoc.

Childe, V. G. (1936). *Man Makes Himself.* London: Watts.

Clark, G. (1946). *From Savagery to Civilization*. London: Cobbett.

Cline-Cole, R. A., et al. (1990). *Wood Fuel in Kono*. Tokyo: United Nations University Press.

Cook, E. (1976). *Man, Energy, Society*. San Francisco: Freeman.

Cottrell, F. (1955). *Energy and Society*. New York: McGraw-Hill.

Daly, H. E. (1974). "Steady-State Economics Versus Growthmania: A Critique of the Orthodox Conceptions of Growth, Wants, Scarcity, and Efficiency." *Policy Science* 5:149–67.

Duncan, O. D. (1975). "Does money buy satisfaction?" *Social Indicators Research* 2:267–74.

Goen, R., and White, R. (1975). *Comparison of Energy Consumption Between West Germany and the United States*. Washington DC: Government Printing Office.

Hawken, P.; Lovins, A.; and Lovins, L. H. (1999). *Natural Capitalism: Creating the Next Industrial Revolution*. Boston: Little, Brown and Company.

Holdren, J. P. (1976). "Technology, Environment, and Well-Being: Some Critical Choices." In *Growth in America*, ed. C. L. Cooper. Westport, CT: Greenwood.

Kuletz, V. L. (1998). *The Tainted Desert: Environmental Ruin in the American West*. New York: Routledge.

Lotka, A. (1925). *Elements of Physical Biology*. Baltimore: Williams & Wilkins.

Lovins, A. B. (1976). "Energy strategy: the road not taken?" *Foreign Affairs* 54:65–96.

Mokhiber, R. (1988). *Corporate Crime and Violence: Big Business Power and the Abuse of Abuse of the Public Trust*. San Francisco: Sierra Club Books.

Nader, L. (1981). "Barriers to Thinking New About Energy." *Physics Today* 34(3):9, 99–102,

Nader, L. (1995). "Energy Needs for Sustainable Human Development-Anthropological Aspects." In *Energy as an Instrument for Social Change*, eds. Jose Goldenberg and T. B. Johansson. New York: United Nations Development Programme Publications.

Nader, L., and Beckerman, S. (1978). "Energy as it Relates to the Quality and Style of Life." *Annual Review of Energy* 3:1–28.

Nader, L., and Milleron, N. (1979). "Dimensions of the 'People Problem' in Energy Research and the Factual Basis of Dispersed Energy Futures." *Energy* 4(5):953–967.

Odum, H. T. (1971). *Environment, Power and Society*. New York: Wiley (Interscience).

Ostwald, W. (1907). "The Modem Theory of Energetics." *The Monist* 17(l): 511.

Perrow, C. (1999). *Normal Accidents: Living with High-Risk Technologies*. Princeton: Princeton University Press.

Sampson, A. (1975). *The Seven Sisters: The Great Oil Companies and the World They Made*. New York: Viking.

Schipper, L., and Lichtenberg, A. J. (1976). "Efficient Energy Use and Well-Being: The Swedish Example." *Science* 194:1001–13.

Soddy, F. (1912). *Matter and Energy*. London: Oxford University Press.

Stanford Research Institute. (1976). "A Preliminary Social and Environmental Assessment of the ERDA Solar Energy Program 1975-2000." Vol. 1, Main Report. Menlo Park, CA: Stanford Research Institute.

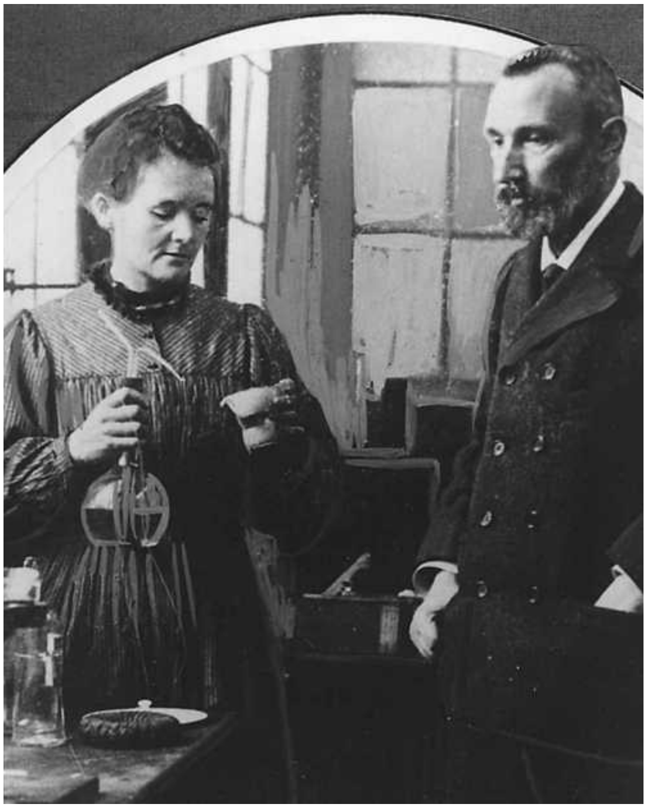
Weizdäker, E. von; Lovins, A. B.; and Lovins, L. H. (1997). *Factor Four: Doubling Wealth—Halving Resource Use*. London Earthscan Publications, LTD.

White, L. T. (1959). *The Evolution of Culture*. New York: McGraw-Hill.

White, L., Jr. (1962). *Medieval Technology and Social Change*. Oxford: Clarendon.

Zonabend, F. (1993). *The Nuclear Peninsula*. Translated from the French by J. A Underwood. New York: Cambridge University Press.

CURIE, MARIE SKLODOWSKA (1867–1934)



Marie Curie, working in the laboratory with her husband Pierre. (AP/Wide World Photos)

Maria Salomee Sklodowska was born to Wladislaw and Bronislawa Sklodowski on November 7, 1867, in Warsaw, Poland. This was a difficult time in Poland's history, since Poland was under Russian control. Nevertheless, Curie excelled in school, and she finished her secondary school education in 1883 with a gold medal. Although her parents were academicians and emphasized the importance of both education and religion, Curie maintained an agnostic view. Her rejection of religion is generally attributed to the loss of her mother and sister while she was still a young child and also may be related to her attraction to science.

Because women were not allowed to pursue higher education in Poland, Curie and her elder sister Bronia decided to earn their advanced degrees in France. While Bronia studied medicine at the Sorbonne, her sister worked as a governess to support her and then joined her in Paris in 1891, enroll in the Faculty of Sciences of the Sorbonne. During this period in France, Marie adopted the French variation of her name. By 1893 Curie had finished first in her class with a degree in physics despite beginning the program with poor French-language skills. In the following year she completed a mathematics degree and met her future husband Pierre a physicist at the School of Industrial Physics and Chemistry. A respected researcher in the fields of magnetism and piezoelectricity, Pierre convinced Marie that they shared an attitude of single-minded devotion to science and they were married in 1895.

The Curies continued to work together in Pierre's lab after their marriage and in 1897 Marie published her first paper, on magnetism. In September of that same year, her daughter Irène was born. For Marie to continue her career, the Curies hired a house servant, and Pierre's father, Eugène, moved in after the death of his wife. This allowed Marie to continue her search for her doctoral thesis topic. At that time the completion of a doctoral thesis by a woman anywhere in Europe was unprecedented.

Curie chose for her dissertation research the new topic of uranium rays, a phenomenon that had only recently been observed by Henri Becquerel. The mystery was the source of the energy that allowed uranium salts to expose even covered photographic plates. Curie's first efforts in the field were systematic examinations of numerous salts to determine which salts might emit rays similar to those of Becquerel's uranium. After discovering that both thorium and uranium were sources of this radiation, Curie proposed the term "radioactive" to replace "uranium rays." She also discovered that the intensity of the emissions depended not on the chemical identity of the salt but on the amount of uranium in the compound. This key observation eventually led Curie to propose that radioactivity was an atomic property rather than a chemical property. Among the samples that Marie studied were two uranium ores that proved to be more radioactive than pure uranium. This observation led her to propose that the ores contained a new element more radioactive than uranium. Gabriel Lippmann presented the paper describing Curie's work to the Académie des Sciences on April 12, 1898.

Although the Curies noted that one equivalent gram of radium released one hundred calories of heat per hour, they were uninterested in the practical implications of this, as they were both devoted to pure scientific discovery. During their work with pitchblende in 1898, the Curies discovered two new radioactive elements, which they named polonium (in honor of Marie's homeland) and radium. By 1902 they had isolated a pure radium salt and made the first atomic weight determination.

On June 25, 1903, Curie defended her doctoral thesis at the Sorbonne, delivering a review of the research in the area of radioactivity. In December of the same year, the Curies were named joint recipients of the Nobel Prize for Physics along with Becquerel. This award was made for the recognition of the phenomenon they called radioactivity. Due to their poor health, the Curies were not able to travel to Sweden to accept the prize until June 1905. Although they were both clearly suffering from radiation sickness, neither of the Curies realized this at the time. In the meantime, Marie gave birth in 1904 to another daughter, Eve.

In 1906 Pierre was awarded a full professorship and position as chair of physics at the Sorbonne and Marie was promised a position as director of the laboratory which the university planned to create for Pierre. However, in April 1906 Pierre was killed when he stepped into the path of a horse-drawn cart. While this event personally devastated Marie, it was a pivotal point in her professional career. She was offered Pierre's chaired position at the level of assistant professor, making her the first woman in France to obtain a professorship and allowing her to both continue her research and financially support her family.

After Pierre's death, Marie was faced with having to present her work without the support and social skills of her husband. Furthermore, she spent numerous years defending her work from William Thomson (Lord Kelvin) who did not believe that radioactivity could be an atomic property. Nevertheless, in 1911 Curie became the first person to receive a second Nobel Prize when she was awarded the Nobel Prize in Chemistry for her discovery of polonium and radium and the isolation of radium.

Curie spent much of the remainder of her life raising money for research. Always politically active, she also worked with her daughter Irène (who later earned her own Nobel Prize in Chemistry) during World War I establishing mobile x-ray services and training workers to perform x-rays on the battlefield. It was not until the 1920s that the issue of the health hazards of radium emerged. Despite her own health problems, Curie was reluctant to accept that radiation could be linked to the illnesses and deaths of so many of her colleagues in the field. Eventually Curie was diagnosed with a severe form of pernicious anemia (caused by years of exposure to radiation) and she died on July 4, 1934.

Margaret H. Venable
T. Leon Venable

See also: Nuclear Energy; Thomson, William.

BIBLIOGRAPHY

Curie, E. (1938). *Madame Curie; A Biography,* tr. V. Sheean. Garden City, NY: Doubleday, Doran.

Giroud, F. (1986). *Marie Curie: A Life,* tr. L. Davis. New York: Holmes & Meier.

Ogilvie, M. B. (1988). *Women in Science*. Cambridge, MA: MIT Press.

Rayner-Canham, M., and Rayner-Canham, G. (1997). *A Devotion to Their Science: Pioneer Women of Radioactivity*. Philadelphia: Chemical Heritage Foundation.

Rayner-Canham, M., and Rayner-Canham, G. (1998). *Women in Chemistry: Their Changing Roles from Alchemical Times to the Mid-Twentieth Century*. Philadelphia: Chemical Heritage Foundation.

Reid, R. (1974). *Marie Curie*. New York: Saturday Review Press/E. P. Dutton.

CURRENT, ALTERNATING AND DIRECT

See: Electricity

D

DAMS

See: Hydroelectric Energy

DEMAND-SIDE MANAGEMENT

WHAT IS DSM?

Demand-side management (DSM) refers to active efforts by electric and gas utilities to assist customers in modifying their use of energy. DSM encompasses a variety of activities designed to change the level or timing of customers' energy consumption. Most discussion of DSM focus on programs that help customers save energy by encouraging them to adopt energy-efficient measures or practices. The most well-known are rebate programs that lower the cost of energy-efficient appliances. Other DSM programs promote changes in the "shape" of a utility's load by shifting demand away from high daytime electricity rates when it is most expensive for utilities to generate electricity.

U.S. utility DSM programs can be divided into seven categories: general information, targeted technical information, financial assistance, direct installation of energy-efficient technologies, performance contracting, load control/load shifting, and innovative tariffs. General information programs aim to increase customers' awareness of their energy use patterns and opportunities to use energy more efficiently. Almost all utilities provide general information, ranging from educational brochures about turning off gas-furnace pilot lights during warm months to bill inserts describing energy-efficient products and services. General information is also distributed through advertisements and by utility representatives.

Targeted technical information programs include audits of customers' current energy use patterns, accompanied by recommendations for ways to use energy more efficiently. Audits are typically offered free of charge although some utilities are experimenting with charging fees.

Financial assistance programs—loans or direct payments—lower the cost of purchasing energy-efficient technologies. Cash payments or rebates have been the most popular type of DSM program. Rebates reduce some or all of the cost of purchasing and installing an energy-efficient technology. Rebates can be fixed payments per unit (e.g., a $100 coupon toward the purchase of an energy-efficient refrigerator) or payments that lower the initial cost of a technology to some predetermined level (e.g., to ensure that the energy saved within three years of purchase will pay for the extra cost of buying an energy-efficient refrigerator). Some utilities offer low-interest loans in place of or in conjunction with rebates. When given a choice, customers generally prefer rebates to loans, so utilities interested in achieving large impacts over a short period of time have devoted a much larger share of DSM budgets to rebate programs than to loans.

Direct installation of energy-efficient technologies programs send utility staff or utility-hired contractors to a customer's premises free of charge to provide energy audits and install pre-selected energy-saving technologies. Because utilities pay the full installation cost, these DSM programs are frequently the most expensive to operate, as measured by the cost of energy saved. Utilities have typically offered installation programs either as a last resort—for example, when there is an imminent threat of supply shortfall—or to serve particular market segments (e.g., low-income residential customers that have proven difficult to reach with other DSM programs.)

Performance contracting are programs in which a third party, often an energy service company (ESCO), contracts with both a utility and a customer to provide a guaranteed level of energy savings. Performance contracting programs involve either competitive bidding, in which ESCOs and customers make proposals to the utility, or "standard offers," in which the utility agrees to pay for energy-saving projects at a fixed price per unit of energy saved. Payment is contingent on verification that the customer actually saves the amount of energy guaranteed in the contract. When an ESCO enters into a performance contract with a utility, the ESCO must recruit utility customers and form a separate contractual relationship with them so that the ESCO can finance and install energy-saving technologies and verify their performance. Performance contracting has mainly involved the public sector (i.e., schools and government buildings), business (i.e., commercial) and industrial, rather than residential, customers.

In load-control/load-shifting programs, utility offers payments or bill reductions in return for the ability to directly control a customer's use of certain energy-consuming devices or to assist the customer in installing a device, which alters the timing of demands on the electric system. In load-control programs, utilities directly control some customer appliances during periods when demand for power is high, such as extremely warm days when increased cooling energy use causes heavy power system loads. Load-control programs rotate groups of appliances (typically water heaters or central air conditioners) on and off for short periods of time, reducing net loads on the power generation system. These programs have usually involved residential customers. Customers can also control loads by adopting load-shifting technologies that allow the timing of the customer's load. Thermal storage, for example, allows a customer to use power when rates are low, such as at night, to generate and store heating or cooling to be used at other times of day when rates are high. "Valley-filling" describes programs that shift (or increase) customer loads to times of day when utility system loads and production costs are low (e.g., during the night).

Innovative tariff programs make it cost effective for customers to reduce or change the timing of energy use. These tariffs include interruptible rates, time-of-use rates, and real-time pricing. An interruptible rate is similar to a load-control program; a customer pays a lower rate in return for agreeing to curtail loads whenever requested by the utility. The customer, rather than the utility, determines which loads to reduce when the request is made. Time-of-use rates set different prices for energy used during different times of day, based on the utility's costs of generating power at those times. Real-time pricing is a sophisticated form of time-of-use rates, in which a utility typically gives customers a forecast of twenty-four hourly energy prices one day in advance. With both time-of-use rates and real-time pricing, customers respond by changing energy use to reduce their costs. Innovative pricing programs have targeted primarily industrial and large commercial customers.

HISTORY OF ELECTRIC UTILITY DSM PROGRAMS IN THE UNITED STATES

The history of DSM in the United States is dominated by the activities of electric utility companies. Historically, electricity service was considered a natural monopoly; it was thought that only one company in a geographic region could efficiently capture the economies of scale offered by electricity generation, transmission, and distribution technologies. In the United States, two primary institutions arose to secure the public benefits associated with increased electrification: publicly-owned municipal utilities were established in some large cities and governed by city councils; privately-owned utilities were also formed, governed by state regulatory authorities. More than 80 percent of the electricity produced and sold in the United States comes from privately-owned utility companies that generate, transport, and distribute power.

For much of their history, electric utilities in the United States promoted new uses of power in order to increase their sales and thus their profits. However, during the 1970s, the dramatic rise in world oil prices and growing concern about the environmental impacts of electricity generation (especially those associated with reliance on nuclear power) led to a new emphasis on conserving energy.

In 1978, the federal government passed the National Energy Conservation Policy Act (NECPA). Among other things, NECPA required utilities to offer onsite energy audits to residential customers. This law acknowleded that saving energy could be cheaper than producing it. We now recognize NECPA as the beginning of modern utility DSM programs. In fact, California and Wisconsin authorized utility DSM programs as early as 1975; these programs were the very first DSM programs, predating NECPA. NECPA encouraged utilities to create, staff, train, and maintain internal organizations devoted to helping customers reduce their electricity use. Prior to this time, utility staff efforts were focused on ensuring reliable service and promoting new uses for electricity.

After only modest increases through the mid-1980s, electric utility DSM programs began to increase dramatically during the late 1980s as a handful of states began to direct their regulated utilities to formally adopt least-cost or integrated resource planning principles. Regulatory efforts to encourage utilities to adopt DSM programs were bolstered by growing evidence of the low cost of technologies that could reduce electricity demand by increasing the efficiency with which energy was used. Energy-efficiency advocates conducted numerous analyses showing that substantial amounts of energy could be saved for much less than the cost of building new power plants, with no change in the level of amenity provided to the customer. Least-cost planning advocates argued that utilities should pursue these demand-side options whenever they were less expensive than the supply-side alternatives they could displace. A number of "market barriers" to consumers choosing cost-effective energy-efficient technologies were identified. DSM programs overcame these barriers by providing information and financial incentives to assist customers in selecting energy-efficient technologies that would lower their energy costs.

THE GROWTH IN UTILITY DSM PROGRAMS

Regulators responded to this evidence by encouraging utilities to increase the size and scope of their DSM programs whenever it could be shown that energy efficiency was a cheaper "resource" alternative than investments in new generating plants. While there had been little utility resistance to the original directives of NECPA in the late 1970s many utilities actively resisted suggestions by their regulators to pursue least-cost planning in general and DSM in particular in the late 1980s.

As political pressure to pursue least-cost planning strategies grew, regulators began to recognize that utility resistance could be traced to longstanding regulatory approaches that had been developed to encourage sales of electricity when the (then) increasing economies of scale meant that increased consumption led to lower costs for all. As a result, utilities had a powerful financial incentive to increase earnings by encouraging customers' consumption and constructing capital-intensive, new power plants to meet growing demand. These incentives meant that most activities to reduce sales were directly contrary to the business interests of the utility.

Two regulatory strategies were developed to overcome utilities' incentive to sell electricity and to insure that investment in programs to reduce customer demand would not be a financial burden on utilities. The first strategy compensates utilities for sales "lost" as a result of cost-effective DSM programs. The second "decouples" revenue from sales by establishing a revenue target that is independent of the utility's sales, and creating a balancing account to compensate for the difference between revenues actually collected and the revenue target. By making total revenues independent of actual sales volumes in the short-run, these approaches took away utilities' incentives to increase loads.

Some states created separate financial incentives for the delivery of superior DSM programs. Three types of incentives have been used. First, utilities earn a profit on money spent on DSM. Second, utilities earn a bonus paid in \$/kWh or \$/kW based on the energy or capacity saved by a DSM program. Third, utilities earn a percentage of the net resource value of a DSM program. Net resource value is measured as the difference between the electricity system production costs that the utility avoids because of the program(s) and the costs required to run the program(s). These new incentives were instrumental in stimulating growth of DSM energy-efficiency programs.

The effect of these efforts to realign the financial interests of utilities with the public interest was a dramatic increase in utility spending on DSM programs.

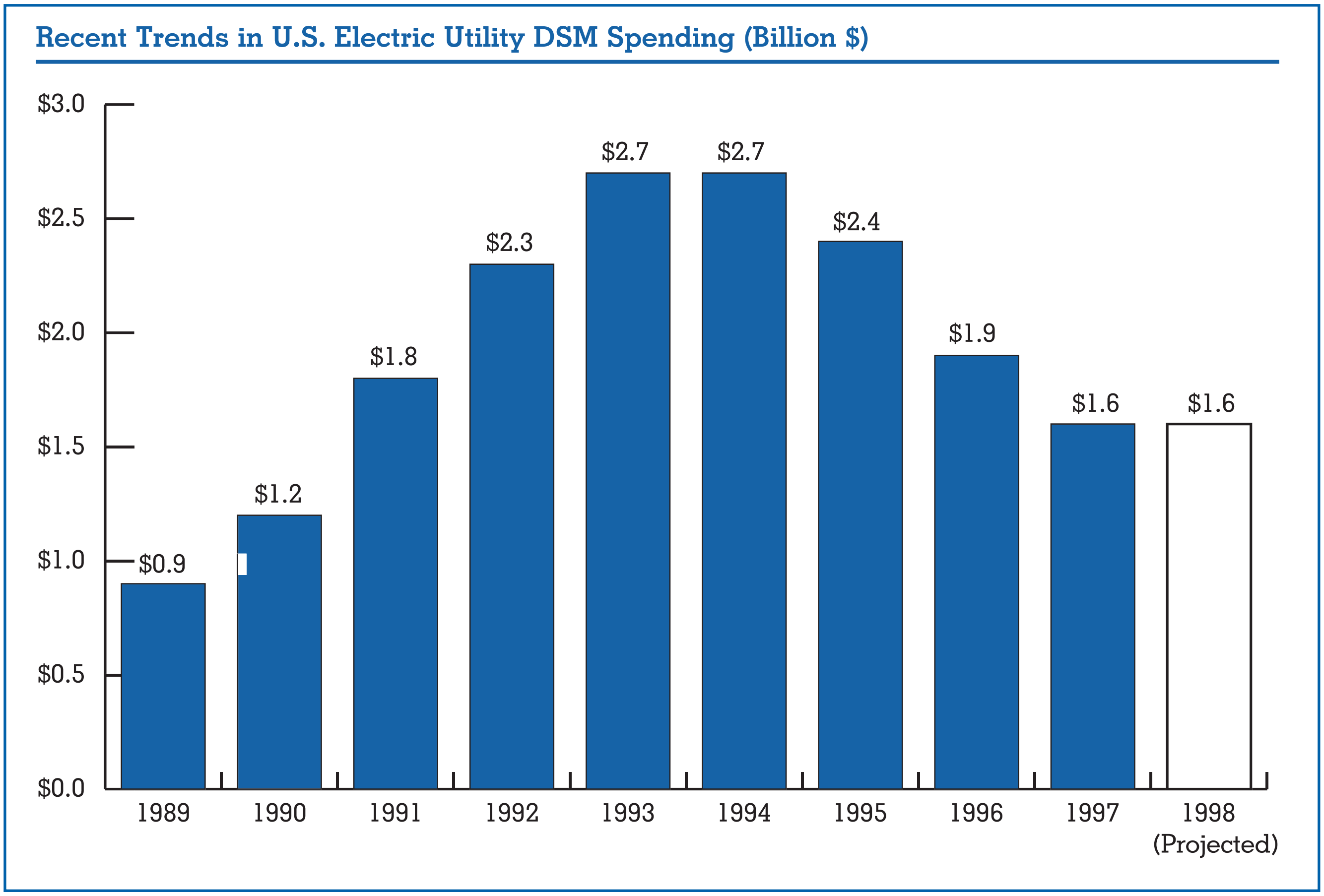


Figure 1.
Trends in U.S. Electric Utility DSM Spending.
SOURCE: U.S. Energy Information Administration

Spending increased so significantly that the U.S. Energy Information Administration began tracking DSM expenditures formally in their annual survey of utility operations starting in 1989. These surveys revealed that DSM spending by electric utilities in the United States had increased from $0.9 billion in 1989 to $2.7 billion in 1994 (see Figure 1). Electric utility DSM programs reached their largest numbers in 1993, accounting for more than $2.7 billion of utility spending or about one percent of U.S. utility revenues.

The record of utility achievements demonstrated that DSM energy efficiency programs were able to save electricity cost effectively. In the most comprehensive analysis to date, an examination of the forty largest energy-efficiency programs targeted to the commercial sector (i.e., office buildings, retail establishments, schools, etc.) showed that these programs had saved energy at an average cost of 3.2 ¢/kWh and that they were highly cost effective when compared to cost of electricity generation they allowed the sponsoring utilities to avoid. However, not all utilities were equally effective in running energy-efficiency DSM programs. The study also found that some utilities, notably those with large DSM programs, had saved energy at cost of less than 2¢/kWh, while others, mostly making a more modest investment, had saved energy at a cost in excess of 10¢/kWh. Electric utility DSM programs save more than 56 million kWh annually—30 percent more than the total growth of electric sales in the United States between 1996 and 1997. DSM programs have enabled utilities to avoid the construction of approximately forty average size coal-fired power plants.

THE FUTURE OF DSM

Although DSM programs have been successful, the U.S. electric utility industry is in the midst of changes that will continue to affect its future. Starting in 1994, states actively began discussing restructuring their electricity industry to allow customers to choose their supplier of electricity and to increase competition among electricity generators. Utilities responded predictably to this threat of competition. First, they have sought regulatory protection for assets whose cost was in excess of current market prices. They also aggressively began cutting costs in all areas, including DSM. According to the Energy Information Administration, total DSM spending in 1997 had declined to $1.6 billion (from a high of $2.7 billion in 1993). By 1998, several states had restructured their electricity industry and many other states are expected to follow this trend.

In a fully restructured industry, utilities become essentially regulated power distribution companies with only an obligation to connect all customers to the power grid but with no obligation to plan and acquire generation to serve all customers. When utilities are relieved of the obligation to serve certain customers, they are also relieved of the obligation to use least-cost planning principles to acquire resources (including demand-side energy efficiency improvements) on behalf of these customers.

Although electricity industry restructuring renders the traditional utility monopoly franchise obsolete, the public purposes that are served by utility DSM programs remain whenever market barriers prevent cost-effective energy-efficiency decisions from being made. Restructuring means that electricity prices set by the market will better reflect the true value of electricity than prices set by regulatory authorities. If successful, elimination of regulatory mispricing would address an important historic market barrier to energy-efficiency. However, it is unlikely that restructuring will address all the market barriers that prevent cost-effective energy-efficiency actions from being taken. For example, the environmental consequences of electricity generation in particular, which are not currently reflected in the market prices paid for electricity, remain a strong argument for continuing energy-efficiency programs.

In the past, utilities were in a unique position to promote public interest in energy efficiency through DSM programs. Utilities had (1) access to low-cost capital; (2) name recognition among customers and acknowledged technical expertise; (3) lack of direct financial interest in promoting particular energy-efficiency products or services; (4) access to detailed information on customer energy-use patterns; and (5) a system for billing customers for services. Whether utilities will retain these desirable features following restructuring is not known. It will depend on decisions regulators make on the organization and rules governing the firms operating in the market.

Many utilities plan to offer DSM programs as a feature to keep and attract customers. Nonutility energy service providers are concerned that ratepayer funding for DSM programs will be used unfairly to subsidize utility development of business opportunities that will be pursued as unregulated profit-making activities. They argue that utility managers already face conflicts of interest when delivering ratepayer-funded DSM programs that historically served broad public interests, while at the same time attempting to maximize shareholder returns by using these programs to keep utility customers from switching to other suppliers or to increase the utility's dominance in local energy-efficiency service markets.

In a restructured electricity industry, utilities are concerned that including DSM program costs in their regulated rates puts them at a competitive disadvantage to other power providers that are not also required to charge for DSM. A surcharge to recover DSM program costs levied on all electricity users, regardless of their suppliers, eliminates this concern. As of 1998, twelve states had adopted these surcharges to continue funding for DSM programs as well as, in some cases, funding for other "public purpose" activities, such as research and development or promotion of renewable energy.

If the utilities can mitigate conflicts of interest, some states are expected to rely on utilities with good past records to continue to administer DSM programs. If local utilities have had poor past performance with DSM or cannot mitigate conflicts of interest, states may consider (1) administration by an existing or newly-created government agency, and (2) administration by an independent, possibly non-profit entity. Both alternatives raise questions of governance and accountability for the administration of funds.

SUMMARY

Demand-side management programs have been a bold experiment in the active promotion of energy efficiency by electric and gas utilities in the United States. By and large, they have demonstrated that market barriers, which constrain consumers' abilities to lower their energy costs, can be successfully and cost-effectively addressed by well-designed and targeted programs. These programs have represented billions of dollars in utility investments that have allowed utilities to avoid even more costly and environmentally damaging power plants. As the utility industry is restructured, many are hopeful that regulators will continue to enact policies that successfully align utility interests with pursuit of cost-effective energy efficiency opportunities.

Joseph Eto

BIBLIOGRAPHY

Battelle-Columbus Division and Synergic Resources Corporation. 1984. *Demand-Side Management, Evaluation of Alternatives.* Palo Alto, CA: Electric Power Research Institute and Edison Electric Institute.

Baxter, L. 1995. *Assessment of Net Lost Revenue Adjustment Mechanisms for Utility DSM Programs.* ORNL/CON-408. Oak Ridge, TN: Oak Ridge National Laboratory.

Blumstein, C.; Krieg, B.; Schipper, L.; and York, C. (1980). "Overcoming Social and Institutional Barriers to Energy Efficiency." *Energy* 5(4):355–72.

Cavanagh, R. 1988. "Responsible Power Marketing in an Increasingly Competitive Era." *Yale Journal on Regulation* 1:331–366.

Energy Information Administration. 1998. *Electric Power Annual 1997, Volume II.* DOE/EIA-0348(97)/2. Washington, DC: Energy Information Administration. October.

Eto, J.; Kito, M.; Shown, L.; and Sonnenblick, R. 1995. *Where Did the Money Go? The Cost and Performance of the Largest Commercial Sector DSM Programs.* LBL-38021. Berkeley, CA: Lawrence Berkeley National Laboratory.

Eto, J.; Goldman, C.; and Nadel, S. 1998. *Ratepayer-Funded Energy Efficiency Programs in a Restructured Electricity Industry: Issues and Options for Regulators and Legislators.* Washington, DC: American Council for an Energy Efficient Economy.

Eto, J.; Stoft, S.; and Belden, T. 1994. *The Theory and Practice of Decoupling.* LBL-34555. Berkeley, CA: Lawrence Berkeley National Laboratory.

Krause, F., and Eto, J. 1988. *Least-Cost Utility Planning, A Handbook for Public Utility Commissioners, The Demand Side: Conceptual and Methodological Issues.* Washington, DC: National Association of Regulatory Utility Commissioners.

Lovins, A. 1976. "Energy Strategy: The Road Not Taken?" *Foreign Affairs* 55(1):65–96.

Moskovitz, D. 1989. *Profits and Progress Through Least-Cost Planning.* Washington, DC: National Association of Regulatory Utility Commissioners.

Nadel, S. 1992. "Utility Demand-Side Management Experience and Potential—A Critical Review." *Annual Review of Energy and the Environment* 17:507–35.

Solar Energy Research Institute (SERI). 1981. *A New Prosperity: Building a Sustainable Energy Future.* Andover, MA: Brick House Publishing

Stoft, S.; Eto, J.; and Kitom M. 1995. *DSM Shareholder Incentives: Current Designs and Economic Theory.* LBL-38059. Berkeley, CA: Lawrence Berkeley National Laboratory.

DEPARTMENT OF . . .

See: Government Agencies

DEREGULATION

See: Regulation and Rates for Electricity

DESULFURIZATION (OIL AND COAL)

See: Refineries

DIESEL, RUDOLPH (1858–1913)

Rudolph Christian Karl Diesel was a German thermal engineer and inventor of the high-efficiency internal-combustion engine that bears his name.

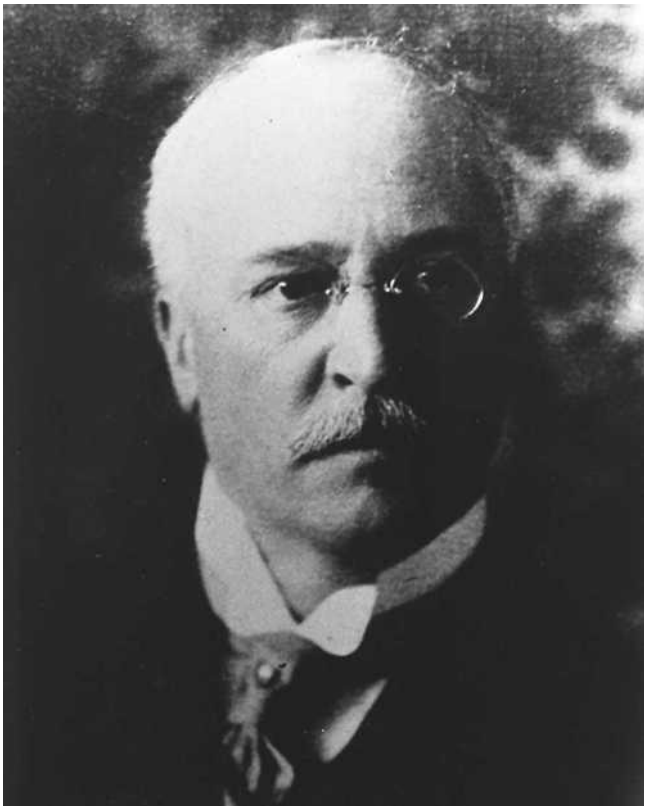
Much of Diesel's life and brilliant career was tragic—from his business failings, to his struggles to translate theory into practice, to his chronic physical and mental ailments. Yet his research and engine prototypes laid the foundation for one of the world's most efficient and widely used fossil fuel engine technologies.

Diesel was born to German Protestant parents. His father, Theodor, a bookbinder, emigrated from Germany to Paris in 1850, and five years later married Nuremberg-born Elise Strobel, a teacher of German and English. The couple had three children—Louise (b. 1856), Rudolph (b. 1858), and Emma (b. 1860)—and were strict disciplinarians. French was spoken in the Diesel home, and Rudolph's mother also taught him English. Rudolph had few friends, but took an interest in technology, reinforced by frequent visits to the Conservatoire des Arts et Métiers, a technical museum in Paris.

In 1870 the Diesels were forced to leave Paris by a government general expulsion order during the Franco-Prussian War, and in November Rudolph went to live with relatives in Augsburg, Germany. He enrolled in the royal commercial school and then studied at the city's industrial school. In a letter to his parents on his fourteenth birthday, Diesel declared his ambition to become an engineer. Perhaps inspired by a "fire piston" at the school (a device that caused a tinder to glow with compressed air), Diesel became intrigued with compression ignition. He was plagued with chronic health problems, especially headaches and insomnia, yet still excelled at his studies. His final exam grades in 1875 were the highest in the school's history.

Diesel's distinguished academic record drew the attention of Professor Karl Max von Bauernfeind of Munich's Techniche Hochschule, who offered Diesel a two-year scholarship. Ultimately Diesel spent four years in the school's mechanical-technical division, where he studied theoretical machine design under Carl von Linde (a leader in that field and in refrigeration science) Moritz Schroeter, and others. Graduating at the top of his class with a civil engineering degree in 1880, Diesel took a job at the Sultzer Brothers factory in Winterthur, Switzerland. Diesel was then hired by the Linde Refrigeration Company in Paris and soon went to work directly for Linde, traveling regularly to consult with clients.

By this time Diesel had become enamored with the "social question"—the social problems and class conflict fostered by industrialization. Perhaps in reaction to his father's earlier strong embrace of magnetic healing, Diesel rejected organized religion in favor of an increasingly popular rational humanism. One of Diesel's key motivations to invent a high-efficiency engine was to help relieve the burdens of the artisan class. Later (in 1903) he published a book called *Solidarismus: Natuürliche wirtschaftliche Erloösung des Menschen (Solidarism: The Natural Economic Salvation of Man)*, in which he called for, among other things, worker-run factories.



Rudolph Diesel. (Library of Congress)

In 1882 Diesel met Martha Flasche, a German. The couple married in November 1883 and for the remainder of the decade lived in Paris, where they had three children: Rudolph, Jr. (b. 1884), Hedy (b. 1885), and Eugen (b. 1889).

Diesel began work on an economical engine as early as 1880. The following year he took out his first patents—for machines to make clear ice. Throughout the remainder of the decade he worked on an ammonia vapor engine and (less rigorously) on a solar-powered engine. In 1889 he moved to Berlin to work as Linde's representative there. In

1890 or 1891 Diesel began to work out the theoretical basis for a constant temperature (isothermic) engine that would later evolve into the diesel engine. Diesel hoped to create a "universal" (flexible-fuel) engine that would operate on the cycle described by Nicolas Leonard Sadi Carnot and thus waste only about 20 to 30 percent of its energy through heat loss. In his model, the first downward stroke of the piston drew air into the cylinder. That air was then compressed to high pressure and temperature with the return (upward) stroke. With the second downward stroke, fuel was introduced at such a rate that the heat generated by its combustion would counterbalance the natural decline in temperature caused by the expansion of the cylinder space.

Diesel received a patent for his engine design in 1893, the same year he published a book on the subject, *Theorie und Kunstruktion eines rationellen Waärmemotors (Theory and Construction of a Rational Heat Engine)*, which was translated into English in 1894. He then persuaded Machinenfabrik Augsburg (Augsburg Engine Works), led by Heinrich Buz, to form a syndicate with Krupp in April 1893 to manufacture a 2-cylinder, 50-horsepower engine. But Diesel and his backers were unable to produce a smoothly running prototype until 1897. For that to happen, Diesel modified many of the fundamentals of his original theoretical design. The working engine operated at much higher pressures (18 to 33 atmospheres); ran on kerosene (instead of any liquefied or pulverized fuel) and at a new fuel-air mix; used compressed air rather than solid injection; and, most importantly, did not operate at constant pressure.

After announcing the success and imminent commercialization of his engine at the June 1897 meeting of the Society of German Engineers, Diesel began seeking licensees throughout the industrialized world. Three German companies bought patent rights, as did several non-German firms and individuals, including brewing magnate Adolphus Busch in the United States. The Augsburg company managed to produce a reliable 60-horsepower model by 1902. But Diesel's foreign licensees continued to struggle, despite drawings and engineering assistance from Augsburg. Busch's company produced a mere 260 engines between 1902 and 1912, when its license expired, the diesel venture having cost the family millions of dollars.

Meanwhile, Rudolph Diesel's growing fame was haunted by a series of personal and business setbacks. Working to the point of exhaustion, he required months of recuperation in a sanitarium from late 1898 to early 1899 and again in 1901–1902. Some of his critics pointedly challenged the originality of his work, claiming that Diesel's engines operated on principals articulated by others, not on the inventor's original concepts. A poor financial manager, Diesel nevertheless maintained a lavish villa in Munich called Jugendstil. In 1913, when he boarded the steamship *Dresden* at Antwerp, bound for England, Diesel faced financial ruin. Sometime during the evening of September 29–30, he disappeared from the ship's deck. His body was recovered ten days later, and all signs pointed to suicide.

Diesel engines—heavier and more expensive to build per horsepower than gasoline engines, but much more durable and cheaper to operate—made rapid inroads in shipping in the 1920s, heavy-duty trucking and construction equipment in the 1930s, railroads in the 1950s, and began to gain ground in passenger automobiles following the energy crises of the 1970s.

David B. Sicilia

See also: Automobile Performance; Carmot, Nicolas Leonard Sadi; Diesel Cycle Engines; Diesel Fuel; Engines; Gasoline Engines.

BIBLIOGRAPHY

Bryant, L. (1976). "The Development of the Diesel Engine." *Technology & Culture* 17:432–446.

Lytle, R. H. (1968). "Introduction of Diesel Power in the United States, 1897–1912." *Business History Review* 42:115–148.

Thomas, D. E., Jr. (1987). *Diesel: Technology and Society in Industrial Germany*. Tuscaloosa: University of Alabama Press.

DIESEL CYCLE ENGINES

The diesel engine is one of the most widely used global powerplants and can be found in almost every conceivable application. From small single-cylinder models to V20 designs, their horsepower can range from as low as 3.73 kW (5 hp) to as high as 46,625 kW

(62,500 hp). Some important applications of the diesel engine include

Light-duty
- cars
- pickup trucks
- riding lawnmowers

Heavy-duty
- heavy-duty trucks
- buses
- locomotives
- industrial power-generating plants
- oilfield exploration equipment
- road-building equipment (e.g., backhoes, excavators, crawler tractors, graders, and bottom dumps)
- agricultural, logging, and mining equipment

Marine
- pleasure craft
- sailboat auxiliary engines
- workboats (e.g., tugs)
- oceangoing merchant ships and passenger liners.

In addition, a wide variety of military equipment, including tanks, armored personnel carriers, HUMVEEs and ships, is powered by diesel engines. The governed speed of diesel engines can range from as low as 85 rpm in large-displacement, slow-speed models, to as high as 5,500–6,000 rpm in smaller automotive type models.

Although today's technologically advanced diesel engine is named after the German Rudolph Diesel, it is a direct result of developmental work that began in the late 1700s when the first internal-combustion engine was constructed. This basic concept was further developed in 1824 by a young French engineer named Sadi Carnot. Other individuals added to this knowledge: Lenoir in 1860 with the first commercial internal-combustion engine; Beau De Rochas in 1862; Otto in cooperation with Langen in 1867; Clerk in 1881; Ackroyd-Stuart in 1890, and finally Diesel in 1892. Since its first practical inception in 1895 by Rudolph Diesel, the diesel cycle engine has been a source of reliable, efficient, long-lasting power.

Both gasoline and diesel engines are available in either a two-stroke- or a four-stroke-cycle design. The fundamental difference between the Otto engine cycle (named after Nikolaus Otto, who developed it in 1876) and the diesel engine cycle involves the conditions of the combustion. In the Otto cycle, the almost instantaneous combustion occurs at a constant volume, before the piston can move much. The pressure goes up greatly during combustion. However, in the diesel cycle, combustion occurs under constant pressure, at least for a time, because the piston moves to increase the volume during the burn to hold the pressure constant.

FOUR-STROKE-CYCLE ENGINE

The four-stroke-cycle internal-combustion engine design is widely employed in both gasoline and diesel engines. The high-speed four-stroke-cycle diesel engine produces superior fuel economy, lower noise factors, and ease of meeting exhaust emissions regulations over its two-stroke-cycle counterpart. In the four-stroke-cycle diesel engine, the concept shown in Figure 1 is used. A total of 720 degrees of crankshaft rotation (two complete revolutions) are required to complete the four piston strokes of intake, compression, power, and exhaust. The actual duration in crankshaft degrees for each stroke is controlled by both the opening and closing of the intake and exhaust valves by the camshaft and will vary among makes and models of engines.

In Figure 1 the engine crankshaft is rotating in a clockwise direction when viewed from the front of the engine. During both the intake and the power strokes, the piston moves down the cylinder, while on both the compression and the exhaust strokes the piston moves up the cylinder. Basically during the intake and exhaust strokes the piston acts as a simple air pump by inducting air and expelling burned exhaust gases from the cylinder. On the compression stroke the upward-moving piston raises the air charge to a pressure typically between 30 and 55 bar (441 and 809 psi) based upon the piston compression ratio and whether the engine is naturally aspirated or employs an exhaust-gas-driven turbocharger to boost the air supply pressure. A net loss in energy (waste heat to the cooling, lubrication, and exhaust systems, and to friction and radiation) occurs during the intake, compression, and exhaust strokes, since only during the expansion or power stroke do we return energy (torque, which is a twisting and a turning force) to rotate the crankshaft.

The diesel engine operates with a much higher compression ratio (CR) than does a gasoline engine, and therefore is manufactured with structurally

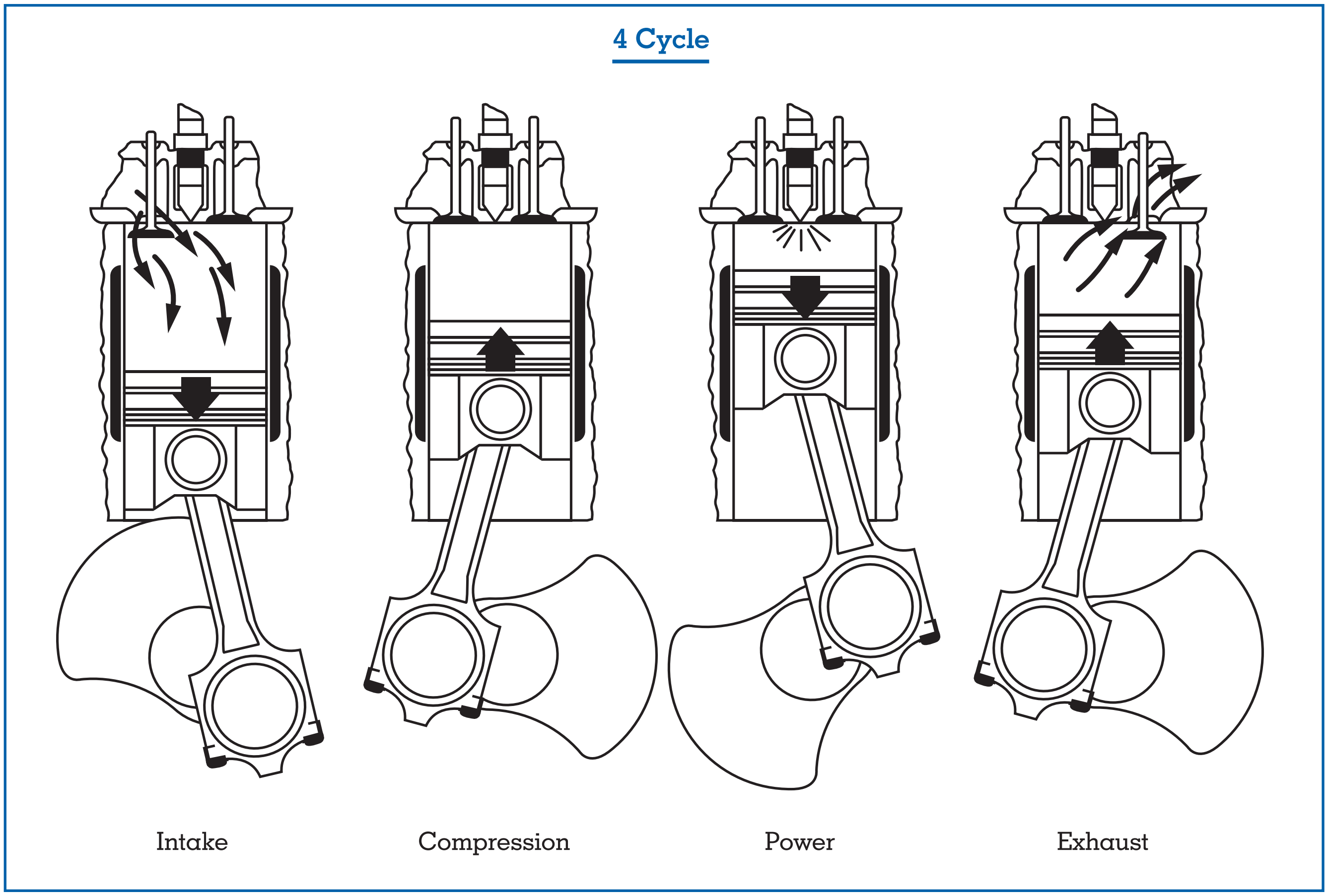


Figure 1.
Sequence of individual piston and valve events for a four-stroke diesel cycle engine.
SOURCE: Detroit Diesel Corporation

stronger components capable of handling this design feature. This higher CR results in a much higher cylinder pressure and temperature; therefore a greater expansion rate occurs when the piston is driven down the cylinder on its power stroke than occurs in a gasoline engine. CR is the difference between the volume of air remaining above the piston while at bottom dead center (BDC), versus that at top dead center (TDC). Typically a gasoline engine will operate with CRs between 9 and 10.5:1, while a DI (direct-injected) diesel CR usually varies between 15 and 17:1. IDI (indirect-injected) engine model CRs usually range between 18 and 23:1. These high CRs result in compressed air temperatures prior to the delivery of fuel from the fuel injector, typically in a range between 649°C to 927°C (1,200°F to 1,700°F). This hot air converts the injected fuel from an atomized liquid to a vapor to permit self-ignition (it establishes a flame front) without the need for a spark plug.

DIESEL CONSTANT PRESSURE CYCLE

Diesel's original concept was for his slow-speed engine to operate on a constant-pressure design throughout the power stroke, obtained by continually injecting both compressed air and fuel. To increase the efficiency of the diesel cycle, his first engines used no cooling system, with disastrous results. Later engines, with cooling systems, corrected this part of the problem but resulted in cylinder heat losses accompanied by frictional, radiated, and exhaust heat losses. In addition, although com-

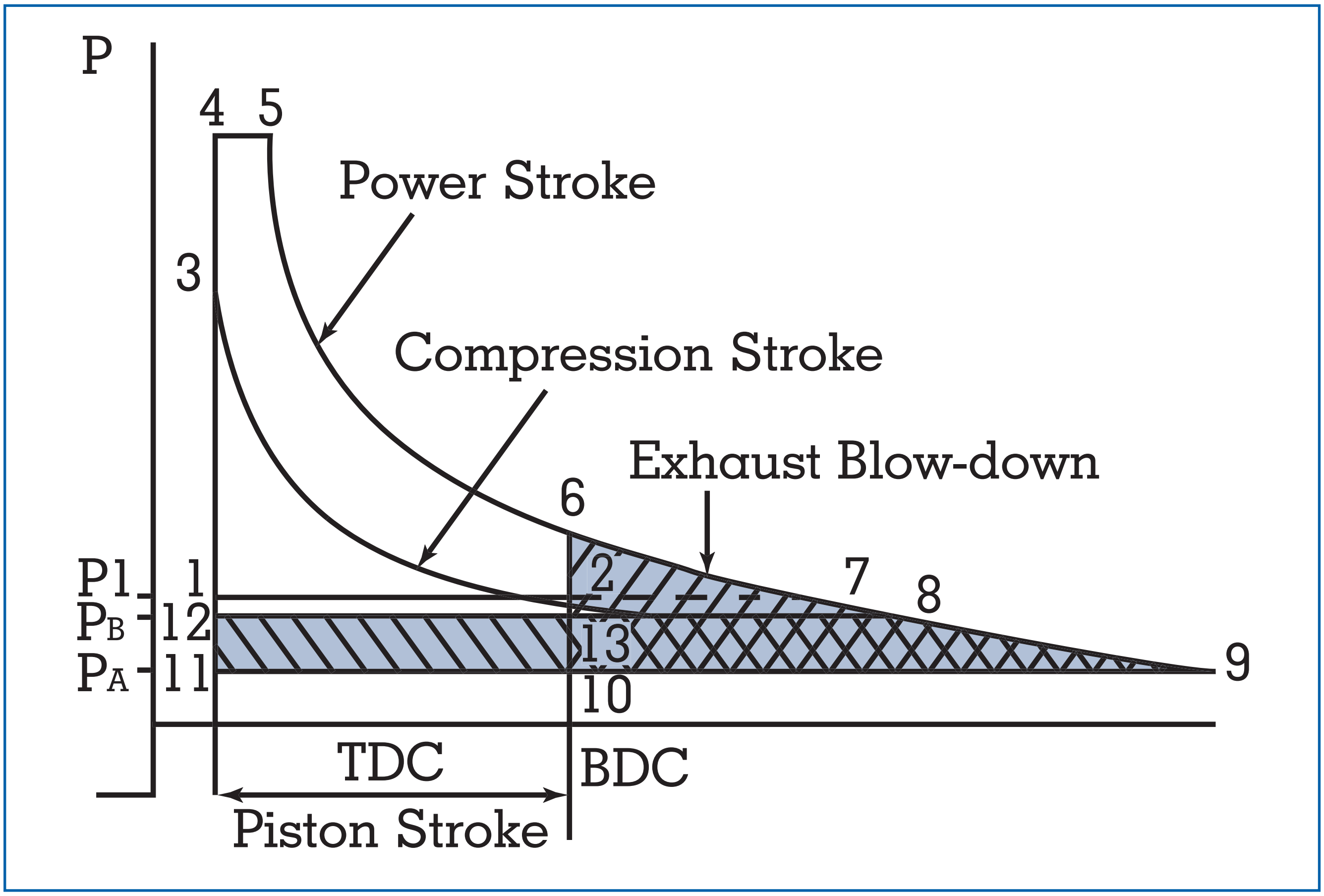


Figure 2.
Pressure-volume curve illustration for a turbocharged and direct-injected high-speed heavy-duty four-stroke diesel cycle engine.

pressed air and fuel were supplied to the cylinder throughout the power stroke, the increasing cylinder volume as the piston moved down on its power stroke was unable to maintain a high enough air temperature and pressure to sustain effective combustion, therefore the air/fuel ratio was not conducive to continued combustion. In today's electronically controlled high-speed diesel engines, fuel is injected for a number of degrees after top dead center (ATDC). This maintains cylinder pressure at a fairly constant level for a given time period, even though there is an increase in clearance volume above the descending piston. Engineers designing and testing engines like to compare the air standard cycles under actual engine performance with corresponding values for highly idealized cycles based on certain simplified assumptions.

PRESSURE-VOLUME CURVE

The energy used and returned to the engine crankshaft/flywheel is illustrated in Figure 2, showing a PV (pressure-volume) diagram for a turbocharged and direct-injected high-speed heavy-duty four-stroke-cycle diesel engine. This schematic simplifies the internal operation of a piston throughout its four strokes. In Figure 3 we show the actual combustion operating principle in graphic form, with the piston at 90 degrees before TDC and at 90 degrees ATDC.

In Figure 2 you can see that from position 1 to 2 the piston moves down the cylinder on the intake stroke as it is filled with turbocharger boost air higher than atmospheric pressure, as indicated in line Pl. Depending on the valve timing, actual inlet valve closure will control the degree of trapped cylinder air

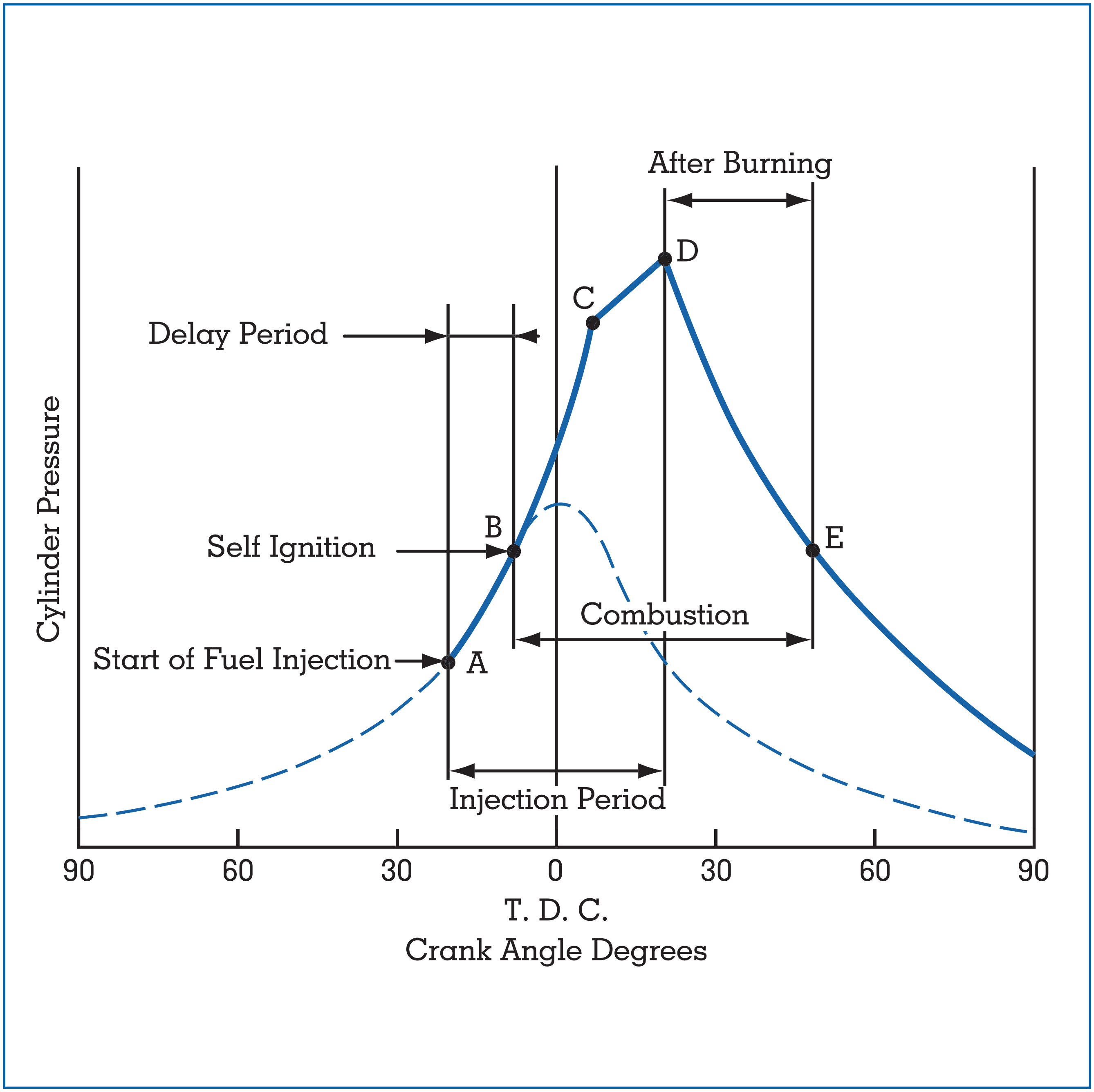


Figure 3.
Example of the sequence of events within the cylinder at 90 degrees BTDC (before top-dead center) and 90 degrees ATDC (after top-dead center).
SOURCE: Zexel USA, Technology and Information Division

pressure. In this example, compression starts at position 2, as the piston moves up the cylinder. Fuel is injected at a number of degrees BTDC (before top dead center), and there is both a pressure and a temperature rise as the fuel starts to burn from positions 3 to 4. As the piston moves away from TDC on its power stroke, positions 4 to 5, the continuous injection of fuel, provide a constant pressure for a short number of crankshaft degrees. From positions 5 to 6, the piston is driven downward by the pressure of the expanding gases. The point at which the exhaust valves open BBDC (before bottom dead center) depends on the make and model of the engine used. The work represented by area 6–7–2 is available to the hot end of the turbocharger (turbine wheel) from the hot, pressurized exhaust gases. Line PA indicates atmospheric pressure along points 9–10–11. The exhaust manifold pressure is shown as line PB and the exhaust gas blow-down energy is represented by points 6–9–10. The exhaust process from the engine cylinder is shown among points 6, 13 and 12, where 6 through 13 is the blow-down period when the exhaust valves open and the high-pressure gases expand into and through the exhaust manifold. From points 13 to 12 the piston moves from BDC to TDC, displacing most of the exhaust gas out of the cylinder. Therefore the potential work of the exhaust gases in this turbocharged engine is represented by the crosshatched areas identified as points 10–11–12–13. The maximum energy to drive the turbine of the turbocharger is that shown in the area identified as points 6–9–10 and 10–11–12 and 13. Ideally during points 6–9–10, if both the cylinder pressure and the turbocharger inlet pressure could both be maintained at equal levels before the piston moves upward from BDC on its exhaust stroke, a system close to ideal would be created when using a pulse turbocharger system.

The higher CR and the fact that diesel fuel contains a higher heat content per gallon or liter (approximately 11%) than does gasoline are two of the reasons why the diesel engine produces better fuel economy and higher torque at the crankshaft and flywheel. Further fuel economy improvements can be attributed to the fact that in the diesel engine, the throttle pedal is used to control the cylinder fueling rate directly. In a gasoline engine, however, manual operation of the throttle (gas pedal) directly controls the volume of air entering the engine intake manifold by restricting the size of the opening.

Diesel engines, however, operate on an unrestricted air flow at all speeds and loads to provide the cylinders with an excess air charge. This results in a very lean air/fuel ratio of approximately 90:1 to 100:1 or higher at an idle speed. At the engine's rated speed (full load maximum power output) the air/fuel ratio will drop to 20:1 to 25:1 but still provide an excess air factor here of 10 to 20 percent. This excess air supply lowers the average specific heat of the cylinder gases, which in turn increases the indicated work obtained from a given amount of fuel. Compare this to most gasoline electronically controlled engines, where the TBI (throttle body injection) or MPFI (multiport fuel injection) system is designed to operate at a stoichiometric or 14.7:1 air/fuel ratio.

TWO-STROKE-CYCLE DIESEL ENGINE

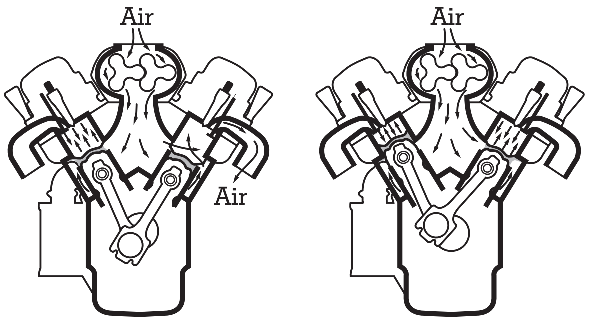
Figure 4 illustrates the basic operational concept for a two-stroke cycle vee-configuration diesel engine where a power stroke is created every 360 degrees of crankshaft rotation versus the 720 degrees needed in the four-stroke-cycle design. The two-stroke design eliminates the necessity for individual intake and exhaust strokes required in a four-stroke cycle engine; therefore it becomes necessary in the two-stroke-cycle engine to employ a gear-driven blower to supply a large supply of low-pressure fresh air for:

- combustion of the injected fuel;
- cooling (approximately 30% of the engine cooling is performedby air flow, while 70% is performed by coolant flow within the engine radiator or heat exchanger);
- scavenging of exhaust gases from the cylinders;
- positive crankcase ventilation.

In the two-stroke-cycle diesel engine the cylinder-head-located poppet valves are to permit scavenging of the exhaust gases. Fresh air for the four functions listed above is force-fed through a series of ports located midway around the cylinder liner. Each liner receives its air supply from an "air box" cast within the engine block. On high-speed diesel engines this air pressure varies between 27.6 kPa and 48.3 kPa (4 and 7 psi) higher than atmospheric throughout the engine speed range. When an exhaust-driven turbocharger is used in conjunction with the gear-driven blower, air-box boost pressures at engine full load typically range between 172 and 207 kPa (25 and 30 psi) above atmospheric.

With reference to Figure 4, each piston downstroke provides power, while each upstroke provides compression of the blower/turbo-supplied cylinder air. The actual number of degrees of both of these strokes will vary based on the specific engine make and model, and its year of manufacture needed to comply with U.S. Environmental Protection Agency (EPA) exhaust emission limits.

Note in Figure 4 that in a high-speed engine, the



Air entering through port to combustion chamber

Air being compressed with exhaust valves closed

Figure 4.
Two-stroke diesel cycle engine principle of operation.
SOURCE: Detroit Diesel Corporation

power stroke begins at TDC and ends at approximately 90 to 92 degrees ATDC (after top dead center), when the exhaust poppet valves start to open by camshaft action. This allows the pressurized exhaust gases to start flowing from the cylinder through the open exhaust valves. As the piston continues moving down the cylinder, the exhaust gas pressure decreases. At approximately 59 degrees BBDC (before bottom dead center), the piston begins to uncover the cylinder liner ports, permitting the now higher air-box pressure (ABP) to enter the cylinder; this is the start of the "scavenging" stroke. Since the ABP is now higher than the exhaust gas pressure, positive displacement of the exhaust gases out of the cylinder takes place. The scavenging process lasts for approximately 118 degrees of crankshaft rotation (59 degrees BBDC and 59 degrees ABDC (after bottom dead center). It is during this 118-degree period that scavenging; cooling of the piston, liner, valves, and cylinder head; and inducting fresh air for combustion purposes occur.

During the piston upstroke the exhaust valves do not close until after the cylinder liner ports have been covered at approximately 59 degrees ABDC by the upward-moving piston. Typically the exhaust valves close between 3 and 5 degrees after port closure. The piston is now on its compression stroke. The start of fuel injection at all speeds is variable, based on the year of manufacture, the make and model of the engine, and whether it is mechanically or electronically controlled. The fuel injection duration typically lasts for 10 to 14 degrees at an idle speed between 500 and 700 rpm beginning at BTDC and ending just at TDC, or a few degrees ATDC based on the engine make and model and its year of manufacture.

DIESEL COMBUSTION TYPES

The majority of existing diesel engines now in use operate on what is commonly referred to as a direct-injection (DI) design (see Figure 5). This means that the high-pressure injected fuel (as high as 30,000 psi or 207 MPa) enters directly into the combustion chamber formed by the shape of the piston crown. In the indirect-injection (IDI) system the injected fuel is sprayed into a small antechamber within the cylinder head. Combustion begins in this small chamber and forces its way into the main chamber, where it consumes the remaining air required for additional combustion. IDI engine designs require use of an electrical glow plug to initiate satisfactory combustion, something that is not required in a DI engine. Use of a glow plug allows the IDI engine to burn a rougher grade of fuel than the DI engine. Fuel is injected at a lower pressure in the IDI; in addition, the larger combustion surface area of the IDI engine creates greater heat losses. This results in the IDI engine consuming approximately 15 percent more fuel than an equivalent horsepower (kW) rated DI engine. Additionally, using a rougher grade of diesel fuel results in higher exhaust emissions than from a DI engine using high-quality, low-sulfur fuel.

THERMAL EFFICIENCY

Rudolph Diesel's original intent was to produce a low-heat-rejection internal-combustion engine without the need for a cooling system. He believed that this would provide less heat losses from the combustion process and provide him with a superior heat, or thermally efficient (TE), design concept. To his chagrin,

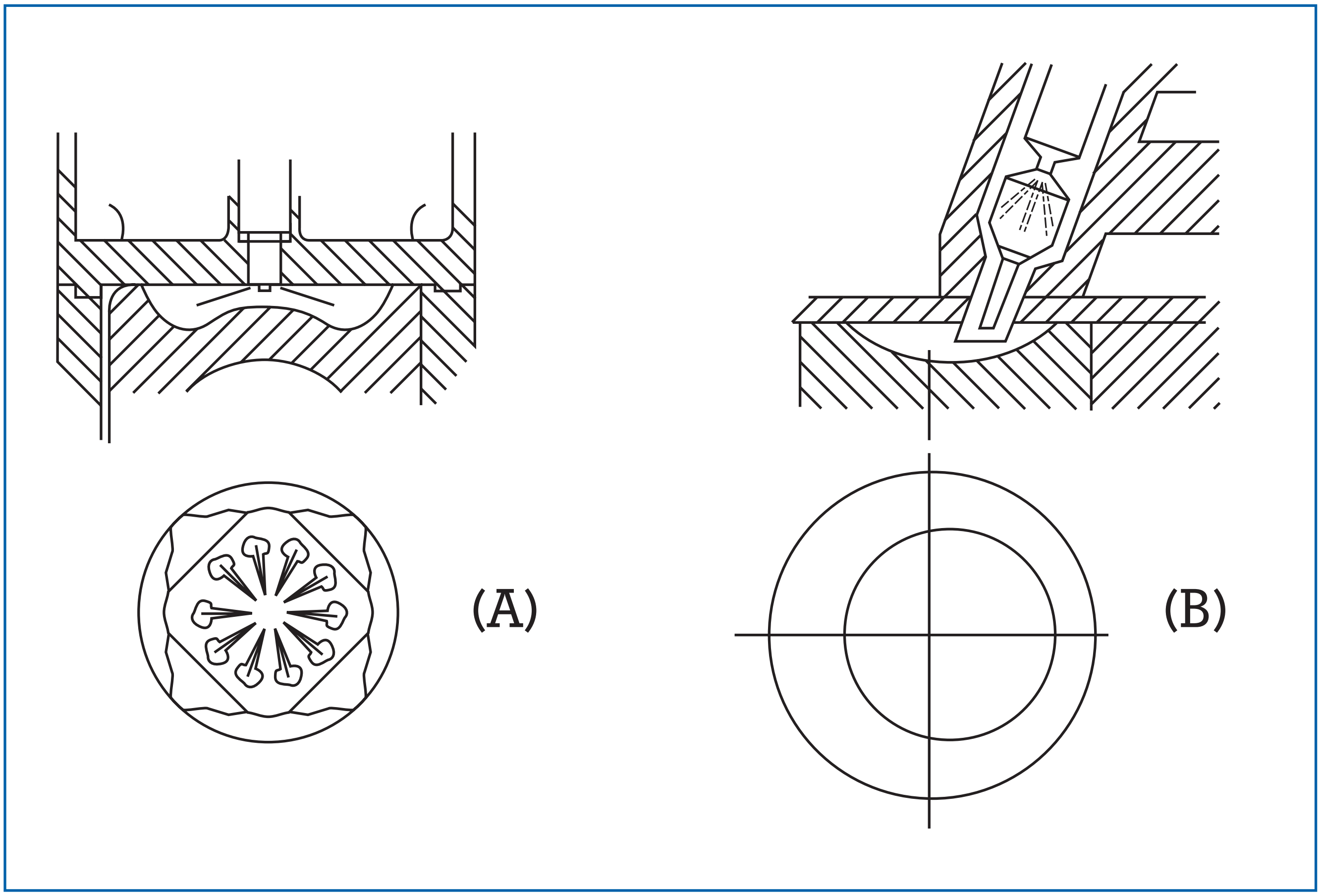


Figure 5.
Principles of (A) DI (direct-injection) and (B) IDI (indirect-injection) combustion chamber designs.

however, he found that this was not a feasible option when his first several test engines failed to perform to plan.

Basic physics involving friction and heat losses prevents the construction of a perfect internal-combustion engine. If friction between moving components could be eliminated, the mechanical efficiency (ME) of the engine would improve. Similarly, if we could eliminate heat losses from the combustion process, we could improve the TE of the combustion process. If no friction or heat losses existed we could design a perfect or ideal engine—one that could provide closer to 100 percent efficiency. A simplified way to consider TE is that for every dollar or hundred cents of fuel consumed by the internal-combustion engine, how much is returned as usable power? Therefore the TE is a comparison of the actual ratio of useful work performed in the engine versus the total energy content of the fuel consumed. Typically gasoline engines today return twenty-eight to thirty-six cents on the dollar. The diesel engine returns approximately forty to forty-three cents on the dollar. It is a measure of how efficiently an internal-combustion engine uses the heat released into the combustion chamber from the fuel to produce mechanical power. Based on the specific make and model of the engine, the cooling and exhaust systems typically account for heat losses of about 23 to 27 percent each; friction losses can range between 7 and 9 percent; while radiated heat from the engine accounts for 3 and 5 percent. Therefore this combination will generally account for a 57 percent heat loss, resulting in a TE of approximately 43 percent in current electronically controlled high-speed diesel engine models. However, many stationary diesel power plants that recapture waste exhaust heat for cogeneration purposes can return TEs in the mid-50-percent and higher ranges. New mechanical and electronic design concepts plus the adoption of ceramic components are just some of the new technologies being adopted not only to improve thermal efficiency but also to drastically reduce exhaust emissions into the atmosphere.

DIESEL COMBUSTION

The diesel engine without the benefit of a spark plug does not generate instantaneous combustion, as occurs within the gasoline engine. Instead, the diesel cycle relies on the high-pressure atomized-injected fuel mixing with the hot compressed air to cause it to vaporize. Once this vaporization occurs, the air/fuel mixture generates a flame front to initiate combustion. This concept creates what is known as "ignition delay" and is one of the characteristics that gives the diesel cycle its unique pinging noise. The longer the ignition delay, the louder the combustion noise (hard combustion) due to the larger volume of injected fuel that collects within the combustion chamber prior to actual ignition. The start of ignition includes the fuel injected prior to this phase and is known as the "premixed flame."

Once the fuel ignites, the remaining fuel being injected has no ignition delay, since it is being sprayed directly into an established flame front. Under full-load conditions, peak cylinder pressures can average between 1,800 and 2,300 psi (12,411 and 15,859 kPa). These tremendous pressures produce the power within the diesel engine during the power stroke, resulting in a higher overall BMEP (brake mean effective pressure—the average pressure exerted on the piston throughout the power stroke), versus the much lower values in a gasoline engine and the higher TE levels of the diesel.

DIESEL EXHAUST EMISSIONS

All internal-combustion engines, due to their inherent design characteristics, are unable to burn the injected fuel to completion. The make, model, year of manufacture, cylinder displacement, speed, and load all affect the percentage of emissions emitted into the atmosphere and the air we breathe. Therefore, major research and development is a continuing effort to clean up the type and quantity of pollutants. The EPA is a government agency charged with setting the limits on all industrial and internal-combustion-engine limits. The European Economic Community as well as Asian and other countries have similar agencies tasked within these same parameters.

Major exhaust emissions from internal-combustion engines targeted by the EPA can be categorized into the following areas:

1. Carbon dioxide (CO_2), although nonpoisonous, does contribute to "global warming." Complete combustion in an internal combustion engine produces CO_2 and water.
2. Carbon monoxide (CO) is a colurless, odorless, and tasteless gas. Inhalation of as little as 0.3 percent by volume can cause death within thirty minutes. The exhaust gas from spark ignition engines at an idle speed has a high CO content. For this reason NEVER allow an engine to run in an enclosed space such as a closed garage.
3. Oxides of nitrogen (NO_x) have two classes. Nitrogen monoxide (NO) is a colorless, odorless, and tasteless gas that is rapidly converted into nitrogen dioxide (NO_2) in the presence of oxygen. NO_2 is a yellowish-to-reddish-brown poisonous gas with a penetrating odor that can destroy lung tissue. NO and NO_x are customarily treated together and referred to as oxides of nitrogen.
4. Hydrocarbons of many different types are present in exhaust gas. In the presence of nitrogen oxide and sunlight, they form oxidants that irritate the mucous membranes. Some hydrocarbons are considered to be carcinogenic. Incomplete combustion produces unburned hydrocarbons.
5. Particulate matter, in accordance with U.S. legislation, includes all substances (with the exception of unbound water) that under controlled conditions are present as solids (ash, carbon) or liquids in exhaust gases.

Diesel engines, due to the combustion processes described herein, tend to have a rougher time meeting some of the specific exhaust emissions standards. The primary cause of combustion noise and the generation of oxides of nitrogen in the diesel engine can be traced to that portion of the combusted fuel that burns as a very rapid premixed flame. On the other hand, the slower-burning diffusion flame (fuel-injected after the start of ignition) is the primary cause of soot and unburned hydrocarbons.

At this time it is not possible to produce a totally soot-free diesel engine because heterogeneous combustion always produces soot. Diesel engine operation, due to local concentrations of overly rich mixtures in the diffusion flame, leads to an increase in the emission of black smoke to a moderate extent even with moderate excess air. The relatively low exhaust gas temperatures of diesel engines create a problem for effective catalytic emission control of hydrocarbons, particularly in light-duty diesel engines.

Reduction of exhaust emissions is being tackled in two ways by engineers, including precombustion and postcombustion technology. One of the most effective methods now being researched and adopted includes use of synthetic fuel made from natural gas. This fuel is crystal clear, and just like water, it has no aromatics, contains no sulfur or heavy metals, and when used with a postcombustion device such as a catalytic converter any remaining NO_x or other emissions can be drastically reduced. Estimates currently place the cost of this fuel at $1.50 per gallon, with availability in 2004 to meet the next round of stiff EPA exhaust emission standards.

Some precombustion technology involves improvements in internal engine hardware components, various engine sensors, and electronically controlled common-rail fuel injection equipment. Other systems now on test incorporate the addition of a small quantity of a reducing agent such as urea (sometimes called carbamide) injected into the combustion chamber, resulting in a chemical reaction that releases ammonia. This in turn converts the exhaust gases into nontoxic levels of nitrogen and water.

The second method used to reduce exhaust emissions incorporates postcombustion devices in the form of soot and/or ceramic catalytic converters. Some catalysts currently employ zeolite-based hydrocarbon-trapping materials acting as molecular sieves that can adsorb hydrocarbons at low temperatures and release them at high temperatures, when the catalyst operates with higher efficiency. Advances have been made in soot reduction through adoption of soot filters that chemically convert CO and unburned hydrocarbons into harmless CO_2 and water vapor, while trapping carbon particles in their ceramic honeycomb walls. Both soot filters and diesel catalysts remove more than 80 percent of carbon particulates from the exhaust, and reduce by more than 90 percent emissions of CO and hydrocarbons.

EPA diesel exhaust emissions limits for 1998 on-highway diesel truck and bus engines in g/bhp-hr (grams/brake horsepower/hour) when using existing 0.05 percent low-sulfur diesel fuel were: hydrocarbons, 1.3; CO, 15.5; NO_x, 4.0; and particulate matters, 0.1. Regulations due to come into effect beginning with the 2004 model year represent approximately a 50 percent reduction in emissions of NO_x and particulate matters, as well as reductions in hydrocarbons.

DIESEL AUTOMOTIVE CONSUMER RELUCTANCE

Despite the superior fuel economy of the diesel engine and its longer life to overhaul versus that for most of its equivalent gasoline counterparts, the general automotive consumer has preferred the choice of the gasoline engine in passenger cars and light trucks. However, within the European Community, the better fuel economy of the diesel accounts for up to 35 to 40 percent of all vehicle sales. In North America, diesel pickup trucks from all of the domestic manufacturers are popular options. To capture a greater percentage of the vehicle market, diesel engines require some technological improvements to bring their overall performance closer to that of the gasoline engine.

The following are the nine basic advantages of gasoline over diesel:

1. Quieter operation (no *ignition delay* or diesel knock; lower peak cylinder pressures and temperatures).
2. Easier starting, particularly in cold-weather operation (gasoline vaporizes at a much lower temperature than does diesel fuel; spark plugs provide instant combustion).
3. Less unpleasant odor, particularly in the exhaust gases.
4. Tends to burn visually cleaner at the exhaust tailpipe, since it operates in a closed-loop electronic mode (oxygen sensors interacting with the powertrain control module) to maintain an ideal air/fuel ratio of 14.7:1.
5. Quicker acceleration (no ignition delay) at lower engine speeds.
6. Can operate at higher speeds (rpm). Less inertia forces due to lighter components.

7. Good fuel economy at steady-state highway cruising speeds.
8. Lower weight, resulting in a higher power-to-weight density.
9. Generally lower production costs.

The following are the four basic advantages of diesel over gasoline:

1. Greater mileage between engine overhaul/repair (more robust).
2. Superior fuel economy (more thermally efficient) particularly at low speeds due to lack of restriction of the air flow; air flow restriction occurs in a gasoline engine through the throttling action of the gas pedal.
3. Lower carbon monoxide levels.
4. Higher crankshaft torque-producing capability.

DIESEL TECHNOLOGICAL ADVANCES

Since the 1985 model year, many heavy-duty high-speed diesel engines have been equipped with turbocharged aftercooled engines using electronic fuel injection controls. By the 1990 year all major high-speed engine original equipment manufacturers (OEMs) in North America employed electronic controls, since adopted by other diesel engine OEMs. Metallurgical advances have provided lighter but stronger engine components, and the use of plastics and fiberglass and aluminum alloys has increased in many external engine components

Four-valve cylinder heads, overhead camshafts, ceramic turbocharger components, crossflow cylinder heads, two-piece cross-head pistons, electronically controlled injectors, and hydraulically actuated electronically controlled unit injectors that do not require a pushrod and rocker arm assembly are all in use on existing engines. Future rockerless valve control engines, nonferrous piston and liner components, and turbocompounding will all improve the thermal efficiency of future engines.

Variable valve timing similar to that now in use in gasoline engines will become more common on all internal-combustion engines. Future valveless engines might employ rotating hollow-type shafts in place of the long-used poppet valves, or an electric solenoid will be used to operate both intake and exhaust valves, once again reducing valve-train frictional losses to improve overall thermal efficiency. Turbocompounding is a process whereby the hot, pressurized exhaust gases, after driving the turbocharger, will be directed to a large expansion turbine geared to the engine crankshaft to return additional energy, which would otherwise be wasted by flowing out of the exhaust system. The result will be a substantial increase in thermal efficiency. Turbocompounding is not yet available in a full-scale production engine, but look for this feature on future engines. Low-flow cooling and lube systems that have been in use for some time have reduced parasitic losses to further improve fuel economy. Future cooling systems will employ ceramic components, permitting the engine to run at a higher coolant temperature and providing a further increase in thermal efficiency.

Robert N. Brady

See also: Automobile Performance; Carnot, Nicolas Leonard Sadi; Combustion; Diesel Fuel; Diesel, Rudolph; Engines; Gasoline and Additives; Gasoline Engines; Government Agencies; Otto, Nikolaus August; Thermodynamics.

BIBLIOGRAPHY

Bosch Automotive Handbook, 4th ed. Cambridge, MA: Robert Bentley.

Brady, R. N. (1996). *Modern Diesel Technology*. Englewood Cliffs, NJ: Prentice Hall.

Brady, R. N., and Dagel, J. F. (1998). *Diesel Engine and Fuel System Repair*, 4th ed. Englewood Cliffs, NJ: Prentice Hall.

Lilly, L. R. C., ed. (1984–1985). *Diesel Engine Reference Book*. Stoneham, MA: Butterworth.

Stinson, K. W. (1980) *Diesel Engineering Handbook*, 12th ed. Norwalk, CT: Business Journals

DIESEL FUEL

Liquid fuels for use in internal-combustion engines are extracted and refined from crude oil, with diesel fuels being part of the middle distillate or kerosene fraction. Kerosene was initially derived from coal pyrolysis. The initial main use of this type of distillate was for the kerosene lamp, which had replaced lamps based on whale oil.

In 1859 at Titusville, Pennsylvania, Edwin Drake drilled the first successful oil well, and by the late 1880s most kerosene was made from crude oil. The search for crude accelerated in 1892, when Charles Duryea built the first U.S. automobile powered by a gasoline-fueled internal-combustion engine. Rudolph Diesel patented a compression ignition engine running on middle distillate at about the same time. While this engine was more fuel-efficient, it proved too complex to manufacture, not gaining in popularity until the middle of the twentieth century.

PRODUCTION OF DIESEL FUEL

Within this encyclopedia is an article covering crude oil refining in greater detail. However, as a brief introduction, Figure 1 gives a typical example of the products extracted and refined from a barrel of crude oil. Diesel fuel averages a little more than 18 percent of the total, or approximately 9.2 gallons from each 45-gallon barrel of crude oil.

To extract the products shown in Figure 1, the crude oil is heated in a distillation furnace. The resulting liquids and vapors are discharged into a distillation tower (also called a fractionating unit). This tower is hottest at the bottom, with the temperature dropping gradually toward the top. The distillation (vaporization temperature) separates the crudes into various fractions according to weight (specific gravity) and boiling point. Distillate runs down through the tower over a series of horizontal trays that are perforated to allow the upflow of vapors. Each tray is cooler than the one below it, thus providing a temperature gradient throughout the height of the tower. As different fraction reaches the tray where the temperature is just below its boiling point, its vapors can condense and change back into liquid, where it can be drawn off if desired.

The very lightest fractions that rise in the tower remain in a vaporized state and are used as a fuel in the refinery. Other fractions condense at various points in the tower according to their boiling points, as shown in Figure 2. Medium-weight liquids including kerosene and diesel fuel remain in the tower middle; consequently they are referred to as middle distillates. Components for diesel fuel, which have a higher boiling point than gasoline, boil in the range of 175°C to 355°C (347°F to 671°F). Following distillation, many of the distillate fractions are processed further to purify them or to convert them into lighter or heavier fractions, according to market demands.

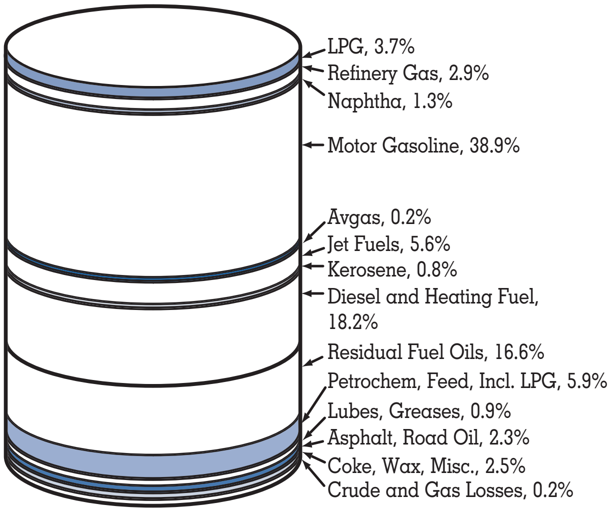


Figure 1.
Typical end-products extracted from a barrel of crude oil.

Quantities of middle distillates, such as diesel fuel, can be increased by several processes. Mild thermal cracking, a process known as visbreaking (viscosity breaking), breaks heavier molecules down with heat to reduce their viscosity. The heaviest fractions from the distillation towers can also be transformed using more severe conditions to “crack” heavy hydrocarbon molecules into lighter ones. Fluid catalytic cracking (“cat cracking”) uses intense heat, low pressure, and a catalytic substance to accelerate these thermal reactions. Hydrocracking employs a different catalyst, slightly lower temperatures, much greater pressure, and a hydrogen atmosphere to convert heavy molecules.

Residues may also be processed by removing carbon. These processes include deasphalting and coking, which produce usable liquid fuel components while rejecting a carbon-rich phase (asphalt or coke).

Diesel fuel makeup can represent various combinations of volatility, ignition quality, viscosity, sulfur level, specific gravity, and other characteristics. Various additives are used to impart special properties

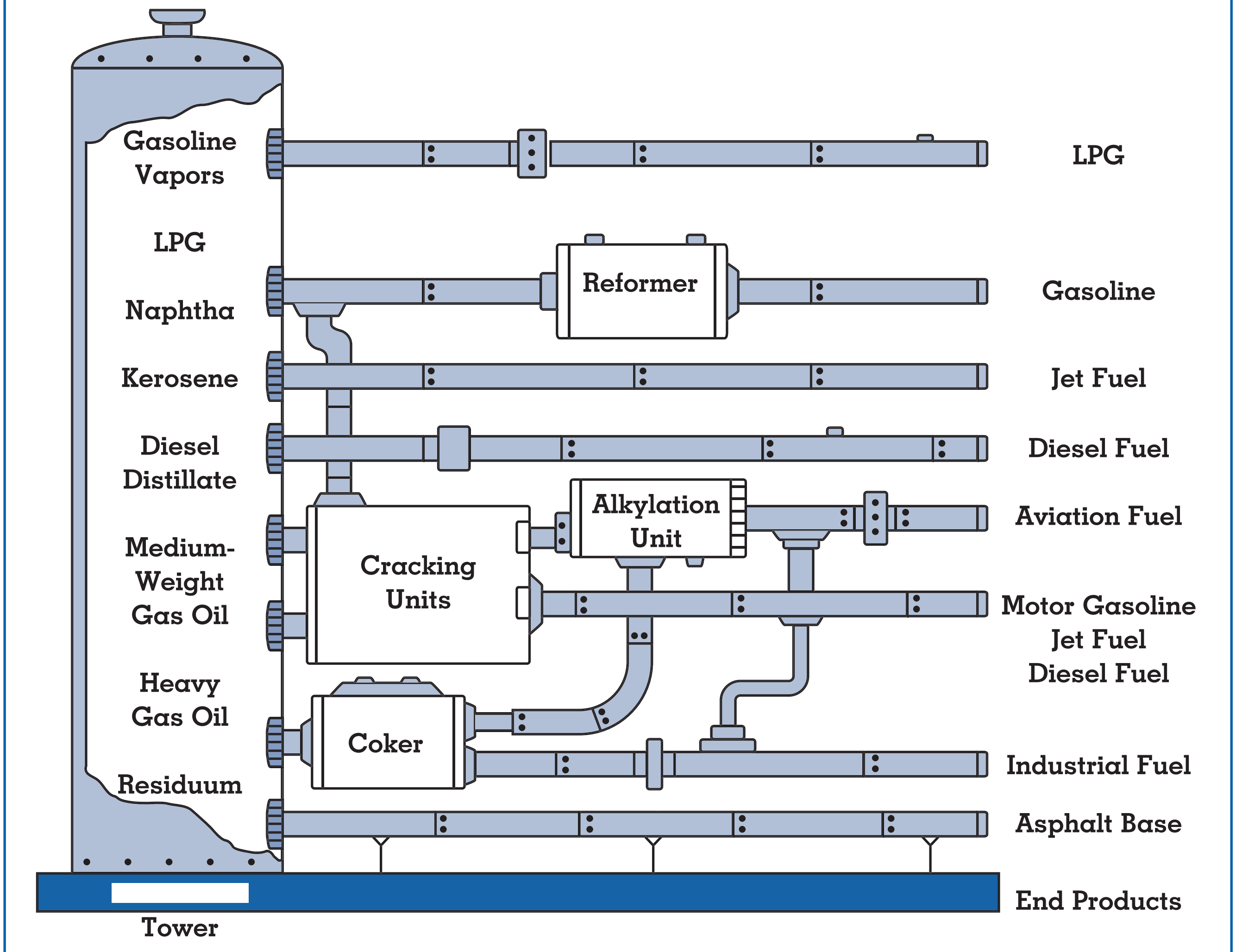


Figure 2.
Simplified arrangement of a crude oil refining system.
SOURCE: Chevron Research and Technology Company, a division of Chevron USA Inc.

to the finished diesel fuel. These can include ignition quality improvers, oxidation inhibitors (stability), biocides, rust preventatives (anticorrosion), metal deactivators, pour point depressants, demulsifiers, smoke suppressants, detergent dispersants, conductivity improvers, dyes, and deicers.

DIESEL FUEL CHARACTERISTICS

All liquid fuels and oils manufactured and sold in the United States since 1919 must comply with the engineering standards and quality specifications published by the American Petroleum Institute

(API). Many of API's standards are used around the world.

The main elemental constituents of diesel fuel are carbon (86% by weight) and hydrogen (13% by weight). Diesel fuel hydrocarbons contain molecules with eight to fifteen carbon atoms. The density of diesel fuel will vary based upon the grade, with a range of 0.815 to 0.855 kg/l (6.80 to 7.13 lb/gal) based upon its API gravity rating. Heating value, measured in British thermal units (Btus), depends on the fuel's heat of combustion and also whether the heat of vaporization in the water formed by combustion of the hydrogen in the fuel is usefully recovered. The higher heating value results if the water vapor is condensed. On the other hand, the lower heating value typifies the more common case, where the uncondensed water vapors are lost with the hot flue gases up the chimney or through the exhaust pipe.

Diesel fuel has a specific gravity that is lighter than water (specific gravity-1.00). Therefore a diesel fuel will float on water. Typical boiling points for diesel fuel will vary between 180°C and 360°C (356°F to 680°F), with the actual 100 percent boiling point (vaporization) being based upon the fuel grade. Its latent heat of vaporization is approximately 42.5 MJ/kg. Diesel fuel ignition temperature will vary based upon its grade, cetane number, and distillation range (vaporization temperature), but will average 250°C (482°F).

DIESEL FUEL GRADES

Diesel fuel oil is graded and designated by the American Society for Testing Materials (ASTM). For high-speed heavy-duty diesel engine, basically only two grades are now considered acceptable for transport use so that the engine exhaust emissions can meet Environmental Protection Agency (EPA) air pollution standards for toxic particulates and sulfur. These are the Nos 1-D and 2-D grade fuel oil classification. Fuel classifications below grades Nos 1-D and 2-D, which are designed for use in larger, slow-speed diesel engines, are not considered acceptable for use in high-speed automotive or truck engines. The No. 1-D fuel (typically API gravity 44) is a lighter distillate than a No. 2-D grade (typically API gravity 38 to 39).

Consequently, all things being equal, an engine operating on the same No. 2-D fuel with greater heat content per gallon will yield superior fuel economy than when operating on a grade No. 1-D fuel. This will generally result in increased engine output, and 5 to 10 percent better fuel mileage. In summary, lower-density fuels as well as winter blended fuels have lower volumetric heat content, lower viscosity, and poor lubrication characteristics.

FUEL PROPERTIES

The individual characteristics that particularly affect the performance of diesel fuels stem from the specific requirements of the diesel engine.

In the internal-combustion engine, air at ambient temperature and pressure is used to evaporate a rather volatile fuel (gasoline), and the combustible mixture is valved into the combustion chamber. Here it is compressed by the piston upstroke, and it is then ignited by a spark plug. The expanding hot gases drive the piston into its downstroke, thus performing work. The most important characteristics of the fuel are volatility (ready evaporation in cold weather but no vapor lock in hot weather) and smooth ignition without any engine knock. The latter quality is measured by the octane number.

In the diesel engine, compression ratio is much higher than in an internal-combustion engine. This adiabatic compression raises the air temperature above the ignition temperature of the diesel fuel. The diesel fuel is sprayed into the combustion chamber at the top of the piston upstroke via a high-pressure injection nozzle. The fuel still has to have enough volatility to vaporize quickly, and there has to be enough heat capacity in the compressed air to supply the heat of vaporization to the fine droplets. If these conditions are not met, as in a cold engine or with heavier fuels, the burning droplets will tend to form smoke and soot. The cetane number is analogous to the octane number in internal-combustion engines, and it is the single most important parameter in judging diesel fuel quality.

Distillation Temperature

This is the temperature at which the liquid fuel will vaporize when injected into the combustion chamber; therefore it becomes an important factor in the ignition-delay period of the fuel. Lower-boiling-range fuels such as No. 1-D are more volatile, while No. 2-D has a lower volatility, therefore requiring higher temperature to vaporize or boil. Due to these factors, higher-volatility fuels such as No. 1-D are preferential in applications where the engine will idle for long periods, such as in city bus/coach installations, or when operating in subzero ambient temperatures.

Boiling Point

The temperature at which the fuel is boiled off or vaporized at the refinery is known as the "end point temperature" listed in ASTM test D86, while ASTM spec D975 uses a 90 percent boiling point or distillation temperature to determine its suitability to vaporize. However, a number of major heavy-duty, high-speed diesel engine manufacturers specify that prior to selecting a diesel fuel you should ensure that a 95 percent distillation temperature is considered to ensure better combustion.

Cetane Number

Diesel fuel uses a numbering system whereby the higher the cetane number, the more volatile the fuel. Since a diesel engine requires that the injected fuel must vaporize from its liquid state and self-ignite from the heat of the air compression, a low-cetane-number fuel will result in a longer ignition delay, hard starting, incomplete combustion, smoke in the exhaust, a noisier engine (hard combustion), and lower horsepower. In cold weather operation, noticeable white smoke at the exhaust stack will last until the engine warms up. In engines equipped with charge-air-cooling to lower the turbocharger air inlet temperature, a low-cetane fuel usually will cause white exhaust smoke during light load operation. In high-speed heavy-duty electronically controlled diesel engines throughout the 1990s and into the new millennium, a minimum fuel cetane number of 45 is generally specified by the engine manufacturer for satisfactory performance. Cetane number is determined by putting the engine through a series of performance tests. The fuel marketer may add special additives to the fuel to improve the cetane number.

Sulfur Content

Stringent EPA exhaust emissions regulations introduced in October 1993 mandate that heavy-duty on-highway vehicle engines *must* use a low-sulfur diesel fuel that contains a maximum sulfur content of 0.05 percent. Prior to this time, the sulfur content was typically limited to 0.50 percent. However, most quality No. 2 brands ranged between 0.23 and 0.28 sulfur. Off-highway and marine engines still use cheaper fuels with a higher sulfur content, but EPA plans on all diesel engines eventually employing low-sulfur fuel to lower exhaust emissions. Diesel fuel sulfur content above 0.3 percent mass will cause premature piston ring wear, cylinder wear, and deposit formations, plus an increase in exhaust particulates. High-sulfur fuel for off-highway equipment can be identified by its greenish appearance due to the addition of a blue dye to identify it from the low-sulfur, honey-colored fuel for on-highway engines. Low-sulfur diesel fuel leaves approximately 0.01 g/bhp-hr sulfate in the raw exhaust, or approximately one-fifth or less lower than regular high-sulfur fuel. Because much of the crude oil imported into the United States has a high sulfur content, refining it to produce low-sulfur diesel fuel requires additional and more expensive procedures in the refining process.

Fuel Viscosity

Besides fulfilling the combustion requirement of the engine, diesel fuel also must act as a lubricant and cooling agent for the fuel injection system components. As the engine heats up, the fuel temperature also will increase resulting in a decrease in fuel viscosity. Therefore, to prevent scuffing and scoring of the fuel injection components, a selected fuel must meet minimum viscosity specifications. Additionally, hot fuel will reduce the metered quantity for a given throttle position, resulting in lowered horsepower. Thus, many current electronically controlled diesel fuel systems employ fuel coolers to maintain fuel operating temperatures at 65°C (150°F) or less.

Cloud Point

This condition is of concern only when equipment operates in subzero ambient temperatures. Since diesel fuel extracted from crude oil contains a quantity of paraffin wax, at some low ambient temperatures this paraffin will precipitate and create wax crystals in the fuel. This can result in plugging of the fuel filters, resulting in a hard or no-start condition. Any moisture in the fuel can also form ice crystals. Cloud point temperatures for various grades of diesel and other fuels should be at least 12°C (21.6°F) below the ambient temperature. In cases where cloud point becomes a problem, a fuel water separator and a heater are employed.

Pour Point

The pour point is the temperature at which the diesel fuel will no longer flow and is typically listed as being 5°C to 8°C (9°F to 14.4°F) lower than the cloud point of –13°C to –10°C (9°F to 14°F).

Flash Point

Due to its low vaporization temperature, diesel fuel is a fairly safe transportable fuel. Flash point has little to do with the fuel's combustibility factor or the performance characteristics of the engine. Rather it is a measure of the temperature at which the fuel oil vapors flash when in the presence of an open flame. Safety in handling and storage are the main points warranting consideration for flash point. A No. 1-D fuel has a flash point of 38°C (100°F), while a No. 2-D fuel flash point is 52°C (126°F). The storage of diesel fuel should always take place in clean and contaminant-free storage tanks. Tanks should be inspected regularly for dirt, sludge, and water that can cause microbial growth. Tanks need to be drained and cleaned regularly if signs of contamination are evident.

Carbon Residue

Carbon residue is expressed as a percentage by weight of the original sample of the fuel, with the amount determined by burning a given quantity in a sealed container until all that remains is carbon residue. The amount of carbon residue left within the combustion chamber of the engine has a direct bearing upon the internal deposits and affects the cleanliness of combustion, particularly the smoke emissions at the exhaust stack.

Ash Content

Diesel fuel may contain ash-forming materials in the form of abrasive solids or soluble metallic soaps. These solids cause wear of injection equipment, pistons, piston rings, and liners as well as increasing engine deposits. Ash content is expressed as a percentage of the weight of the original test sample of the fuel when burned to completion in an open container.

Water Content

All diesel fuels tend to contain trace water, expressed in parts per million (ppm). With the very high fuel injection pressures now used in electronically controlled diesel engine, fuel-filter/water separators are widely used, since water allowed to circulate freely through the injection system can result in seizure of components and erosion of injector orifice holes, and in extreme cases the high compressibility factor of water can blow the tip off of the fuel injector.

SYNTHETIC DIESEL FUEL

Engine manufacturers and oil refiners are researching and developing a synthetic blended diesel fuel. The many advantages of diesel power can be greatly improved by reducing the exhaust emission levels to comply with ever stricter EPA-mandated levels.

Substantial benefits can be achieved through the use of synthetic diesel fuels that burn much cleaner. Research has shown that current diesel engines with electronics technology can burn up to twenty times cleaner on this type of fuel versus existing fuel blends. This synthetic fuel is made from natural gas and can run in today's engines without modification. The fuel is crystal clear. It has no aromatics or sulfur, and it contains no heavy metals. With no sulfur, the addition of a catalytic converter to the exhaust system can further clean up the nitrous oxides and other emissions to where the diesel is as clean as its gasoline counterpart. Fuel refiners estimate that they can produce this synthetic diesel fuel for less than $1.50 per gallon. It will assist diesel engine manufacturers in meeting the next round of strict EPA exhaust emissions limits.

SUMMARY

In summary, diesel fuel with very low to no sulfur content is now possible with chemical and technological advances. Along with catalytic converters, electronic fuel systems, and sensors, the diesel engine for the new millennium will be capable of complying with ever more stringent EPA exhaust emissions. The diesel engine will continue to serve as the main global workhorse for all of the many thousands of different applications of its power cycle.

Robert N. Brady

See also: Automobile Performance; Combustion; Diesel, Rudolph; Diesel Cycle Engines; Engines; Fuel Cell Vehicles; Gasoline and Additives; Gasoline Engines; Government Agencies; Hybrid Vehicles; Kerosene; Synthetic Fuel.

BIBLIOGRAPHY

Bosch Automotive Handbook, 4th ed. Cambridge, MA: Robert Bentley

Brady, R. N. (1996). *Modern Diesel Technology*. Englewood Cliffs, NJ: Prentice-Hall.

Brady, R. N., and Dogel, J. F. (1998). *Diesel Engine and Fuel System Repair*, 4th ed., Englewood Cliffs, NJ: Prentice-Hall.

Lilly, L. R. C., ed. (1984–1985). *Diesel Engine Reference Book*. Stoneham, MA: Butterworth.
Stinson, K. W. (1980). *Diesel Engineering Handbook*, 12th ed. Norwalk, CT: Business Journals.

DI-METHYLETHER

See: Alternative Fuels and Vehicles

DIRECT CURRENT

See: Electricity

DIRECT CURRENT MOTOR

See: Electric Motor Systems

DISTRICT HEATING AND COOLING

Thermal energy delivered to a building from an outside source is known as district heating and cooling, which can range in size from small systems serving two or three buildings to networks serving entire cities. District heating and cooling is widely used in developed countries throughout the world and offers numerous advantages over individual building apparatus, including greater safety and reliability, reduced emissions, and greater fuel flexibility, particularly in using alternative fuels such as biomass or waste.

The earliest examples of district heating were Roman hypocausts, a type of hot-air furnace often adapted to warm several buildings in close proximity, such as the three temples at Carnutum (Vienna). The hypocaust and other Roman technologies were reintroduced during the Renaissance, serving primarily as starting points for improvements. Meanwhile, city fathers in Chaudes Aigues, a small town in the volcanic Cantal region of southern France, had by 1322 levied a tax on several houses heated by a natural hot spring channeled through open trenches dug in the rock. The history of this system, which still operates warming 150 residences, includes the introduction of wooden pipe, later replaced by plastic conduits. Accounts of this system appeared in numerous architectural works and may have been the inspiration for the proposed introduction of district heating in London in 1622 by Dutch polymath Cornelius Drebbel. This scheme was primarily intended to distribute heat for cooking and warming and thus reduce air pollution, caused by individual coal stoves. Unfortunately, Drebbel's patron, Prince Charles, was more concerned with wooing the Spanish Infanta, and Drebbel used his talents for other purposes, including building the first working submarine.

Although Roman engineers almost exclusively used hot air for heating, they extensively employed hot water in public baths. This technology was also reborn in the late sixteenth and early seventeenth centuries by Sir Hugh Plat, Solomon DeCaus, and others during what some have called the Rosicrucian Enlightenment.

Steam and hot water were also used to heat extensive horticultural nurseries in England and Sweden. A French engineer, Bonnemain, used hot water to heat several large residences, and in 1785 a Bonnemain-type system was used to heat the three large buildings of Count Potemkin's Taurida Palace in St. Petersburg. A competitor of James Watt's, Matthew Murray, heated his house in 1804 by piping steam from his nearby factory, and in 1808 a Scottish engineer, Robertson Buchanan, wrote that "a number of neighboring buildings might be served with one boiler." In 1826, the religious utopian community of Old Economy, Pennsylvania, used the waste heat from its steam engine to heat several other buildings by means of buried steam pipes.

Despite a strong awareness of the potential advantages of district heating in the early nineteenth century, widespread adoption did not occur until technology was developed to handle the numerous problems associated with heat distribution. The Great

The main thermae of Sbeitlas is an example of an early hypocaust. (Corbis Corporation)

Exhibition of 1851 marked one of the positive turning points, with a separate boiler plant providing steam to power exhibits in the enormous Crystal Palace, a 600-meter-long glass structure. A similar exhibition, complete with separate boiler house, was held in New York two years later.

District heating began to appear more frequently in institutional complexes, which were becoming more numerous in both Europe and America. A new state capital complex in Springfield, Illinois in 1867 was served by a steam system, and the 1876 Croydon Asylum in the outskirts of London used hot water. Another large steam system was built for the American Centennial Exposition of 1876 in Philadelphia.

In March 1877, New York inventor Birdshill Holly introduced the Holly Steam Combination System to provide heat and power using steam distributed through an underground piping network. The first Holly Steam system was installed in Lockport, and within five years was installed in nearly fifty cities in North America and Europe. Although not all of the early systems were successful, and many were later abandoned for various reasons, the Holly System serving the downtown district of Denver, Colorado began service in November 1880 and has been operating ever since. A variant of Holly's design has operated in New York City since March 1882, providing reliable steam service to a large portion of Manhattan.

Developers in the late nineteenth century also recognized the need for artificial cooling in their customers' buildings. The New York Steam Company offered steam-driven absorption chillers as early as 1886, and four years later a district cooling company began operating in Denver by distributing chilled brine through underground pipes. Several other cities had such service by the 1920s, with systems distributing either brine or ammonia, often serving meat-packing and similar industries. Many electric companies also made ice in their central plants and distributed it to their customers for domestic refrigeration. During the 1930s, large, chilled water refrigeration plants were installed at Rockefeller Center in New York city and at the U.S. Capitol in Washington. In 1940 Southern Methodist University installed a central chiller plant and district cooling system at its campus in Dallas. Many institutional and industrial users installed large district cooling systems, and in 1962 commercial district cooling service was started in downtown Hartford, Connecticut, the first of many such systems.

Despite the early synergies between electric generation and district heating in America, the tremendous growth in generator size made it impractical to locate plants close to urban areas. Coal shortages during World War I (primarily caused by government takeover of railroads) sparked a movement to locate power plants close to coal mines, since it was thought to be more economical and even more environmentally desirable to transport electricity via wires. Small district heating systems that distributed heat from a power plant measured in kilowatts could not compete against plants making tens or even hundreds of megawatts of power. Many systems, however, benefited from managers who recognized the advantages of district heating and located plants where they could still serve heating customers.

At about the time that many American systems were being dismantled, district heating began to be incorporated into European cities such as Paris and Hamburg. The new Soviet Union studied various heating technologies throughout the world and adopted district heating to warm its new cities. In Iceland, the pipe for a planned geothermal system for the capital of Reykjavik was still in Copenhagen harbor when the Nazi invasion stopped the shipment. The United States eventually assumed responsibility for supplying Iceland, and calculated that it was more efficient to provide and transport twenty-four miles of steel pipe rather than deliver oil for Icelandic furnaces.

Many European cities, such as Rotterdam, incorporated district heating while rebuilding after World War II. Investigators from the American Strategic Bombing Survey were surprised to discover that aerial bombs had little effect on high-temperature hot-water piping in German factories. Within a short time such systems appeared on American military bases and, later, on the many colleges and universities opened in the post-war period. A Danish engineer working at the American air base in Thule, Greenland, invented a better way to prefabricate heating pipe. He returned home to start a company that was able to take advantage of the oil embargo that hit Denmark especially hard in the early 1970s.

Since Denmark imported nearly all of its energy in the form of petroleum, the Danish government adopted an ambitious plan to become energy self-sufficient. This was accomplished over the next twenty years through a combined effort to increase energy efficiency and to make use of alternative fuels, especially renewable resources such as wind and biomass. District heating played a large role in this effort, allowing previously wasted heat from factories and power plants to be used for useful heating. By 2000, more than 55 percent of Danish residences were connected to district heating networks. The environmental advantages of district heating have also become of importance due to global warming concerns. Utilizing the energy wasted in industrial and electric generation plants avoids burning additional fuel in individual buildings.

District energy systems primarily serve commercial and institutional buildings in the United States, although some systems serve multi-unit residential buildings. According to U.S. Department of Energy survey data, about 10 percent of commercial building floor space used district heating and about 4 percent used district cooling as of 1995. Steam was the predominant type of energy produced and distributed, accounting for about three-quarters of distributed energy. The median size system serves about two million square feet of floor space.

Overall, there were about six thousand district energy systems operating in the United states in 2000. They collectively provide over one quad of energy annually—about 1.3 percent of all energy used in the United States. Most systems serve institutions such as colleges, hospital complexes, and military installations, that is, they serve a number of buildings owned by a single organization. Utility systems that sell heating and/or cooling to separately owned buildings account for only around 16 percent of the district energy provided in the United States.

About 10 percent of the district energy systems are part of combined heat and power systems where both electric energy and useful thermal energy are produced. The electric generation capacity of these systems totals about 3,500 MW, about 0.5 percent of total installed electric generation capacity in the United States. However, the use of combined heat and power along with district energy distribution is growing, with new systems installed in recent years in Philadelphia, Penn., Trenton, N.J., St. Paul, Minn., and elsewhere.

Morris A. Pierce

DOMESTIC ENERGY USE

Domestic energy involves the production and consumption of heat, light, and power for a variety of activities. Heat provides space and water heating as well as cooking. Light, both artificial and natural, is necessary for indoor activities. Power is necessary for moving water, transporting people, fuel, and goods, and for numerous other domestic activities, including building and cultivation. For much of human history, energy came from just two sources: the sun and musclepower, both human and animal. At some point, fire became domesticated and over time became the largest and today, including its form of electricity, almost the only source of domestic energy. At the beginning of time, domestic energy was rather crude but generally uniform. Today it remains virtually unchanged in some areas, intensely sophisticated in others, and by no standard uniform. Even advanced developed cultures with similar living standards use vastly different amounts; for instance, on a per capita basis northern Europeans use half the energy as their North American cousins living in a similar climate.

As the philosopher Thomas Hobbes noted, historically human life has often been nasty, brutish, and short. For most of human history, tending the domestic sphere would have been quite sufficient to cause this misery. Communities, and later cities, first appeared during times when agricultural production exceeded the demands of the agricultural population. Most, perhaps all, of these cities were in mild climates, such as the Fertile Crescent, the Mediterranean basin, and their counterparts in America and Asia. The word “climate” itself is derived from the Greek word for latitude, and Aristotle and other Greek philosophers believed that the narrow band of latitude they lived in (along with its undiscovered counterpart in the Southern Hemisphere) were the only areas where humans could survive. Water was moved by muscle, windows let in light during the day, and torches, tapers, oil lamps and the glow of a fire provided it at night. Cooking could be done with wood confined in a ring of stones, but charcoal braziers were quite common, along with a remarkable variety of cooking and baking devices dating from preclassical times, often to bake bread. One example was the tannûr, which was conical and partly sunk into the earth. It is the only form of furnace mentioned in the Old Testament and is precisely detailed in Jewish law. Artificial warming could have been provided by a fire or a charcoal brazier, although Greek architects occasionally incorporated passive solar heating into their designs. In several locales fuels such as chaff, straw, reeds, rushes and dried dung were (and still are) commonly used. Warm bathing was virtually unknown among the Greeks, who considered it effeminate.

The Romans were the first real technologists, and as their empire expanded to the north, their heating systems and gravity-powered water aqueducts followed. The Roman hypocaust—a central heating system with an undeground furnace and tile flues to distribute the heat—was widely used and adapted to local fuels, such as coal in Britain. Charcoal braziers were commonly used for domestic heating, but stoves also appeared in certain areas. Emperor Julian wrote in 363 C.E. that he did not care for the Parisian stoves during a visit there, but was almost suffocated by fumes from a brazier used to warm his sleeping room. The Roman architect Vitruvius wrote the architectural treatise *De Architectura* which contains extensive descriptions of warming techniques. Romans also loved public baths, which along with bread and circuses were used to mollify the masses during hard times. Although there are mixed opinions regarding the fuel efficiency of the hypocaust, the sheer number of them exhausted forests from the Italian peninsula and required fuelwood to be imported from North Africa by ship. After the fall of the Roman Empire, the era known as the Dark Ages swept over Europe. Roman heating technology fell into disuse, but historians of climate generally see this period as being rather warmer than normal. The arrival of the “Little Ice Age” in the fourteenth century led to a rather intensive period of technological innovation, marked most prominently by the widespread adoption of the chimney to replace the communal fire, whose smoke escaped through a hole in the roof. The chimney soon led to improved fireplaces, and in an historical instant society changed, as a household could contain and manage multiple fires. In the 1370s William Langland's "The Vision of Piers the Plowman" assessed the chimney bitterly:

> Woe is in the hall each day of the week.
> There the lord and lady like not to sit.
> Now every rich man eats by himself
> In a private parlor to be rid of poor men,
> Or in a chamber with a chimney,
> And leaves the great hall.

This social upheaval was probably not matched until our own era, when families used to bonding around a single television set were wrenched apart by cheap portable sets that allowed individuals to watch in their own rooms. Once started, however, the process of social division was irreversible and indeed accelerated, as technology improved, at least for the upper classes. Chimney historian Leroy Dresbeck argues that chimneys may have promoted the art of love more than troubadors, but the Italian gentleman Octavian seems to have not minded the old way of communal heating, when as an overnight guest he would choose a place near one or another of the women of the house "by whom he was sometimes well received and sometimes got his face scratched."

One notable exception was a small village in the volcanic Cantal region of southern France. By 1322, several residences in Chaudes Aigues were being taxed for using hot water from a geothermal hot spring to heat their houses. The water was channeled to each house through a trench cut in the volcanic rock, which was covered with stone flooring in the living areas of the house. This system is still in operation and today heats 150 houses through a network of plastic piping. Apart from this notable exception, most improvements in domestic energy technology seem to have been aimed at the wealthier classes. For example, Elizabethan lawyer Sir Hugh Plat invented methods to heat greenhouses and generate steam to clear the skin of gentlewomen. He had not forgotten lesser mortals, however, and also invented a better candle, which would burn more than 120 hours, and a "coal ball," which was made from coal mixed with other ingredients to reduce smoke and increase fuel efficiency. His coal balls had the further advantage of being able to be manufactured by disabled veterans and orphaned children, which would reduce the public dole in addition to saving one-third of fuel consumed in England.

Plat was only one of several inventors active during what one historian has called the "Rosicrucian Enlightenment" in early Stuart England. Another of these was Cornelius Drebbel, a Dutchman who in 1622 proposed a district heating system in London to allow residents to warm their houses and cook without creating smoke. Drebbel also demonstrated an artificial cooling apparatus to the king and others in Whitehall Palace, who were driven out into the summer heat due to the intense cold. Among Dribbel's other inventions was the first working submarine. A contemporary German, Franz Kessler, corresponded with his London comrades and in 1618 published the first heating book, simply titled *Saving Fuel*. This book was illustrated with a number of stoves commonly used throughout Europe and discusses the various economies of the different types. He also predicted, quite accurately, that as heating apparatuses became more efficient, their use would become much more widespread. His book was translated into French the following year, but Kessler is remembered now more for his paintings and optical telegraph.

This early enlightenment came to an end in the chaos of the Thirty Years' War and the English Civil War, but the work of these men was revived later in the seventeenth century as growing national economies and populations required ever greater resources. The search for better heating apparatuses became critical as a bitter cold wave swept over Europe in the early eighteenth century. Technologists responded to the need, and in 1713 Nicholas Gauger published *La Méchanique de Feu*, which described a number of methods to improve heating systems. By 1720 Paris was suddenly warm, and newspapers in both London and Paris advertised firms that could heat rooms of any size "without the least suffocation." Warmth came with a price, however, and the streets of Paris were soon cluttered with thousands of carts laden with wood and sawyers plying their trade.

The great immigration to America began at this time, and English colonists found themselves surrounded by vast forests that could fuel "Christmas fires every day." Benjamin Franklin in the 1730s noted that Pennsylvania Germans burned one-fourth to one-fifth the fuel consumed by English colonists and had fewer children die during epidemics. He made a careful study of fireplace and stove technology, primarily relying on Gauger's book for basic information, and in 1744 advertised details of his new Pennsylvania Fire Place. Franklin's work was the most extensive investigation of heating systems up to that time, examining fuel, comfort, environmental effects, and health in some detail. He chose not to patent his inventions as he did not want to benefit from a product designed for the general good. Perhaps his great contributions were to debunk the English notion that cold houses were both healthy and inevitable, and to advance the idea that properly applied technology could deliver not only better warmth but also better health. Franklin's work had enormous impact on improving domestic heating systems throughout America and Europe.

Although open fireplaces were still common in houses for both heating and cooking, the stove came into widespread use during the nineteenth century. Part of this was due to the increasing number of manufacturers who could produce a quality product at low cost, but probably more important was the disappearance of wood as a domestic fuel over the course of that century, in America, just as it had in England three centuries earlier. This had probably been inevitable. Within two decades of the day the first English landed in Boston, the surrounding fuel supply had been burned up. One of the reasons why Massachusetts acquired Maine in 1677 was to tap its seemingly inexhaustible forests. Even Harvard University in 1800 had its own ship to bring fuelwood for its students and faculty from its land in Maine. As Americans moved westward across the continent, they were sure that the forests would last forever. In fact they did not last long at all, and only the discovery of coal in large quantities in Wyoming made it possible to build and operate the transcontinental railroad after the American Civil War. One of the most telling stories of this is Laura Ingall Wilder's *The Long Winter*, in which she describes living in a small town in South Dakota during the winter of 1880–1881. There were no trees (her six-year-old sister had never seen one), and the community was entirely dependent on coal delivered by railroad. During a particularly severe blizzard the tracks were blocked for several weeks, and Laura and her family burned straw and furniture to maintain a minimum amount of heat.

Coal could not be burned easily in an open fireplace, and the transition from wood to coal created enormous demand for coal stoves and other heating technology. The most promising was steam, where the boiler could be placed in a separate room (or even building) to allow coal to be kept away from the living area of a house. Steam apparatus was initially very expensive, and its initial markets were quality housing in larger cities. As the technology improved, particularly for safety devices, the prices dropped and they found a wider audience. One of the other advantages of steam was that it could also cook food. Steam kitchens had been thoroughly explored by Count Rumford early in the nineteenth century, although they were primarily limited to institutions, hotels, and large restaurants. The first residential steam heating apparatus used steam from a kitchen boiler and was described to the Royal Society in 1745 by Colonel William Cook.

On a larger scale, Lockport, New York, inventor Birdsill Holly in 1877 resurrected Drebbel's 1622 plan to provide district heating and cooling in London. Although Holly's district steam system became widespread in the commercial, institutional, and industrial marketplaces, his first systems included steam supply to residences for domestic use, including space and water heating and cooking. A variety of domestic steam appliances were invented. Demonstrations of steam-cooked meals were held for reporters in several cities, yet the idea never caught on. One that did was to use gas for these same purposes, which at that time meant gas manufactured in a local gas works by heating coal or coke and capturing the gaseous residue. Manufactured gas had been used since the late eighteenth century for street lighting as well as for interior illumination in a small number of commercial and larger residential buildings, but in the 1850s became more widespread in cities as clean whale oil became scarce and expensive. Kerosene was later to become widely used as a domestic lighting fuel, but gas made significant inroads into the residential market in many communities.

Although manufactured gas was too expensive for space heating, the small quantities necessary for domestic cooking and water heating provided a desirable market, and several local gas companies actively pursued it. Often a second meter was installed for the heating and cooking service, which was charged at a lower price. Manufacturers of coal stoves responded by developing a line of "summer stoves" that were primarily for cooking and water heating when the larger, winter heating stove was not necessary. Before long, salesmen from kerosene companies such as Standard Oil were also selling kerosene stoves for the same purposes, creating an intensely competitive market in some communities.

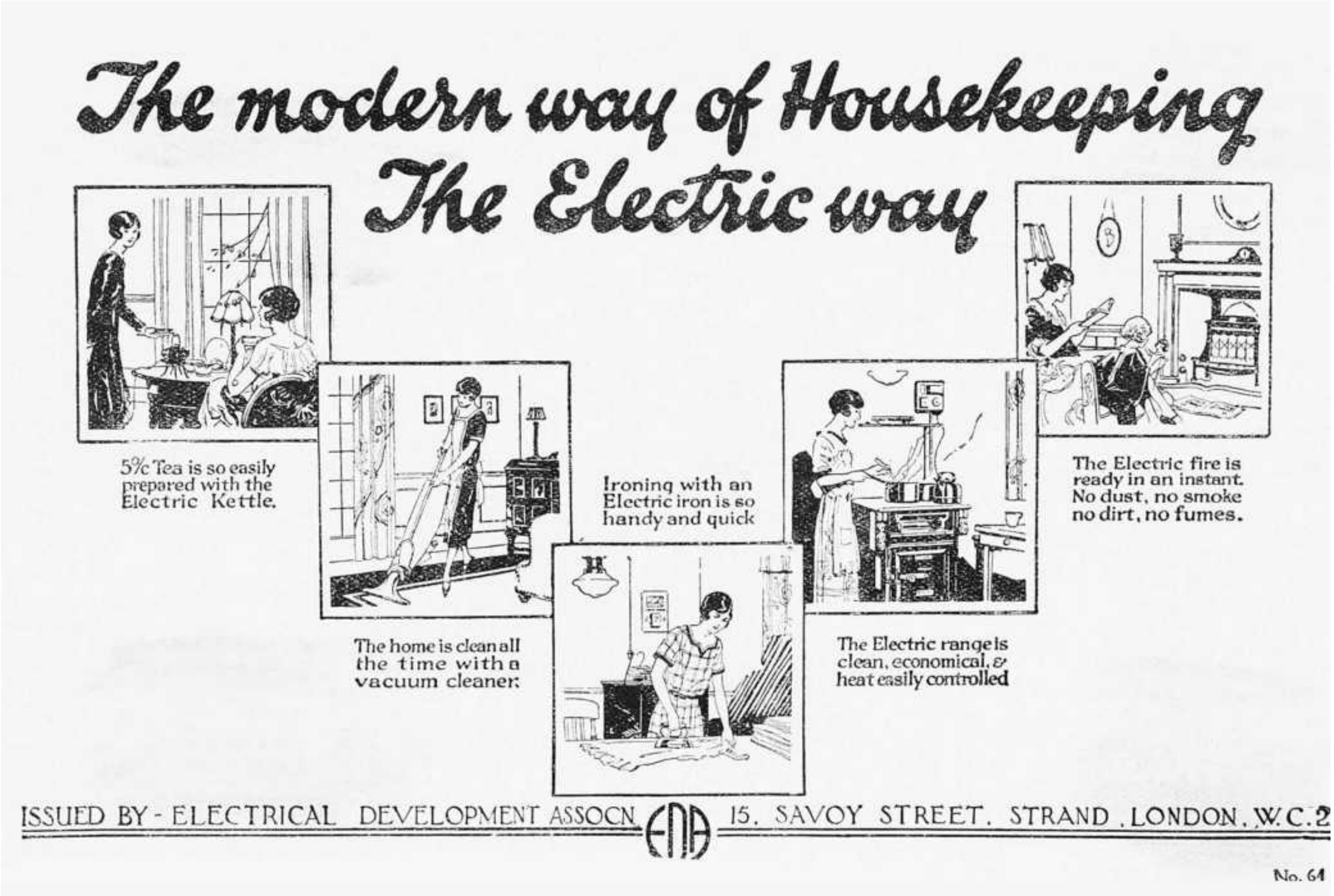
While this market struggle was going on, Thomas Edison was installing his first electric light plant, on Pearl Street in New York City. It began service in 1882, but it took a long time before electricity showed up in most houses. In 1907 only 7 percent of American households were electrified, and many of these probably retained some gas lighting as well for competitive reasons. Twenty years later, half of American households were electrified, and gas lighting was rapidly becoming only a memory. Much has been written about the struggle between alternating-current and direct-current electrical systems, but the much larger (and still ongoing) struggle has been gas and electricity for domestic cooking, heating, and refrigeration. Lighting was won very quickly by the incandescent lamp, later joined by fluorescent and other types to improve efficiency. Electric appliances such as toasters, irons, vacuum cleaners, and washers also prevailed in the market, often despite the early efforts of many local electric companies to discourage their use. Several local electric companies first became aware that housewives were using electric irons when power was shut off during the day (a normal occurrence in residential areas, since lights were not in use) and numerous complaints poured in.

Domestic refrigeration prior to World War I meant an icebox cooled with ice delivered by wagon and later by truck. Both electric and gas models appeared during the 1920s, and initially gas refrigerators were quite popular, since they were quiet and less expensive to operate. In 1930 only 5 percent of American households had a refrigerator, but within twenty years 80 percent had one. Although gas refrigerators are still used in recreational vehicles, hunting cabins, and other areas where electricity in unavailable, it is worthwhile to remark on their superb reliability. The American Consumer Product Safety Commission in 1998 issued a safety notice on a potential hazard from Servel gas refrigerators dating back to 1933, of which a large but unknown number were still in service. Once electric refrigerators were made quiet and efficient, they easily captured the market.

There still is intense competition between gas and electricity for the large ovens and ranges market. While electricity has completely prevailed in smaller appliances, including the relatively recent microwave oven, clothes dryers and domestic water heaters are also intensely competitive, but household heating by natural gas has become prevalent in most regions of the country where it is available. One of the interesting domestic energy sidenotes of the twentieth century was the "all-electric house," which was promoted as the ultimate in cleanliness and efficiency. Unfortunately such houses included electric space heating, which is not only the most expensive heating fuel but also the reason why many utilities built nuclear power plants. Domestic electric space heating is still prevalent in Norway, Québec, and areas of the northwestern United States with large amounts of hydroelectric power, but it remains to be seen how it will manage in an age of deregulated electric markets. Air conditioning also has become more widespread, not only in southern climates but also in any region with warm summer temperatures. Despite ongoing attempts by the gas industry to market gas residential cooling, this has largely been conceded to the electric utility industry.

Domestic water supply also has changed dramatically since the mid-1850s. Prior to that time water was collected from a source outside the house and delivered in buckets or other containers. A farmhouse without a spring or natural well probably would have a hand pump and later a windmill-powered pump, which became very popular in the American West. In urban areas a household water supply would have become more widely available, and over time transitioned from an intermittent supply, which might have provided water four hours a day, to a constant supply and eventually to a constant-pressure supply. Constant and generous sources of clean water led to indoor flush toilets; lush, green lawns in desert areas; and washing the family car. The availability of new forms of energy also led to other significant domestic improvements, including the telephone and the automobile.

It is easy enough to describe and even quantify domestic energy use as it has changed over the centuries and particularly since 1900, but it is much more difficult to assess how these changes have affected society, culture, and individuals for better or for worse. Ruth Schwartz Cowan argued in *More Work for Mother* (1983) that this new domestic technology resulted in the disappearance of domestic ser-



An early twentieth century advertisement from the Electrical Development Association (United Kingdom) promotes the use of electricity for housekeeping. (Corbis Corporation)

vants and higher standards of cleanliness that the lady of the house had to maintain. Yet even in 1950 a single working person would have been hard pressed to prepare his or her own meals and likely would have lived in a boardinghouse that provided one or more meals each day, or eaten in restaurants. Today meals are relatively easy to prepare and without much effort can even be quite elegant. It has, in fact, become so easy that many men now view domestic cooking as a fun thing to do. Probably this statement alone summarizes the changes that energy and technology have wrought in the household.

Morris A. Pierce

See also: Air Conditioning; Air Quality, Indoor; Appliances; Building Design, Residential; Coal Consumption of; Consumption; Edison, Thomas Alva; Electric Power Transmission and Distribution Systems; Heat and Heating; Insulation; Lighting; Natural Gas, Consumption of; Petroleum Consumption; Refrigerators and Freezers; Thompson, Benjamin; Water Heating; Windows.

DRILLING RIGHTS

See: Property Rights

DRIVETRAINS

The principal role of the automotive drivetrain is to transfer power from the engine output shaft to the drive wheels of the vehicle. Among its other functions are to multiply engine torque for improved vehicle performance, to operate the engine at a point offering good fuel economy, to enable operation of the vehicle in reverse, and to allow the engine to continue operation at idle speed while the vehicle is stationary. The transmission makes all of these functions possible. The drivetrain also incorporates a differential, so that when turning the vehicle, the outside drive wheel can rotate faster than the inside wheel. Finally, the drivetrain must accommodate the relative motion that occurs between the engine/transmission and the drive wheels. This requirement is fulfilled by the universal joint.

TRANSMISSION

In the absence of drivetrain losses, the power available to the drive wheels is equal to the power delivered by the engine. Power is proportional to the product of torque (the turning effort around the center of rotation) and rotational speed (typically measured in revolutions per minute, or rpm). The inevitable drivetrain losses are accounted for by a drivetrain efficiency. Losses in the transmission, and also the differential, arise from the friction of gears, bearings, and seals, and from churning of the gear lubricant. Churning losses depend on lubricant viscosity, which for a given lubricant varies with its temperature. In addition, the contemporary automatic transmission experiences flow losses in the torque converter and parasitic losses from the transmission pump used to provide the pressurized oil that controls shifting automatically.

Drivetrain efficiency depends on the rotational speed of individual components in the drivetrain and on the torque being transmitted. With a manual transmission, maximum drivetrain efficiency can be as high as 95 percent, but with an automatic transmission it falls closer to 85 percent. The passenger-car transmission seldom runs at its maximum efficiency point, however. For both transmission types, drivetrain efficiency trends toward zero as the transmitted torque approaches zero. Note the following:

$$\text{wheel torque} \times \text{wheel rpm} = \text{drivetrain efficiency} \times \text{engine torque} \times \text{engine rpm}$$

The wheel torque is equal to the product of the rolling radius of the drive wheels and the propulsive force acting on the wheel axle to move the vehicle forward. Thus:

$$\text{propulsive force} \times \text{wheel radius} \times \text{wheel rpm} = \text{drivetrain efficiency} \times \text{engine torque} \times \text{engine rpm}$$

This expression indicates that for a given delivered engine power and drivetrain efficiency, represented by the right side of the second equation, the propulsive force can be increased by turning the drive wheels at a slower speed. The transmission does this by altering the speed ratio between the engine and the drive wheels.

This ability of the transmission to change speed ratio is also important for improving fuel economy during normal vehicle driving. Some of the power developed within the cylinders of a traditional gasoline engine never reaches the output shaft. Rather, it is spent in overcoming mechanical friction in the engine and in pumping the fresh air-fuel mixture into the cylinders and the exhaust products out of the cylinders. Generally, the friction losses are decreased by operating the engine at a lower rotational speed. At a given rotational speed, the pumping losses are decreased by operating the engine with a more open inlet throttle valve. The transmission allows the engine to satisfy the road-load power requirement of the vehicle at a combination of rotational speed and throttle opening that offers good fuel economy commensurate with satisfactory engine response to a sudden demand for increased power.

The gearbox of a manual transmission houses a selection of gears of different diameter. The input/output speed ratio of a pair of meshed gears varies inversely with their diameters. The diameter of meshing gears is proportional to the number of teeth on each gear. The driver of the vehicle uses the gear-shift lever to select which gears mesh. The typical manual transmission for a passenger car has three to six forward speeds from which to choose, with five being the most common at the close of the twentieth century.

A three-speed version is represented in Figure 1. As shown, the transmission is in neutral—that is, it is transmitting no torque to the output shaft. The input and output shafts share a common axis and are paralleled by a countershaft. Drive gears on the input and output shafts are always meshed. The output shaft is splined so the gears it carries can be slid axially along it. In low gear, which is selected for starting a vehicle from rest and which gives the highest input/output

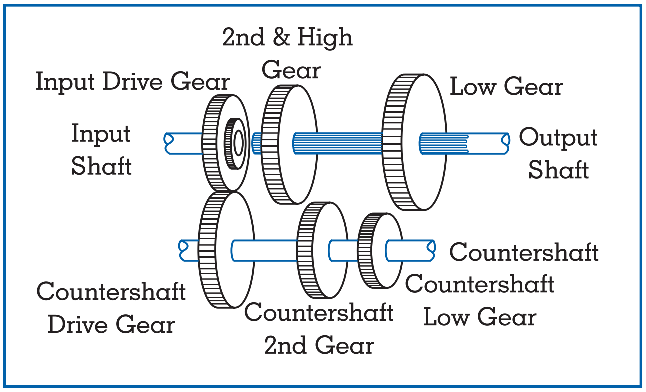


Figure 1.
Gearbox schematic for a three-speed manual transmission.

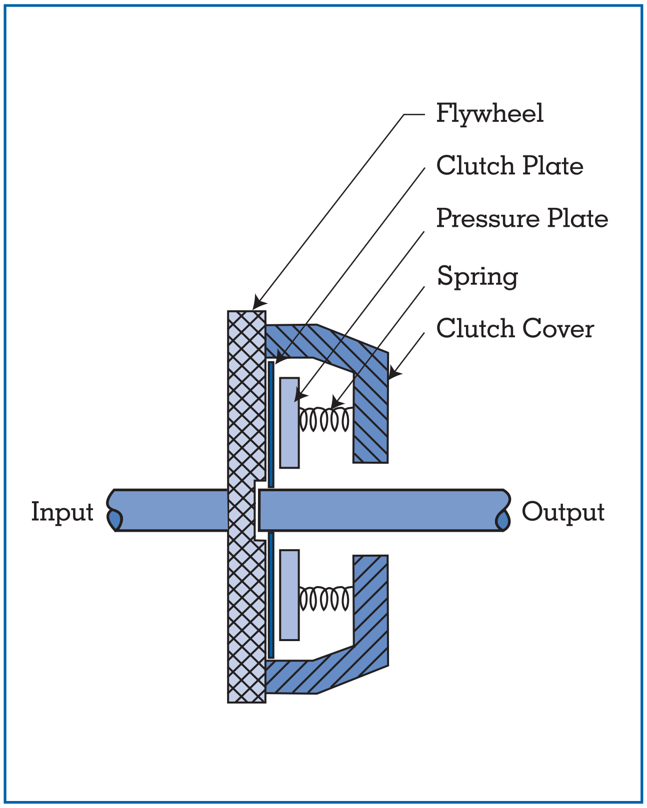


Figure 2.
Schematic of an automotive clutch.

speed ratio (greatest multiplication of engine torque), the low gear on the output shaft is slid axially to the left to mesh with the low gear on the countershaft. In second gear, the meshed low gears are disengaged and second gear on the output shaft is moved to the right to mesh with second gear on the countershaft. The gear at the left end of the output shaft contains teeth that can be engaged with matching teeth added to the input drive gear. In high gear, used for vehicle cruising, high gear on the output shaft slides to the left, locks onto the input drive gear, and engine torque is transmitted directly through the gearbox without multiplication.

When the transmission is shifted into reverse, an idler gear (not shown in Figure 1) is interposed between appropriate gears on the countershaft and output shafts to reverse rotation of the output shaft. When the vehicle is stationary with the engine running, the transmission is shifted into the neutral condition of Figure 1.

During gear shifting, the driver depresses a clutch pedal to disconnect the engine from the transmission input until the newly chosen gears are meshed. A disengaged clutch is shown schematically in Figure 2. Input torque is supplied by the engine flywheel, which is attached to the crankshaft. A clutch plate is mounted parallel and in close proximity to the flywheel. The clutch plate is coated with a friction lining. On the opposite side of the clutch plate is a pressure plate.

When the driver disengages the clutch by pushing down on the clutch pedal, a lever system (not shown) moves the pressure plate to the right against compression springs, as illustrated. When the driver engages the clutch by releasing the clutch pedal, the springs press the pressure plate and clutch plate against the flywheel, transmitting engine torque to the transmission input shaft. The friction linings on the clutch plate slide until engagement is completed. This prevents jerking of the vehicle during gear shifts.

Most automatic transmissions in production retain a gearbox, but the mechanical clutch is replaced by a rotating fluid unit that eliminates the need for a driver-operated clutch. In the gearboxes of most automatic transmissions for passenger cars, planetary gearing replaces the countershaft arrangement of Figure 2. A planetary gear set, illustrated in Figure 3, is comprised of a sun gear and a ring gear with intervening planet pinions, which are connected together by a planet carrier. The speed ratio across a planetary set depends on which of its three gear elements is prevented by a clutch from rotating, and also on which of the remaining two gear elements is the input. If the planet carrier is fixed, the ring and the sun rotate in opposite directions. If either the ring or the sun is fixed, the remaining elements rotate in the same direction.

The fluid unit in the typical automotive automatic transmission is called a hydrodynamic drive because it transmits power solely by dynamic fluid action in a closed recirculating path. Hydrodynamic drives are often classified into fluid couplings and torque converters. Both were patented in Germany in 1905 by Herrmann Foettinger. The original application was in marine vessels. In the 1930s, hydrodynamic drives began to appear in transit buses, where freedom from manually shifting a transmission was particularly desirable for a driver who had such other duties to perform as fare collection. Hydrodynamic transmissions also saw limited application in military vehicles during World War II.

The first passenger-car automatic transmission to see widespread use was the Hydramatic from General Motors. It used a fluid coupling, an example of which is illustrated in Figure 4. The fluid coupling has a torus-shaped split housing in which vanes are set radially to the axis of rotation. The impeller, on the right, is fastened to the engine. The turbine, on the left, is fastened to the gearbox input shaft. The vanes in the rotating impeller transfer engine torque into the transmission fluid as it is pumped outward. That torque is transferred by the swirling impeller discharge flow to the turbine vanes as the fluid flows radially inward to reenter the impeller.

Although the torque delivered by the turbine is equal to the engine torque absorbed by the impeller, their rotational speeds are not equal. When the engine is operating and the vehicle is stationary, there is 100 percent slip in speed from impeller to turbine, and the efficiency across the coupling is zero. If the vehicle is stationary with the engine idling, this provides the infinite output/input speed ratio of a disengaged clutch in a manual transmission, thus facilitating elimination of the clutch pedal. However, when the driver's foot is removed from the brake, the torque of the idling engine passes through the transmission to cause vehicle creep. During constant-speed cruising at moderate to high road speed, the slip in rotational speed between impeller and turbine is as little as 2 to 3 percent, yielding a fluid-coupling efficiency of 97 to 98 percent.

The torque converter soon replaced the fluid coupling as the hydrodynamic device of choice in the automatic transmission. A schematic of a simple torque converter appears in Figure 5. As in the fluid coupling, transmission fluid circulates around a torus. However, the torque-converter flow path contains an additional member: a row of reactor vanes. The reactor is mechanically connected to the stationary housing of the torque converter through a one-way clutch. This clutch allows the reactor to rotate around the axis of the torque converter when the fluid forces impose on it a torque in one direction, but to lock it against rotation when the torque acting on it is in the opposite direction. In contrast to the vanes in the fluid coupling, torque-converter vanes in the impeller, turbine, and stator are all carefully curved out of the plane of the rotational axis.

The distinguishing performance characteristic of the torque converter, in contrast to the fluid coupling, is that it is capable of multiplying torque. Torque multiplication is made possible by vane curvature and the presence of the reactor. When the converter is stalled—that is, the turbine and the reactor are stationary—the torque delivered to the gearbox is typically 2

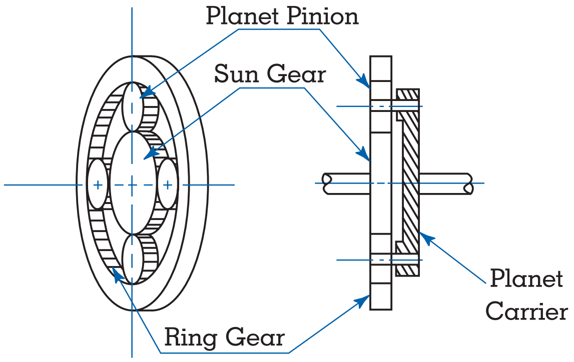


Figure 3.
Planetary gear set.

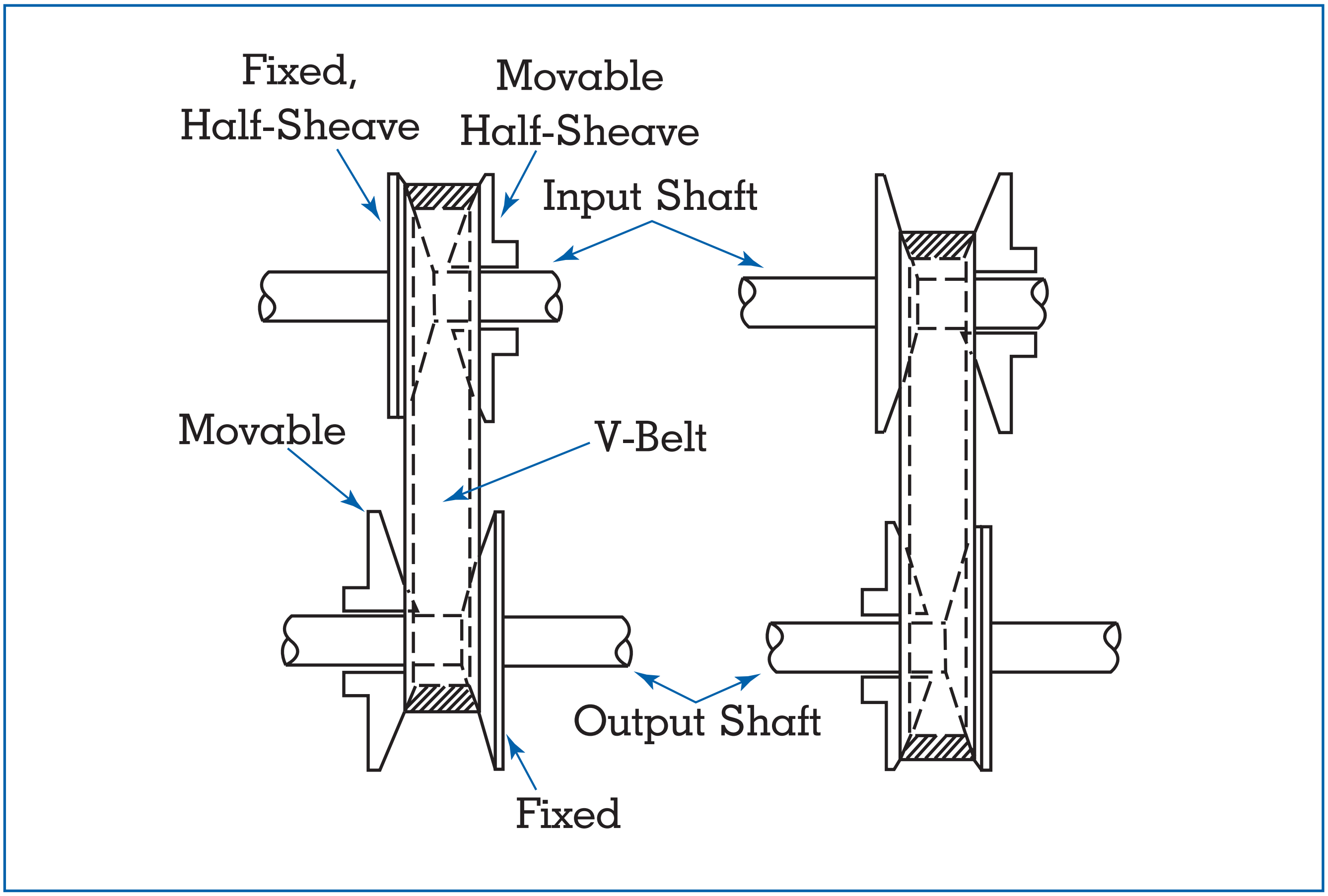


Figure 4.
Fluid coupling.

to 2.5 times the input torque from the engine. This torque multiplication normally allows a given level of vehicle acceleration performance to be achieved with one less step in the gearbox than if either a mechanical clutch or a fluid coupling were used. Since 1950, passenger-car torque-converter transmissions of the type described have usually employed gearboxes with two to five forward gears, with four most common at the end of the twentieth century.

As a vehicle is accelerated from rest, the initially stationary turbine accelerates toward impeller (engine) speed, and torque multiplication falls steadily. In a typical automotive converter, by the time the turbine has reached about 85 percent of the impeller speed, the torque ratio across the converter has dropped to unity and the one-way clutch allows the reactor to rotate. The speed ratio at which this occurs is known as the coupling point. Beyond the coupling point, the rotating reactor can no longer redirect the flow of torque-converter fluid and so cannot multiply input torque. The torque converter then acts like a fluid coupling, with efficiency peaking as high as 97 to 98 percent as a result of slip between impeller and turbine. In modern automatic transmissions, a torque-converter clutch is added that mechanically links the turbine to the impeller at this condition to prevent the slip responsible for such efficiency losses. Actually, slight slip is often allowed in the torque-converter clutch to minimize transmission of engine torque pulses to the vehicle drive wheels.

For many years, gear shifting in the automatic transmission has been controlled hydraulically, typically in response to vehicle speed and the position of the engine inlet throttle. In recent times that control has been reassigned to an electronic computer, which facilitates greater smoothness and flexibility in shift control.

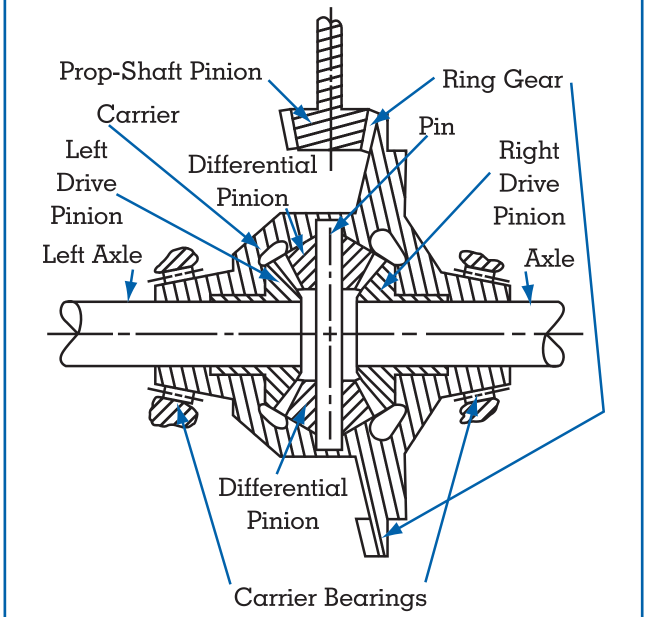


Figure 5.
Torque converter.

The substantial forces required to actuate gearbox clutches are still developed hydraulically, however.

Since 1985, more than three-quarters of new cars sold in the United States have had automatic transmissions. In fact, many licensed drivers in the United States have never learned to shift a manual transmission. In Europe, the high cost of gasoline has helped to retain the manual transmission because it is generally conceded to offer somewhat better fuel economy, although interest in automatic shifting there is growing. The congested driving encountered in Japan provides a greater incentive for the automatic transmission there than in Europe.

Practically all of the automatic transmissions sold in recent years have been of the torque-converter/gearbox type described, but there is growing interest in other approaches. Generally these have been of the continuously variable transmission (CVT) type. It is expected that by eliminating step changes in engine speed during shifting, as occurs with a gearbox, driving will be more pleasant. Also, eliminating the discrete speed steps of the gearbox with a CVT is expected, on average, to allow the engine to operate more efficiently. Countering those positions, the gear shifts of a traditional torque-converter/gearbox automatic have been smoothed until nearly imperceptible by evolutionary improvements that have included electronic shift control. Further, the increasing number of speeds in the gearbox has helped to diminish the margin of fuel-economy improvement attributable to a continuously variable ratio. It is noteworthy that Daimler and Benz equipped their first mass-produced gasoline-engine-powered vehicle with a CVT in 1886.

One type of CVT that has recently seen limited production use is the belt and variable sheave arrangement. Modern research on this approach started during the 1960s in the Netherlands. In that country, van Doorne makes a transmission of this type that uses a belt comprised of steel blocks joined by a steel band. On the left, spacing between the sheave halves has been decreased on the input shaft and increased on the output shaft to decrease the rotational speed of the output shaft. On the right, sheave-half spacings have been reversed to drive the output shaft faster than the input shaft. Because of load limitations on the belt, this CVT is suitable only with small engines.

A second type of CVT relies on rolling contact between two metal surfaces. The surfaces require lubrication by a special type of oil. General Motors worked on a toroidal transmission of this type in the 1920s and 1930s, but it lost out to the hydrodynamic type. The toroidal traction drive was actively pursued by Perbury in England, as well as by others, during the 1970s and 1980s. As of the start of the twenty-first century, it is receiving renewed attention by several developers striving for improved fuel economy.

In a toroidal traction drive, toroidal input and output disks face one another, separated by a number of rollers that contact the toroid surfaces. The rollers are mounted so that they can be tilted to vary the radius from the centerline of the disks where they contact the toroids, and therefore determine the input/output speed ratio of the rotating disks. A substantial axial force must be applied to the disks to prevent the rollers from slipping on the disk surfaces. To avoid excessive losses when the torque transmitted is low, this force needs to be modulated in proportion to the torque transmitted.

DIFFERENTIAL

As a vehicle turns a corner, the wheel on the outside of the turn must rotate faster because it travels farther, and the inside wheel must rotate slower than if the vehicle were traveling straight ahead. When the two wheels are only there to support the vehicle and not to deliver a propulsive force, this is not a problem because the wheels need not be mechanically connected. However, if they are drive wheels producing propulsive force from an engine, a way must be found to deliver torque to both wheels as they turn at different speeds, even though the input torque from the engine is supplied at a single speed.

In France in 1769, Nicholas Cugnot designed a self-propelled three-wheeled artillery gun tractor powered by a steam engine. He avoided the problem of propulsion in turns by delivering torque only to the single front wheel, as in a modern child's tricycle. However, instability of the tricycle on rough terrain necessitated transition to a four-wheel configuration. To manage its problem of propulsion in turns, only one of the two rear wheels was powered. With only the one wheel driving, though, traction proved inadequate in some situations. To improve traction, both rear wheels were driven, but then in sharp turns one of the wheels had to be manually disconnected from the power source. Finally, in about 1827, the Pecqueur idea of differential gearing was conceived to deliver power to both driving wheels while allowing them to rotate at different speeds in turns. Today the differential is so common that the early problem of propelling a four-wheeled vehicle in turns is forgotten by most.

In a rear-wheel-drive automotive differential, the differential is preceded by a "final drive," which consists of a pair of gears—a small pinion gear fixed to the prop shaft and meshing with a larger ring gear. These gears are beveled so the rotational axis is changed from the fore-and-aft orientation of the prop shaft to the transverse orientation of the rear axles. The difference in gear diameters causes the axle speed to be reduced while the delivered torque is correspondingly multiplied, as suggested by the first equation above. The input/output speed ratio typically falls between 2.5 and 4.5. Some such form of final drive is necessary because the overall speed ratio required from engine to drive wheels cannot normally be provided by the transmission alone.

Fixed to the ring gear and therefore rotating with it about the axle centerline is a carrier. Two identical beveled pinion gears are free to rotate about a pin fixed in the carrier. Each of these differential pinions engages both the right and the left drive pinions, which are fixed to the left and the right axles, respectively, but are free to rotate within the carrier.

When the vehicle drives straight ahead, the differential pinions do not rotate about the pin on which they are mounted. However, they tumble with the pin, end over end, about the axle centerline because the pin is fastened to the carrier and its ring gear. In this condition, the drive pinions, and with them their respective axles, rotate with the carrier and drive the vehicle forward.

In a turn, the carrier continues to tumble the differential pinions and their mounting pin about the axle centerline. However, if the vehicle turns to the left, the right drive pinion rotates about its centerline in the carrier in the forward direction. This increases the rotational speed of the right drive wheel, which is equal to the rpm of the carrier *plus* the rpm of the right drive pinion in the carrier. The rotating right drive pinion also makes the differential pinions turn on their mounting pin in opposite directions. This causes the left drive pinion to rotate about its centerline in the carrier at a speed equal and opposite to that of the right drive pinion. Now the rotational speed of the left drive wheel equals the rpm of the carrier *minus* the rpm of the left drive pinion in the carrier, and the difference in drive wheel speeds necessary to avoid wheel slip during the turn is achieved.

The differential can introduce a problem when driving on a slippery road because the torques on the two differential pinions are always equal, thus delivering equal torques to the drive wheels. In the extreme, if the tire of one drive wheel rests on ice and therefore lacks traction, the differential allows it to spin freely while the opposite drive wheel, and the vehicle itself, remain at rest. Special differential designs have been devised to overcome this problem. Traction is also improved in some vehicles through the application of four-wheel drive, whereby additional shafts and gears are employed to distribute engine power to all four wheels of the vehicle.

UNIVERSAL JOINT

The drive wheels of an automobile follow the irregularities of the roadbed. Although deflection of pneumatic tires absorbs some of these irregularities, the suspension system, incorporating springs and shock absorbers, is the primary means for smoothing out the ride for the passengers. The suspension system connects the drive-wheel axles to an assembly including the body, engine, and transmission. Thus as the vehicle moves forward, the road-induced vertical oscillations of this assembly, and therefore the transmission output shaft, differ from those experienced by the drive wheels.

Using a rear-wheel-drive vehicle as an example, a prop shaft delivers power from the transmission output, near the middle of the vehicle, to the differential that drives the wheels at the rear. If this prop shaft is designed to be stiff, which is normally the case, it could not be rigidly attached to the transmission output at one end and the differential at the other because of the differences in vertical movement between the drive wheels and the chassis. Typically, two universal joints are inserted into the drivetrain to accommodate this situation.

Following the operating principle of the most commonly used universal joint, the input shaft and the output shaft both terminate in yokes that are oriented in mutually perpendicular planes. The branches of each yoke are pinned to a cross connector so that each yoke can pivot about its beam of the cross. This mechanism was employed in the sixteenth century by Italian mathematician Geronimo Cardano, who used it to maintain a shipboard compass in a horizontal plane, regardless of the movement of the ship. Consequently it is often called a Cardan joint.

The seventeenth-century British mathematician Robert Hooke showed that if the input and output shafts do not share the same rotational axis, a uniform input rotational speed is transformed into a nonuniform rotational speed on the output shaft. He further showed how to avoid this characteristic by correctly phasing two joints in series. Because he acquired a patent on this mechanism, the Cardan joint is also known as the Hooke joint.

In a front-wheel-drive car, the drive wheels experience not only the road-induced vertical motion of the rear wheels but also must rotate about a vertical axis to accommodate steering. Several different configurations of constant-velocity universal joints have been developed to manage such motion. These constant-velocity joints are larger and more expensive than the joint described above.

Charles A. Amann

See also: Automobile Performance; Electric Vehicle; Fuel Cell Vehicles; Gasoline Engine; Hybrid Vehicles.

BIBLIOGRAPHY

Beachley, N. H., and Frank, A. A. (1980). “Principles and Definitions for Continuously Variable Transmissions, with Emphasis on Automotive Applications.” ASME paper 80-C2/DET-95.

Society of Automotive Engineers. (1979). *Universal Joint and Driveshaft Design Manual*. Warrendale, PA: Society of Automotive Engineers.

Zingsheim, E. W., and Schall, M. M. (1955). “Progress of Torque Converters.” *SAE Transactions* 63:84–97.

DYNAMITE

See: Explosives and Propellants

ECONOMICALLY EFFICIENT ENERGY CHOICES

If a person has the choice of installing oil, gas, or electric systems to heat a house and believes that any one of the three would perform equally well, the system that is cheapest is the efficient choice. If, however, the individual compares heating with an oil furnace to heating with a wood-burning stove, monetary cost may not be the only consideration.

Wood may be cheaper and oil more convenient. If two people are confronted by the same information and one chooses wood while the other chooses oil, both decisions may be economically efficient in the sense of maximizing the utility, or satisfaction, of the decisionmaker.

People differ in their preferences and in the value they put on their time. For consumer choices in particular, the entire list of qualities of services provided by different energy sources can be important. For example, both kerosene lamps and electric lights can be used to illuminate a home. Most households shifted quickly to electricity once it became available, and few would shift to kerosene today even if kerosene for lamp fuel were free. The lighting provided differs in more ways than cost.

Inputs and outputs usually can be valued according to market price. The opportunity cost of any choice is the best opportunity that has to be given up to make that choice. For the firm, the opportunity cost of oil or natural gas is measured by the price the firm must pay for it. That does not imply that the firm will always choose the cheapest fuel, because the costs of using the fuel must also be considered. Even in a situation where electricity is many times as expensive per British thermal unit (Btu) as coal, the firm may choose to buy electricity to operate lights, motors, and computers.

FINDING THE OPTIMAL MIX OF INPUTS

Energy can almost always be replaced in part by other inputs. For example, a steam pipe can be insulated more heavily or an industrial process can be modified to use more labor and less energy. Economic efficiency does not imply minimizing the use of energy or any other input, but rather finding the appropriate mix of inputs. The economically efficient level of inputs is reached when the last dollar spent on energy yields the same amount of benefits as the last dollar spent on labor or materials or any other input.

In discussions of economic efficiency, the concept of decreasing marginal rate of substitution plays a crucial role. This simply means that for a great many different activities, it becomes increasingly difficult to substitute one input for another as one continues to make substitutions. For example, a household with access to both electricity and natural gas probably will use gas for heating the house, electricity for lighting and refrigeration, and might choose to use either one for cooking, heating water, and drying clothes. It is possible to shift all of those activities to either energy source, although electric heat is expensive in most applications and gas refrigerators and lighting are rarely used when electricity is available. The substitutions become increasingly expensive as one moves to one extreme or the other. If the quality of the service offered by either fuel is identical, then the most efficient mix of energy sources would be the cheapest mix, which would depend on the relative prices of natural gas and electricity.

Firms face similar substitution possibilities in many of their activities and especially in industrial processes. Large commercial and industrial firms can even substitute electricity that they generate themselves for some or all of their purchased electricity. In the choice between purchasing and generating electricity, the two are perfect technical substitutes—that is, one can be substituted for the other without encountering a diminishing marginal rate of substitution. However, the cost of generation will increase as the firm tries to cover its own peak loads.

An investment in new equipment generally must be made before one energy source can be substituted for another. If new investment is required, the decision to switch fuels is not lightly made in response to fuel price fluctuations, particulary if they are viewed as temporary. In such cases, energy choices are most easily made when the activity is in the planning stage. For example, in some places new homes constructed in the late 1970s were not allowed to have connections to natural gas lines because the policy of the federal government was based on the assumption that gas reserves would soon be exhausted. New houses that were built with electric resistance heating systems required major investments before they could be converted to natural gas when the misguided policy was abandoned.

ADJUSTING FOR THE TIMING OF COSTS AND BENEFITS

In comparing the economic performance of energy alternatives, it is essential to take account of the times at which costs are incurred and benefits received. For example, in the case of a railroad line that crosses a range of mountains, it is possible to save fuel on every trip by tunneling under the mountains instead of traveling over them. Does this mean that the mountain crossings should be replaced by a tunnel? Suppose that the tunnel is on a lightly traveled route, costs $1 billion to construct, and will last indefinitely. If it saves $1 million per year in fuel and other operating costs on an ongoing basis, is the project economically efficient? Ignoring the opportunity cost of capital, the sum of the annual savings would eventually (after one thousand years) equal the cost of construction. But the resources devoted to its construction have an opportunity cost and the project is not cost-effective when the opportunity cost is taken into account.

Discounting makes it possible to compare costs incurred at one time with costs and benefits received at another taking into account the opportunity cost of capital. In the absence of such a procedure, one cannot compare alternatives that differ in the timing of their costs and benefits.

A dollar that will be received a year from today has a “present value” of $1 divided by (1+r), where r is the discount rate, which is equal to the opportunity cost of capital; and a dollar that will be received two years from today has a present value of $1 divided by (1+r)(1+r) or $(1+r)^2$. A payment that is to be received t years from today must be divided by $(1+r)^t$. If the opportunity cost of capital is fairly high, savings that will be realized many years from today will be heavily discounted. For example, if r is 10 percent, the present value of a dollar that will be received seven years from today is about 51 cents. If a dollar will be received twenty-five years from today, its present value is not even a dime. The total value of the tunnel that saves $1 million per year indefinitely is only $1 million divided by r. If the opportunity cost of capital is 10 percent, the tunnel is worth only $10 million. The economy will not prosper if it sinks $1 billion in building a tunnel that will generate only $10 million of benefits.

While this example is constructed to be an extreme case, it illustrates the importance of not being misled that a long-lasting stream of returns necessarily means that a capital investment will be profitable. Returns from energy savings to be received far in the future will have a low present value unless some mechanism works persistently to raise future energy prices at a rate that is commensurate with, or exceeds, the discount rate.

After the average crude oil price increased from $3.18 per barrel in 1970 to $21.59 in 1980, many analysts forecast skyrocketing energy prices for the remainder of the century. The “middle price path” of the U.S. Energy Information Administration in 1979 projected a nominal price of $117.50 per barrel in 1995! Such forecasts seemed to be soundly based not only in recent experience but also in the economic theory of exhaustible resources. As a consequence, U.S. industries invested heavily in energy conservation measures, with the result that industrial consumption of energy decreased from 31.5 quads in 1973 to 27.2 in 1985. Some of this investment was probably not warranted on economic efficiency grounds because prices ceased to rise after 1981, and even plummeted to $10 per barrel in 1986.

Whether the most energy-efficient equipment is also the most economically efficient depends on the circumstances. For example, a truck operator may be able to cut the fuel costs in half by spending $100,000 to replace an old truck with a new, energy-efficient model. But if a long-haul trucker spends $30,000 per year on fuel, the half saved is a significant amount. But if the truck is used mainly to shuttle containers between a port and nearby warehouses, the annual fuel bill might be $5,000 and the half saved ($2,500 per year) not enough to justify spending $100,000 for the new truck.

Similarly, an electric motor can use electricity that costs more than the motor during a year of continuous operation. Even if the motor is in perfect condition, it may be cost effective to replace it with a new motor that is a few percentage points more efficient at converting electricity into work. In many applications, however, an electric motor operates only a few hours per year. In such cases, the cost of the electricity is negligible relative to the cost of a new motor, so that even a large gain in energy efficiency is not worth the cost.

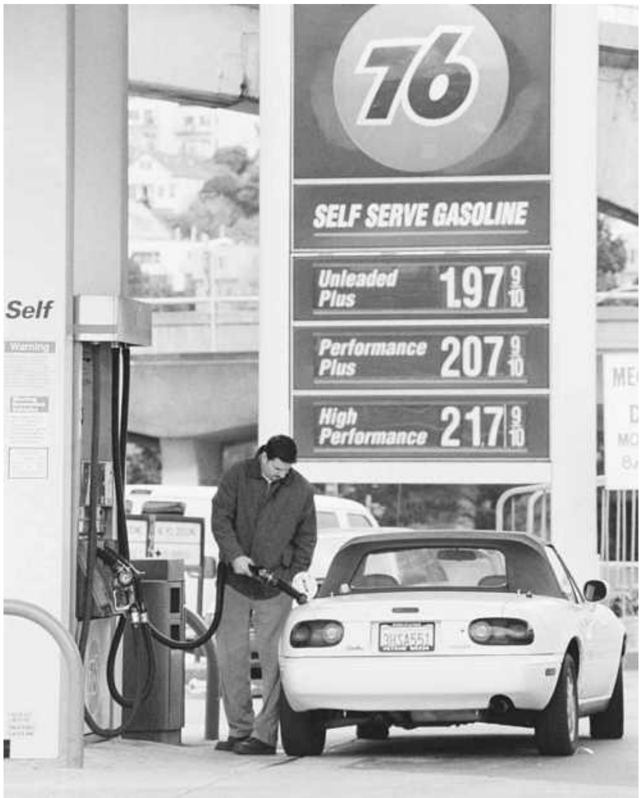
TAKING INTO ACCOUNT ALL INPUTS

As can be seen from the above examples, one characteristic of the concept of economic efficiency is that it takes account of all inputs, not just energy. Even if one were interested only in conserving energy, the economic approach would guide one to use labor, capital, and other inputs to conserve the greatest possible amount of energy for the budget. Of course, the economic approach is not usually associated with minimization of any one input or maximization of any one output, but rather with the minimization of costs for a given level of benefit or maximization of net benefits.

PRICES GUIDE DECISIONS

Because economic efficiency is simply a description of the rules by which individuals and firms can gain the most of what they want, another characteristic of economic efficiency is that firms and individuals do not need orders or special incentives to induce them to pursue it. They are simply acting in their own interest as they see it.

All of the information required for firms and individuals to pursue economic efficiency is conveyed by the price system. The price of natural gas relative to coal conveys information to all potential users of the two fuels about their relative scarcity. The price of the output relative to the inputs conveys information to potential producers about whether an activity will be profitable to expand. Decisionmakers can then pursue activities to the point where the benefit of expanding any activity or of any input in any activity is equal to the cost of that expansion. The price system is especially valuable because it conveys subtle information that is otherwise very difficult to ascertain or to factor into the analysis. Attempts to allocate particular inputs by rules or bureaucratic orders in wartime or in other controlled economies have invariably proved extremely inefficient, at best, and often disastrous.



A motorist fills his tank at a Union 76 service station where two out of three grades of gasoline cost more than $2 per gallon (March 7, 2000). West Coast fuel prices on the spot market shot up in reaction to the highest oil prices in nearly a decade. (Corbis-Bettmann)

PRICES MAY FAIL TO REFLECT EXTERNAL COSTS

This powerful effect of prices in conveying information throughout the economic system naturally leads to the question of whether the information conveyed about inputs and outputs generally, and about energy specifically, is accurate. Critics have indicated various ways in which energy prices can be misleading. One classic problem is that of external costs. For example, unregulated coal mining pollutes streams with acid drainage from underground mines and silt from unreclaimed surface mines. In such a world, the market price of coal, to which firms and individuals react in making their decisions, is too low because it does not include such damages. One possible solution to this problem is to assign ownership of the stream to someone (anyone), who would then charge polluters for the damage done. The cost and market price of coal would then incorporate the formerly external costs.

Generally, the United States has not followed this approach. Instead, regulations have been promulgated to specify either the production techniques that must be used to eliminate or lessen the external costs, or the permissible levels of emissions of pollutants. These regulations have helped to clean the environment and also have increased the cost of energy, but few economists would claim that the existing set of regulations leads to the same behavior and the same efficiency that a perfected set of prices would. One particular difficulty is that regulations rapidly become obsolete as technology and markets change, whereas prices adjust to changed circumstances and exert pressure for behavior to adjust accordingly.

Another characteristic of the economic-efficiency concept is that it does not require arbitrary decisions by the analyst about, for example, how coal should be evaluated compared with natural gas. The question of whether 1 Btu of coal is equal to 1, or perhaps 1/2, Btu of natural gas is answered directly by the market. The weightings of the marketplace, revealed in relative prices, vary with scarcity, cost of production, technology, and human preferences. Decisionmakers do not need to think about the underlying reasons, however. They need to know only current prices (and make their best guesses about future prices).

Prices can fail to reflect true social cost for reasons other than externalities. Factors such as taxes, subsidies, monopolies, and fear of expropriation also can cause prices to diverge from marginal social cost.

NET ENERGY ANALYSIS

The characteristics of economic efficiency noted above are considered advantages by most economists and disadvantages by a small group of critics, many with an ecological orientation. The most politically influential challenge to the concept of economic efficiency comes from "net energy analysis" (NEA). This type of analysis attempts to convert all inputs and outputs into weighted energy equivalents in the hope that the resulting project appraisals will be more stable and consistent than those provided by economic analysis. Arriving at weights for different forms of energy and energy equivalents for labor and capital has proved to be difficult and controversial. Moreover, some analysts question what use can be made of the results of NEA if they differ from those of economic analysis.

AREAS OF AGREEMENT AND DISAGREEMENT AMONG ECONOMISTS

Within the mainstream of economics, no serious challenges to the conventional analysis of economic efficiency have been sustained. Judgments differ on the extent to which market prices may need to be adjusted to compensate for externalities or other imperfections. For example, does global warming (presumed to result from emission of carbon dioxide and other greenhouse gases) pose a serious enough threat that the prices of all fuels containing carbon should be raised by imposing a tax equal to the amount of the damage done? If so, how much is that amount? Note that the disagreements about this issue reflect real gaps in knowledge about the effects of carbon dioxide emissions, not disagreement about the concepts and analysis.

Similarly, economists generally agree that analyses involving time require discounting according to the standard formulas, but disagree regarding which discount rate should be used. The basic issue is that using a high discount rate is equivalent to saying that benefits or costs that are expected far in the future do not receive much weight in decisions made today. Why worry about the costs of global warming if they will not be felt for a century or more? At any reasonable discount rate, the value of a dollar received a century from today is negligible. Some critics have argued that we owe something to future generations and therefore should value their preferences as highly as our own.

The supply of capital is limited, however. If an investment that yields a 2 percent return is adopted because it yields benefits to future generations, but if it consumes capital that could have yielded a 20 percent rate of return to an investor, the capital stock will grow more slowly. To maximize growth of physical capital and personal income, the highest-yielding investments should be chosen.

William S. Peirce

See also: Economic Externalities; Efficiency of Energy Use; Efficiency of Energy Use, Economic Concerns and; Efficiency of Energy Use, Labeling of; Energy Economics; Environmental Economics; Green Energy; Industry and Business, Energy as a Factor of Production in; Industry and Business, Productivity and Energy Efficiency in; Risk Assessment and Management; Subsidies and Energy Costs; Supply and Demand and Energy Prices; Taxation of Energy; True Energy Costs; Utility Planning.

BIBLIOGRAPHY

Gilliland, M. W. (1975). "Energy Analysis and Public Policy." *Science* 189:1051–1056.

Hayek, F. A. (1945). "The Use of Knowledge in Society." *American Economic Review* 35:519–530.

Huettner, D. A. (1976). "Net Energy Analysis: An Economic Assessment." *Science* 192:101–104.

Mikesell, R. F. (1977). *The Rate of Discount for Evaluating Public Projects.* AEI Studies 184. Washington, DC: American Enterprise Institute for Public Policy Research.

Peirce, W. S. (1996). *Economics of the Energy Industries.* Westport, CT: Praeger.

Pindyck, R. S., and Rubinfeld, D. L. (1998). *Microeconomics,* 4th ed. Upper Saddle River, NJ: Prentice-Hall.

U.S. Department of Energy, Energy Information Administration (1979). *Annual Report to Congress.* Washington, DC: U.S. Government Printing Office.

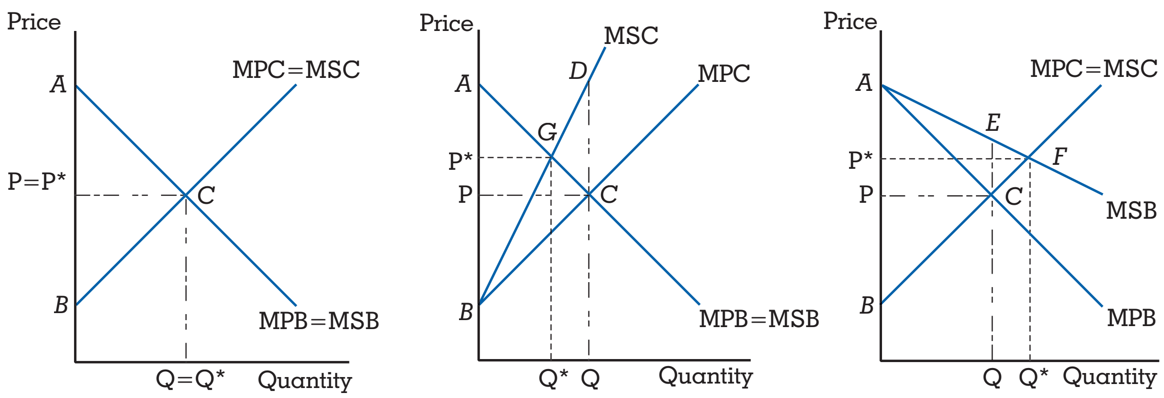
ECONOMIC EXTERNALTIES

An economic externality exists whenever the well-being of some individual is affected by the economic activities of others without particular attention to the welfare of that individual. For example, smog-related illnesses such as bronchitis and exacerbated cases of childhood asthma have been blamed, to some extent, on the emissions of nitrogen oxides from automobiles and large fossil-fuel-burning power plants. These illnesses have high treatment costs that are not incorporated in the related electricity-production and oil-consuming activities of the power plant and transportation industries, and must therefore be borne by the affected third parties. Air pollution sort is a classic example of an economic externality, and is called a negative externality because it has external costs.

Many environmental problems arise from externalities of energy exploration, production, refining, distribution, and consumption. This is especially so for fossil fuels. Air pollution, global warming and climate change, and acid rain are due mainly to emissions of carbon, sulfur, and nitrogen oxides associated with the burning of fossil fuels. Coastal and marine degradation, wildlife habitat destruction, and the availability and quality of fresh water can be blamed to some extent on oil spills, drilling for oil and gas, coal mining, and the underground storage of oil and gasoline. The nuclear power industry deals constantly with toxic-chemical and hazardous-waste issues. Some of these important environmental problems and others (e.g., deforestation and desertification) can also be attributed to the changing patterns of, and increases in, population, land use, transportation, and industry. Energy plays a significant role, and energy-environment externalities often have strong socioeconomic and environmental welfare effects.

Market forces determine much of energy production and use. Associated externalities are often beyond the capacity of the market to resolve. To understand how energy externalities impose costs on society, one must first understand how markets allocate resources (including energy) efficiently. Figure 1a shows a typical demand and supply diagram for a commodity (e.g., coal) or service. For many goods, the demand curve reflects marginal private benefit (MPB) and the supply curve reflects marginal private cost (MPC), since commodities usually are produced and consumed privately. The demand (or marginal benefit) curve is downward-sloping to reflect the fact that people will pay less for additional units of a good as they consume more of it. The demand curve also shows people's willingness to pay for a good, and so the downward slope means that as the price of the good decreases, people are willing to buy more of it. Thus the demand curve shows the amount of a good that is demanded at each price. Similarly, the supply (or marginal cost) curve shows the amount that is produced at each price. The upward slope of the supply curve reflects increasing costs of production, and also that producers are willing to supply more at higher prices.

Market Equilibrium and External Effects



(a) Market Equilibrium (b) Negative Externality (c) Positive Externality

Figure 1.

The areas under the curves represent benefits from consuming, and costs of producing, the commodity. These benefits and costs increase as more of the good is consumed or produced. Benefits are higher than costs up to the point where MPB equals MPC, and thereafter costs are higher. Therefore net private benefits are maximized when MPB equals MPC, and Q units of the good are demanded and supplied at a price of P. The area enclosed by triangle ABC in Figure 1a represents maximum net private benefits. Social net benefits are maximized when MPB and MPC are identical to marginal social benefit (MSB) and marginal social cost (MSC), respectively. Markets efficiently allocate resources to achieve this outcome, and there is market failure whenever divergence exists between MPC and MSC, and/or between MPB and MSB. Market failure is caused by many factors, including:

- externalities;
- imperfect markets—when markets are not competitive;
- incomplete markets—when property rights are not well defined to enable exchange;
- public goods—goods that are indivisible and may be free to some consumers;
- imperfect information—when costs and benefits are not fully known by all;
- nonconvexities—when MSC is shaped so that it crosses MSB at several points.

Social net benefits must be used in considering how energy externalities impose costs on society. In Figure 1, private market forces promote production and consumption of Q units at a price of P, and social net benefits are maximized when MSB equals MSC with production of Q* units at a price of P*. In Figures 1b and 1c, triangles ABG and ABF represent maximum attainable social net benefits, respectively. Market production and consumption of Q units provide social net benefits equal to area ABG less area GCD in Figure 1b, and area ABF less area ECF in Figure 1c. Thus externalities impose costs on society by making it impossible to gain maximum social net benefits. In Figure 1b we can see that a negative externality such as energy-related pollution implies private market production of too much energy and pollution. Similarly, Figure 1c shows that a positive externality such as plant growth enhancement by carbon dioxide emissions from the burning of fossil fuels implies that too little energy and positive externalities are produced. In both cases, the market price for energy is too low. In reality, Figure 1b best represents the case of energy externalities, since the pollution effects outweigh the plant growth enhancement effects (i.e., the resultant external effect of energy production and use is negative).

The solution to the problems posed by energy externalities is to internalize the externality, so that the external costs and/or benefits are included in the transactions and other activities involved in the production and consumption of energy. For example, public policy that uses a tax to raise energy prices and/or restrict energy production to socially desirable levels would solve energy externality problems. Energy is an important factor of production, and any policy that affects energy price or quantity ultimately affects the entire economy.

The problems that energy externalities present are further complicated by the nature of energy pollution, other market failure issues, and inappropriate government intervention. Internalizing the externalities of air pollution and other emissions is sound in theory, but in practice, quantifying pollutants and their impacts and equitably dealing with the problem is very difficult. Energy pollutants are emitted from both stationary and mobile sources and may accumulate in the environment. Stationary sources (e.g., power plants) are generally large and few, are run by professional managers, and provide simple local pollution patterns. Mobile sources (e.g., cars) are abundant, are and run mainly by individuals, and complicate local pollution patterns. Pollutants that accumulate in the environment can affect future generations, and introduce intergenerational equity problems and issues. For example, carbon dioxide emissions in excess of the absorptive capacity of the environment can accumulate. This is because carbon dioxide is not a true pollutant because it is essential for plant life and is absorbed by plants and the oceans. Although carbon dioxide is the main greenhouse gas blamed for global warming, there is considerable scientific uncertainty regarding global warming and climate change, and there are critical measurement issues in determining both market and external costs and benefits.

The impact of energy pollutants on the environment can be local, regional, or global. As the zone of influence of pollutants extends beyond local boundaries, the political difficulties of adopting and implementing control measures are compounded. For example, sulfur oxides are regional pollutants, while excess carbon dioxide is a global pollutant. Carbon dioxide pollution policy thus requires international cooperation, but sulfur oxide policies may require only national policy. The United States has a program for trading sulfur emissions, and Japan taxes sulfur oxides, but there seems to be little progress with international attempts to control carbon dioxide. Developing nations fear that participating in carbon dioxide emission reduction programs will retard their economic development. This introduces international income distribution issues, but in many cases intranational income distribution issues also must be addressed. In short, the nature of energy pollutants, uncertainty and measurement issues, income distribution effects, intergenerational equity, economic development, and political difficulties make public policy regarding energy externalities very challenging.

Samuel N. Addy

See also: Air Pollution; Climatic Effects; Environmental Economics; Environmental Problems and Energy Use; Ethical and Moral Aspects of Energy Use; Government and the Energy Marketplace; Historical Perspectives and Social Consequences; Industry and Business, Energy as a Factor of Production in; Market Imperfections; Subsidies and Energy Costs; Supply and Demand and Energy Prices; True Energy Costs.

BIBLIOGRAPHY

Baumol, W. J., and Oates, W. E. (1988). *The Theory of Environmental Policy*. Cambridge, Eng.: Cambridge University Press.

Folmer, H.; Gabel, L. H.; and Opschoor, H., eds. (1995). *Principles of Environmental and Resource Economics*. Cheltenham, Eng.: Edward Elgar.

Goldemberg, J. (1996). *Energy, Environment, and Development*. London: Earthscan.

Kahn, J. R. (1998). *The Economic Approach to Environmental and Natural Resources*, 2nd ed. Orlando, FL: Dryden Press.

Landsberg, H. H., ed. (1993). *Making National Energy Policy*. Washington, DC: Resources for the Future.

Portney, P. R., ed. (1990). *Public Policies for Environmental Protection*. Washington, DC: Resources for the Future.

Tietenberg, T. H. (1996). *Environmental and Natural Resource Economics*, 4th ed. New York: HarperCollins.

ECONOMIC GROWTH AND ENERGY CONSUMPTION

Energy is a vital ingredient to economic growth. This has been recognized at least as long as economic statistics have been compiled by government, and probably for much longer than that. Perhaps the best example of the fundamental role that energy plays in large, complex national economies is found in the

1973–1974 oil embargo, when oil-producing nations of the Middle East restricted supply and prices rose fourfold in a space of a few months. The resulting chaos in the oil-consuming economies of the industrialized West was widely considered to be a direct result of the embargo. In the United States alone, Gross Domestic Product—an accepted measure of economic activity—fell in 1974, after two decades of steady growth. The high cost and scarcity of oil was seen as the primary cause.

ENERGY AS AN INPUT TO PRODUCTION

What makes energy and economic growth go hand-in-hand? Traditionally, economists since Adam Smith have discussed the major inputs to economic activity as being land, labor, and capital. While very descriptive of the agrarian economies of the seventeenth and eighteenth centuries, the growth of industrial nations in the nineteenth century can be seen in retrospect to have been the result of a fourth major input, energy. Energy can be seen simply as the ability to multiply the work of laborers exponentially. Where the agrarian society had to make use of horses and mules for transportation services, the industrial economy could take advantage of the miracle of the internal combustion engine, which, when powered by gasoline, could lower the costs and increase the availability of transportation by orders of magnitude. Where once laborers did their jobs with scythes, shovels, and other tools, energy enabled them to increase their outputs tremendously by powering great machines such as tractors, cranes, and pile drivers. Power for illumination allowed the growth of multiple "shifts," greatly increasing the output that could be produced over a given period of time.

ENERGY INTENSITY

The ratio of energy consumption to economic activity is referred to as the "energy intensity" of an economy. Energy intensity may also be measured at lower levels of aggregation, such as at the industrial or transportation sectors of an economy. In general, as nations move into a more industrialized state, they find that their energy intensity greatly increases, as the demands of a more complex economy require a greater amount of energy per unit of output. Another way to think about the role of energy is that the ability to harness it technologically allows an economy to greatly increase its economic potential (e.g., to expand its production possibilities frontier). An example of such growth might be that of the United States during the late nineteenth and early twentieth centuries, as it began the transformation that would make it the world's leading industrial power, powered by its then-abundant supplies of petroleum. Another example would be China during the two decades starting in the late 1970s, as it moved away from a totally state-controlled economy to a partially market-driven system. During this period, China experienced one of the highest rates of growth in the world, driven largely by its ability to harness its huge coal reserves in the production of electricity and for transportation services.

1997 Rank	*Country*	*Energy Intensity (Thousand Btu per 1990 U.S. Dollar of Gross Domestic Product)*	
		1980	*1997*
1	United States	17.77	13.84
2	Japan	7.78	6.36
3	Germany	9.37	7.60
4	France	8.94	7.42
5	Italy	7.02	6.48
6	United Kingdom	11.70	9.07
7	China	109.63	45.53
8	Canada	22.05	18.59
9	Brazil	10.73	13.26
10	Spain	8.82	8.06

Table 1.
Ten Largest World Economies' Energy Intensity, 1980–1997

THE EFFECT OF ECONOMIC MATURITY ON ENERGY USE

As nations become more economically mature, two effects are typically seen. One, the rate of economic growth necessarily slows, as the base of economic activity expands and opportunities for easy expansion become more scarce. Two, the use of energy becomes more efficient as consumers and manufacturers become more knowledgeable about its use, and technological progress enables economic output to be produced with less energy input.

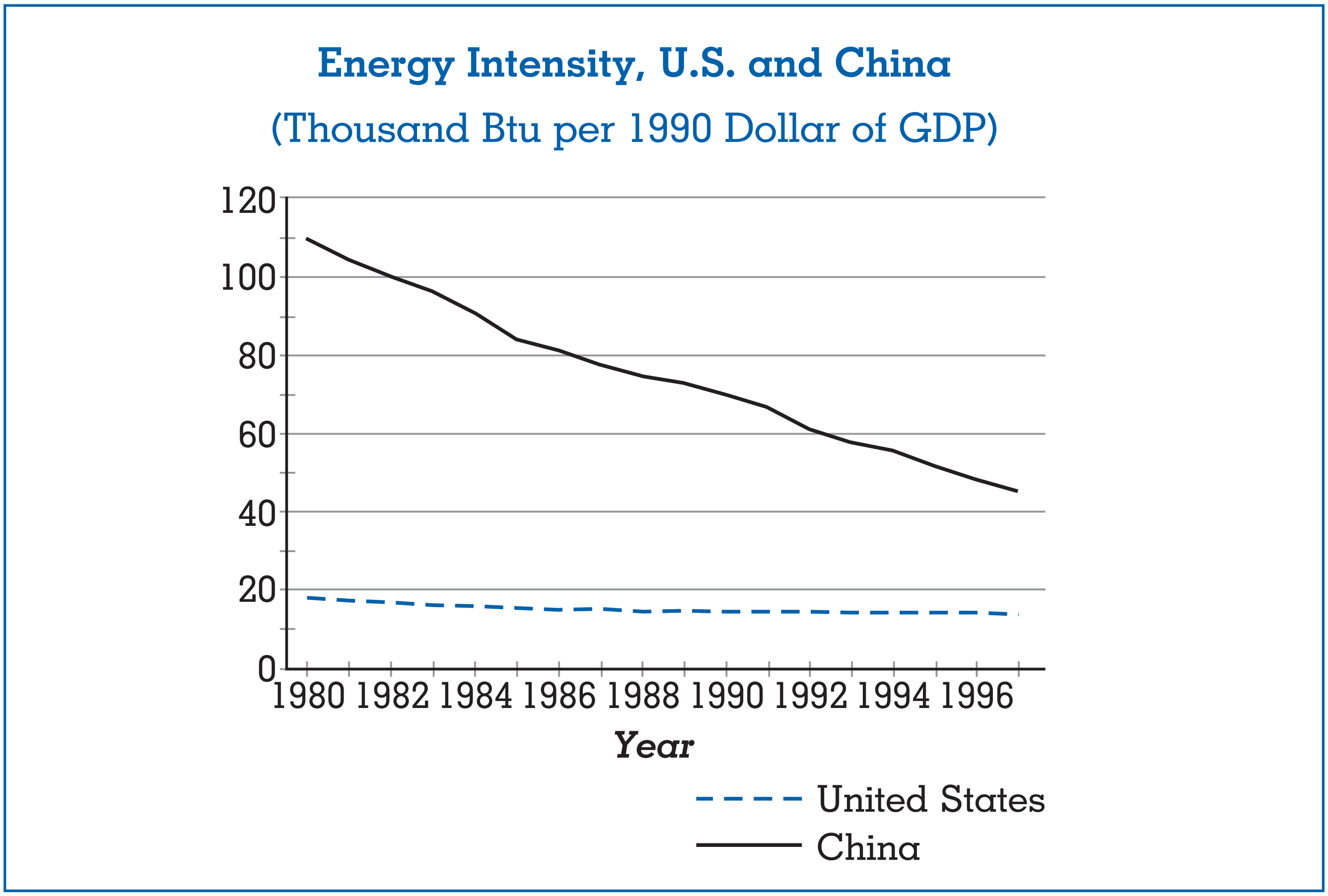


Figure 1.

The post–World War II experience of North America and Western Europe provides an excellent example of both of these effects. Between 1949 and 1973, the year of the oil embargo, the United States expanded its output of goods and services at an annual rate of 4.1 percent; Great Britain showed a growth rate of 3.0 percent. The oil embargo caused a major structural shift in these energy-consuming nations. It took more than a decade for these countries to return to the economic growth rates they had enjoyed before 1973. Not until after 1983 did Great Britain return to consistent economic growth of 3 percent or more. In the United States, the post-embargo period was characterized by economic growth rates that were more moderate than the pre-embargo period, especially after a second round of oil price increases in 1979–1980. But the second effect also became evident, as energy efficiency, stimulated by the extraordinary rise in energy prices, became a hallmark of the industrialized nations. In the United States, energy intensity fell more than 30 percent between 1970 and 1986; in France, the decline was 21 percent between 1970 and 1990.

FACTORS DECREASING ENERGY INTENSITY

More efficient automobiles were large contributors to the decline in overall energy intensity, as both consumers and government regulators took steps to increase the efficiency of cars by making them lighter, smaller, and equipped with more efficient engines. Homeowners also contributed to the reduction in energy intensity, by lowering thermostats, using more efficient lighting, and purchasing more efficient furnaces, air conditioners, and electric appliances. Some of these behaviors were mandated by laws and regulations forcing manufacturers to market appliances that consumed less energy, or to label their

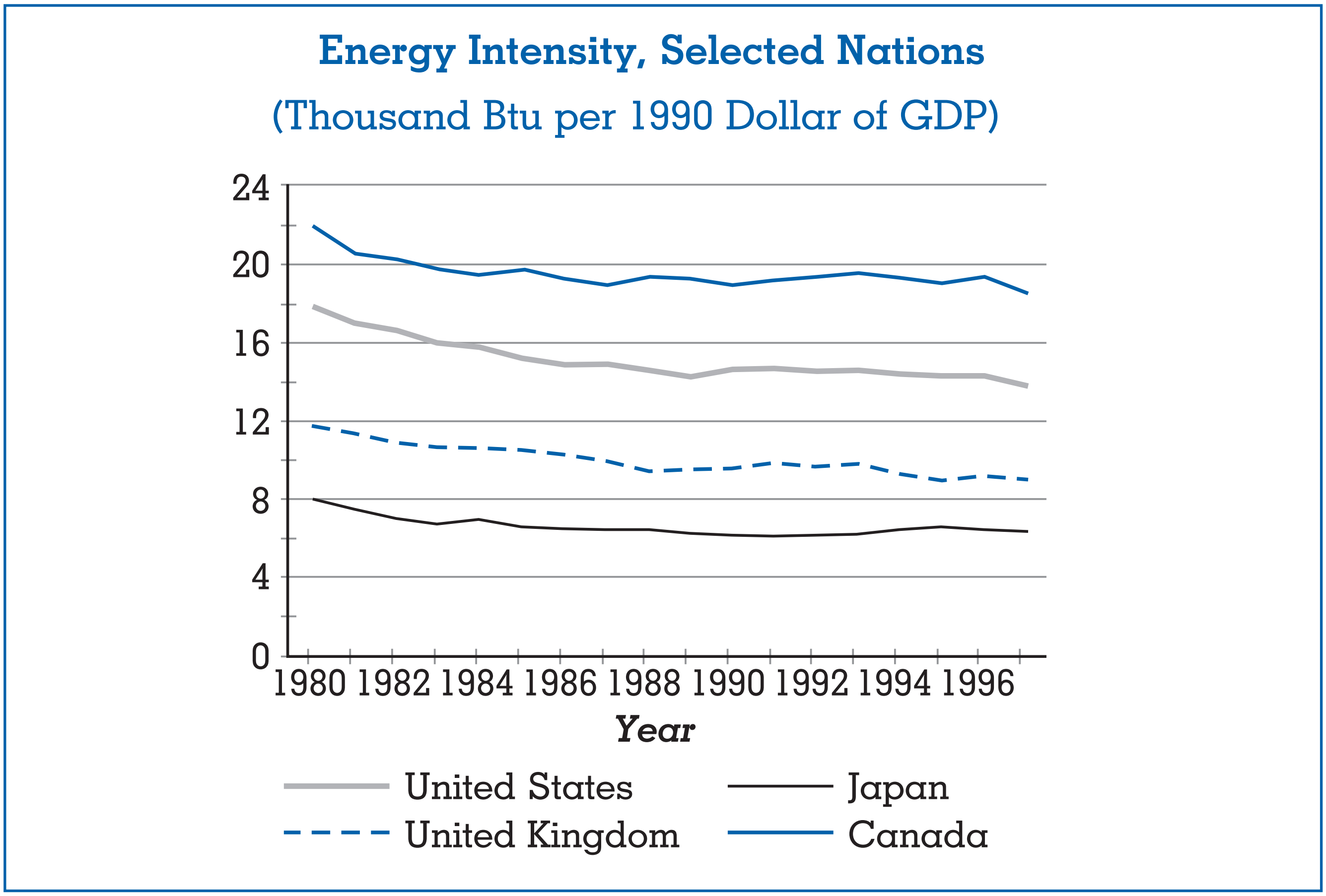


Figure 2.

average energy consumption. But most were the result simply of supply and demand stimulating long-term changes in habits that reduced the consumption of a high-priced economic input, energy, in favor of other, lower-priced inputs.

THE WORLD'S LARGEST ENERGY USERS

Past patterns of economic growth shape current patterns of energy consumption. The largest energy-consuming nations tend to be those whose economies are also the largest, such as the United States, Japan, and Western Europe. In 1997, the world's ten largest economies—ranging from the United States to Spain, and representing more than three quarters of the world's gross domestic output—also accounted for 58 percent of the world's energy consumption (Table 1). The energy intensities of these nations varied widely, however, ranging from less than 6,400 British thermal units (Btu) per 1990 U.S. dollar of Gross Domestic Product (GDP) in Japan, the most energy efficient of the world's largest economies, to more than 45,000 Btu per GDP dollar in China (Figures 1 and 2). In large measure, these variations reflected the progress of each nation in reaching economic maturity. It also reflected the relative abundance of energy resources within these countries. Japan, for instance, has very little indigenous energy supplies, and is highly dependent on energy imports. China, on the other hand, is both a developing country and one with large domestic reserves of coal. Japan, along with many of the world's consuming nations, has also discouraged consumption by taxation; in 1998, regular leaded gasoline in Japan cost approximately $2.80 per gallon, of which about $1.70 was taxes. By contrast, the same gallon cost about $1.05 in the United States, only about $0.35 of which was due to taxes.

THE ENVIRONMENTAL IMPLICATIONS OF ECONOMIC GROWTH AND INCREASING ENERGY CONSUMPTION

Economic growth and increasing energy consumption are not always considered an unalloyed benefit. There are significant environmental consequences to energy consumption, including increased concentration of carbon gases in the atmosphere, emissions of sulfur dioxide (that cause acid rain) and nitrogen oxides (precursors to smog), water pollution caused by oil spills, and land issues related to coal mining and other energy production. Debates about how to ameliorate these effects inevitably include discussions of the economic impacts of such amelioration, and the effect on economic well-being of higher energy taxes or outright bans on consumption or production of certain kinds of energy. Non-polluting energy sources such as hydroelectricity, solar and wind energy, or, more controversially, nuclear-based electricity-for which there are considerable concerns about safety and waste disposal—have been discussed as long-term alternatives to the more traditional fossil-based fuels. While coal, oil, and natural gas have clearly been indispensable to the growth of today's modern industrial economies, it is also a certainty that the supplies of these energy forms are ultimately limited, and that future economic growth will depend on a long-term transition to other energy sources.

OUTLOOK

Some analysts have posited that, given the wide variation in energy intensities of the world's economies, we can look forward to a time when economic growth and energy consumption are effectively uncoupled. Moreover, this uncoupling can make it possible to reduce the environmental consequences of economic growth at a relatively low cost, because such growth will not mean an inevitable rise in the consumption of fossil fuels. Certainly the success of Japan and Western Europe in achieving low energy intensities is evidence that further reduction is possible in other parts of the world.

On the other hand, some of the differences in energy intensities reflects infrastructure that may be difficult to change. Some nations, such as the United States and Canada, are large energy consumers because their climate requires above-average heating and cooling, and their vast size makes transportation a more difficult and energy-intensive activity. China, which has similar characteristics with respect to climate and size, has the additional burden of being a largely agrarian, labor-intensive economy that still uses energy in relatively inefficient ways because it has not yet achieved the technological progress characteristic of the developed countries. Until more efficient means of heating, cooling, and transportation are found, and greater progress is made toward use of more efficient technologies in the developing countries, economic growth will likely continue to require additional energy consumption, but perhaps less so than in the past.

The possibilities of new sources of energy, such as energy from hydrogen, may also some day become economical, and help to uncouple fossil fuel consumption from economic growth

Scott B. Sitzer

See also: Energy Intensity Trends; Environmental Problems and Energy Use; Industry and Business, Energy as a Factor of Production in.

BIBLIOGRAPHY

Clark, J. G. (1991). *The Political Economy of World Energy.* Chapel Hill and London: University of North Carolina Press.

International Energy Agency. (1994). *Energy in Developing Countries.* Paris: OECD Publications.

International Energy Agency. (1999). *Energy Prices and Taxes: Quarterly Statistics, Second Quarter 1999.* Paris: OECD Publications.

Jorgenson, D. W. (1978). "The Role of Energy in the U.S. Economy." *National Tax Journal* 31(3):209–220.

Jorgenson, D. W. (1998). *Energy, the Environment, and Economic Growth.* Cambridge, MA: MIT Press.

U.S. Department of Energy, Energy Information Administration. (1998). *International Energy Annual 1996.* Washington, DC: Author.

EDISON, THOMAS ALVA (1947–1931)

Thomas Alva Edison, is the archetype of American ingenuity and inventiveness. He played a critical role in the early commercialization of electric power. He designed the first commercial incandescent electric light and power system and his laboratory produced the phonograph, a practical incandescent lamp, a revolutionary electric generator, key elements of motion-picture apparatus, and many other devices. He was owner or co-owner of a record 1,093 U.S. patents

"Tom" Edison was born in the small town of Milan, Ohio, the son of middle-class parents. He was educated at home by his parents rather than at the local school, where he was thought to be of low intelligence. As a boy he showed an early proclivity for chemistry experiments and for turning a profit, first peddling vegetables, then newspapers. When the Civil War began, Edison was exempted from service because of deafness in one ear; he became a telegrapher, and one of the fastest operators in the corps of generally brash, swaggering men who ran the railroad telegraph systems. He drifted around the country, winding up in New York in 1869, searching for a job. During these years he tinkered with some contrivances, but these had led nowhere.

Then, hanging around looking for a job in a New York brokerage office during one frenzied day of trading, he jumped in to fix the telegraphic stock ticker, which had broken down. He was hired on the spot at a large salary but stayed only a few months, leaving to form a company devoted to the business of invention. The company's first product, an improved stock ticker, was sold to his previous employer for the astounding sum of forty thousand dollars, and Edison set up operation with a staff of fifty men in Newark, New Jersey.

During its first years, Edison's company devoted itself to the manufacture of stock tickers and to improvements in telegraphic equipment but then spread into other areas, offering to provide inventions as ordered. In 1876 the laboratory moved to Menlo Park, New Jersey, and the staff eventually grew to about one hundred. Over the years, a number of men who worked with Edison at Menlo Park achieved fame as inventors and scientists in their own right. This group included Nikola Tesla, the inventor of the alternating current electric induction motor; John Fleming, the inventor of the vacuum tube diode; William Dickson, the inventor of the first sound movie; Arthur Kennelly, a discoverer of the ionosphere; and Edward Acheson, the inventor of carborundum.

During the following years a brilliant series of inventions at his laboratory earned Edison the appellation "Wizard of Menlo Park." As early as 1878, the mere announcement that Edison intended to produce a practical electric light was sufficient to cause the price of gas illumination stock to fall sharply.

Some of Edison's commercial inventions were produced solely to break the monopolies of patents already granted. Many others represented improvements or changes of known devices; these included Edison's electric light and dynamo and his quadraplex telegraph and improved telephone transmitter. This does not detract from the importance of his work, because in the cases of the electric light and dynamo, in particular, his work led to commercially practical devices that were widely adopted. Although some inventions, such his motion picture apparatuses, were not the result of his work alone, but the result of the joint efforts of the staff of the laboratory, Edison's contribution as leader in these projects cannot be ignored.

One case in particular vividly delineates Edison's own individual original genius. In 1877 he accidentally discovered that he could obtain an audible sound from a mechanical arm touching a rapidly rotating tin disk he had inscribed with a spiral series of dots and dashes representing telegraph signals. He then sketched and gave his machinist for construction a drawing of the first phonograph. The device worked at first try. There had never before been even a description of a machine for recording and replaying sound.

Edison's greatest contribution was the design of an electric power distribution system to provide power to factories and homes. He challenged the commonly held but mistaken idea that an unavoidable loss of half the power would occur in the generator and designed a power station, a revolutionary electric generator, a system of radiating power lines to consumers and electric meters to measure consumption. In 1882 he opened the first electric power station, on Pearl Street in lower Manhattan. He correctly calculated that he would have to manufacture bulbs for less than forty cents to make a profit; he achieved this goal in the fourth year of operation. After he had recouped his initial losses, he sold out of the business to support his other activities.

Edison often portrayed himself as a tough businessman whose sole interest was the profit to be made from an invention. However, he was not an outstanding businessman, and he made a number of obvious blunders during his career. He led a bitter, losing fight against the adoption of ac power distribution; he did little to keep some brilliant assistants; he made financial miscalculations; and he seemingly did not understand marketing and failed to meet his customers' desires. Nor was Edison a man of science; as a matter of fact he showed little interest in scientific matters. Edison was interested in the creation of new technology, and the close connection of technology and science was not as evident in the nineteenth century as it is now. He later admitted that while conducting his experiments he had had no understanding at all of the nature of electric current. Indeed, his own reminiscences make it clear that his real motivation had not come from the desire for profit or knowledge but came instead from the creative urge to invent. That urge drove Edison for the rest of his life. Although he produced relatively little after the turn of the century, and lived to be an old man, he never completely stopped working.

Leonard S. Taylor

See also: Electricity; Electric Power, Generation of; Electric Power Substations; Electric Power Transmission and Distribution Systems; Lighting; Tesla, Nikola.

BIBLIOGRAPHY

Brittain, J. E. (1977). *Turning Points in American Electrical History*. New York: IEEE Press.

Israel, P. (1998). *Edison: A Life of Invention*. New York: John Wiley & Sons.

Norman, D. A. (1999). "Backing the Right Technical Horse." *IEEE Spectrum* 36(2):5–62.

Wilson, M. (1954). *American Science and Invention*. New York: Bonanza Press.

EFFICIENCY OF ENERGY USE

Although energy efficiency was already heavily emphasized in the 1970s as a key strategy for energy security, more recently it has also been recognized as one of the most cost-effective strategies for reducing environmentally harmful emissions. Energy efficiency is more than just a resource option such as choosing between coal, oil, or natural gas. It curbs demand rather than increasing supply, and thus provides additional economic value by preserving the resource base and reducing pollution.

For specific applications, we can calculate the ratio of the measure of the goods or services provided to the energy input required. For example, in the transportation sector, energy efficiency is based on miles per gallon for personal vehicles, seat-miles per gallon for mass transit, and ton-miles per gallon for freight transportation.

For the entire economy, with its countless services and inputs, economists usually define the "service" or economics efficiency as the entire GDP (Gross Domestic Product) divided by E, the annual total primary energy used: Economic Efficiency = GDP/E.

Economists also track the reciprocal, E/GDP, which is called energy intensity. For example, the energy intensity of the United States in 1998 was 91 quads/$8.5 trillion in 1996 dollars (1996$), which divides out to be 10,700 Btu/$. (Note: 1 "quad" = 1 Q = 1 quadrillion Btu = 10^{15} Btu.)

Measuring energy efficiency gains for the entire economy is not a precise science since the population continues to expand, new technologies continue to be introduced, and there is great variability in the behavior of individuals using technology. Nevertheless, trends in economic effiency and energy intensity best reflect the impact of energy efficiency improvements.

ORIGINS OF ENERGY EFFICIENCY

The increased availability of energy fueled the Industrial Revolution. The United States became the world's largest oil producer, and the new fossil fuels were abundant and modestly priced. A technology's energy efficiency was not a key part of capital investment decisions. Energy-efficient technology as a priority ranked well behind improved performance.

Energy intensity declined from 60,000 Btu/$ (1992$) in the 1850s to 13,000 units in 1995. There were rapid drops in the 1860s as the switch was made from wood to more efficient coal, and, starting in the 1920s, as the switch was made to even more efficient oil and gasoline. The 1973 OPEC oil embargo and the next eleven years of rising energy prices triggered the final dip. The overall drop by a factor of 4.6 in 145 years corresponds to a steady annualized efficiency gain of 1.1 percent.

Before the OPEC embargo, there was no Department of Energy, and energy efficiency was not considered to be a government responsibility. Other aspects of energy were understood to be appropriate for government support. For example, research and development (R&D) on futuristic power supply technologies such as fission nd fusion was funded by the Atomic Energy Commission. From fixed year (FY)1948 through FY1972, in 1999 constant dollars, the federal government spent about $22.4 billion for nuclear (fission and fusion) energy R&D and about $5.1 billion for fossil energy R&D. The government also had a role in electrification as an economic development strategy. The entire rural electrification effort, including the federally subsidized Power Marketing Administration is still a major government program today. But it took an OPEC embargo to convince Americans to create a Department of Energy (DOE) in 1974 and to use public funds for efficiency research and development.

RECENT TRENDS IN U.S. ENERGY INTENSITY

The relative lack of importance of energy prices changed dramatically with the OPEC oil embargo. Even though energy prices were still a small fraction of total costs, people and businesses began to make energy-efficient capital-investment decisions in expectation of higher prices.

From 1974 through 1992, Congress established several complementary energy-efficiency and energy-conservation programs. By the 1980s, this concern over finite resources had dissipated as higher prices encouraged greater innovation in efficiency and in resource recovery. Energy efficiency had become a cost-saving, "demand-side management" tool that helped to avoid expensive power plant construction. The DOE's 1995 report, *Energy Conservation Trends,* states that energy efficiency and conservation activities from 1973 through 1991 curbed the pre-1973 growth trend in primary energy use by about 18 Q, an 18 percent reduction.

By the late 1980s, concerns over air pollution began to play a role in the government rationale for energy efficiency, which in the 1990s was followed by concern over global warming. Neither concern had a strong impact on energy use. Since 1985, national energy use has climbed about 20 Q, reaching a record high of 92 Q in 1999. From FY1980 through FY1999, the DOE spent $7 billion on energy efficiency R&D, which accounted for about 10 percent of all energy supply R&D.

E/GDP experienced a steep decline during the eleven OPEC years (1974 through 1985) and a recent equally steep decline starting in 1997. The latter drops may be associated with the rapid growth of the U.S. economy and with the explosive growth in information technology and the Internet. From 1960 through 1973, energy prices were low and there was almost no improvement in E/GDP. Similarly, after the collapse of OPEC in 1985, prices were again low, energy policy wavered, and E/GDP leveled off. The overall drop from 18 in 1973 (the year before the embargo) to 10.5 in 1998 is a drop of 57 percent, and corresponds to a steady gain of 2.2 percent/year for 26 successive years.

Thus, improved energy efficiency can be credited with energy savings of $232 billion in 1999. The arithmetic is as follows. A drop to 57 percent corresponds to a savings of 43 percent; but roughly one-third of the gain (about 15%) came from structural change as we switched from a smokestack to a service economy, and only two-thirds (about 30%) came from a true increase in efficiency. So pure efficiency has reduced our energy intensity only to 70 percent (not 57%). Had our efficiency stayed frozen at its 1973 value, we would now use more energy by the factor 1/0.7 = 1.43, that is, we would use 43 percent more energy for every dollar of GNP. Our 1999 energy bill was $540 billion, and would have been larger by 43 percent, which is $232 billion. The fraction that can be attributed to government intervention in the marketplace is highly debatable. Clearly, the marketplace would have made the energy efficiency improvements anyway; yet, a considerable portion would not have taken place without government intervention.

It is interesting to compare this huge annual saving of $232 billion with two other 1999 expenditures. The total non-military discretionary federal budget was $300 billion; therefore, efficiency savings pay for three-fourths of our entire civilian discretionary budget. Efficiency savings also equate to a large percentage of the U.S. Social Security budget, which in 1999 was $392 billion.

As technology develops steadily, there follows a corresponding decline of E/GDP, averaging about 1 percent/year. This can be accelerated in the marketplace by new fuels, new technologies, and innovations in existing technology. Government intervention—such as efficiency labels, performance standards for buildings and equipment, tax incentives, utility policy, and voluntary agreements with industry—which is usually implemented during periods of rising energy prices—can further accelerate the decline in E/GDP.

INTERNATIONAL COMPARISONS OF ENERGY INTENSITY

The E/GDP for the United States has sloped steadily downward from 18,000 to 11,000 Btu/$. Europe and Japan are typically only half as energy-intensive as the United States. An explanation is that, during their development, Western Europe and Japan were petroleum-poor compared to the United States, so energy use was perceived to imply imports (and risk of supply disruption) and trade deficits. Thus, they adopted tax policies to conserve energy. The United States took the opposite path; to stimulate economic growth, domestic oil and gas production was subsidized.

Among E/GDPs for developing countries or regions, the most notable is that of China, reaching 110,000 to 120,000 Btu/$ until 1976, but then declining steadily 5.2 percent/year for 21 years to 40,000 in 1997. This two-thirds drop shows the striking potential savings for other developing economies. The former Soviet Union (FSU) had a steady but inefficient economy until 1989. After that the FSU's rise in E/GDP is mainly because of the collapse of GDP. Eastern Europe comes next. It started off indistinguishably from the Soviet Union, made small improvements until 1989, and then rapid improvements as it adopted market economies. It is expected that the FSU curve will also soon turn down as its GDP picks up. Although India's efficiency trends are not currently in the right direction, developing countries, and particularly the FSU, have a high potential for cost-effective efficiency gains. Well below India come the industrialized countries, with the United States at the top and Japan at the bottom.

ENERGY SAVINGS IN THE BUILDINGS SECTOR

Energy is used in buildings to provide a variety of services such as lighting, space conditioning, refrigeration, hot water, and electronics. In the United States, building energy consumption accounts for slightly more than one-third of total primary energy consumption. Percentages reported for energy consumption and related carbon emissions in all four sectors are based on the Energy Information Administration's *Annual Energy Outlook 2000,* DOE/EIA-0383(2000), December 1999. By 2010, significant changes are expected to occur that will affect how buildings are constructed, the materials and systems used to build them, and the way in which buildings are maintained and used. A wide array of technologies can reduce energy use in residential and commercial buildings. Using sensors and controls to better manage building energy use, and improving building design and construction materials to maximize the thermal resistance of the building shell can also significantly reduce building energy requirements. For example, a cool white roof can reduce air-conditioning energy use by 20 percent.

Appliances have shown very dramatic improvements in energy efficiency, and perhaps the most impressive efficiency gains have come in improving refrigerators. In what follows, we show that refrigerators' efficiency gains are due to the interplay of regulatory and technological advancement. Two energy regulatory innovations (appliance labels, soon followed by standards) and a major technological innovation (blown-in foam insulation) led to the change from an annual energy use growth of 7 percent/year to a drop of 5 percent/year.

In the 27 years between the 1974 peak annual usage of 1,800 kWh and the 2001 federal standard of 450 kWh, refrigerator energy use dropped to one-quarter of its former use, even as the average volume grew from 18 cu. ft. to 20 cu. ft. This corresponds to a compound annual efficiency gain of 5.1 percent. As for economic savings, by the time 150 million refrigerators have reached year-2001 efficiency, compared to 1974, they will save 200 billion kWh/year, which corresponds to the output of forty huge (1 GW) power plants, and to one-third of the nuclear electricity supplied last year in the United States. Consumers will save annually $16 billion annually in electric bills, but their net savings will be only $10–11 billion, because there is a cost premium for the improved refrigerator (typically repaid by bill savings in three years). This $16 billion annual electricity saving from refrigerators alone roughly matches the entire $17 billion wholesale annual value of all United States nuclear electricity.

This surprising equality arises because an efficient appliance saves "expensive" electricity at the meter, at an average retail price of 8 cents/kWh; whereas one kWh of new wholesale supply is worth only 2–3 cents at the power plant. Thus, even if electricity from some future new remote power plant is "too cheap to meter," it still must be transmitted, distributed, and managed for 5–6 cents/kWh. It is impossible to disentangle the contribution of standards and of accelerated improvement in technology, but clearly the combination has served society well.

ENERGY SAVINGS IN THE INDUSTRIAL SECTOR

The industrial sector is extraordinarily complex and heterogeneous. It includes all manufacturing, as well as agriculture, mining, and construction. In the United States, industrial energy consumption accounts for slightly more than one-third of total primary energy consumption. Recent data show non-manufacturing industries such as agriculture and construction have maintained their energy use growth rate while that for manufacturing has dropped.

Still, the manufacturing sub-sector accounts for about 70 percent of industrial-sector energy consumption. Nonenergy-intensive manufacturing accounts for an increasing share of energy use; for example, electronic equipment is expected to have a growth rate twice that of the manufacturing sector as a whole. The most energy-intensive (in terms of energy used per dollar of output) manufacturers are iron and steel, pulp and paper, petroleum refining, chemicals, and cement; together, these industries account for about half of the primary energy consumed in the industrial sector.

Of the end-use sectors, the industrial sector—especially in its more energy intensive industries—has shown the greatest and the fastest energy efficiency improvements. For example, over the past quarter century, the U.S. steel industry has reduced its energy intensity by nearly 50 percent; the cement industry has improved its fuel efficiency nearly 30 percent since 1975 although, since 1986, the energy intensity improvements have slowed somewhat. The energy use per pound of product in the chemicals industry has fallen at an average of 2 percent per year, and its energy efficiency continued to increase during periods when energy price was stable or falling (though less steeply when the price was falling). By using more of its former waste products for energy, the pulp and paper industry increased its purchased fuel efficiency by nearly 45 percent from 1972 to 1994.

Clearly, the more recent industrial energy efficiency gains are due more to technological progress than to energy prices. Unlike the buildings and transportation sectors, industry has adopted some supply-side energy-efficient technologies that reduce emissions without necessarily reducing energy demand. These include more efficient use of by-product fuels and retrofitting boilers for combined heat and power. On the electricity demand side, some generic improvements, such as high-efficiency motors and advanced motor system drives and controls, have applications in almost all types of industry.

Other energy-efficient technologies are more sector-specific. For example, in the future, the steel industry, even with increases in recycling, will need to make some steel from ore. A new cokeless steelmaking process could cut energy use 30 percent relative to a blast furnace by going directly from solid ore to steel. This "smelt reduction" technique could also increase the industry's productivity, as its investment costs and operating costs are much lower. There are many reduction opportunities of this order of magnitude in various industries, but the single biggest "bang for the buck" is in more efficient heat and power systems.

Combined Heat and Power (CHP) systems, also called "cogeneration" systems, generate electricity (or

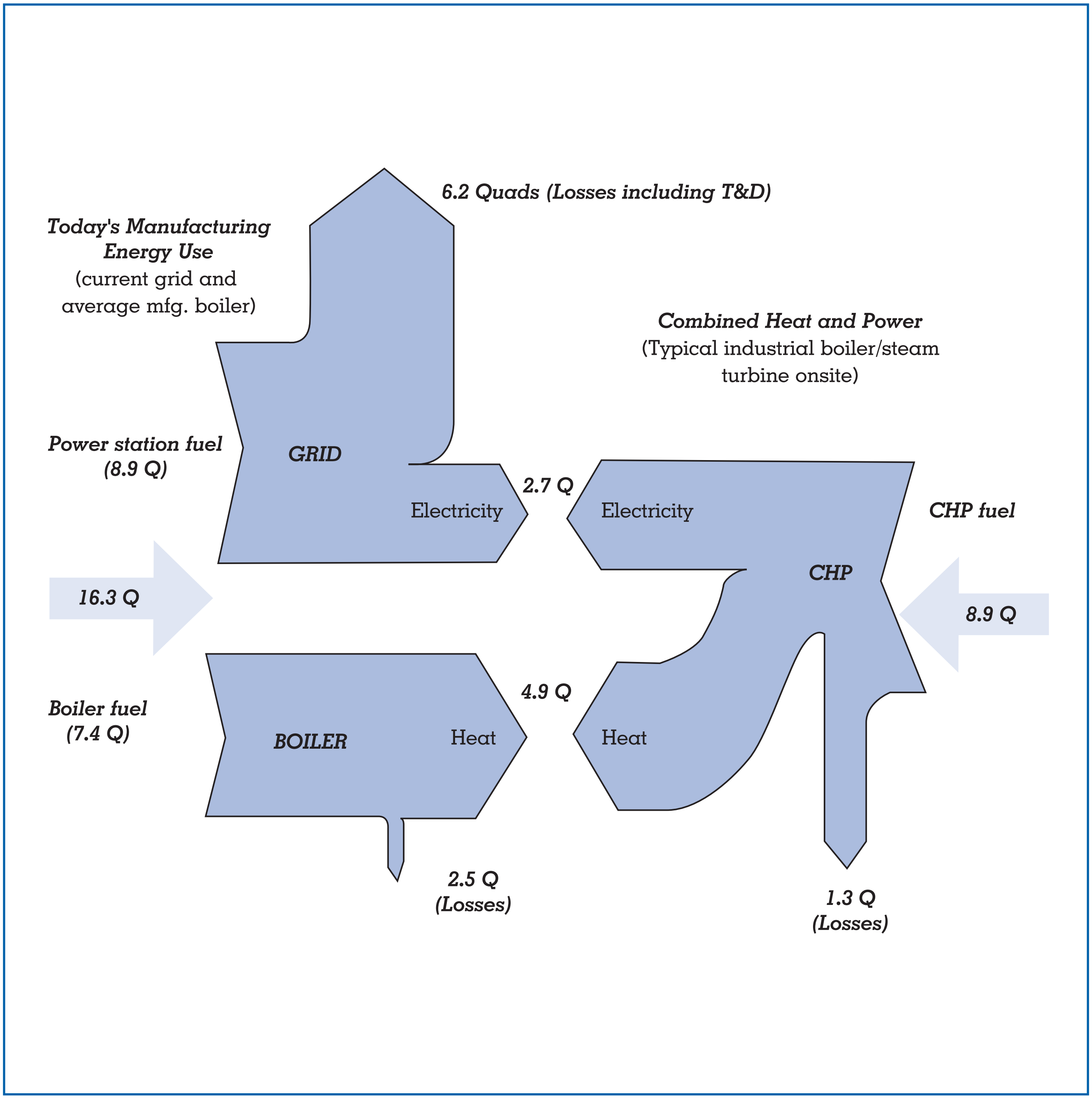


Figure 1.
GRID + Boiler SHP vs. Industrial Gas Turbine CHP, which provides manufacturers with electricity and steam using 45 percent less fuel than SHP.

mechanical energy) and heat simultaneously at the point of use. Figure 1 shows that in 1994, manufacturers used 7.4 quads to generate electricity and, together with the on-site steam produced from separate boilers, required 16.3 quads of fuel, for a system thermal efficiency of 46.5 percent. If produced jointly as CHP at 85 percent efficiency (clearly achievable based on the previous figure), the total fuel requirements would be only 8.9 quads, nearly 50 percent less. Replacing much of industrial Separate Heat and Power (SHP) with CHP by 2010 is not so far-fetched. According to one source (Kaarsberg and Roop, 1999),

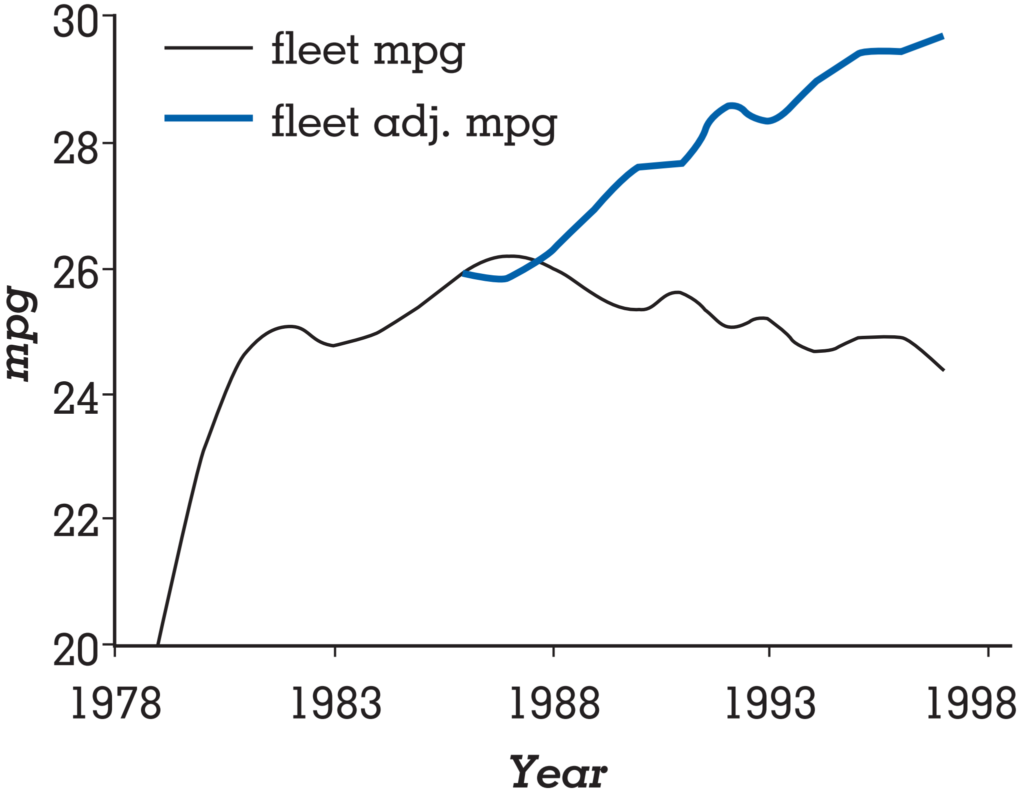


Figure 2.
Fuel economy for the new vehicle "fleet," including light trucks (actual mpg and mpg adjusted for mass, power, and with light truck fraction frozen at 29 percent (1986 value).

more than 75 percent of the industrial thermal capacity installed today will be retired by 2010.

ENERGY SAVINGS IN THE TRANSPORTATION SECTOR

Transportation accounts for about one-quarter of total U.S. primary energy consumption. Since 1986, average new car horsepower has increased nearly 40 percent. The average miles per gallon (mpg) of new light-duty vehicles, new cars, and light trucks combined has not changed significantly since 1982. In the absence of new efficiency standards, carmakers' technical improvements respond to consumer demands for roomy, powerful vehicles. Between 1986 and 1997, the average fuel economy for new passenger cars increased by less than 2 percent (from 28.2 to 28.6 mpg), while the horsepower (hp) per weight increased by 27 percent (from 3.89 to 4.95 hp/100 lb), and weight further grew by 9 percent. The fuel economy of the entire fleet (including a growing fraction of light trucks) decreased of 6 percent (0.5% /year) over this period. Figure 2 shows our calculation of what the fleet fuel economy could have been if carmakers had focused on reducing fuel use rather than increasing power, weight, and size.

PARTNERSHIP FOR A NEW GENERATION OF VEHICLES (PNGV)

The PNGV is a government-industry (General Motors, Ford, DaimlerChrysler) research partnership. One of the most highly publicized PNGV goals is to triple the fuel efficiency of a car (with a prototype by 2004) while preserving safety, performance, amenities, recyclability, and holding down costs. As a result of PNGV, federal government R&D in advanced automotive technologies has been reorganized and redirected toward this ambitious goal. There are PNGV programs in advanced materials; electric drives, including power electronics; high-power energy storage devices; fuel cells; and high efficiency low-emission diesel engines.

A key element of the PNGV is the development of an advanced hybrid-electric vehicle. Energy losses from conventional engine idling or running at part load are eliminated in hybrid vehicles-which are available today. Efficiency is doubled with such hybrid propulsion systems. One-hundred units of fuel needed in today's new car (averaging 28.6-mpg) produces the same amount of drive power as 50 units needed by the electric hybrid. Even today's relatively efficient gasoline spark-ignition internal combustion (IC) engine loses 84 units of energy per 100 units of fuel in. The most easily reduced of these losses are the "standby" losses, which account for 11 percent on average (or up to 20% under increasingly typical congested conditions). These losses occur when the engine is either idling or running at far less than 100 percent load. Standby losses are so high because the engine is oversized to allow for acceleration and therefore runs almost entirely at less-efficient part load. Other losses, in descending order, are exhaust, radiator, engine friction and pumping losses, and accessories (electrical system, pumps, etc., but not including major seasonal loads such as air conditioning). Once the power is delivered to the drive train, it is finally completely dissipated in braking and in rolling and wind resistance.

The 2x electric hybrid suffers no standby engine losses. It features a less powerful, more fuel-efficient IC engine/generator running at full load (or off) and uses a battery-powered electric motor to boost acceleration. Thus, the IC engine is not oversized. It has only two modes: maximally efficient full load, or "off." Engine power typically is delivered not to a drive shaft, but to a generator, on to a storage battery, and then to electric motors on the wheels. While the engine is off, the battery powers these motors. When the engine is on, the battery is either being recharged or is boosting acceleration. The hybrid configuration also helps to reduce driving losses. Instead of friction-braking, an electric hybrid uses its electric motor as a generator to recapture braking energy and charge the battery. Thus, braking losses are reduced by 70 percent. They would only be eliminated if the batteries and generator were 100 percent efficient and all four wheels could recapture the braking energy.

Energy losses also occur as the car is propelled. The 16 units of power are dissipated in rolling resistance and aerodynamic drag (and in today's cars, in friction-braking). The next challenge to reach 3x is to reduce the weight and rolling losses with lighter materials and reduce drag with sleeker designs. A major obstacle to reaching the PNGV affordability goal is the high cost of advanced, lightweight body and tire materials. In such a path, about half the savings are due to the propulsion system and half to the improved (lighter-weight, more aerodynamic) envelope.

This is only one of the many "paths" to achieve 3x efficiency. Other ways to eliminate standby losses being investigated by the PNGV program include advanced diesels, direct injection stratified charge gasoline engines, and fuel cells. With these options, the engine's efficiency could double without going to a hybrid configuration.

It is too soon to judge the PNGV program. Surprisingly, Honda and Toyota were the first companies to introduce 3x efficiency hybrids into the U.S. marketplace in 1999 and 2000, respectively. These two companies were not part of the PNGV program, and thus the early introduction shows that the marketplace could efficiently develop and market efficient vehicle technology without the aid of government. Whether the PNGV participants can catch up and market far superior hybrid vehicles is yet to be seen.

AN OPTIMISTIC CONCLUSION

Above, we said that (apart from crises like the eleven OPEC years) energy intensity E/GDP falls about 1 percent year (E is inversely proportional to energy efficiency, η). Here we are concerned with trends in United States, and later, world energy use, so we write:

$$\begin{aligned} E &= E/GDP \times GDP \\ &= Const/\eta \times GDP \end{aligned}$$

Then to stabilize E, η must rise annually not by merely 1 percent, but fast enough to cancel our desired annual growth in GDP (or gross world product GWP), that is, about 3 percent. We have discussed six examples of why this is achievable, given sufficient motivation to do so:

1. For the last three years, in the United States, the annual growth in η has averaged not 1 percent,

but 3.5 percent; that is, while GDP has surged nearly 4 percent/year, energy use has leveled off. It is still unclear how much of this gain is a real trend from increased productivity and the explosive growth of information technology and the Internet (particularly business-to-business e-commerce). But if even part of this gain continues, it is very good news for reducing carbon emissions.

2. During the eleven OPEC "crisis" years, it is well known that η grew 3 percent/year, of which 1 percent was "structural" (moving from a "smokestack" to a service economy) and 2 percent was a pure efficiency gain. This annual 2 percent is measured for our whole stock of energy using equipment, most of which has a service life longer than the 11-year "experiment." Thus, cars last 12 years, refrigerators 15, buildings 50, and so on. So this gain in the stock must lag the gain in new products, or (re-worded) the rate of improvement of new products must lead that of the stock. Arthur Rosenfeld and David Bassett (1999) have crudely estimated this lead/lag correction, and find that new products improved 5 percent/year.
3. Our refrigerator discussion showed that under appliance standards, refrigerator energy use has been dropping more than 5 percent/year for 27 successive years. Yet the payback time for the improving technology has stayed at two to five years. This suggests that significant steady gains can be kept up for a very long time.
4. During the OPEC years, auto fuel economy improved 7 percent/year. After correcting for the 40 percent increase in power, we see an adjusted gain from 1975 through 1997 of 4 percent/year. If the PNGV 3x car at 80 mpg is a significant fraction of the new car fleet by 2010, this rate of improvement will have been sustained for 35 years.
5. Combined Heat and Power (CHP) is generally 1.5-2 times more efficient than separate heat and power. It grew nearly five-fold in the United States in the years between the enactment of CHP incentives with the passage of the Public Utilities Regulatory Policy Act of 1978 (PURPA) and today. During some of that period it grew at more than 15 percent per year and now accounts for 9 percent of electricity generation.

THE WORLD'S NEED FOR ENERGY IN 2100

Next we estimate the world's need for energy in 2100 (E_w), under three scenarios, where the symbol α will denote the annual gain in energy efficiency η:

1. "BAU" (Business as Usual): $\alpha = 1$ percent/year, its historic non-crisis rate.
2. "No Regrets": $\alpha = 2$ percent/year; that is, the world runs scared of climate change.
3. "De-Materialization": $\alpha = 3$ percent/year; a bit less than the actual rate for the United States for 1997 to 2000.

For this brief discussion, we factor world energy in 2100 as

$$E_w\ (2100) = \text{Pop.} \times e^{-100\alpha} \times \left.\frac{\text{Watts}_P}{\text{Cap.}}\right| 2000 \qquad (1)$$

where $\text{Pop.}_{2100} \approx 10$ billion and $\alpha = -(d/dt)(E/GDP) = 1$ to 3 percent per year. Population by 2100 will probably level off at slightly under 10 billion, and $\text{Watts}_p(2000)$/capita is the rate of primary energy use today, considered a satisfactory goal by the majority of people in developing countries today. For $\text{Watts}_p(2000)$/capita we propose 5 kW (that is, 5 kW-years of energy for the year 2000), corresponding to that of Western Europe today. Thus, we assert that a poor African or Indian today would happily aspire to a year-2100 standard of living, health, transportation, and so on equal to that of Germany or Scandinavia today (even if they have few SUVs).

We also note a 1985 study (by Goldemberg et al.) showing that the best then-available technology could yield a Western European lifestyle at only 1 watt of primary energy per capita.

The exponential factor arises because, in 100 years, this 5 kW/capita will drop:

$$\begin{aligned} E_w\ (2100) &= 10^9 \times (5\text{kW}_p/\text{Cap.}) \times e^{-100\alpha} \\ &= 50\ \text{TW}_p \times e^{-100\alpha} \qquad (2) \end{aligned}$$

Next, we note that world primary power today is slightly over 10 TW, so we can write:

$$E_w\ (2100) = 4\ E_w\ (2000) \times e^{-100\alpha} \qquad (3)$$

Table 1 shows how many 2000 Worlds of power

Scenario	α*(%/year)*	$e^{-100\alpha}$	$E_w(2100)$
BAU	1%	1/e	$1.5E_w(2000)$
No Regrets	2%	$(1/e)^2$	$0.5E_w(2000)$
De-Materializing	3%	$(1/e)^3$	$0.2E_w(2000)$

Table 1.
Number of Today's Worlds of Additional Energy Needed in 2100

we must construct by 2100. Many authors ignore the sensitive dependence on α and discuss the need for a huge program of technology development and construction. Our view is that in any given year it is cheaper and easier to improve efficiency by 2 percent than to add 2 percent to the world's entire energy supply. Since this can be done purely by investments that produce net (life-cycle) savings, we call it a "No Regrets" scenario. The third line of Table 1 corresponds roughly to the current α for the United States. If this were to continue, and spread globally, gross world product can grow at a healthy annual 3 percent, and energy use could level off at today's rate. If we raised α only to 2 percent we would need to grow supply by only 1 percent/year.

Of course, we in the present developed world will not give up "our" present 7–8 TW, but Equation 3 and Table 1 show that under our "No Regrets" scenario, 10 billion people in developing countries need add only 0.5 new "worlds" of energy supply to provide them with an attractive current Western European standard of living. This is what we labeled above as an "Optimistic Conclusion," but it is a continuous and enduring challenge to the inventiveness of technologists and policy-makers of the twenty-first century.

Arthur H. Rosenfeld
Tina M. Kaarsberg
Joseph J. Romm

See also: Cogeneration; Hybrid Vehicles.

BIBLIOGRAPHY

Goldemberg, J.; Johansson, T. B.; Reddy, A. K. N.; and Williams, R. H. (1985). "Basic Needs and Much More in One Kilowatt per Capita." *Ambio* 14(4–5):190–200.

Kaarsberg, T. M., and Roop, J. M. (1999). "Combined Heat and Power in Industry: How Much Carbon and Energy can Manufacturers Save?" *IEEE Aerospace and Electronic Systems Magazine*, pp. 7-11, January

Rosenfeld, A. H., and Bassett, D. (1999). "The Dependence of Annual Energy Efficiency Improvement on Price and Policy." International Workshop on Technologies to Reduce Greenhouse Gas Emissions. Crystal City, VA, May 1999, Plenary IV. Paris, France: International Energy Agency. <http://www.IEA.org/workshop/engecon/>.

U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy. (1997). *Scenarios of U.S. Carbon Reductions Potential Impacts of Energy Technologies by 2010 and Beyond.* <http://www.ornl.gov/ORNL/Energy_Eff/labweb.htm>.

EFFICIENCY OF ENERGY USE, ECONOMIC CONCERNS AND

The large increases in energy prices during the 1970s encouraged extensive interest in reducing energy use and in using energy resources efficiently. The common belief resulting from the energy crisis was that private markets would not provide an adequate supply of energy and would not conserve the use of energy resources. A particular concern was that households and businesses have insufficient incentives to invest in energy-saving technologies. A simplified investment model illustrates this point. The characteristic of any investment is that an initial commitment of funds is made with the expectation of a future payoff. Funds received in the future are of less value than identical funds today; hence investors must discount future cash flows.

Textbooks on investment present a simple model where the net present value (NPV) of an investment equals annual future revenues (R) summed and discounted at the rate r, minus the initial investment cost, I. Using t as a time subscript to denote different years, the equation is

$$\text{NPV} = -I + \sum R_t/(1 + r)^t$$

Revenues are summed from an initial period throughout the economic lifetime of the investment. The business decision rule is that an investment is profitable when its net present value is positive. If revenues accrue at a constant rate, continuously and forever, the equation becomes simpler. In the inequality $I < R/r$, an investment is profitable if annual revenue, R, divided by the discount rate, r, exceeds the initial investment cost. For instance, assume an initial investment of $100 yields annual net revenues of $11 and the discount rate is set equal to the interest rate for obtaining credit of 10 percent. The net present value of the investment (R/r) is $11/.10, or, $110, which exceeds the initial investment. The investment is profitable, and businesses and households who find 10 percent an acceptable annual rate of return are inclined to make such an investment in an efficient market.

In the above equation, r can indicate the internal rate of return on an investment. Suppose that an investment in an energy-saving technology cost $100 and reduces energy costs by $20 per year indefinitely. The reduction in costs is comparable to net revenues received. The above equation can be modified as follows: I equals R/r, where the values of I and R are specified and the value of r is computed. Hence $100 equals $\$20/r$, and r equals 0.20, or 20 percent. The internal rate of return on the $100 investment is 20 percent per year. An investment is generally profitable when its internal rate of return exceeds the (interest rate) cost of obtaining credit. The investment is attractive when its internal rate of return exceeds the investor's hurdle rate, which may vary depending on the riskiness of the investment, and on the rate that can be earned from alternative uses of the investment funds.

Another investment performance measure is the payback period, which, in its simplest form, is the number of years until the net revenue from an investment equals the initial cost. In the above example of a $100 investment cost and a $20 annual payoff, the investment pays off in five years. The payback investment criterion, a crude rule of thumb, is to accept the investment if it continues to return revenue well after the payback time is reached, without any offsetting costs. This criterion has less appeal than investment decision rules based on net present value and rates of return. Investments with a short payout period may yield a high net value or internal rate of returns, but this result is not inevitable. Investments with a low return for the first few years but a very high return in later years can have a long payout period but still offer a high net return. The payoff period has some merit in considering highly risky investments where risk is a function of time. Where the entire initial investment is at risk—such as in a country with a politically unstable government—recovering an initial investment as soon as possible is important. Note, however, that payback is essentially a break-even measure, not a measure of profitability.

Households and businesses use energy jointly with technologies to produce energy services. In almost every application, consumers have a choice between highly efficient technologies that cost more initially but have lower energy costs, and less efficient technologies that have lower initial costs but higher operating costs. For instance, electric heat pumps use differing amounts of electricity to produce space heating and cooling, depending on their efficiency. The standard investment model indicates that rational consumers will invest in the more efficient heat pump if the present value of energy saving exceeds the higher initial investment cost of the more efficient unit, other factors being equal.

Data obtained from the U.S. Energy Information Administration illustrate the trade-off between efficient heat pumps and currently purchased models. In Table 1, the current popular model heat pump uses 6,973 kilowatt-hours (kWh) per year on average to produce heating and cooling, whereas the new and efficient model uses 5,279 kWh per year. Energy efficiency is sometimes interpreted as a simple technical coefficient, such as the amount of energy required to perform a unit of work. Using this technical definition, the efficient heat pump is necessarily more energy-efficient than the current model. Whether this energy efficiency investment makes good economic sense is another matter.

The efficient heat pump reduces energy use by 1,676 kWh per year on average. Is the efficient model heat pump a good investment? Suppose the incremental cost of the efficient unit, as compared with the less efficient unit, is $1,000, and electricity cost 10 cents per kWh. With this price of electricity, the efficient heat pump reduces electricity costs by $167.60 per year. Taking a simplified approach for purposes of illustration and assuming that each unit lasts indefinitely and has no repair, maintenance, or replacement costs, and ignoring possible tax effects, the internal rate of return may be calculated as $\$1{,}000 = \$167.60/r$, which is 16.76 percent per year. If the household can borrow money at, say, 10 percent per year and earn 16.76 percent, the investment makes economic sense. If we assume a 10 percent discount rate, the present value of the investment is $1,676, which exceeds the initial investment cost. The net present value is $676, which indicates that the investment is feasible.

	Current Model		Efficient Model	
	Annual kWh	**Efficiency**	**Annual kWh**	**Efficiency**
Heating	4742	7.5	3984	9.4
Cooling	2297	12	1802	15.3
Annual kWh	6973	—	5297	

Efficiency data are seasonal energy efficiency ratings (SEER) for cooling and heating; seasonal performance factors (HSPF) for heating.

Table 1.
Illustrative Electricity Use In Current and Efficient Residential Electric Heat Pumps

SOURCE: Arthur D. Little, September, 1998. *EIA Residential and commercial Building Technologies—Reference Case.*

The technical literature on the economic return to energy-efficient investments is vast and yields two main conclusions: First, energy-efficient investments frequently offer a positive net present value, or alternatively, a high internal rate of return; second, energy-efficient technologies often fail to achieve a significant market share, at least for the first few years after introduction. For instance, J.G. Koomey, A. H. Sanstad, and L. J. Shown (1996) conclude that consumers fail to purchase energy efficient light bulbs, even though such purchases offer very high internal rates of return. Paul Ballonoff (1999) challenges their estimates and argues that their estimated internal rates of return are the result of erroneous calculations.

There is an enormous controversy about whether consumers and businesses undertake an efficient level of investment in energy-saving technologies. The view of energy conservationists is that energy-efficient investments offer consumers abnormally high rates of return, but consumers still refuse to make such investments. Furthermore, government efforts are required to encourage energy-efficient purchases in these inefficient markets. The alternative view, often associated with mainstream economists, is that consumers make reasonably efficient choices. Taking this view, additional government regulations are more likely to impose market inefficiencies than reduce them.

In the illustration of the electric heat pump, the simple internal rate of return is 16.76 percent. We observe that some households purchase the efficient model, but others purchase the current model. Are these households irrational, or are there simple explanations? Many economists are concerned whether energy-efficient investments make good economic sense. There are numerous explanations for the refusal of households to buy the more efficient heat pump. The cost of borrowing money is not 10 percent, as assumed above, but for some households the cost of borrowing could be a credit card rate, such as 18 percent per year. At this rate, the NPV of the investment is negative. In the above illustration, the price of electricity is 10 cents per kilowatt-hour, but this is an average cost, not the marginal cost of incremental electricity use, which may be much lower. If the incremental cost of electricity is only 5 cents per kilowatt-hour, the annual energy cost saving is $83.80. The internal rate of return is 8.38 percent. Some households find this rate attractive and purchase the efficient heat pump; others do not. Furthermore, this example ignores other system costs, or assumes them to be the same as for the current system, which may not be realistic if the efficiency gains are achieved through newer, less proven, technology.

The alternative estimate of net present value assumes that households are identical; but households have important differences. The energy saved by the efficient heat pump depends on climate and household conditions. Heat pumps have a lower payoff in very cold regions than in regions with moderate climates. Efficient heat pumps are an attractive investment in some regions; but in other regions, the current model is a better choice. Residential structures that are "tight" offer lower benefits to the efficient heat pump. The net present value of the efficient heat pump may be positive for some households but negative for others.

Technologies that offer increased efficiency tend to be the newest technologies that are entering the market. The adoption rate for many new technologies and products may increase very slowly over time, even when such purchases appear to offer high internal rates of return. The adoption rate of new technologies and products tends to be low at first, but as information becomes more widespread over time, the adoption rate of superior technologies increases. Information about a technology or product is a broad concept. Consumers must first become aware of the new technology. Consumers must also become aware of the performance and cost features of the new technology. Acquiring this information is time-consuming and expensive, but is required if a new technology is to capture a large share of the market. New energy-efficient technologies do not immediately achieve a large market share even when their estimated NPV is positive. Moreover, most new technologies and products that are not energy-related also do not immediately achieve a large market share when their NPV is positive.

The energy-efficient heat pump is likely to achieve a small market share for at least its first few years after introduction. Similarly, other energy-efficient technologies, as well as nonenergy related technologies, are likely to achieve a small market share after they enter the market.

The initial rejection of the efficient heat pump by many consumers may be well founded or not. In addition to the sound economic rationales for rejection, there may be market-impediment explanations. Consumers may have imperfect information and be unaware of the energy savings of new, efficient technologies. The transaction cost of acquiring information and making an efficient choice may be just too high. Because of these impediments, households often fail to make investments that would actually save them money over time.

In summary, the main points in this controversy are first, the contention by energy conservation proponents that numerous investment opportunities currently exist to reduce energy costs. Further, these investment opportunities would meet the investment criteria noted above. Although there are few challenges to this contention, it is not the center of controversy. Conservationists contend that the failure of private markets to make these investments is indicative of serious inefficiencies. Some economists dispute this contention and provide explanations of why normal, well-functioning markets may defer on energy-efficient investments.

Ronald J. Sutherland

See also: Auditing of Energy Use; Capital Investment Decisions; Conservation Supply Curves; Economically Efficient Energy Choices; Economic Growth and Energy Consumption; Efficiency of Energy Use; Efficiency of Energy Use, Labeling of; Energy Economics; Environmental Economics; Government and the Energy Marketplace; Industry and Business, Energy as a Factor of Production in; Industry and Business, Productivity and Energy Efficiency in; Supply and Demand and Energy Prices.

BIBLIOGRAPHY

Arthur D. Little, Inc. (1998). *EIA-Technology Forecast Updates-Residential and Commercial Building Technologies-Advanced Adoption Case*. Washington, DC: Energy Information Administration.

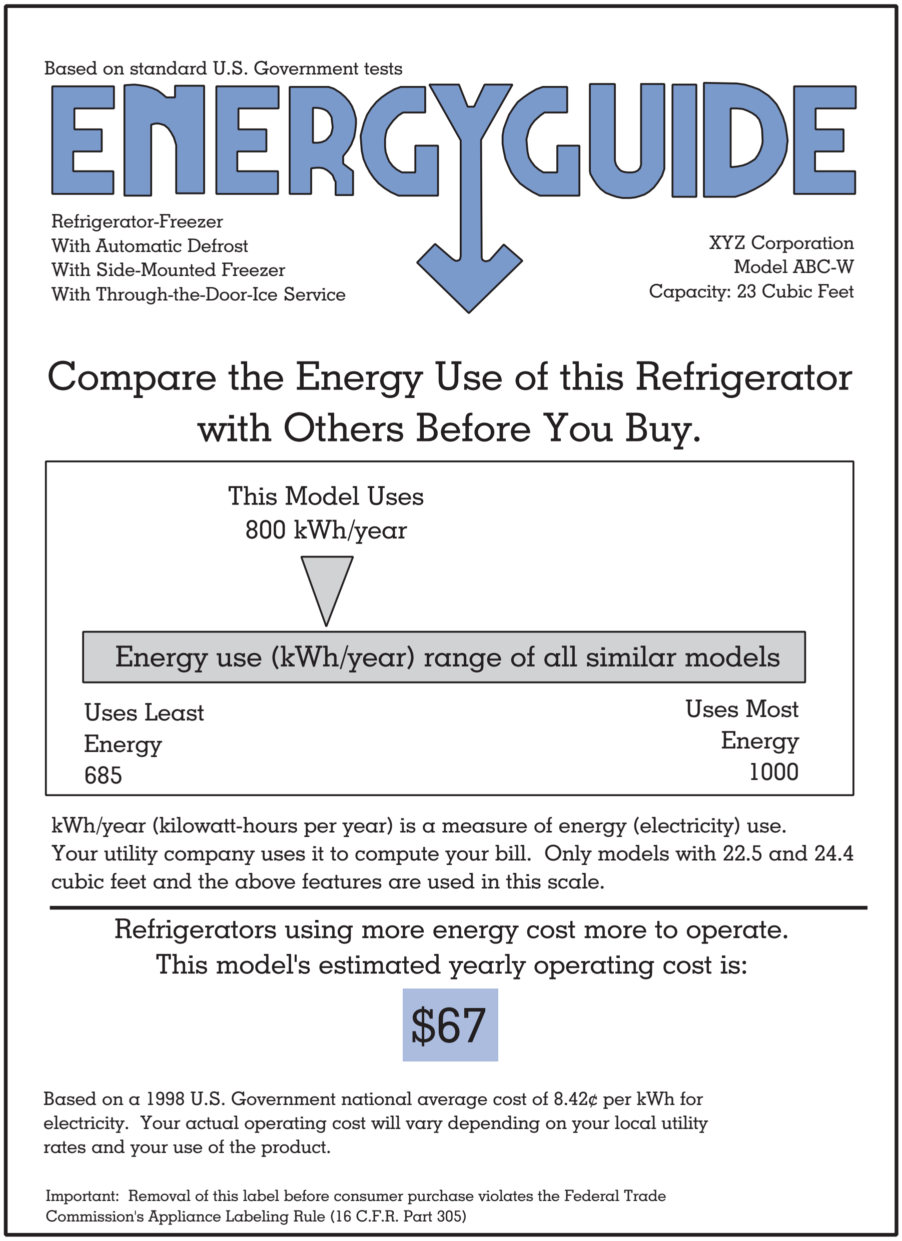
Ballonoff, P. (1999). "On the Failure of Market Failures." *Regulation: The Cato Review of Business and Government* 22(2):17–19.

Koomey, J. G.; Sanstad, A. H.; and Shown, L. J. (1996). "Energy-Efficient Lighting: Market Data, Market Imperfections, and Policy Success." *Contemporary Economic Policy* 14(3):98–111.

EFFICIENCY OF ENERGY USE, LABELING OF

The U.S. Federal Trade Commission (FTC) issued the Appliance Labeling Rule in 1979 in response to a directive from Congress in the Energy Policy and Conservation Act of 1975 (EPCA). In addition to mandating that the FTC promulgate this labeling Rule, EPCA directed the U.S. Department of Energy (DOE) to establish energy conservation standards for residential household appliances and to develop and maintain test procedures by which members of the appliance industry could measure the efficiency or energy use of these products.

The Rule requires manufacturers of most major household appliances to show energy information about their products on labels so consumers purchasing the appliances can compare the energy use or efficiency of competing models. Without this energy use information on EnergyGuide labels, purchasers would have no way to assess the energy efficiency of appliance products and thus would not be able to include the information as a criterion in their purchasing decisions. At the time it was published, the Rule applied to refrigerators, freezers, dishwashers, clothes washers, water heaters, window air conditioners, furnaces, and boilers. In 1987 the FTC included central air conditioners and heat pumps, and in 1989 the FTC added a requirement for a simple disclosure for fluorescent lamp ballasts. In 1993 and 1994, the



Based on standard U.S. Government tests

ENERGYGUIDE

Refrigerator-Freezer
With Automatic Defrost
With Side-Mounted Freezer
With Through-the-Door-Ice Service

XYZ Corporation
Model ABC-W
Capacity: 23 Cubic Feet

Compare the Energy Use of this Refrigerator with Others Before You Buy.

This Model Uses
800 kWh/year

Energy use (kWh/year) range of all similar models

Uses Least Energy 685

Uses Most Energy 1000

kWh/year (kilowatt-hours per year) is a measure of energy (electricity) use. Your utility company uses it to compute your bill. Only models with 22.5 and 24.4 cubic feet and the above features are used in this scale.

Refrigerators using more energy cost more to operate.
This model's estimated yearly operating cost is:

$67

Based on a 1998 U.S. Government national average cost of 8.42¢ per kWh for electricity. Your actual operating cost will vary depending on your local utility rates and your use of the product.

Important: Removal of this label before consumer purchase violates the Federal Trade Commission's Appliance Labeling Rule (16 C.F.R. Part 305)

Figure 1.
Sample label for refrigerator-freezer.
NOTE: The required dimensions of a full-sized label are 5½ inches (width) by 7⅜ inches (height).

FTC added requirements for the disclosure of water use for certain plumbing products, the disclosure of energy-related information for light bulbs and fluorescent lighting products, and the disclosure of energy-efficiency information for pool heaters. The FTC exempted other products listed in EPCA from labeling requirements—such as stovetops and ovens, clothes dryers, television sets, space heaters, humidifiers, and dehumidifiers—because it did not believe that the energy use among different models was great enough to be a potentially significant factor for consumers considering purchasing the products.

To comply with the Rule, manufacturers of all covered major household appliances must place a black-and-yellow "EnergyGuide" label on each of their products that shows the energy consumption or efficiency of the labeled product (Figure 1). The labels also must show a "range of comparability" bar or scale (published regularly by the FTC) that shows the highest and lowest energy consumption or efficiencies for all appliance models that have features similar to those of the labeled one. In addition to the energy consumption or efficiency disclosure, the labels must show the products's estimated annual operating cost, based on a specified national average cost for the fuel the appliances use. Manufacturers must derive the efficiency, energy use, and operating cost information from the standardized tests that EPCA directed DOE to develop. The information on the EnergyGuide label also must appear in catalogs from which the products can be ordered.

Appliance manufacturers must attach the labels to the exterior surface of their products or use hang tag labels, which must be attached in such as way as to be easily viewed by a consumer examining the product. Manufacturers of furnaces, central air conditioners, and heat pumps also must attach EnergyGuide labels to their products showing the products's efficiency and the applicable range of comparability, but the labels need not contain operating cost information. As with appliances, the information on the EnergyGuide also must appear in catalogs from which the products can be ordered.

The requirements for products other than major home appliances vary depending upon the product. Manufacturers of fluorescent lamp ballasts must disclose an encircled "E" on ballasts and on luminaires containing ballasts, as well as on packaging for both. The "E" signifies compliance with DOE's energy conservation standards for those products. Manufacturers of showerheads, faucets, toilets, and urinals must disclose, on the products and their packaging and labeling, the water usage of their products in terms of gallons and liters per flush, per minute, or per cycle. Manufacturers of certain incandescent bulbs, spot and flood bulbs, and screw-base compact fluorescent bulbs must disclose, on packaging, the light output in lumens, energy used in watts, voltage, average life, and number of bulbs. They also must explain how purchasers can select the most energy-efficient bulb for their needs. Manufacturers of certain tube-type fluorescent bulbs must disclose on packages an encircled letter "E" and a statement that the "E" logo means the bulb meets U.S. federal minimum efficiency standards.

The FTC can assess penalties under the Rule against manufacturers for violations of the above requirements. The Rule also states that energy-use-related representations regarding covered products, including print and broadcast advertisements, must be based on the DOE test procedures.

James G. Miller

See also: Economically Efficient Energy Choices

BIBLIOGRAPHY

Appliance Labeling Rule, Code of Federal Regulations, 1979. Vol. 16, part 305.

Energy Policy and Conservation Act of 1975, U.S. Code. Title 42, secs. 6291–6309.

EINSTEIN, ALBERT (1897–1955)

Albert Einstein the twentieth century's most renowned scientist, was born in Ulm, in the kingdom of Warttemberg, now part of Germany, the son of Hermann Einstein, a small businessman, never very successful, and Pauline Einstein (née) Koch. In 1881 Maria, his only sibling, was born. In 1880 the family moved to Munich, where Einstein attended public school and high school, always doing well. (The story that he was a poor pupil is a myth, probably caused by his dislike of formal education.) In those years he also received private violin lessons and, to comply with legal requirements, instruction in the elements of Judaism. As a result of this inculcation, Einstein went through an intense religious phase at about age eleven, following religious precepts in detail and (he later told a friend) composing songs in honor of God. A year later, this phase ended abruptly and forever as a result of his exposure to popular books on science, to "the holy geometry book" (as he called it) on Euclidean geometry, to writings of Kant, and more.

In 1895 Einstein took the entrance examination at the Federal Institute of Technology (ETH) in Zurich but failed because of poor grades in literary and political history. In 1896, after a year of study at a high school in Aaratu Switzerland, he did gain admission, however. In that year he gave up his German citizenship and became stateless, in 1901 he became a Swiss citizen.

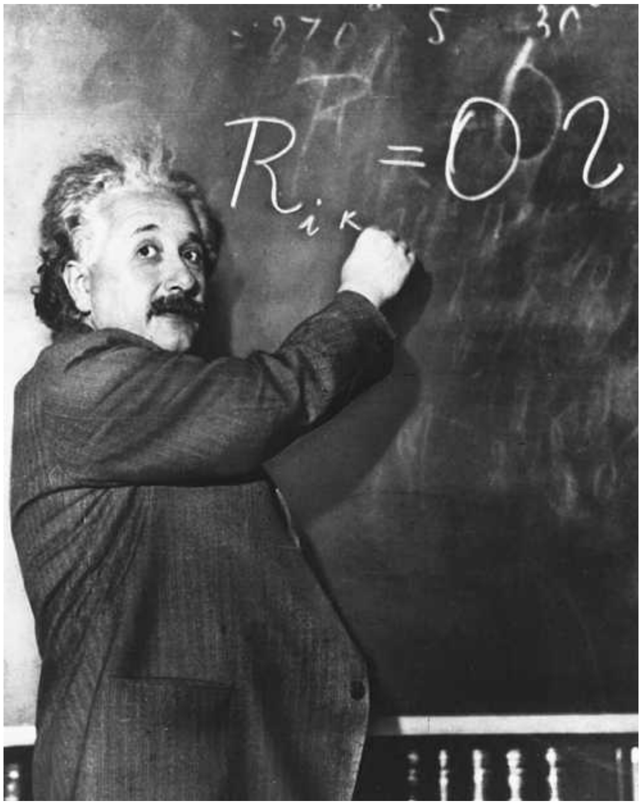
During his next four years as an ETH student, Einstein did not excel in regular course attendance, relying far more on self-study. In 1900 he passed his final examinations with good grades, which qualified him as a high school teacher in mathematics and physics. For the next two years he had to be satisfied with temporary teaching positions until in June 1902 he was appointed technical expert third class at the Patent Office in Berne.

In January 1903 Einstein married Mileva Maric, (of Greek-Catholic Serbian descent), a fellow student at the ETH. In 1902 the couple had a daughter out of wedlock, Lieserl, whose fate remains unknown, and after marriage they had two sons, Hans Albert (1904), who became a distinguished professor of hydraulic engineering in Berkeley, California, and Eduard (1910), a gifted child who became a student of medicine in Zurich but who then turned severely schizophrenic and died in a psychiatric hospital.

In 1914 the Einsteins separated, and they divorced in 1919. Thereafter Einstein married his cousin Elsa Einstein, who brought him two stepdaughters. He had several extramarital affairs during this second marriage.

None of Einstein's first four papers published between 1901 and 1904 foreshadowed his explosive creativity of 1905, his *annus mirabilis*, in which he produced: in March, his proposal of the existence of light quanta and the photoelectric effect, work for which in 1922 he received the Nobel Prize; in April, a paper on the determination of molecular dimensions, which earned him his Ph.D. in Zurich; in May, his theory of special relativity; in September, a sequel to the preceding paper containing the relation $E = mc^2$. Any one of these papers would have made him greatly renowned; their totality made him immortal.

Only after all these publications did Einstein's academic career begin: *privatdozent* in Berne, 1908; associate professor at the University of Zurich, 1909, the year of his first honorary degree (Geneva); full professor at Karl Ferdinand University, Prague, 1911; professor at the ETH, 1912; professor and member of the Prussian Academy of Sciences, Berlin, 1914–1932, where he arrived four months before the outbreak of World War I.



Albert Einstein. (AP/Wide World Photos)

In 1915 Einstein cosigned his first political document, "Manifesto to Europeans," in which all those who cherish European culture were urged to join in a League of Europeans (never realized). Far more important, in that year Einstein completed his masterpiece, perhaps the most profound contribution to physics of the twentieth century: his general relativity theory, on which he had been brooding for the previous eight years. In the special theory all laws of physics have the same form for any two observers moving relatively to each other in a straight line and with constant, time-independent, velocity. In the general theory the same is true for all kinds of relative motion. This demands a revision of Isaac Newton's theory of gravitation. Space is curved, Einstein now asserted, the amount of curvature depending on how dense matter is at that place; matter determines by its gravitational action "what shape space is in."

The superiority of Einstein's over Newton's theory became manifest in 1915, when Einstein could for the first time explain an anomaly in the motion of the planet Mercury (advance of the perihelion), known observationally since 1859. He also predicted that light grazing the sun bends by a factor of two larger than predicted by Newton's theory.

In 1916 Einstein completed his most widely known book "on the special and the general theory of relativity, popularly explained," wrote the first paper on gravitational waves, and became president of the *Deutsche Physikalische Gesellschaft.* In 1917 he became ill, suffering successively from a liver ailment, a stomach ulcer, jaundice, and general weakness, but nevertheless he managed to complete the first paper on relativistic cosmology. He did not fully recover until 1920.

In November 1919 Einstein became the mythical figure he is to this day. In May of that year two solar eclipse expeditions had (in the words of the astronomer Eddington) "confirm[ed] Einstein's weird theory of non-Euclidean space." On November 6 the president of the Royal Society declared in London that this was "the most remarkable scientific event since the discovery [in 18461 of the predicted existence of the planet Neptune."

The next day the *Times of London* carried an article headlined "Revolution in Science/New Theory of the Universe/Newtonian Ideas Overthrown." Einstein had triumphed over Newton (who, of course, remains a stellar figure in science). The drama of that moment was enhanced by the contrast with the recently concluded World War I, which had caused millions to die, empires to fall, and the future to be uncertain. At that time Einstein emerges, bringing newlaw and order. From that time on the world press made him into an icon, the divine man, of the twentieth century.

At about that time one begins to perceive changes in the activities of Einstein, now in mid life. He began writing nonscientific articles. In 1920 he was exposed to anti-Semitic demonstrations during a lecture he gave in Berlin. At the same time, Jews fleeing from the East came literally knocking at his door for help. All that awakened in Einstein a deepened awareness of the Jewish predicament and caused him to speak up and write in favor of Jewish self-expression by means of settling in Palestine, creating there a peaceful center where Jews could live in dignity and without persecution. Thus he became an advocate of what may be called moral Zionism, though he never was a member of any Zionist organization.

The 1920s also was the period of Einstein most extensive travels. In 1921 he paid his first visit to the United States, for to raise funds for the planned Hebrew University, being honored on the way, including being received by President Warren G. Harding. In 1922 his visit to Paris contributed to the normalization of Franco-German relations. Also in that year he accepted membership in the League of Nations' Committee on Intellectual Cooperation. In June Walter Rathenau, foreign minister of Germany, a Jew and an acquaintance of Einstein, was assassinated. After being warned that he, too, might be in danger, Einstein left with his wife for a five-month trip abroad. After short visits to Colombo, Singapore, Hong Kong, and Shanghai, they arrived in Japan for a five-week stay. The press reported that, at a reception, the center of attention was not the empress, everything turned on Einstein.

On the way back, they visited Palestine. In introducing Einstein at a lecture, the president of the Zionist Executive said: "Mount the platform that has been awaiting you for two thousand years." Thereafter Einstein spent three weeks in Spain. In 1925 he journeyed to South America, lecturing in Buenos Aires, Montevideo, and Rio de Janeiro. Apart from three later trips to the United States, this was the last major voyage in Einstein's life.

All these multifarious activities took a lot of Einstein's energies but did not keep him from his physics research. In 1922 he published his first paper on unified field theory, an attempt at incorporating not only gravitation but also electromagnetism into a new world geometry, a subject that was his main concern until the end of his life. He tried many approaches; none of them have worked out. In 1924 he published three papers on quantum statistical mechanics, which include his discovery of so-called Bose-Einstein condensation. This was his last contribution to physics that may be called seminal. He did continue to publish all through his later years, however.

In 1925 quantum mechanics arrived, a new theory with which Einstein never found peace. His celebrated dialogue with Niels Bohr on this topic started at the 1927 Solvay Conference. They were to argue almost until Einstein's death without agreeing.

In 1928 Einstein suffered a temporary physical collapse due to an enlargement of the heart. He had to stay in bed for four months and keep to a salt-free diet. He fully recuperated, but he stayed weak for a year. The year 1929 witnessed his first visit with the Belgian royal family, leading to a life-long correspondence with Queen Elizabeth.

Einstein had been a pacifist since his young years, but in the 1920s his position became more radical in this respect. For example, in 1925 he, Gandhi, and others signed a manifesto against obligatory military service, and in 1930 another, supporting world government. In that year and again in 1931 Einstein visited the United States. In 1932 he accepted an appointment as professor at the Institute for Advanced Study in Princeton, originally intending to divide his time between Princeton and Berlin. However, after he and his wife left Germany on December 10 of that year, they would never set foot in Germany again—in January 1933 the Nazis came to power. Though remaining pacifist at heart, Einstein was deeply convinced that they could be defeated only by force of arms.

Because of the new political situation, Einstein changed his plans, arriving on October 17, 1933, in the United States to settle permanently in Princeton. Whereafter he left that country only once, in 1935, to travel to Bermuda to make from there application for permanent residency in the United States. In 1940 he became a U.S. citizen.

Also in his new country Einstein remained a prominent figure. In 1934 he and his wife were invited by Franklin and Eleanor Roosevelt and spent a night at the White House. Einstein remained scientifically active—he wrote, in fact, some good papers, but nothing as memorable as in his European days.

In 1939 Einstein wrote to Roosevelt to draw his attention to possible military use of atomic energy. His influence on these later developments was marginal, however. In 1943 he became consultant to the U.S. Navy Bureau of Ordnance but was never involved in atomic bomb work. In 1944 a copy of his 1905 paper on special relativity, handwritten by him for this purpose, was auctioned for six million dollars as a contribution to the war effort. (It is now in the Library of Congress.)

After the war he continued to speak out on political issues, such as his open letter to the United Nations urging the formation of a world government, and his frequent condemnations in the press of Senator Joseph McCarthy's activities. After the death of Chaim Weizmann, the first president of Israel, Einstein was invited but declined to be his successor.

In 1948 Einstein was found to have a large intact aneurysm of the abdominal aorta. In 1950 he wrote his testament, willing his papers and manuscripts to Hebrew University (where they are now). On April 11, 1955, he wrote his last letter, to Bertrand Russell, in which he agreed to sign a manifesto urging all nations to renounce nuclear weapons. On April 13 Einstein wrote an incomplete draft for a radio address that ends: "Political passions, aroused everywhere, demand their victims." On the afternoon of that day his aneurysm ruptured. On April 15 he entered Princeton Hospital, where he died on April 18 at 1:15 A.M. His body was cremated that same day. The ashes were scattered at an undisclosed place. The following November his first great-grandson was born.

Abraham Pais

See also: Nuclear Energy; Nuclear Energy, Historical Evolution of the Use of; Nuclear Fission; Nuclear Fusion.

BIBLIOGRAPHY

Pais, A. (1982). *Subtle Is the Lord: The Science and the Life of Albert Einstein.* Oxford, Eng.: Oxford University Press.

Pais, A. (1997). *Einstein Lived Here.* New York: Oxford University Press.

ELASTIC ENERGY

A material is said to be elastic if it returns to its original shape after being deformed. Elastic energy is energy contained by an object as a result of deforming it from its relaxed position. A rubber band used to hold a stack of papers together and a trampoline are elastic. Deforming something requires application of a force. A person pulls on a rubber band to stretch it; a gymnast pushes down on a springy trampoline to deform it. The act of doing work to deform an elastic material produces elastic energy in the material. This elastic energy can then be used to do work as the deformed material returns to its original shape.

Springs made from coils of wire are used to store elastic energy which is a form of potential energy. The most common type of spring is cylindrically shaped, with coils evenly spaced and of the same diameter. A ballpoint pen uses a coil spring to hold the point in place for writing and to return it to the case for protection.

The energy stored in a spring depends on the strength of the spring and the deformation that may be either an extension or compression from its relaxed

Multiple exposure photography is used to illustrate the movement of a pole vaulter. (Corbis-Bettmann)

condition. For many springs, the deforming force is directly proportional to the deformation; doubling the force doubles the deformation. For example, if a 10 N force stretches a spring 0.01 m then a 20 N force will stretch it 0.02 m. Such a spring is said to obey Hooke's law, and the relation between force and deformation is written $F = kx$, where F is the deforming force in newtons (N), x is the deformation in meters (m), and k is the spring constant in newtons/meter (N/m). The stiffer the spring, the larger the spring constant. The elastic energy of a spring obeying Hooke's law is given by $E = \frac{1}{2}kx^2$. The energy depends strongly on the deformation because doubling the deformation (x) quadruples the energy. In other words, a spring stretched 0.1 m will have four times the elastic energy as when it was stretched 0.05 m.

Some watches and clocks, not powered by batteries, use springs to store energy for their mechanical mechanism. Usually, the springs are made by winding wire in a flat spiral. Rather than storing energy by stretching or compressing, the spirals wound tighter. Rewinding a watch by turning a knob amounts to tightening the spirals of the spring inside the watch. Energy released when the spring unwinds is used to run the clock mechanism. The design is such that it takes only a few seconds to store the energy that is then released over twenty-four hours or more.

Elastic energy is not limited to springs. In fact, it is an important property that material scientists consider in selecting materials for products. A bent diving board, a drawn bow, and a planted vaulting pole are just a few of the many examples of elastic materials. It is not easy to calculate the elastic energy for these materials applications, but it is easy to see how energy is involved. A diver does work on the board by jumping on it, a hunter does work on the bow by pulling on the strings, and a vaulter does work on the pole by bending it. Elastic energy is stored as a result. The diver recovers the elastic energy of the board and rises upward when the board relaxes. The arrow recovers the elastic energy of the bow when the hunter releases the string. And the vaulter recovers the elastic energy of the pole and rises upward.

Joseph Priest

See also: Potential Energy.

BIBLIOGRAPHY

Hobson, A. (1995). *Physics: Concepts and Connections*. Upper Saddle River, NJ: Prentice-Hall.

Serway, R. A., and Faughn, J. S. (1995). *College Physics*. Philadelphia, PA: Saunders College Publishing.

ELECTRIC CIRCUIT

See: Electricity

ELECTRICITY

Electrical and electronic devices and machines have become an integral part of contemporary life, ranging from household appliances and computers to huge industrial machines. When home and business owners pay the monthly bill from the electrical power company for the use of all of these items, they are paying for energy very conveniently delivered over electrical wires from the power company.

Although the delivery of electricity to homes and businesses has been possible only during the past century, static electricity was observed by the ancient Greeks over 2,000 years ago. They also noticed natural magnets, called lodestones, found near the town of Magnesia. These were important discoveries because scientists now know that electricity and magnetism are intimately related. Magnetism is used by power companies to produce the electricity used every day.

ELECTRIC CHARGES AND ELECTRIC FORCE

If the humidity is low, it is common to experience a shock when touching a metal doorknob after walking across a carpeted floor. This is static electricity, and it can be studied by rubbing a hard rubber rod on some fur, and then touching the rod to a small metal ball that is suspended on the end of a silk thread. The ball quickly bounces away from the rod and is repelled by it, as shown in Figure 1a. The rod and the ball are electrified or charged with the same type of (net) charge. Further, when a glass rod is rubbed with silk and touched to a second similar ball, the second ball is repelled by the glass rod, shown in Figure 1b. But when the two balls are brought near each other (without touching), they are clearly attracted to each other as shown in Figure 1c.

In the mid-1700s Benjamin Franklin proposed that there are only two kinds of electrical charge, which he called plus and minus. He defined the net charge on the rubber rod to be negative, and the charge on the glass rod to be positive. Further, charges of the same kind are repelled from each other, while opposite charges are attracted to each other. The amount of electrical charge (often represented by the letter q or Q) is measured in coulombs (abbreviated as C) in the Standard International system of units, called S.I. units.

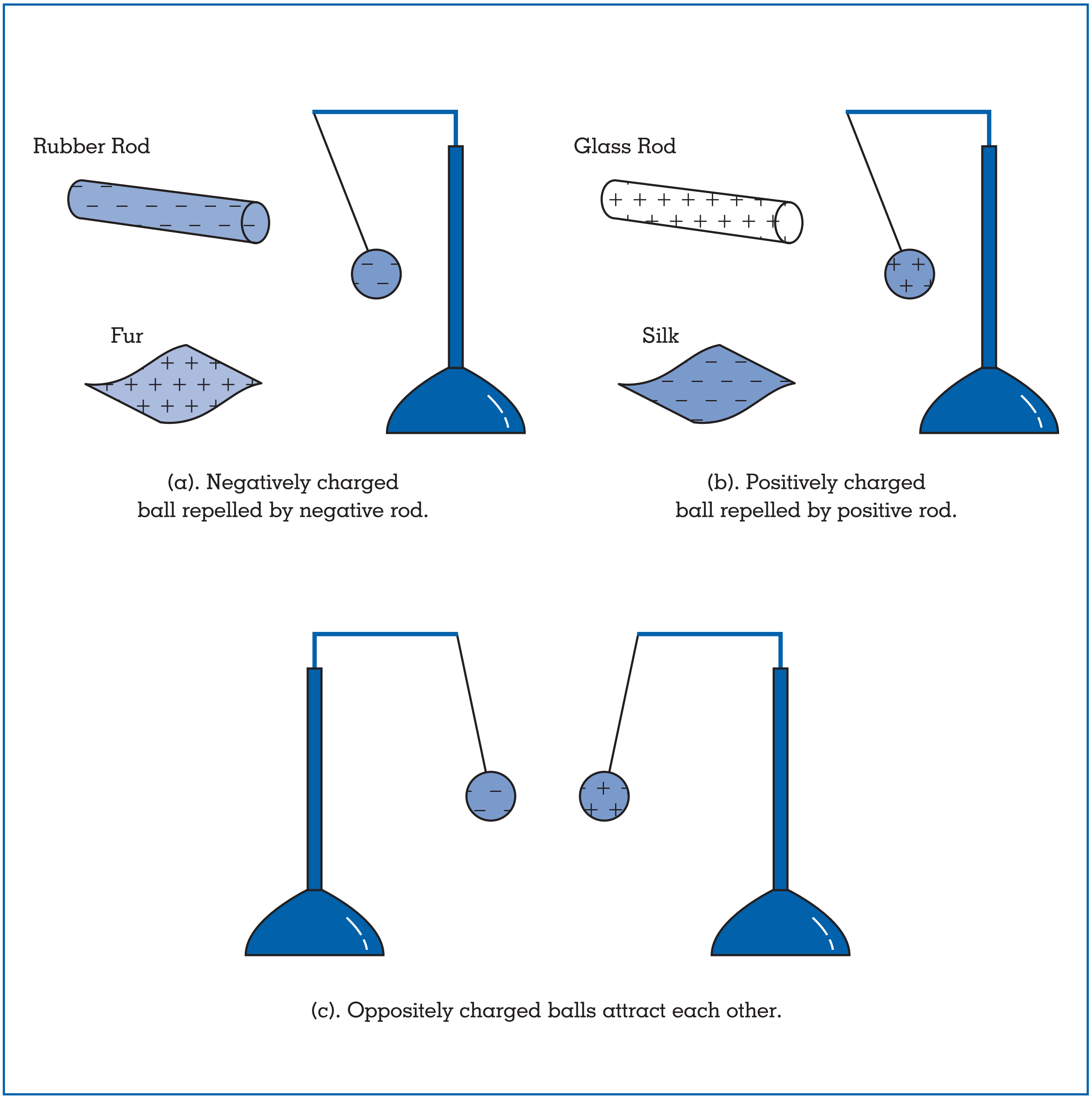
CHARGES IN THE STRUCTURE OF AN ATOM

Electrons particles in matter with a negative charge were discovered around 1900. All matter is made of tiny atoms packed closely together. The structure of the atom was proposed to be like a tiny solar system, with negative electrons revolving in orbits around a very tiny positive nucleus (see Figure 2). The charge ("e") of the electron is a certain fixed amount: only a millionth of a trillionth of 0.16 coulombs (or 0.16×10^{-18} C). The total negative charge of all of the electrons in an atom is exactly the same (but of opposite sign) as the charge of the positive nucleus; thus an atom taken as a whole is normally uncharged or "neutral."

Since a piece of matter (e.g., a piece of rubber or copper) is made up of a great many neutral atoms, the piece is itself normally neutral or uncharged. When a rubber rod is rubbed with fur, some electrons are pulled from the fur onto the rod, giving it extra negative charge (the fur is then deficient of electrons so it is positively charged).

ELECTRIC CIRCUITS AND CURRENT FLOW

There were few applications of electricity (or of magnetism) before the invention of the battery by Alessandro Volta in 1800. A battery can cause charges to move for long periods of time. The movement or



(a). Negatively charged ball repelled by negative rod.

(b). Positively charged ball repelled by positive rod.

(c). Oppositely charged balls attract each other.

Figure 1.
Static electricity experiments.

"flow" of electrical charge is called a current. A basic battery cell consists of two different metals (called electrodes) immersed in an acid or salt solution (called an electrolyte). Through chemical interactions, one electrode develops extra electrons and becomes negatively charged, while the other develops a deficiency of electrons and becomes positively charged; these are respectively labeled the negative (-) and positive (+) battery cell terminals.

Work (W) is done by a battery whenever it pushes a positive charge (+q) away from the (+) terminal (through space outside the battery) to the (-) terminal. The "potential" or "emf" or "voltage" (V) of a battery is defined as the work done in this process divided by

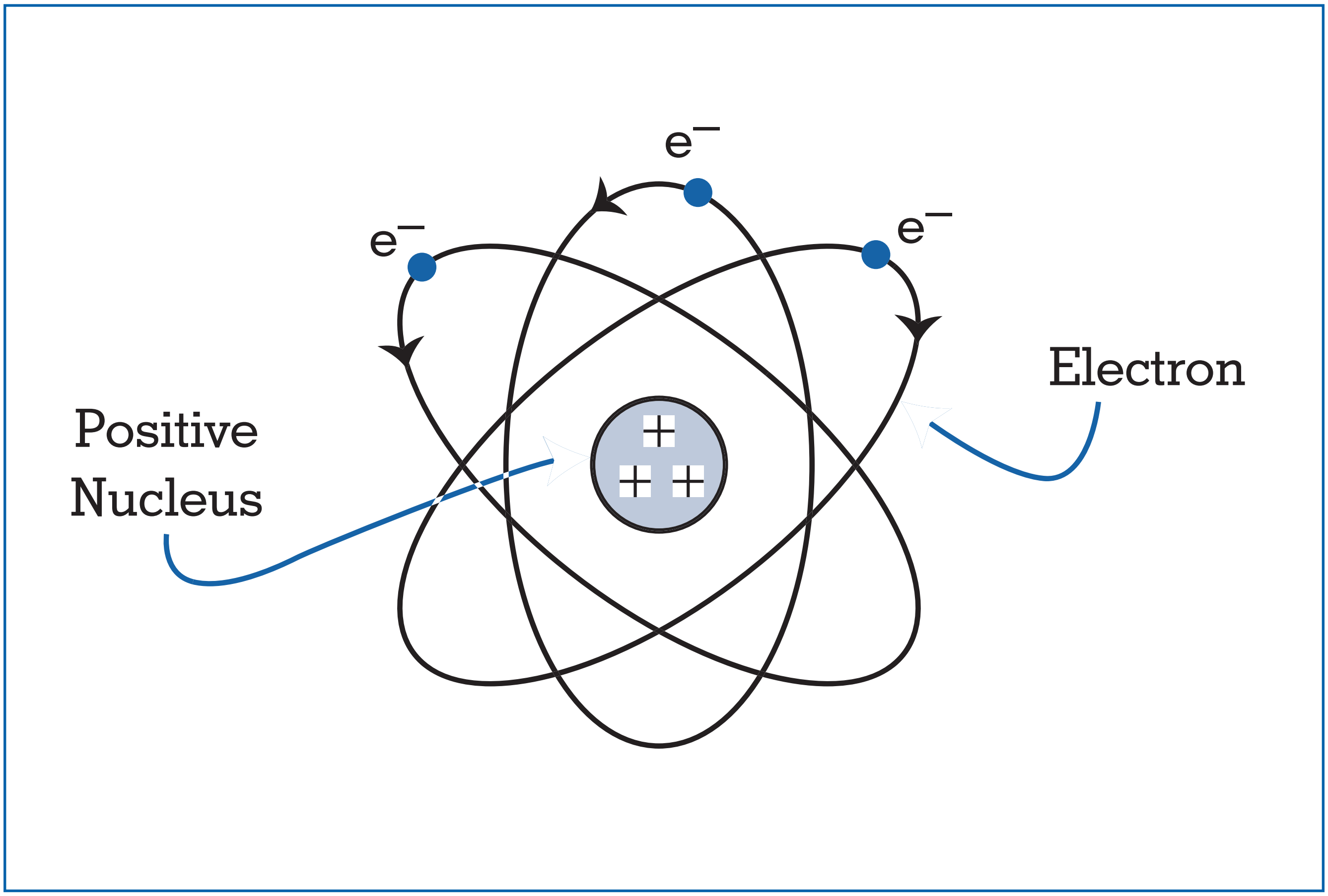


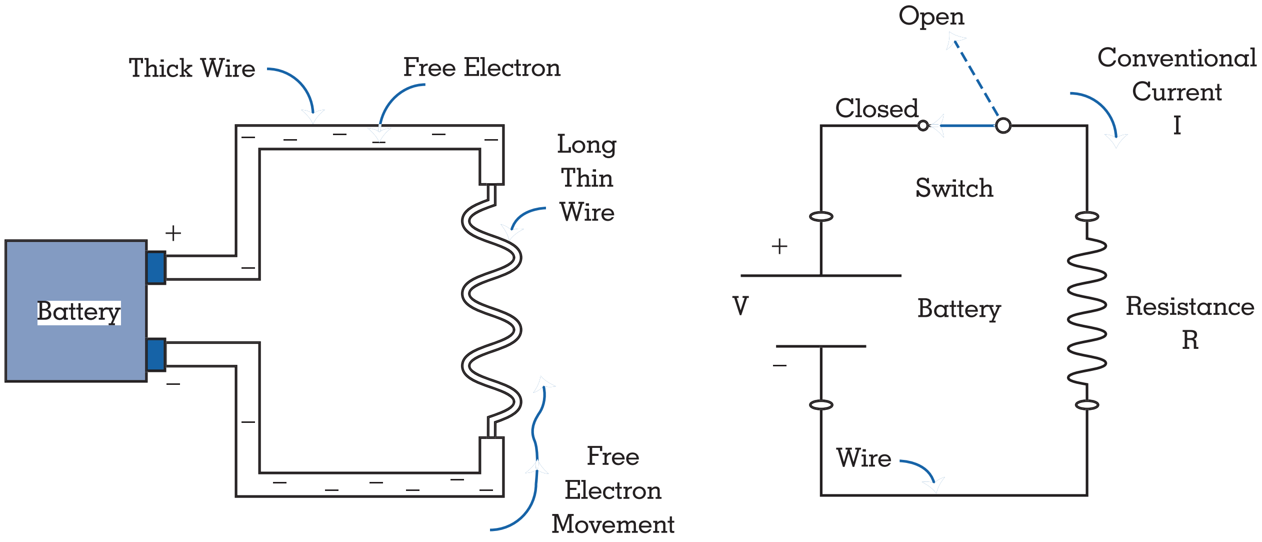
Figure 2.
Model of an atom with three electrons.

the charge: V= W/q. In S.I. units the voltage V is in volts V to honor Volta. A battery cell often used in flashlights is the "carbon-zinc dry cell." If unused and fresh, the carbon-zinc cell has a voltage of about 1.5 V. Many flashlights use two carbon-zinc cells placed end-to-end or in "series." Strictly speaking, the two cells together constitute a "battery," and the voltage of this battery is 3.0 V. The common 12 V "lead-acid" battery used in automobiles has six cells connected in series with it; each lead-acid cell has a voltage of 2.0 V.

Suppose a long thin metal wire is connected by a pair of thick wires between the terminals of a battery. This is a basic "electric circuit" as shown in Figure 3a. In all metals, each atom permits roughly one of the outer electrons to move quite freely in the material; these are called the "free electrons." In contrast, all electrons of the atoms of good electrical "insulators," such as glass, rubber, and air, are tightly bound to the atoms and are not free to move through the body of the material. The free electrons are repelled from the negative battery terminal and attracted toward the positive terminal, so that a continuous movement of charge (an electrical current) results around this complete path or "circuit." But there are frictional effects (electrons bumping into atoms as they move along) that resist the movement or flow of electrons through the wires, especially in the thin wire. The frictional effects result in the electrical "resistance" (R) of the wire, and cause the wires to heat up, especially the long thin wire.

An important property of this or any electrical circuit is the rate that charge moves past a place in the circuit (e.g., out from or into a battery terminal). The electrical current (I) is defined to be the charge (Q) that flows, divided by the time (t) required for the flow: I = Q/t. In S.I. units the current (I) is in amperes (A).

In the early 1800s, Georg Ohm studied the effect



(a). Drawing of an electrical circuit.

(b). Electrical circuit drawn with conventional electrical symbols.

Figure 3.
Basic d.c. electrical circuit.

on current of changing the battery voltage as well as the length, cross-sectional area, and material of the wire in an electrical circuit like Figure 3a. He found that the current flow (I) was proportional to the voltage (V) and inversely proportional to the resistance (R) of the wire to this flow. The resistance (R) is quite constant for a given piece of metal wire; Ohm's law states that we may write that I=V/R. The S.I. units of resistance are ohms (Ω). The resistance to current flow of a wire is larger if the wire is longer, but smaller if the wire is thicker (larger cross-sectional area).

The smallest diameter copper wire generally permitted within the walls for U.S. household wiring is called no. 14 wiring; such wire has a diameter of about 1/16 inch (1.63 mm). A 10 foot length (about 3 meters) of no. 14 copper wire has a resistance of about 0.025 ohms; note that a 100 foot length (about 30 meters) has a resistance of 0.25 ohms.

Figure 3b shows the basic electrical circuit of Figure 3a as it is drawn using conventional electrical symbols. Note the symbol for a battery. The connecting wires with negligible resistance are drawn as lines, while an element with significant resistance is drawn as a zigzag line. An electrical switch is a length of metal that can be moved so that a space of air appears in the circuit; the air space has a large resistance (perhaps billions of ohms) so that virtually no current flows through it when it is "open" (as represented by the dashed line in Figure 3b. When the switch is closed the circuit path is complete and current flows around the "closed circuit." It is conventional to draw the current flow direction as from the (+) battery terminal to the (-) terminal. This is the direction that (+) charges would move if they were moving within the wire, and this was assumed to be the case from about 1800 to 1900. It is now known electrons move in the opposite direction, but the old conventional direction is almost universally used. In any case, the current flows in only one direction in this circuit; this situation is called "direct current" or "d.c."

ELECTRICAL POWER AND ELECTRICAL ENERGY

As electrical current flows through a wire of significant resistance, the wire is heated and the wire gives off heat energy. The wire in an incandescent light bulb is heated so much that it glows and gives off light energy as well as heat energy. This energy is supplied by a battery or other source of emf. Further, "power" is the rate that energy is being supplied, or that work is being done: power = energy/time, or power = (work done)/time. In S.I. units power is measured in watts. Using the definition of voltage and of current one can show that when a current flows from a battery (or any source of emf), the battery is delivering power at the rate of the voltage times the current: $P = VI$. For example, suppose a 115 V battery is supplying 0.87 amperes of current to a light bulb connected to the battery; then the power being used by the light bulb is equal to (115V) × (0.87 A) = 100 watts. This particular bulb is a "100 watt light bulb." Note that because 1,000 watts equals one kilowatt, this bulb is using a tenth of a kilowatt or 0.10 kilowatts (0.10 kW) of power.

Since power = energy/time, it follows that energy = power × time; that is, energy supplied is equal to the power multiplied by the time-interval. The unit of electrical energy unit used by the electrical utility company is (kilowatts) × (hours), or kWh. Suppose a 100 W bulb is turned on for one day or 24 hours. The energy used by the bulb is (0.10 kW) × (24 h) = 2.4 kWh. A typical electrical energy cost in the United States might be 8 cents per kWh, so the cost for using this bulb for an entire 24 hour day would be (2.4 kWh) × (8 cents/kWh) = 19.2 cents.

Note that electrical energy used (and paid for) depends not only on the power consumption by a device or appliance, but also the length of time that it is used in a month. Therefore people are sometimes surprised to discover that there is a modest electric bill even after they have been gone from home on vacation and "everything" had been turned off. However, a number of devices around the home were probably operating, such as electric clocks, the electric motor on the furnace, the refrigerator, and the freezer. There are also many applicances operating on standby mode, such as televisions, answering machines, and cordless appliance. These are sometimes called "phantom power" devices.

MAGNETS, MAGNETIC FIELD, AND ELECTROMAGNETS

Magnets also repel or attract each other. A magnet always has both a north pole (N) and a south pole (S). Again, "opposites attract and likes repel" so that two magnets with north poles close to each other are repelled from each other, even without touching. The influence in the region near a magnet is often pictured as a "magnetic field"; invisible "magnetic field lines" are imagined, and often drawn in a diagram of a magnet. Figure 4a illustrates a steel bar "permanent magnet" with its magnetic field lines.

In the early 1800s H. C. Oersted discovered that an electric current flowing through a coil of wire produces a magnetic field, as shown in Figure 4b. A long coil of wire, often wound on an iron core to enhance the magnetic field, is called a "solenoid" or "electromagnet" and attracts iron or steel objects, just like a permanent magnet.

ELECTRIC GENERATOR

Shortly after Oersted's discovery, Michael Faraday found that changing the number of magnetic field lines within a coil of wire induces an emf or voltage in the coil so that a current flows if there is a closed circuit. Thus not only can a current produce a magnetic field, but a changing magnetic field can generate a current. Figure 5a shows the basic idea of an electric generator. A coil consisting of a single turn of wire is pictured; this is twirled (rotated) in front of a bar magnet so that the magnetic field lines pass through the coil, first from one side and then the other side. The changing magnetic field within the coil results in an alternating emf or voltage that is measured between the terminals A and B, and a current that reverses directions (i.e., an alternating current or a.c.) flows through the attached "load" resistance.

The alternating voltage V_{AB} measured between terminals A and B is graphed in Figure 5b. If the coil is turned at 60 cycles per second, then the frequency of alternation is 60 cycles per second (called 60 Hz in S.I. units). This is the standard frequency for commercial a.c. power in the United States (in Europe the standard is 50 Hz). By increasing the number of turns (or loops) in the coil, or by increasing the magnetic field strength, it is possible to increase the amount of emf or voltage generated (e.g., the peak voltage V_{peak} illustrated in Figure 5b). A typical power company generator might generate 10,000 V_{rms}. This

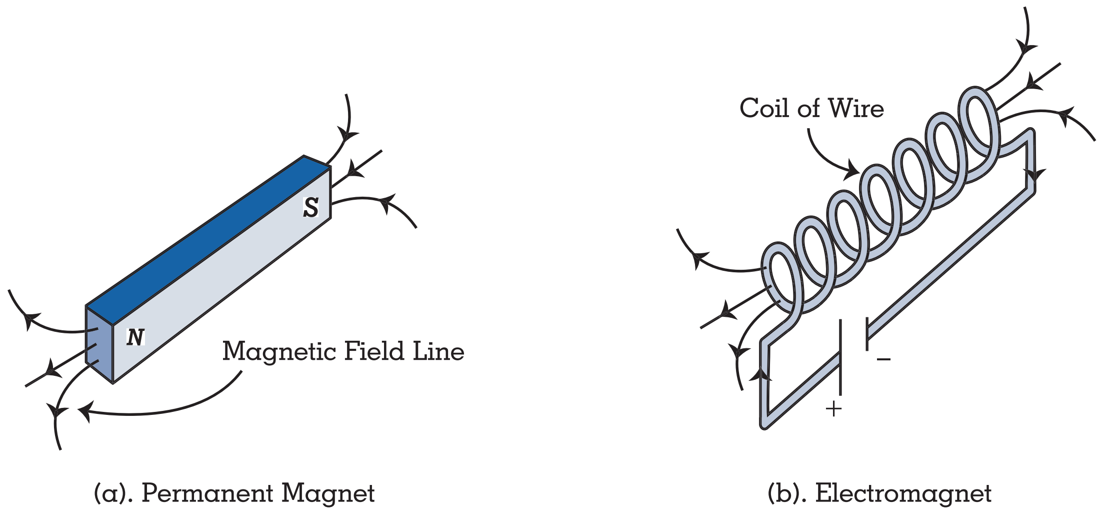


Figure 4.
Magnets with associated magnetic field lines.

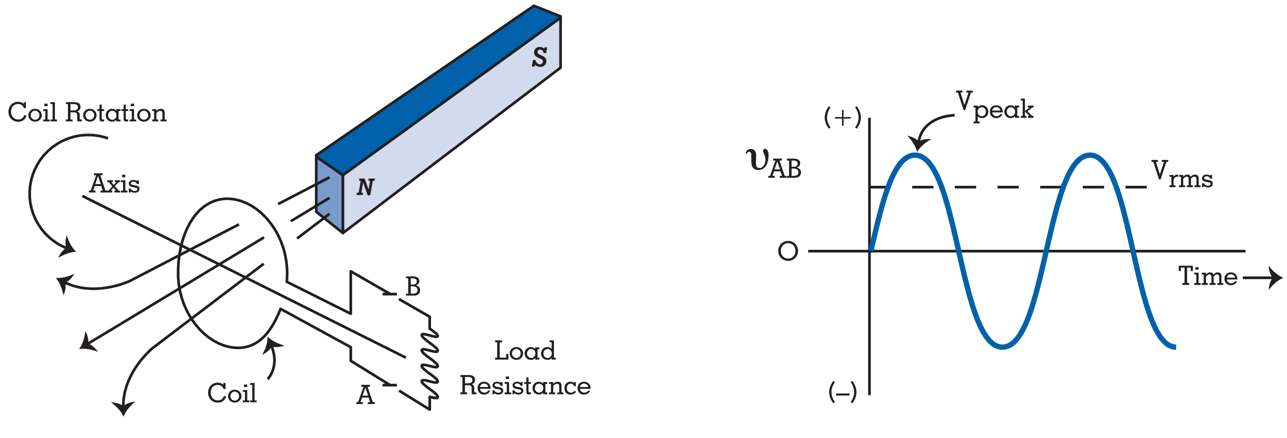
is much too large to be safe for household use, so the generated voltage must be "transformed" down to about 115 V_{rms} that is present at a regular household outlet. (But first it is actually transformed up to much higher voltage to reduce energy lost in transporting the electrical energy.)

Note that a.c. voltages (also a.c. currents) are usually measured and quoted as "r.m.s." The r.m.s. value is 70.7 percent of the peak value. This is done so that the formula to calculate electrical power in the a.c. case is the same as for the d.c. case stated earlier: power is the voltage times the current (provided both are rms values).

It is important to realize that it requires effort to turn the coil of an electrical generator if current is being supplied by the generator; that is, work must be done to turn the generator coil. Conservation of energy requires that the energy used to turned the coil (i.e., mechanical energy input to the generator) is at least as much as the electrical energy produced by the generator in a given period of time. Electrical generators have relatively high efficiency, so that the electrical power output is perhaps 90 percent of the mechanical energy input, with the remaining 10 percent lost in various heating effects in the generator. Most commercial (utility company) electrical generator coils in the United States are turned by steam engines that burn coal to obtain the energy to generate the electricity.

TRANSFORMERS AND DELIVERY OF ELECTRIC POWER

Alternating voltage (called a.c. voltage) can be quite readily changed to a different value through the use of an electrical transformer. Note that an electrical transformer will *not* transform a d.c. voltage to another d.c. voltage value; this is a principal reason why commercial electricity in the United States (and



(a). Rotating coil in magnetic field of a magnet.

(b). Plot of a.c. voltage from generator.

Figure 5.
Basic a.c. generator and voltage.

most of the world) is almost always a.c. rather than d.c. A transformer consists basically of two coils of wire, with one coil often wound on top of the other. Electrical voltage is supplied to the coil called the "primary" coil, while voltage is output from the "secondary" coil. Then the basic transformer relationship is that the ratio of the secondary voltage to primary voltage is the same as the ratio of secondary-coil-turns to primary-coil-turns.

Suppose 10,000 V_{rms} (e.g., from an electrical power plant generator) is input to a transformer where the number of secondary turns is 75 times more than the number of primary turns. Then the voltage from the secondary will be 75 × 10,000 V_{rms} = 750,000 $V_{rms.}$ This very high voltage level is actually produced by the secondary of the transformers at a modern commercial electrical power plant; the 750,000 V_{rms} is connected to the high voltage transmission lines (thick wire cables) that are used to transport the electrical power over long distances (perhaps hundreds of miles). The high voltage is used to reduce the energy lost to heating of the transmission wire resistance by the electrical current. The power lost to heating of the transmission wire is proportional to the square of the current flowing, so that the losses mount rapidly as the current flowing increases. But the power transmitted is equal to the transmission line voltage times the current flowing. For a given power, the current flowing will be much less (and energy lost in the transmission process is much less) if the transmission line voltage is very large. At the output end of the transmission line, a transformer is used again to step the voltage down to a level appropriate for local distribution (e.g., 2,400 V_{rms}), and a neighborhood transformer is finally used to step the voltage down to 220V_{rms} and 110 V_{rms} for use in homes and businesses.

Dennis Barnaal

See also: Electric Power, Generation of; Electric Power Transmission and Distribution Systems; Electricity, History of; Transformers.

BIBLIOGRAPHY

Aubrecht, G. (1995). *Energy*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall.

Faughn, J.; Chang, R.; and Turk, J. (1995). *Physical Science*, 2nd ed. Fort Worth, TX: Harcourt College Publishers.

Giancoli, D. (1998). *Physics*, 5th ed. Upper Saddle River, NJ: Prentice-Hall.

Priest, J. (1991). *Energy: Principles, Problems, Alternatives*, 4th ed. Reading, MA: Addison-Wesley Publishing Co.

ELECTRICITY, HISTORY OF

Electrical effects were known to the ancients through the attraction that amber, when rubbed, had for lightweight objects. They were also aware of the seemingly unrelated phenomenon of lightning. There is the further possibility that a form of electric battery was used for electroplating in Mesopotamia, but the evidence is meager and, even if true, there were no long-term consequences.

The term “electric” comes from the Greek word for amber and was coined by William Gilbert in his book on magnetism, published in 1600. Gilbert showed that other materials had this same attractive property and made the important observation that it was quite different from the attractive property of magnetism.

Although the subject was not abandoned, serious experimentation arguably began with the work of Francis Hauksbee, Stephen Gray, and Charles Dufay in the early decades of the eighteenth century. By rubbing glass rods to generate electric charge, and using silk threads as conductors, they were able to develop a basic understanding of conduction and to stimulate thinking about the nature of electricity.

Two inventions in the 1740s changed the electrical scene dramatically. One was the frictional machine, which made it possible to generate continuous streams of electricity relatively easily; the other was the condenser, or Leyden jar, which made possible the storage and sudden discharge of substantial quantities of electric charge.

Rubbing a simple glass rod can produce on the order of 50,000 volts, but the capacitance (the ability to store electric charge) of the surface is low and the energy available in a discharge has been estimated as less than 0.001 joule. Electrostatic machines could generate somewhat higher voltages, the ultimate being the large plate machine constructed by Martinus van Marum in Haarlem in 1785. It produced sparks up to 60 centimeters long, which (based on the size of the electrodes) translates into about 330,000 volts. Since van Marum’s machine also had a substantial prime conductor (an arrangement of brass tubes and spheres that could store the charge), the energy available in a discharge could be as much as 10 joules. With 100 large Leyden jars this increased to 30,000 joules (about 8 watt-hours).

The apparatus available to a typical experimenter would of course be more modest, but with even one or two Leyden jars, energy levels in the tens of joules are easily possible.

Benjamin Franklin was in the fortunate position of beginning his electrical experiments in 1747, just as these new devices were becoming available. He was quick to discover that the energy they delivered was capable of a substantial physiological effect, but they also made it possible to study electricity in a more systematic matter. Franklin conceived of electricity as a fluid made up of particles that were attracted to ordinary matter in a fashion similar to gravity. He also explained the action of the Leyden jar as due to electrical fluid on one side of the glass being held in place by attraction to matter that had lost its electrical fluid on the other side. This was a serviceable theory (especially when the concept of repulsion was added a few years later by Franz Aepinus) that was the basis for the mainstream of experimental work that followed.

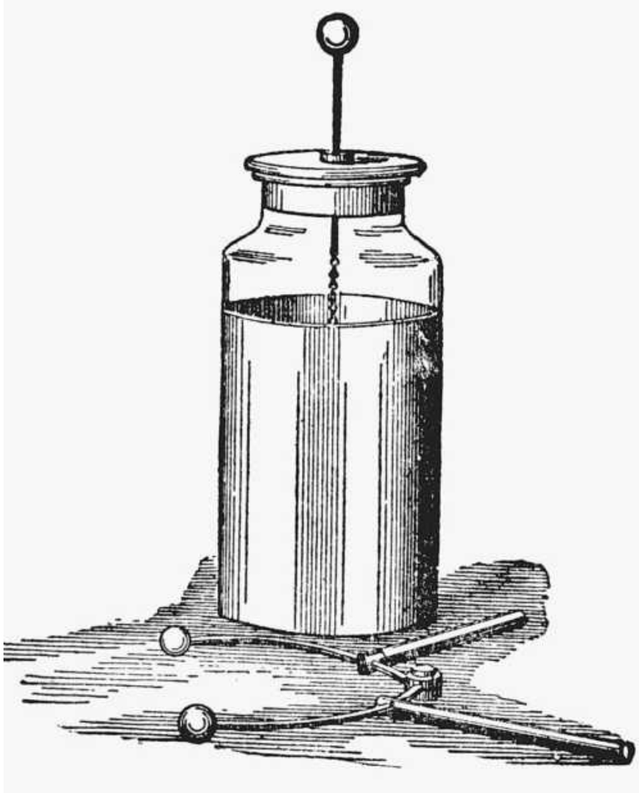
During the remainder of the eighteenth century, electricity was a popular subject for public demonstrations, but the scope of experimental work was limited. Franklin suggested a means for testing the hypothesis that lightning was electrical by using pointed conductors to draw charge from the atmosphere. This was done successfully in France (using a long iron rod) in 1752 and possibly by Franklin (using a kite) somewhat later. Much inconclusive work was done on biological and medical effects of electricity. There were some chemical experiments, including the decomposition of water. In 1785 Charles Coulomb demonstrated that the electrostatic force varied inversely as the square of the distance.

An article by Alessandro Volta in the British journal *Philosophical Transactions*, based on investigations of the previous year, stimulated a whole new range of electrical experiments and discoveries that would be important to energy history. He found that a "pile" of pairs of zinc and silver discs, separated by moistened pads, produced a continuous current of electricity. In an alternative arrangement, the liquid was contained in a series of cups. Plates of silver and zinc were placed in the cups and linked together in series from one cup to the next. Volta's announcement immediately let to a host of extensions and variations on his basic scheme, the ultimate being the Great Battery of the Royal Institution in 1810. This battery consisted of two hundred troughs, each with ten pairs of four-inch square copper and zinc electrodes. Although the energy content of these was not large, the fact that it could be discharged in a controlled fashion over time led to a number of new applications. Most important in the first few years were those in the area of chemistry. Numerous compounds were broken up and several new elements were identified. Furthermore, the basis was laid for electrical theories of the chemical bond.

These early experimenters quickly found that battery discharge was hindered by a process called polarization, or the buildup of deposits on electrodes. Thus began a long period—which continues to the present—to find a combination of materials that will efficiently convert chemical to electrical energy. It would be even better, of course, if the process were reversible, so that electrical energy could be stored in chemical form and then released when needed. A practical form of storage battery was devised by Gustav Planté in 1859, using lead and lead-oxide electrodes in sulfuric acid.

The energy available from nineteenth-century batteries varied depending on the materials used and the form of construction. The voltage from a single cell ranged between one and two volts, currents were typically a few amps, and time of discharge (more difficult to estimate) could be measured in hours. The implication is therefore that a single cell might be good for a few watt hours. This was not enough for economical "power" applications, but it was sufficient for the telegraph in the 1840s and the telephone in the 1870s.

In the 1880s there was great interest in trying to develop storage cells efficient enough and with enough capacity to act as load equalizers for the emerging power industry, and at the end of the century there was further pressure for a design that would power an electric car. This was a problem that remained intractable not only through the nineteenth century but through the twentieth as well.


The Leyden jar and discharger used to collect electric fluid in a bottle half filled with water (undated engraving). (Corbis Corporation)

Several electrical scientists in the early part of the nineteenth century, influenced at least in part by their understanding of German *naturphilosophie*, expected forces of nature to be intimately connected to each other, and some of them spent extraordinary amounts of time looking for the relationship. One of these was a Dane, Hans Christian Oersted, who, after an exhaustive series of experiments, in 1820 found that electricity could indeed produce a magnetic effect. Further experiments by Michael Faraday demonstrated, in 1821, that by proper orientation of an electric current and a magnetic field it was possible to produce continuous motion in what soon would be called a motor. It took an additional ten frustrating years for him to prove what he instinctively felt to be true, that, in a fashion inverse to what

Oersted had discovered, magnetism somehow could produce electricity. He called his device an electromagnetic generator.

For the next thirty-five years motors and generators were little more than classroom demonstration devices, limited as they were by the magnetic fields of permanent magnets. Generators were employed in a few specialized cases, for instance in some isolated lighthouses where large magneto machines were used to power arc lamps, notably by the Société l'Alliance in France, but this was the exception. The best of the machines were bulky, weighing on the order of 2,000 kilograms, and produced a little over two kilowatts at an efficiency of under 20 percent.

A breakthrough came in 1867 with publications by Charles Wheatstone, Werner Siemens and S. Alfred Varley that they had, independently, invented generators in which both the armature and the field magnets were electromagnets. Similar constructions were described by others but did not lead to practical consequences. These generators were, in other words, "self-excited." This was accomplished because the core of the field coils kept a small amount of residual magnetism that produced a small current in the moving armature, part of which was fed back into the field coils, which very rapidly reached their full magnetic strength. Improvements in armature and field design led to further efficiencies. These were successfully applied by the Belgian Zenobe Gramme, who by 1872 was manufacturing commercial dynamos (self-excited generators). His first machines weighed about half as much as the Alliance magnetos and produced more electricity at higher efficiency. In 1876, direct comparisons of magneto generators (Alliance) and dynamos (Gramme and Siemens) showed that a dynamo of as little as one tenth the weight, and little more than one tenth the cost, for the same input horsepower had more than twice the efficiency of the magneto.

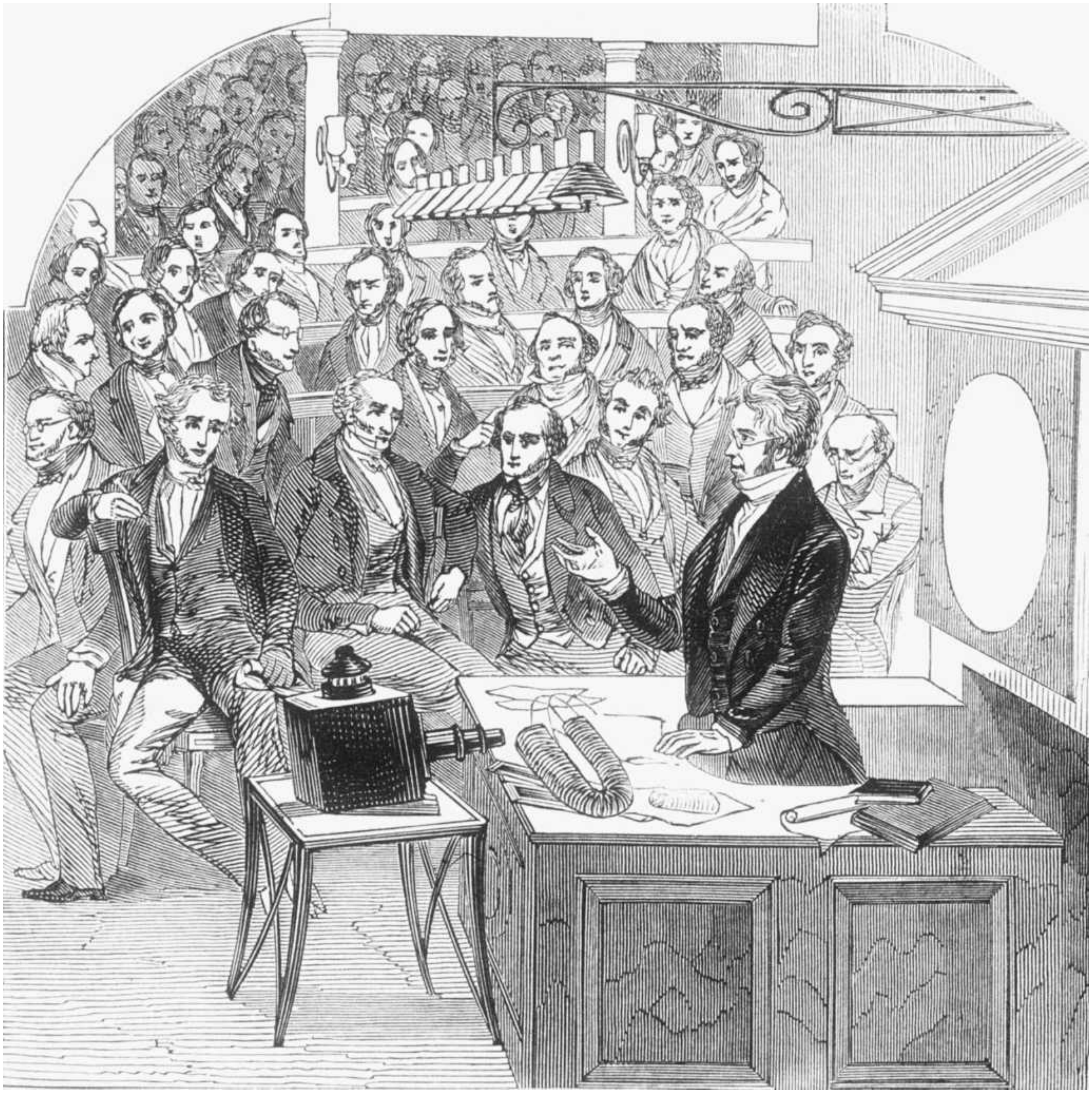
The new dynamos made arc lighting economically feasible, especially for street lights. Initial installations were made in Paris in 1877, London at the end of 1878, San Francisco in 1879, and New York in 1880. The other principal commercial application in those early years was for electroplating.

The success of arc lighting in turn stimulated the search for a less intense incandescent lamp. It was Thomas Edison who, after more than a year of exhaustive effort, won this race. But for our purposes it is important to note that at the same time he was seeking improvements in the dynamo. Others reasoned that the best system was one in which the internal resistance of the dynamo equaled the external resistance, since this would maximize the amount of energy provided to the external load. Edison, however, was quick to point out that this was hardly a bargain because it limited the efficiency to 50 percent. He therefore designed generators with low internal resistance and was able to achieve efficiencies in excess of 90 percent. His first successful dynamos, with which he lit his first lamps, produced about five kilowatts. In 1881, the "Jumbo" dynamo, which he shipped to the Paris electric exhibition, produced approximately 120 kilowatts.

The 1880s saw a phenomenal growth in the number of (mainly incandescent) electric lighting systems, especially in the United States. After 1886, street railways also came into use. Electric lights could run on either alternating (ac) or direct current (dc), but motors for street cars or any other purpose needed dc. There was also the possibility that truly practical storage batteries would be developed that could serve as load equalizers for dc systems. Finally, there was evidence that for the same voltage, alternating current was more dangerous.

No effective storage battery appeared, however, and the safety argument was for most people not a determining factor. Therefore, when Nikola Tesla invented a practical induction motor in the late 1880s, the advantages of dc became inconsequential. Alternating current could easily be transformed from low to high voltage and back, and at high voltage it could be transmitted over long distance with relatively low loss. The "battle of the systems" could be said to have been officially over with the merger of the Edison company and Thomson-Houston in 1892, and especially with the adoption of ac for the planned Niagara Falls generating plant in 1893. Westinghouse built ten two-phase, 25-Hz alternating current dynamos based on several Tesla patents for the initial Niagara power station; two were installed in 1895, the final one in 1900. Each was rated at 5000 horsepower. Electricity was transmitted at the generated 2200 volts to local industrial plants and at 11,000 volts (converted to three phase) to Buffalo.

In the United States the early decades of the new century were dominated by ever-increasing size: steam-turbine generators with capacities that reached over a thousand megawatts, and transmission voltages (the highest ones now often dc) above 700,000 volts. But after mid-century these "economies of



Members of the Royal Institution attend a lecture given by Michael Faraday on magnetism and light (London, England, 1846). (Corbis-Bettmann)

scale" were proving not to be so economical after all, and there was a decided move toward smaller gas-turbine co-generation plants.

Another legacy of the late nineteenth century was identification of the electron by an appropriate interpretation of the Pieter Zeeman effect in 1896, and more especially by J. J. Thomson's experiments the following year. Starting in the 1920s, physicists devised machines that could accelerate electrons (or protons) to increasingly higher energies and then cause them to collide with themselves or with atoms of various elements. This proved to be a very effective means for studying the fundamental properties of matter, with the result that higher and higher energies

William Shockley (seated), John Bardeen (standing, left), and Walter H. Brattain doing transistor research at Bell Telephone Laboratories (New York, 1948). (Corbis Corporation)

were sought. By the end of the century, bunches of electrons in the Stanford Linear Accelerator were being raised to the level of 50 billion electron volts. On the average, in continuous operation, this means a power level of tens of kilowatts. Individual bunches, operating over very short periods of time, attain power levels approaching a million billion kilowatts.

On a more practical level, Edison's observation in 1881 that the filaments of his lamps emitted negatively charged particles of carbon led to John Ambrose Fleming's invention of a two-element "valve" in 1904 and Lee de Forest's three-element "audion" in 1906. De Forest thought that gas in his audion tube was essential to its operation, but by 1912 several people realized that, to the contrary, a critical feature was a really good vacuum. This made possible the rapid development of effective vacuum tube amplifiers for wireless sets, for long-distance telephone lines (transcontinental service opened in 1915), for radio receivers (after 1920), and eventually for television, radar and microwave systems. As rapid switches vacuum tubes were also critical to the development of early computers.

Of special interest is the fact that with a good vacuum, these tubes, in properly arranged feedback circuits, could "oscillate" and produce radio waves. By the mid-twenties they were competing effectively with the popular high-frequency alternating dynamos, the largest of which generated 200 kilowatts. A typical high-power radio station was, however, operated at 50 kilowatts.

In the 1940s, higher-frequency devices—notably the magnetron and the klystron—produced radiation in the microwave region, which was especially useful for radar. By the early postwar period, these could easily generate continuous signals in the range of several kilowatts, with pulses in the megawatt region.

Even higher frequencies were achieved by the maser, conceived by Charles Townes in 1951, and by the laser, first demonstrated by Theodore Maiman in 1960. The maser, operating in the microwave region, has been used mainly in amplifiers. Its optical- (and near optical-) frequency sibling, the laser, has found a variety of uses, ranging from eye surgery to weapons, where the amount of energy that can be released is important. Some versions have operated in continuous fashion at one kilowatt and above, or with very short pulses up to 100 megawatts.

The study of electrons trapped in matter (commonly termed "solid state") led eventually to the invention of the transistor in 1947 by Walter Brattain, John Bardeen, and William Shockley at Bell Laboratories, and then to the integrated circuit by Robert Noyce and Jack Kilby a decade later. Use of these devices dominated the second half of the twentieth century, most notably through computers, with a significant stimulus to development being given by military expenditures.

Another way of considering the changing "face" of electricity is to look at the professional organizations it has spawned. Through much of the nineteenth century electricity was produced at low energy levels (mostly by batteries) and employed in operations (mainly the telegraph) that were much more complicated mechanically than electrically. The self-excited dynamo and the incandescent lamp changed all that, and by the mid-1880s schools of electrical engineering began to appear in American colleges. This shift can be seen in Britain where the Society of Telegraph Engineers, founded in 1871, became the Institution of Electrical Engineers (AIEE) in 1888. In the United States, where there was no prior engineering society (although there were journals), some like-minded practitioners joined together to form the American Institute of Electrical Engineers in 1884. The significance of the new wireless-electronic technology became evident with formation of the Institute of Radio Engineers (IRE) in 1912.

Merger of the AIEE and the IRE in 1963 into the Institution of Electrical and Electronics Engineers (IEEE) preserved in its name a sense of the earlier bifurcation of the profession, but in fact the lines had become quite blurred and the numerous constituent societies drew and continue to draw from both traditions. At the end of the century, the IEEE was easily the largest professional organization in the world with membership over 350,000. Its thirty-six constituent societies provided an indication of the degree to which electricity has pervaded the modern world, ranging from "Communications" and "Lasers & Electron Optics" to "Neural Networks" and "Engineering in Medicine & Biology," from "Power Electronics" and "Vehicular Technology" to "Aerospace and Electronic Systems" and "Robotics & Automation." Significantly, the largest constituent group was the Computer Society.

Bernard S. Finn

See also: Cogeneration Technologies; Edison, Thomas Alva; Electricity; Electric Motor Systems; Electric Power Transmission and Distribution Systems; Matter and Energy; Regulation and Rates for Electricity; Siemens, Ernst Werner von; Tesla, Nikola; Thomson, Joseph John; Townes, Charles Hard; Turbines, Gas; Turbines, Steam; Volta, Alessadro; Wheatstone, Charles.

BIBLIOGRAPHY

Adams, E. D. (1927). *Niagara Power: History of the Niagara Falls Power Company, 1886–1918.* Niagara Falls, NY: The Niagara Falls Power Company.

Bowers, B. (1982). *A History of Electric Light and Power.* Stevenage, UK: Peter Peergrinus, Ltd.

Finn, B. (1991). *The History of Electrical Technology: An Annotated Bibliography.* New York: Garland Publishing, Inc.

Heilbron, J. L. (1979). *Electricity in the 17th and 18th Centuries: A Study of Early Modern Physics.* Berkeley: University of California Press.

Hughes, T. P. (1983). *Networks of Power: Electrification in Western Society, 1880–1930.* Baltimore: Johns Hopkins University Press.

King, W. J. (1961–1962). "The Development of Electrical Technology in the 19th Century." *U.S. National Museum Bulletin* 228. Papers 28–30:231–407.

McMahon, A. M. (1984). *The Making of a Profession: A Century of Electrical Engineering in America.* New York: IEEE Press.

Riorden, M., and Hoddeson, L. (1997). *Crystal Fire: The Birth of the Information Age.* New York: W. W. Norton & Company.

ELECTRICITY REGULATION AND RATES

See: Regulation and Rates for Electricity

ELECTRIC LEAKAGE

See: Electricity

ELECTRIC MOTOR SYSTEMS

Electric motors are everywhere. These ubiquitous devices come in a wide variety of sizes and power outputs, ranging from a fraction of a watt to huge multikilowatt applications. Tiny ones operate computer disk drives, power windows/mirrors, and windshield wipers; moderate-sized ones run appliances such as fans, blenders, electric shavers, and vacuum cleaners; a large drive pumps, elevators, sawmills, and electric trains and vehicles.

There has been not only growth in the total number of electric motors (more standard appliances in use), but also a proliferation in their use for new, novel applications. Both trends will continue to increase demand for the electricity to run electric motors. In the United States, electric motors are responsible for consuming more than half of all electricity, and for the industrial sector alone, close to two-thirds. Since the cost of the electricity to power these motors is enormous (estimated at more than $90 billion a year), research is focused on finding ways to increase the energy efficiency of motors and motor systems.

EARLY DEVELOPMENT

Electric motors are devices that convert electrical energy into mechanical energy. Devices that do the opposite—convert mechanical energy into electrical energy—are called generators. An important example is power plants, where large gas, steam, and hydroelectric turbines drive generators to provide electricity. Since generators and motors work under the same principles, and because construction differences are minimal, often generators can function as motors and motors, as generators, with only minor changes.

The early development of electric motors and generators can be traced to the 1820 discovery by Hans Christian Oersted that electricity in motion generates a magnetic field. Oersted proved the long-suspected promise that there is indeed a relationship between electricity and magnetism. Shortly thereafter, Michael Faraday built a primitive electric motor, which showed that Oersted's effect could be used to produce continuous motion.

Electric motors consist of two main parts: the rotor, which is free to rotate, and the stator, which is stationary (Figure 1). Usually each produces a magnetic field, either through the use of permanent magnets, or by electric current flowing through the electromagnetic windings (electromagnets produce magnetism by an electric current rather than by permanent magnets). It is the attraction and repulsion between poles on the rotor and the stator that cause the electromagnet to rotate. Repulsion is at a maximum when the current is perpendicular to the magnetic field. When the plane of the rotator is parallel to the magnetic field (as in Figure 1) there is no force on the sides of the rotator. The left side of the rotator receives an upward push and the right side receives a downward push; thus rotation occurs in a clockwise direction. When the rotator is perpendicular to the magnetic field, forces are exerted on all its sides equally, canceling out. Therefore, to produce continuous motion, the current must be reversed when reaching a vertical direction. For alternating-current motors this occurs automatically, since the alternating magnetic attraction and repulsion change directions 120 times each second; for direct current-motors, this usually is accomplished with a device called a commutator.

Before useful electric motors could be developed, it was first necessary to develop practical electromagnets, which was primarily the result of work done by Englishman William Sturgeon as well as Joseph Henry and Thomas Davenport of the United States. In 1873 Zénobe-Théophile Gramme, a Belgian-born electrical engineer, demonstrated the first commercial electric motor. A decade later Nikola Tesla, a Serbian-American engineer, invented the first alternating-current induction motor, the prototype for the majority of modern electric motors that followed. Magnetic fields today are measured in units of teslas (symbol T), to acknowledge Tesla's contribution to the field.

MOTOR TYPES

The electrical power supplied to electrical motors can be from a direct-current (dc) source or an alternating-current (ac) source. Because dc motors are more

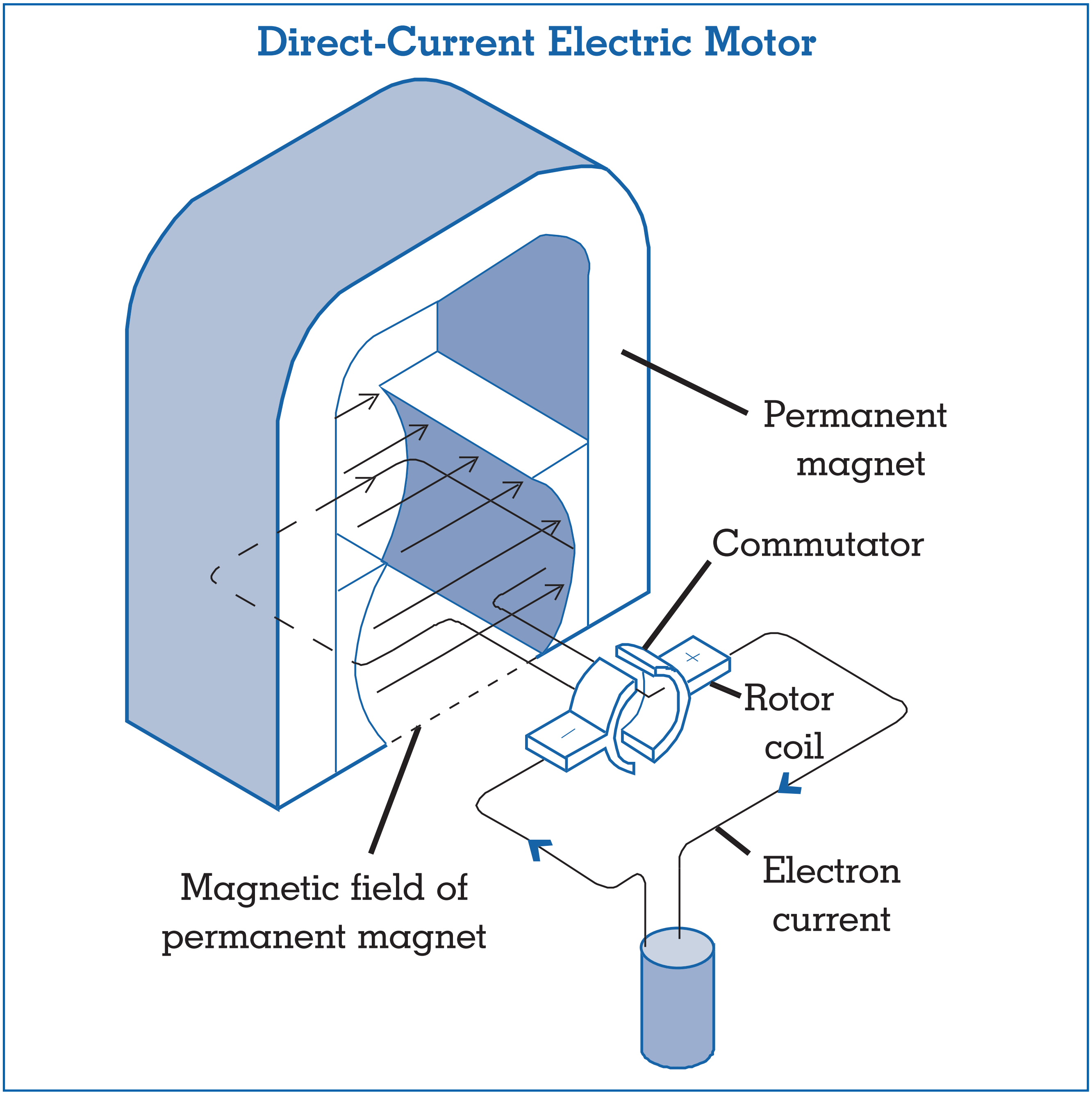

Direct-Current Electric Motor

Figure 1.
The attraction and repulsion between the poles on the rotor and stator cause the electromagnet to rotate. Direct-current motors usually need commutators to achieve continuous motion.

expensive to produce and less reliable, and because standard household current is ac, ac motors are far more prevalent. However, the growing demand for portable appliances such as laptop computers, cordless power drills, and vacuums ensures a future for dc motors. For these portable applications, the alternating current charges the battery, and the direct current of the battery powers the dc motor.

Based on the way magnetic fields are generated and controlled in the rotor and the stator, there are several additional subclassifications of direct- and alternating-current motors.

Direct Current Motors

Direct-current motors are classified as separately excited motors, series motors, shunt motors, and compound motors. The field winding of a separately excited motor is in a circuit that is energized by a separate dc source; the field winding is not physically connected to the armature circuit (containing the armature winding).

For series, shunt, and compound motors, only one power supply is needed. In a series motor the field winding is connected in series with the power supply and the armature circuit. In a shunt motor the field winding is connected across, or in parallel with, the dc supply, which energies the armature winding. In a compound motor, also known as a cumulative compound motor, the field winding is physically in parallel with the armature winding circuit and is magnetically coupled to a coil or a winding in series with the armature winding.

Direct current motors are most appropriately used in applications where a dc power supply is available or where a simple method of speed control is desired. The fans used in automobile heating and air conditioning systems are driven by direct-current motors.

Alternating-Current Motors

Alternating-current motors are classified as induction motors or synchronous motors. Faraday found that a stationary wire in a magnetic field produced no current. However, when the wire continues to move across magnetic lines of force, it produces a continual current. When the motion stops, so does the current. Thus Faraday proved that electric current is only produced from relative motion between the wire and magnetic field. It is called an induced current—an electromagnetic induction effect.

Induction motors usually entail insulated wiring windings for both the rotor and the stator, with the stator connected to an external electric power source. Between the narrow gap of the stator and the rotor, a revolving magnetic field is established. A current can be established only when the waves of the rotor and stator windings are not in phase—not at a maximum simultaneously.

The induction motor is the most common motor in industrial applications, and are also very prevalent for smaller applications because of their simple construction, reliability, efficiency, and low cost. Historically they have been found in applications that call for a constant speed drive, since the alternating-current power supply is of constant voltage and frequency; however, the continual development of more powerful and less inexpensive solid-state electronic devices has allowed for electronic inverters (which control voltage and frequency) to more accurately control the speed and torque of induction motors, thereby matching the control performance of a motor. Since induction motors are less expensive, more compact, more efficient, and more reliable (better voltage overload capabilities), it is likely that induction motors will continue to replace motors in most applications.

Synchronous motors operate like induction motors in that they rely on the principle of a rotating magnetic field, usually produced by the stator. Synchronous motors differ in that the rotor generates a constant unidirectional field from a direct-current winding powered by a direct-current source. This field interacts with the rotating field. To get around the need for direct current-power, the stator can be constructed from permanent magnets so that the permanent magnetic field and rotating field can be synchronized. Synchronous motors are most useful for low-load applications where constant speed control is crucial, such as in phonographs, tape recorders, and electric clocks.

ADVANCES IN PERFORMANCE, EFFICIENCY, AND RELIABILITY

Innovations in designs and materials led to continual advances in the performance, efficiency and reliability, of electric motors throughout the Twentieth century. The best results for motor designers occur when starting with the function at hand and working back toward the power source, optimizing each element along the way, foremost being the improvement of the end use of mechanical energy. If end use mechanical energy is curtailed, the demand for power generated from an electric motor declines, and consequently the demand for electricity to run the motor. For a factory or warehouse conveyer belt system, a streamlined design, better bearings, and lighter components can yield far greater energy savings than replacing a standard-efficiency motor with an energy-efficient model.

Performance

There are few shortcuts. If you want a powerful electric motor, it is going to have to be large and entail an extensive amount of copper windings. That is why it is much more cost-effective to rebuild many large (more than 100 horsepower) industrial motors than to replace them with new motors.

Efficiency

In theory, electric motors can be more than 95 percent efficient. Since electric motors can to convert almost all the electrical energy into mechanical energy, it partly explains the continued growth of electrical technology at the expense of competing technologies. It is widely believed that the electric motor will eventually replace even the internal-combustion engine (in which only about 25 percent of the heat energy is converted to mechanical energy) once the costs of better-performing battery and fuel cell technologies decline.

In practice most electric motors operate in the 75 to 90 percent range, primarily because of core magnetic losses (heat losses and electric current loss), copper resistive losses, and mechanical losses (friction in the bearings, and windings, and aerodynamic drag). At low speeds core magnetic losses are greatest, but as speeds get higher, core magnetic losses decline, and copper resistive losses become more dominant. The mechanical losses do not vary much, remaining fairly constant at both low and high speeds.

There were no standards for energy-efficient motors until the National Electrical Manufacturers Association (NEMA) developed design classifications for energy-efficient, three- phase induction motors in 1989. This standard was made the national minimum efficiency level by the Energy Policy Act of 1992, which went into effect in October 1997. Manufacturers responded to this higher efficiency level by reducing losses through the use of better materials, improved designs, and precision manufacturing.

Aside from the efficiency of the motor itself, energy efficiency is very dependent upon proper sizing. While the efficiency of a motor is fairly constant from full load down to half load, when a motor operates at less than 40 percent of its full load, efficiency drops considerably, since magnetic, friction, and windage losses remain fairly constant regardless of the load. Moreover, the power factor drops continuously as the load drops. The problem is most discernible in small motors.

The obvious answer is to properly size the motor for an application. However, properly sizing a motor is difficult when motors are required to run at a wide spectrum of loads. In many cases a decrease in the motor's speed would reduce the load while maintaining the efficiency.

A major hurdle to greater efficiency is the constant-speed nature of induction and synchronous motors. Nevertheless, considerable advances have been made in improving motor speed controls that essentially better optimize the motor speed to the task at hand, resulting in substantial energy savings, decreased wear of the mechanical components, and usually increased productivity from the user.

Usually the lowest first cost solution is to use multispeed motors with a variety of torque and speed characteristics to match the different types of loads encountered. The more costly and more energy-efficient choice is to use electronic adjustable speed drives that continuously change the speed of AC motors by controlling the voltage supplied to the motor through semiconductor switches. Energy efficiency is also achieved by converting the 60-hertz supply frequency to some lower frequency, thereby enabling induction motors to operate at slower speeds, and thus consuming less energy. Slower speeds may be desirable for many applications such as fans and conveyor belts. There are several different types of electronic adjustable-speed drives, yet no one technology has emerged as superior to all others. That is because the multitude of different motors, different sizes (horsepower) and speeds, and control requirements make it difficult for one control system technology to be superior for all applications.

The cost premium for a motor equipped with speed control can be substantial, sometimes costing twice that of a single-speed motor. But the energy savings from speed control can be substantial, especially for fan and pump systems. Electronic adjustable speed drives continue to become more attractive because the costs of microelectronics and power electronics technologies continue to fall as performance and energy efficiency improve.

Reliability

Electric motors have proven to be very reliable and continue to become more reliable because of better materials and designs. However, because of the excellent reliability record, long life cycles, and the lower first cost of rebuilding motors instead of purchasing new, more energy-efficient models, it will be decades before energy-efficient motors significantly penetrate the market.

Even when the time comes to make a purchasing decision, an energy-efficient motor purchase is not a certainty. Sometimes an energy-efficient motor will be the economically efficient choice; at other times, not. The capital investment decision is based on the cost in relation to performance, efficiency and reliability. Moreover, the decision depends on the application and the amount of time the motor is in operation. It can be the major component of a product (drill or mixer), or a minor component (computer disk drive); it can be the major component cost of a product (fan), or it can be a minor component cost (stereo tape deck); it can run almost constantly (fan, pump, and machinery), or only a few minutes a day (vacuums and power tools). For example, contractors purchase circular saws almost solely based on performance and reliability. Time is money, and since the saw is operating only a few minutes a day and the contractor is often not responsible for the electricity costs to run the motor, energy efficiency is not a consideration; performance and reliability are what matter most. On the other hand, an industrial user, who runs huge electric motors twenty-four hours a day to work pumps, machinery, and ventilation equipment, is very concerned with energy efficiency as well as performance and reliability.

John Zumerchik

See also: Batteries; Capital Investment Decisions; Consumption; Economically Efficient Energy Choices; Electricity; Electric Power, Generation of; Faraday, Michael; Fuel Cells; Fuel Cell Vehicles; Magnetism and Magnets; Oersted, Hans Christian; Tesla, Nikola.

BIBLIOGRAPHY

de Almeida, A. T., and Greenburg, S. (1997). "Energy Efficient Technologies: Electric Motor Systems Efficiency." In *CRC Handbook of Energy Efficiency*, ed. F. Kreith and R. E. West. Boca Raton, FL: CRC Press.

Beaty, H., and Kirtley, J. (1998). *Electric Motor Handbook.* New York: McGraw-Hill.

Electric Power Research Institute. (1992). "Electric Motors: Markets, Trends and Applications." EPRI Report TR-100423. Palo Alto, CA: Author.

Nadel, S.; Shepard, M.; Greenburg, S.; Katz, G.; and Almeida, A. (1992). *Energy-Efficient Motor Systems: A Handbook on Technologies, Programs and Policy Opportunities.* Washington, DC: American Council for an Energy Efficient Economy.

ELECTRIC POWER, GENERATION OF

INTRODUCTION

Electric power systems can be thought of as being comprised of three important sectors: generation, transmission, and distribution. For most utilities, generation capital equipment costs account for approximately 50 percent of total plant in costs. Generation also accounts for close to 75 percent of total operation and maintenance expense.

Generation is the production process center of the power industry. This production process is multifaceted and starts with the conversion of primary energy, such as fossil fuels, uranium, and the kinetic energy of water, to electrical energy. The process by which this primary energy is converted to electricity varies depending upon the prime mover, or technology, of the power generator. Mainstream generation technologies include hydroelectric facilities, internal combustion or combustion turbine facilities, and steam generation facilities. Alternative electric generation can include prime movers powered by the wind, sun, or some other renewable fuel such as biomass or solid waste.

Hydroelectric facilities use the kinetic energy of falling water to turn a water turbine to create electricity. These facilities usually have limited technical applicability and are located in geographic regions that meet certain elevation, water level, and stream flow requirements. The advantage of hydroelectric facilities is that they are virtually free of fuel costs. Their disadvantage, in addition to their limited geographic applicability, is that they have relatively high capital costs.

Historically, internal combustion engines and combustion turbines have been considered unique and limited types of generation facilities. These are similar in many respects to engines used in the auto-

Interior of electrical power plant. (JLM Visuals)

motive and aeronautics industries. Both technologies burn one or a combination of various fossil fuels (oil, diesel, propane, or natural gas) to create mechanical energy to turn electric generators. The advantage of these technologies is that they have relatively low capital costs. Since these technologies are run on fossil fuels, their disadvantage rests with their relatively high and potentially volatile operating costs.

The traditional power generation work horse has been the steam generator. This technology uses fossil fuels (coal, natural gas, and oil) to heat water in a boiler to create steam. The steam, in turn, drives an electricity generator. Nuclear power is a special case of the steam generator, using uranium and nuclear fission to create steam. While these facilities have high capital costs, their operating costs are lower than their nearest competitor, combustion turbines.

On the fringe of electricity generation technologies are alternate-fuel generators, including solar photovoltiac and thermal applications, wind powered applications, and the use of waste agricultural byproducts such as rice hulls or bagasse, and even garbage. These technologies have traditionally played a small role in the overall generation portfolio of utilities in the United States, given their relatively high cost and limited capacity. Most of these technologies have been promoted under the auspices of research and development or within the context of relatively unique niche applications.

GENERATION PLANNING AND OPERATION

In the past, generation planning consisted of developing and maintaining a portfolio of facilities to meet the various types of electricity loads that occur in any given hour, across any given day, in any given season. Reliability tended to be the most important planning

Power plant exterior. (JLM Visuals)

consideration, followed closely by cost. Thus, generation planning strategies consisted of constructing and operating enough power plants to meet demand on a cost effective basis. In many instances, having the ability to meet sudden surges in demand entailed constructing and maintaining large capacity reserve margins that remained idle during large parts of the year.

Since load varies considerably across hour, day, and year, utilities have traditionally segmented their generation facilities into three classifications: baseload generation, intermediate or cycling generation, and peaking generation. Baseload generators are typically steam generation facilities used to service minimum system load, and as such are run at a continuous rate. While these units are the most efficient to operate, they are costly to start up from a cold shut down (designed for continuous combustion); therefore, they are usually run at a near-constant rate. Intermediate load plants are typically older steam units or combustion turbines brought on line during periods of forced or planned outage of baseload units. Intermediate units can also be thought of as units that bridge the dispatch of baseload and peaking units during periods of unusually high demand. These units can be older and less efficient than baseload units. Peaking units are typically combustion turbines that have the ability to generate electricity immediately, and serve temporary spikes in demand such as during a heat wave when residential and commercial air conditioning demands begin to surge.

In the past, electric utilities dispatched generating units to meet demand on a lowest-to-highest cost basis. This form of dispatch is commonly referred to as "economic dispatch." The marginal or incremental cost of dispatching units is traditionally the benchmark used to rank order available generators. These marginal costs, in the very short run, are typically

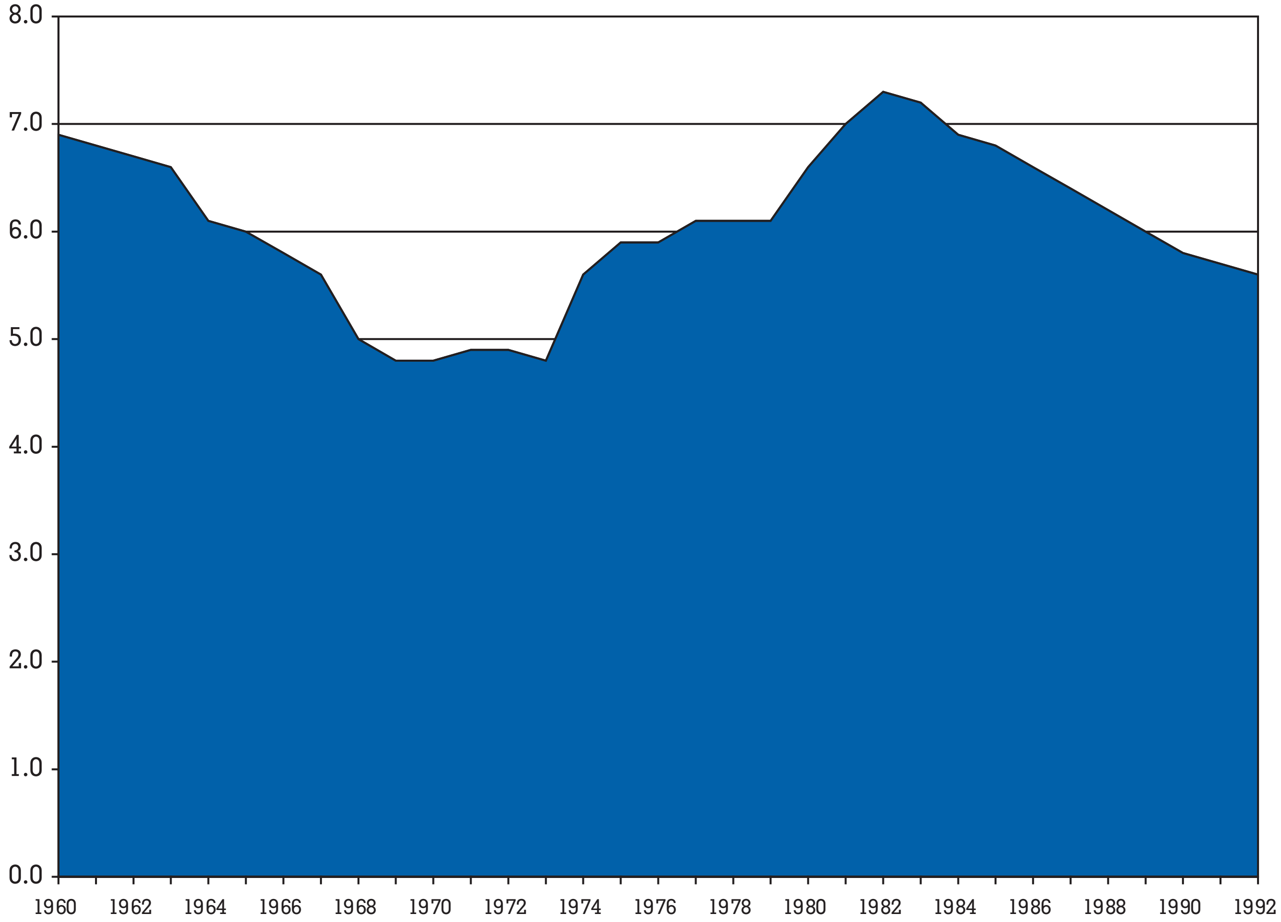


Figure 1.
Real Electricity Prices (1995 cents per kWh)
SOURCE: Energy Information Agency. (1995). *Electric Power Annual.* Washington, DC: U.S. Department of Energy.

associated with changes in fuel costs and other variable operating and maintenance (O&M) costs. Historically, baseload units, which are almost always large coal, hydro, or nuclear units, had the lowest incremental costs and were dispatched first to meet load. As load increased during the day, or across seasons, less efficient intermediate or cycling units, which generate electricity at slightly higher costs, were brought on line. Higher cost peaking units would be the last types of units brought on line, for example, during a heat wave with a resulting large demand for air conditioning. The cost of the last dis-

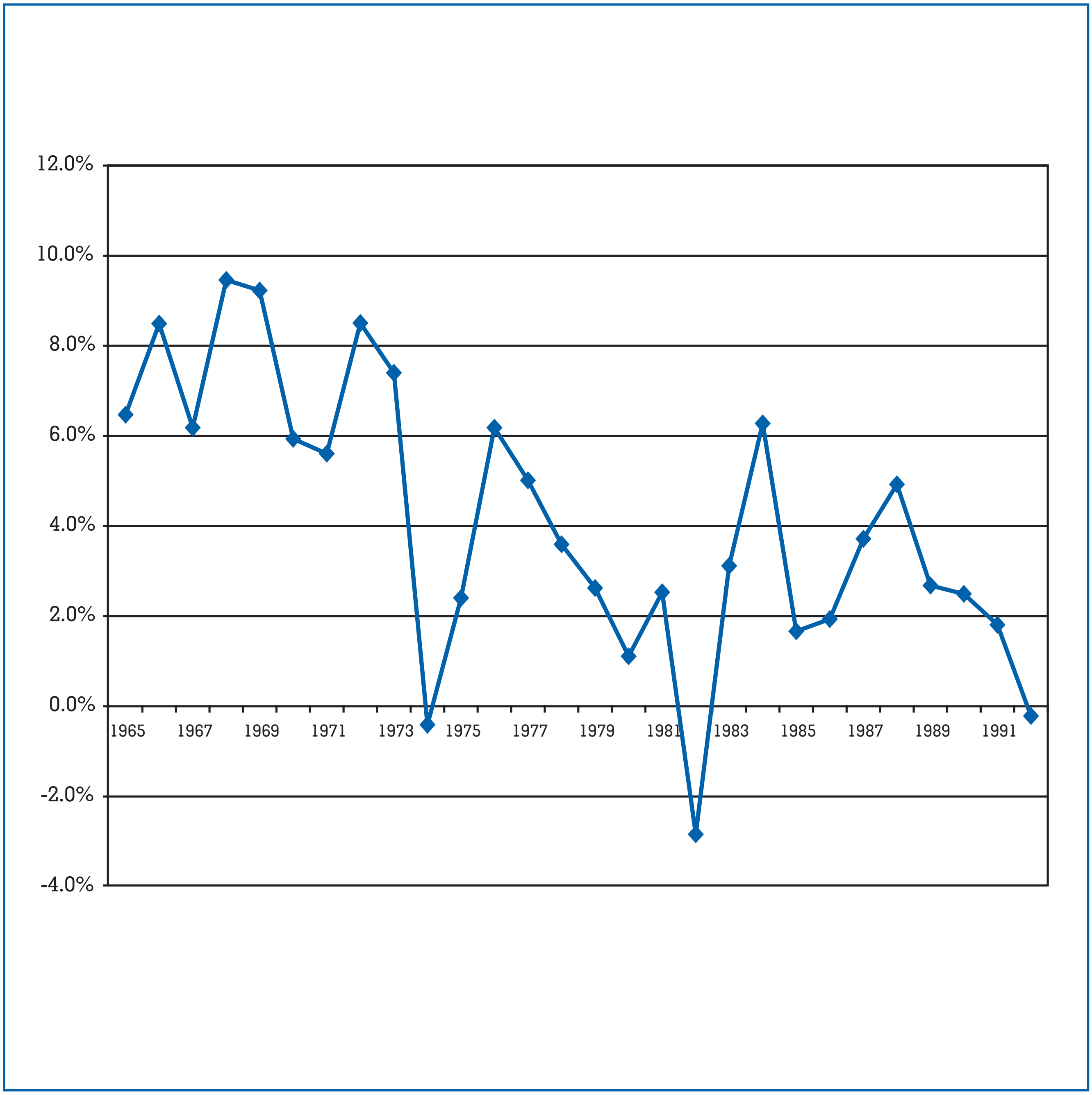


Figure 2.
Annual Rates of Electricity Demand Growth
SOURCE: Energy Information Agency. (1995). *Electric Power Annual.* Washington, DC: U.S. Department of Energy.

patched unit would therefore define the system marginal costs, often referred to as the system "lambda."

Given the importance and relative size of the electricity generation sector, shifts in the costs of constructing and operating electric power plants can have considerable influence on final rates of electricity for end users. Figure 1 shows the historic trend of real, or inflation-adjusted, electricity prices over time. Noticeable in the graph are the spikes that occurred in the early 1970s and again in the early 1980s, when utilities in the industry were faced with a number of almost insurmountable challenges leading to the undermining of their unique natural monopoly cost advantages.

The history of electricity generation planning can be broken into two distinct periods: one period proceeding, and one period following the energy crisis of 1973. Prior to 1973, electricity generation planning was a relatively straightforward endeavor. During this period, forecasted increases in load were met with the construction of new generation facilities. Utilities typically tried to meet this load with the most cost effective generation technology available at the time. As shown in Figure 2, the annual rate of electricity demand prior to 1973 grew at an annual average rate between 6 and 10 percent. This constant, significant growth placed many utilities in the position of having multiple construction projects ongoing at any given time.The period following the energy crisis of 1973 dramatically changed the generation planning process for utilities. During this time the industry was plagued by high inflation and interest rates, high fuel prices, financial risk, and regulatory uncertainty. These uncertainties increased costs, which in turn had a deleterious affect on electricity demand. Dramatic electricity price increases, resulting from the volatile operating environment of the post–1973 environment, stifled the growth of electricity demand and set strong incentives to end users to conserve electricity. As a result, utilities found themselves with considerable excess capacity that quickly became technologically and economically obsolete. This excess and uneconomic capacity, in combination with eventual emergence of new technologies and increased competition, resulted in an undermining of the natural monopoly justification for electric utility regulation.

THE ECONOMIC REGULATION OF POWER GENERATION

Historically, the electric power industry was characterized as being a natural monopoly. Natural monopolies typically occur in industries with very large fixed capital costs and relatively low operating costs. The cost characteristics of these industries tend to make them the most efficient producers in a given regional market. However, since these natural monopolies face no competition, they have the ability, if left unchecked, to charge prices that could be considerably above costs. Industries with large infrastructure requirements, such as telecommunications, water and wastewater, natural gas, and electric power, have historically been considered natural monopolies.

Many industrialized nations grapple with the unchecked power that infrastructure industries can have in any given market. In these instances, government has two public policy options. First, government can expropriate, or nationalize, these industries. Here the government takes over power generation ownership and operates the industry in the public interest by providing service at a reasonable (government determined) price.

Under the second policy option, the government can maintain private ownership and regulate firms operating in the public trust. This has been the unique policy option exercised within the United States for a greater part of the past century. Despite some municipal and federal government ownership of electricity generation facilities, much of electric generation capacity is investor-owned, privately controlled electric companies. Figure 3 shows the electricity generation capacity ownership percentages by type of entity.

Beginning in the 1920s, an extensive set of electric power industry regulation arose based upon the notion that this industry, like others, is a natural monopoly. An additional rationale for power industry regulation has been that electricity, like so many other regulated utility industries, is imbued with the public interest. Perhaps one of the greatest influences on power generation over the past half century has been the role of government and its public policies.

Since electric power moves within and between states, this regulation has roots in both federal and state jurisdictions. Federal intervention in electric power markets has its origins in the Federal Power Act, the Public Utility Holding Company Act, and the Rural Electrification Act. State regulation has evolved from state statutes, constitutions, and other legal precedents. At the federal level, electric power sales are regulated by the Federal Energy Regulatory Commission (FERC), while at the state level, regulation is directed by state Public Utility Commissions (PUCs).

Regulatory bodies at both the federal and state level attempt to ensure that electric power is provided economically, and in a safe and reliable manner. The primary method of electricity regulation has been rate of return, or cost-based, regulation. Here regulators set the rates utilities are allowed to charge their customers. This cost-based regulation allows

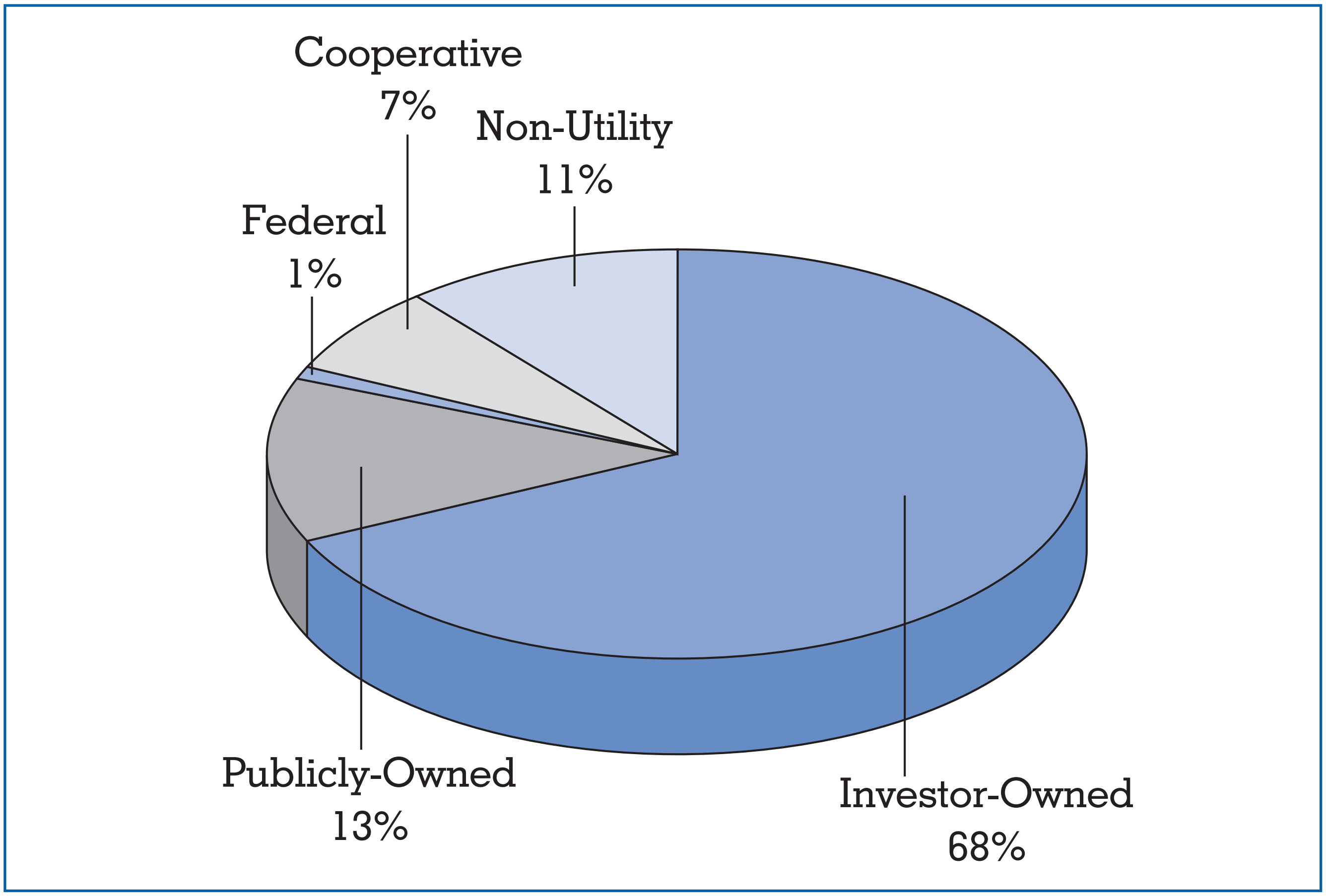


Figure 3.
Ownership Percentages of Total U.S. Generating Capacity
SOURCE: Energy Information Agency. (1995). *Electric Power Annual*. Washington, DC: U.S. Department of Energy.

utilities to recover their prudently incurred costs and also earn a reasonable rate of return on their investments. In return, utilities are granted exclusive franchises and monopoly service privileges

In the period from the 1930s to the early 1970s, regulation in the power industry was relatively uneventful. The energy crisis of the 1970s and early 1980s, however, dramatically changed the regulatory environment for electric power generation. During this period, state and federal government became exceptionally proactive in both generation planning and operation.

At the federal level, Congress passed the National Energy Act of 1978, which was composed of five different statutes: (1) the Public Utilities Regulatory Policy Act (PURPA), (2) the National Energy Tax Act, (3) the National Energy Conservation Policy Act, (4) the Power Plant and Industrial Fuels Act (PPIFA), and (5) the Natural Gas Policy Act. The general purpose of the National Energy Act was to ensure sustained economic growth during a period in which the availability and price of future energy resources was becoming increasingly uncertain. The two major themes of the legislation were to: (1) promote the use of conservation and renewable/alternative energy and (2) reduce the country's dependence on foreign oil.

While all aspects of the National Energy Act affected the electric power industry, PURPA was probably the most significant. PURPA was designed to encourage more efficient use of energy through industrial cogeneration. These cogenerators produce electricity through the capture of waste steam from their production processes. As a result of

PURPA, a whole new class of electricity generation facilities emerged, commonly referred to as "qualifying facilities" or QFs. PURPA required utilities to interconnect and purchase power from any QF at a rate not to exceed the utility's avoided cost of generation. This legislation opened the door to competition in the generation portion of the industry by legitimizing and creating a market for non-utility generation.

State regulation during this period continued the trend of promoting nonutility sources of generation. While PURPA was passed by the federal government, it was the responsibility of state regulators to set the rates at which utilities were required to purchase cogenerated or QF power. In response to the energy crisis of the period, state regulators began to set overly generous rates for QF power to stimulate conservation and alternative sources of electricity in a period of uncertainty. As a result of the generous rates and guaranteed market for nonutility generated power, generating capacity from nonutilities increased from close to 2 percent in 1978 to over 8 percent a decade later.

Regulators also began to require utilities to subject themselves to competitive bidding when they had a need for additional capacity. In addition, regulators began to require utilities to investigate other alternatives to the construction of new generation facilities, including the evaluation of demand-side management, or energy conservation measures, as a means of meeting future load growth. As a result of both of these policies, the fundamental premise of utility regulation and generation planning came under fire as more and more cost-effective, reliable, and alternative means of meeting electricity needs began to emerge.

ELECTRICITY GENERATION AND POWER INDUSTRY RESTRUCTURING

The structural and institutional environment for electric generation began to change dramatically in the late 1980s and throughout the 1990s as more and more competitive providers of electricity began to emerge. By the early 1990s, policy makers were actively discussing the possibility of restructuring the industry by introducing competition into the generation portion of the business. Throughout the 1990s, the terms "restructuring," "deregulation," and "competition" became virtually synonymous.

The passage of the Energy Policy Act of 1992 (EPAct) is considered the watershed federal legislation opening the door to complete power generation competition. This legislation allowed the Federal Energy Regulatory Commission (FERC) to order utilities to "wheel" or transport power over their transmission lines on behalf of third parties on an open access and nondiscriminatory basis. In subsequent years, FERC passed Order 888 and Order 889, which established the rules and institutions under which interstate or wholesale competition would be allowed. This wholesale competition was restricted to customers that were bulk power customers buying on behalf of other customers such as municipal utilities, rural cooperatives, and other IOUs. Retail competition, that is, competition for residential, commercial, and industrial customers, soon followed.

The origin of retail competition has run almost parallel to wholesale restructuring initiatives. State restructuring initiatives began initially in California and were soon adopted in New England. Both regions of the country were suffering from exceptionally high retail rates that were, in some cases, double the national average. Ratepayers, typically industrial ratepayers, appealed to regulators to allow competitive forces, rather than continued regulation, to discipline electric power generation and power markets.

The advent of competition has virtually transformed the industry in every aspect, including its name. In the not too recent past, the industry was referred to as the "electric utility industry." Today, given its significantly wide and numerous participants, it is more appropriate to refer to the industry as the "electric power industry." This new power industry has new power generation and sales participants with names such as qualifying facilities, exempt wholesale generators, merchant facilities, small power production facilities, power marketers, and sales aggregators.

The Mechanics of the Restructuring Process

Restructuring is the process of completely reorganizing the electric power industry. The generation portion of the industry will become more competitive, while the transmission and distribution portion of the industry will remain under regulation. While many specific aspects of restructuring differ between different states, and between federal and state jurisdictions, there are three common transition procedures.

The first transition procedure requires vertically integrated, former electric utility companies to unbundle, or separate, their electric power generation and energy sales operations from other utility operations. This separation is required to prevent former utilities from using their monopoly transmission and distribution assets in an anticompetitive manner to benefit their generation and sales operations. This divestiture, or separation, can be either physical or functional. Under physical divestiture, deregulated utilities are required to sell either all or a portion of their generating assets; the assets are "physically" removed from the former utility's control. Under functional divestiture, utilities are simply required to establish separate corporate affiliates, with stringent rules of conduct between regulated and unregulated companies. Most states opt for functional divestiture.

The second transition procedure establishes independence for the transmission system. This procedure is also required to ensure that a monopoly asset, in this case transmission, is not used in an anticompetitive manner. Two institutional structures are currently being debated for this transmission independence: an Independent System Operator (ISO), or a Transmission Company (Transco). The ISO transmission governance structure is typically associated with a multiutility, nonprofit association. Transcos are typically associated with a single-utility, for-profit, governing board. At the end of the 1990s, the ISO was the more prevalent of the two governance structures, with the Transco proposals gaining in number and popularity.

The third transition procedure defines the rules under which competitive suppliers of electricity can compete for end users. There are two polar models that are often debated for power market organization: the direct access (or bilateral contracts) regime, and the Poolco regime. Under direct access, consumers enter into direct contracts with competitive suppliers of electricity, and competitive providers of electricity enter into contracts with, and pay an access fee to, the local (regulated) distribution company for the use of local power lines.

A Poolco regime is a centralized market structure consisting of an ISO and a competitive Power Exchange (PX), where the ISO handles the physical deliveries and coordination of power flows within a regional power system, and the PX handles all the transactional issues associated with system power sales. In the Poolco regime, regional power market competitors submit bid prices and capacity offers into the competitive PX. Load from local distribution companies, representing all electricity end users, are then aggregated by the Poolco. Hour-ahead bid prices are used to construct a least cost dispatched and an hourly supply curve, and an hourly market equilibrium price is determined at the point at which the PX-determined supply curve intersects total regional aggregated demand. Least cost dispatch information is then transmitted from the PX to the ISO that controls all system coordination and security issues.

Many states have debated the efficiency and equity of both the direct access and Poolco market structures. Like many other restructuring transition issues, final policy decisions tend to be some hybrid or amalgamation of both approaches. Alternatively, some states have moved forward with a more centralized process (i.e., Poolco), with gradual implementation of more disaggregated trading regimes (i.e., direct access) at a later date. Since restructuring rules and laws are promulgated at the state level, it is very likely that market structures will be evolving and moving targets well into the early part of the twenty-first century.

THE SHIFT IN TECHNOLOGICAL PARADIGMS

One of the most dynamic factors underlying changes in the power industry has been technology. From the early days of the industry, designing, constructing, and operating more efficient generating units has been a priority. For a good part of the early to mid-twentieth century, the electric power industry, like other major capital intensive manufacturing industries, was one of the leading sectors of the economy in terms of technical innovation and productivity growth.

The amount of heat input, measured in British thermal units (Btu's), needed to generate a kilowatt hour of electricity with steam turbines decreased by almost 40 percent between 1925 and 1945, and by 35 percent during the period 1945 to 1965. During this period, scale became an important factor in power generation planning. Bigger was clearly better, and remained the premier planning paradigm for the utility sector of the industry until the mid to late 1980s. Larger plants usually entailed larger thermal efficiencies, which in turn reduced costs. However, gains in thermal efficiencies tapered throughout the 1970s. As the gains disappeared, so too did the ability to offset the exogenous economic changes in costs that occurred during the energy crisis. The only nonfossil technology of promise during this period, nuclear power, fizzled under the pressure of cost acceleration and inflation, rapidly increasing safety regulations, imprudent management, and regulatory and financial uncertainty. The accident at Three Mile Island in 1979 all but assured the industry that it had run out of large-scale technological innovations in power generation.

However, out of the ashes of the technological failures of the 1970s and early 1980s came a new technological innovation that dramatically changed the nature of the power industry. The experiences of the decade showed that the industry needed a technology that was flexible, modular, could be constructed quickly, and had minimal environmental impacts. Advances in the aerospace industry made it possible to deliver combustion turbine technologies that met the requirements of a new power generation environment.

Ironically, throughout the early 1980s, it was the nonutility generation portion of the industry that began to aggressively adopt the new efficient combustion turbine and combined cycle applications of the new natural-gas fired technologies. Widespread nonutility deployment of these technologies was the direct result of PURPA and the guaranteed market for nonutility generated power. Combined cycle plants, in particular, were rapidly preferred technologies for onsite generation at large nonutility generation facilities throughout the United States. The rapid deployment of these small, modular, and highly efficient facilities was an underlying technological rationale for introducing competition into generation markets.

Combined-cycle plants were in many ways an extension to the idea of cogeneration. These plants were effectively natural gas-fired combustion turbines with additional waste heat recycling unit—thus, a combined cycle of electric generation. The first stage generates gas-fired electricity from a turbine, while the second stage captures the waste heat to run a second-stage electric generator. Clearly, market participants with small scale power generation construction experience, like industrial cogenerators, can develop and operate projects of this nature. With these technologies, utilities need not be the only party participating in power generation construction and operation.

The popularity of combined-cycle units has increased dramatically over the past several years in both the utility and nonutility generation of electricity. As shown in Figure 4, in the year 2000, 3 percent of total generating capacity consisted of combined cycle technology, while traditional steam generating capacity comprised 69 percent of total. By the year 2020, however, these percentages shift dramatically in favor of combined cycle technologies with over 20 percent of total generating capacity invested in this technology.

The widespread adoption of these combined cycle and combustion turbine units represents a technology paradigm shift from large central station generation to more modular, flexible generating units. Under the new planning paradigm, size is less important than flexibility, fuel availability (natural gas), and location to load center. While some scale is still presumed to have benefits under this new paradigm, it is not the foremost consideration that it was a decade before.

The newest paradigm in power industry is known as distributed generation (DG) or, more generally, distributed energy resources (DER). Here, small scale power generation and storage equipment is located at the distribution—not transmission—level of interconnection. DER/DG includes such technologies as reciprocating engines, micro-turbines, fuel cells, and small solar photovoltiac (PV) arrays. While many of these technologies are relatively expensive now, future deployment, as well as changes in competitive generation market conditions, can make a number of applications cost-effective. DER could usher in a new level of competition much like its predecessor, the combined cycle technology, did a decade earlier.

CONCLUSIONS

At the beginning of the 1990s, the power industry was considered an old and tired industry in the United States and global economy. However, changes stimulated by the forces of new technology, environmental consciousness, and public policies promoting competition have brought about a renaissance in the power generation portion of the electric power industry. Like other large-scale manufacturing industries, the power

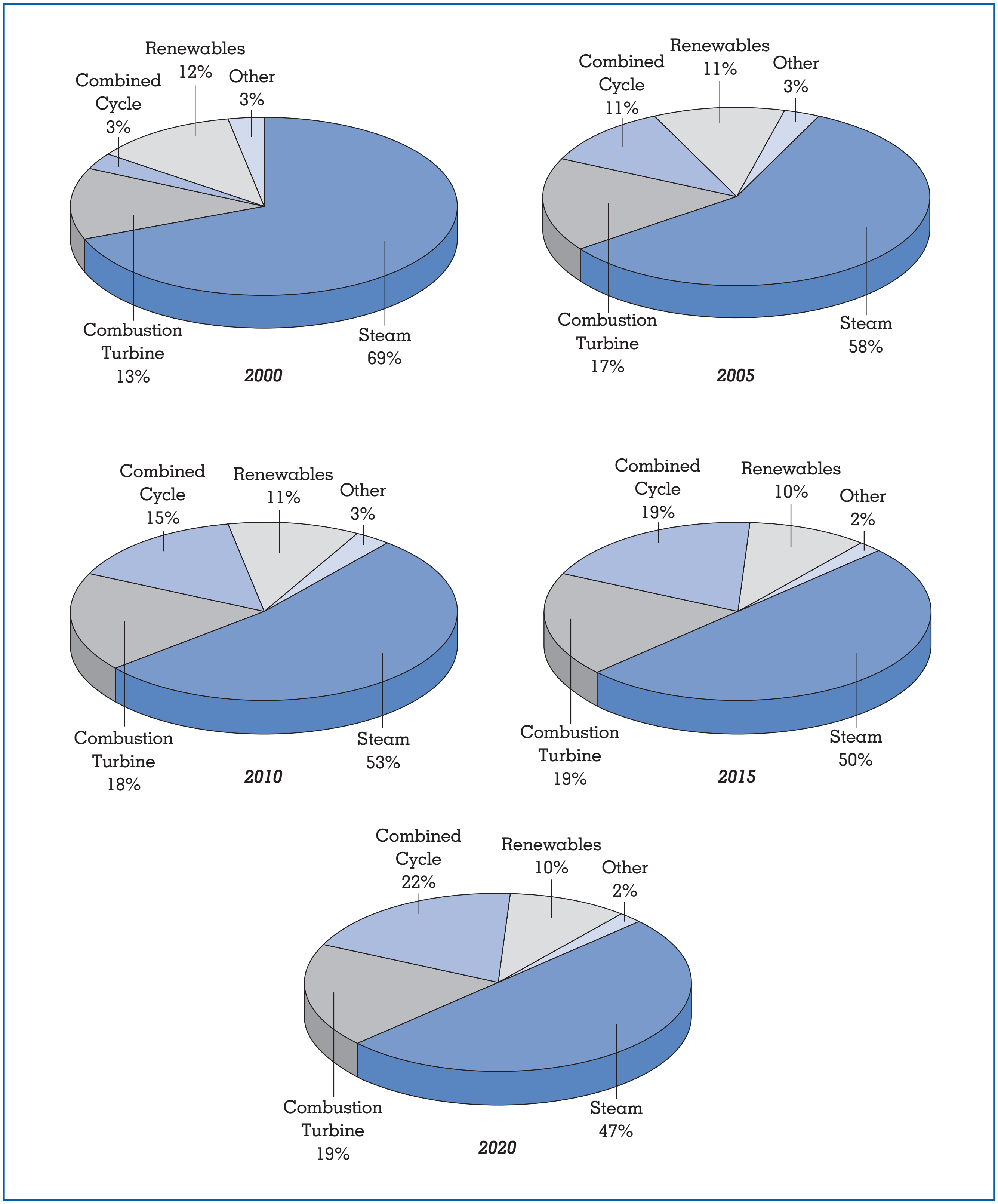


Figure 4.

Projected Changes in Total Capacity Percentages by Prime Mover

SOURCE: Energy Information Agency. (1995). *Electric Power Annual.* Washington, DC: U.S. Department of Energy.

industry has restructured and retooled, taking advantage of informational, technological, and managerial innovations. Competition, choice, and changes in the way power is generated, delivered, and sold to end users should continue this trend well into the next century.

David E. Dismukes

See also: Cogeneration; Demand-Side Management; Engines; Hydroelectric Energy; Market Transformation; Supply and Demand and Energy Prices; Turbines, Gas; Turbines, Steam.

BIBLIOGRAPHY

Energy Information Administration. (1993). *The Changing Structure of the Electric Power Industry*, 1970–1991. Washington, DC: U.S. Department of Energy.

Utility Data Institute. (1994). *Electric Utility Power Plant Construction Costs*. Washington: DC: Utility Data Institute.

ELECTRIC POWER, SYSTEM PROTECTION, CONTROL, AND MONITORING OF

Protection is the branch of electric power engineering concerned with the principles of design and operation of equipment (called "relays" or "protective relays") which detect abnormal power system conditions and initiate corrective action as quickly as possible in order to return the power system to its normal state. The quickness of response is an essential element of protective relaying systems—response times of the order of a few milliseconds are often required. Consequently, human intervention in the protection of system operation is not possible. The response must be automatic, quick, and should cause a minimum amount of disruption to the power system.

THE NATURE OF PROTECTION

In general, relays do not prevent damage to equipment; they operate after some detectable damage has already occurred. Their purpose is to limit, to the extent possible, further damage to equipment, to minimize danger to people, to reduce stress on other equipment, and above all, to remove the faulted equipment from the power system as quickly as possible so the integrity and stability of the remaining system is maintained. There is a control aspect inherent in relaying systems which complements the detection of faults and helps return the power system to an acceptable configuration as soon as possible so that service to customers can be restored. There is also a vital need to constantly monitor the power and the protective systems to analyze operations for correct performance and to rectify errors in design, application, or settings.

Reliability, Dependability, and Security

Reliability is generally understood to measure the degree of certainty that a piece of equipment will perform as intended. Relays, in contrast with most other equipment, have two alternative ways in which they can be unreliable. They may fail to operate when they are expected to, or they may operate when they are not expected to. This leads to the two-pronged definition of "dependability," the measure of certainty that the relays will operate correctly for all faults for which they are designed to operate and "security," the measure of certainty that the relays will not operate incorrectly for any fault.

Zones of Protection

Relays have inputs from several current transformers (CTs) and the zone of protection is bounded by these CTs. While the CTs provide the ability to detect a fault inside the zone, circuit breakers (CBs) provide the ability to isolate the fault by disconnecting all of the power equipment within the zone. Thus, a zone boundary is usually defined by a CT and a CB. When the CT is part of the CB it becomes a natural zone boundary. When the CT is not an integral part of the CB, special attention must be paid to the fault detection and fault interruption logic. The CT still defines the zone of protection, but communication channels must be used to implement the tripping function. Figure 1 shows the zones of protection in a typical system.

Relay Speed

It is, of course, desirable to remove a fault from the power system as quickly as possible. However, the relay must make its decision based upon voltage

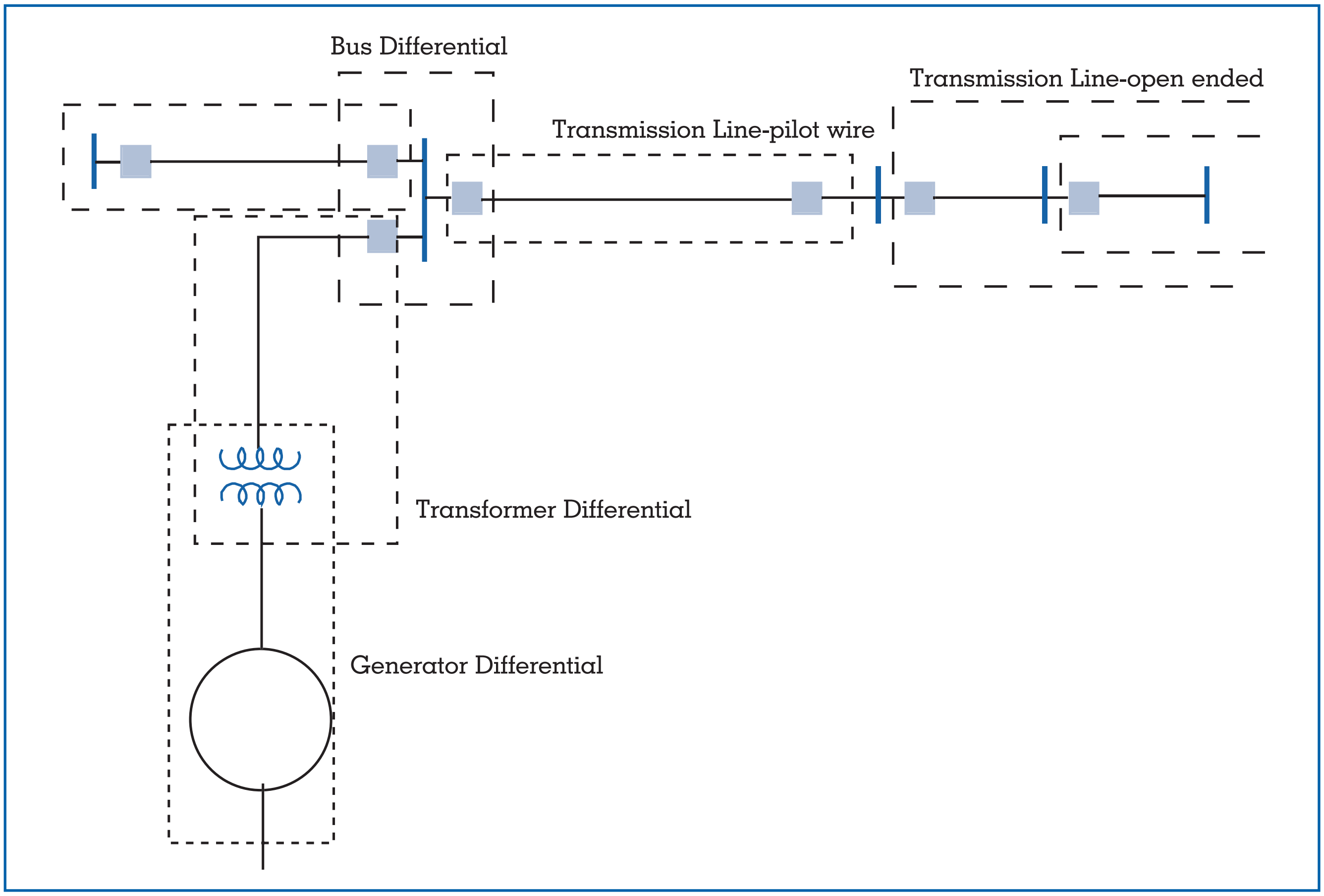


Figure 1.
Zones of protection.

and current waveforms which are severely distorted due to transient phenomena which follow the occurrence of a fault. The relay must separate the meaningful and significant information contained in these waveforms upon which a secure relaying decision must be based. These considerations demand that the relay take a certain amount of time to arrive at a decision with the necessary degree of certainty. The relationship between the relay response time and its degree of certainty is an inverse one and is one of the most basic properties of all protection systems.

Although the operating time of relays often varies between wide limits, relays are generally classified by their speed of operation as follows:

Instantaneous—These relays operate as soon as a secure decision is made. No intentional time delay is introduced to slow down the relay response.

Time-delay—An intentional time delay is inserted between the relay decision time and the initiation of the trip action.

High-speed—A relay that operates in less than a specified time. The specified time in present practice is 50 milliseconds (3 cycles on a 60 Hz system)

Ultra high-speed—This term is not included in the present relay standards but is commonly considered to be operation in 4 milliseconds or less.

Primary and Backup Protection

The main protection system for a given zone of protection is called the primary protection system. It operates in the fastest time possible and removes the least amount of equipment from service. On extra-

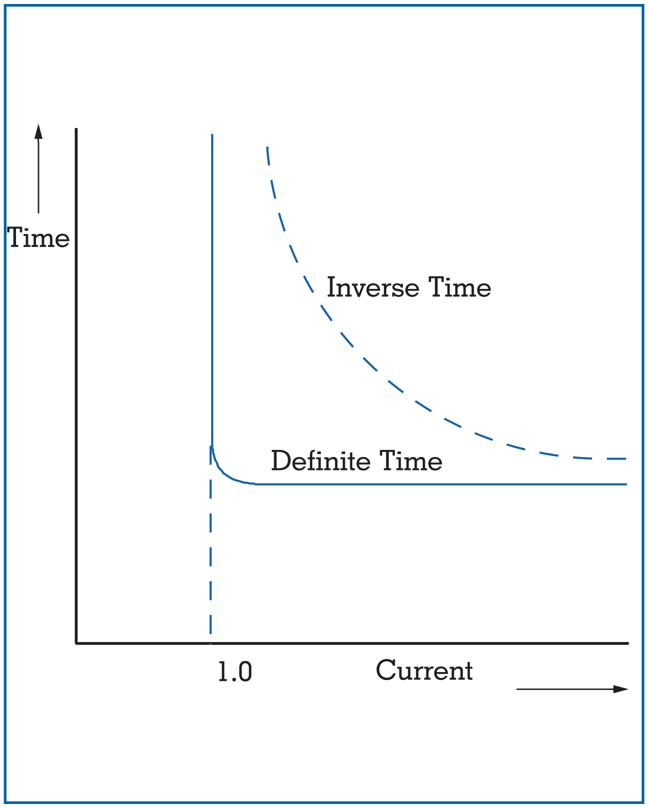


Figure 2.
Level detector relay.

high-voltage systems (230 kV and above) it is common to use duplicate primary protection systems in case any element in one primary protection chain fails to operate. This duplication is therefore intended to cover the failure of the relays themselves. One may use relays from a different manufacturer, or relays based on a different principle of operation, to avoid common-mode failures. The operating time and the tripping logic of both the primary and its duplicate system are the same.

It is not always practical to duplicate every element of the protection chain. Particularly on lower voltage systems, backup relaying is used. Backup relays are slower than the primary relays and, generally, remove more system elements than may be necessary to clear a fault. They may be installed locally, that is, in the same substation as the primary relays, or remotely.

RELAY OPERATING PRINCIPLES

In general, as faults (short circuits) occur, currents are increased and voltages decrease. Besides these magnitude changes, other changes may occur. Relay operating principles are based upon detecting these changes.

Level Detection

This is the simplest of all relay operating principles. Any current above, or voltage below, a set level may be taken to mean that a fault or some other abnormal condition exists inside the zone of protection. Figure 2 shows a definite time and an inverse time overcurrent relay.

Magnitude Comparison

This operating principle is based upon the comparison of one or more operating quantities. The relay will operate when the phasor division between the two or more circuits differs beyond the normal operating parameters. In Figure 3, I_A and I_B may be equal or at a fixed ratio to each other.

Differential Comparison

This is one of the most sensitive and effective methods of providing protection against faults and is shown in Figure 4. The algebraic sum of all currents entering and leaving the protected zone will be close to zero if no fault exists within the zone and will be the sum of I_1 and I_2 if a fault exists within the zone. A level detector can be used to detect the magnitude of this comparison or a special relay such as a percentage differential or harmonic restrained relay is applicable. This is the most common protective device used for generators, motors, buses, reactors, capacitors, etc. Its only drawback is that it requires currents from the extremities of a zone of protection which may require excessive cable lengths or a communication system.

Phase Angle Comparison

This type of relay compares the relative phase angle between two alternating-current quantities. It is commonly used to determine the direction of a current with respect to a reference quantity. Normal power flow in a given direction will result in the phase angle between the voltage and the current varying around the power factor angle (e.g., 30°) while power in the reverse direction will differ by 180°. Under fault conditions, since the impedance is primarily the inductance of the line, the phase angle of the current with respect to the voltage will be close to 90°.

Distance Measurement

This type of relay compares the local current with the local voltage. This is, in effect, a measurement of

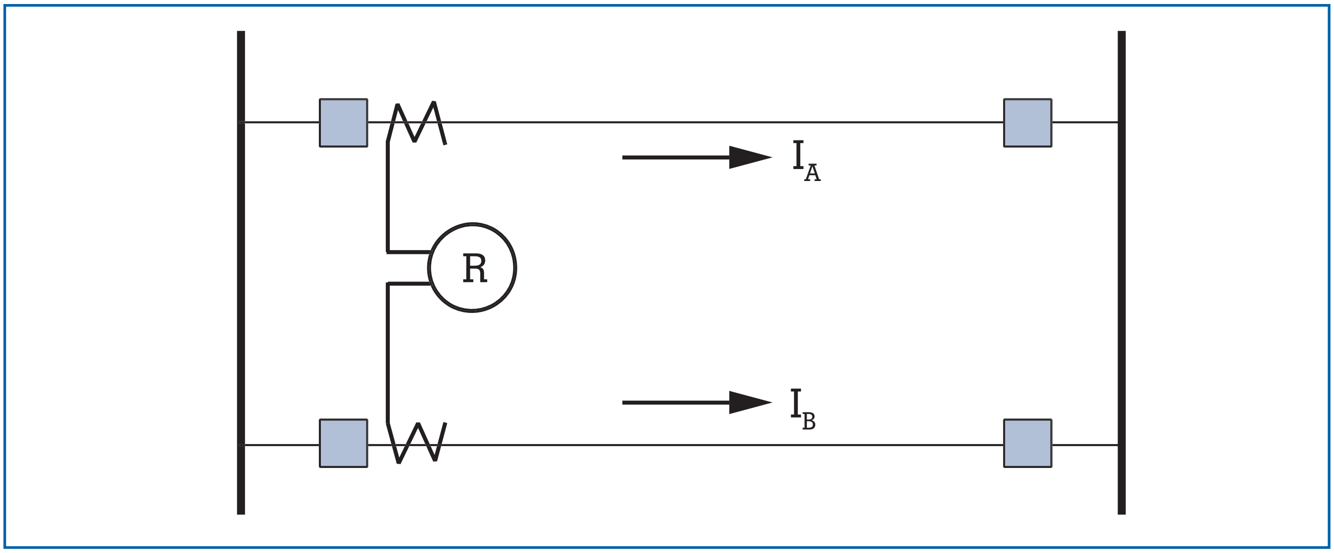


Figure 3.
Magnitude comparison relaying for two parallel transmission lines.

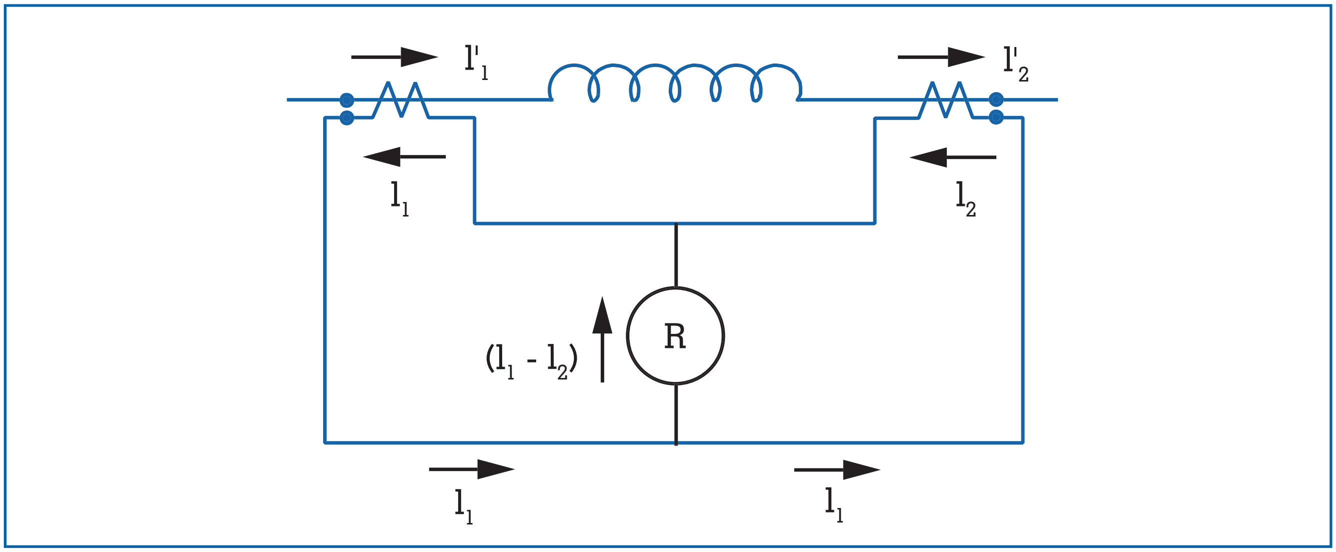


Figure 4.
Differential comparison principle applied to a generator winding.

the impedance as seen by the relay. An impedance relay depends on the fact that the length of the line (i.e., its distance) for a given conductor diameter and spacing determines its impedance. This is the most commonly used relay for the protection of high voltage transmission lines. As shown in Figure 5, zones can be identified as "zone one" which provides instantaneous protection to less than 100 percent of the associated line segment, and zones two and three which cover more than the line involved but must be delayed to provide coordination.

Harmonic Content

Currents and voltages in a power system usually have a sinusoidal waveform of the fundamental power system frequency plus other normal harmon-

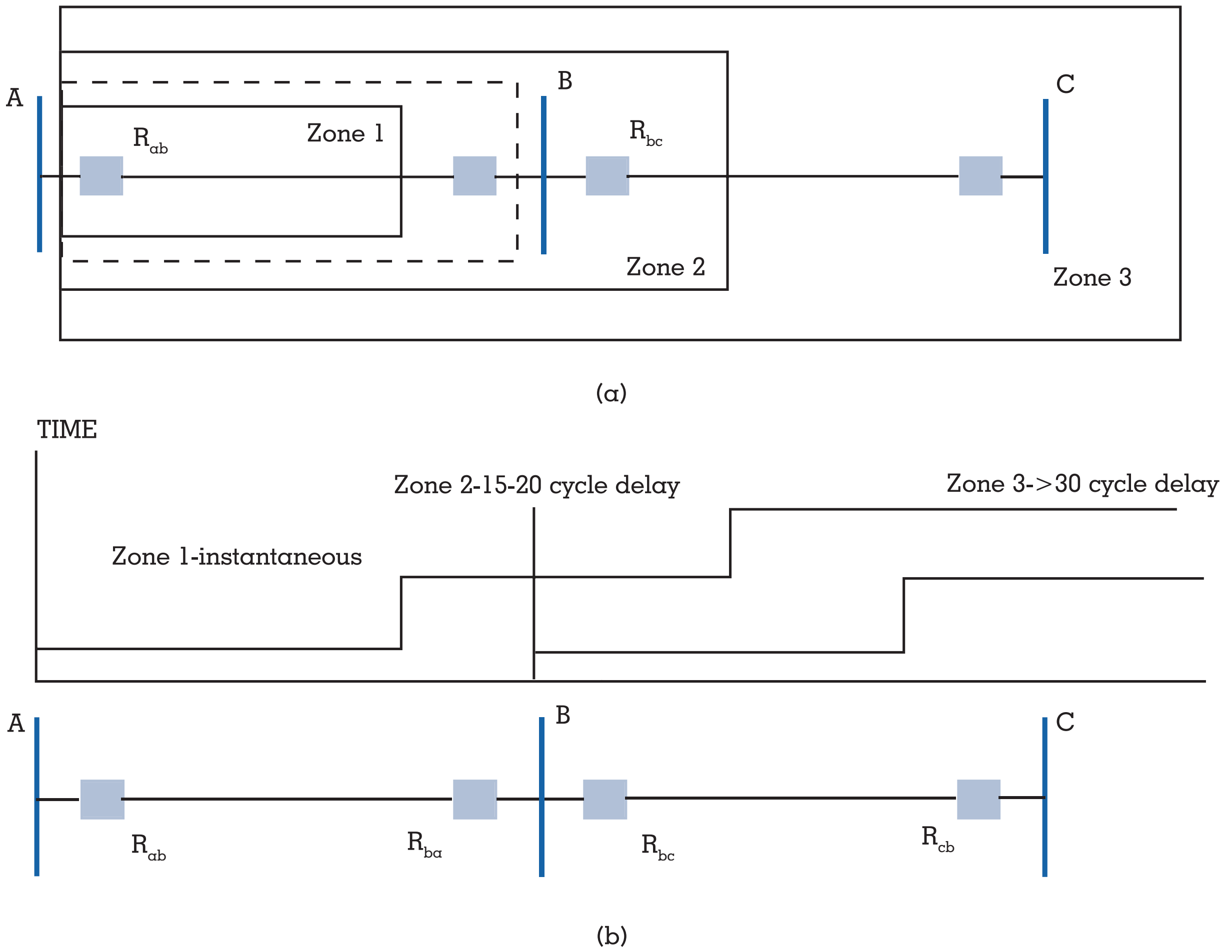


Figure 5.
Three-zone step distance relaying to protect 100 percent of a line and back up the neighboring line.

ics (e.g., the third harmonic produced by generators). Abnormal or fault conditions can be detected by sensing any abnormal harmonics that accompany such conditions.

Frequency Sensing

Normal power system operation is at 50 or 60 Hz depending upon the country. Any deviation from these values indicates that a problem exists or is imminent.

RELAY DESIGN

The following discussion covers a very small sample of the possible designs. Specific details must be obtained from the manufacturers.

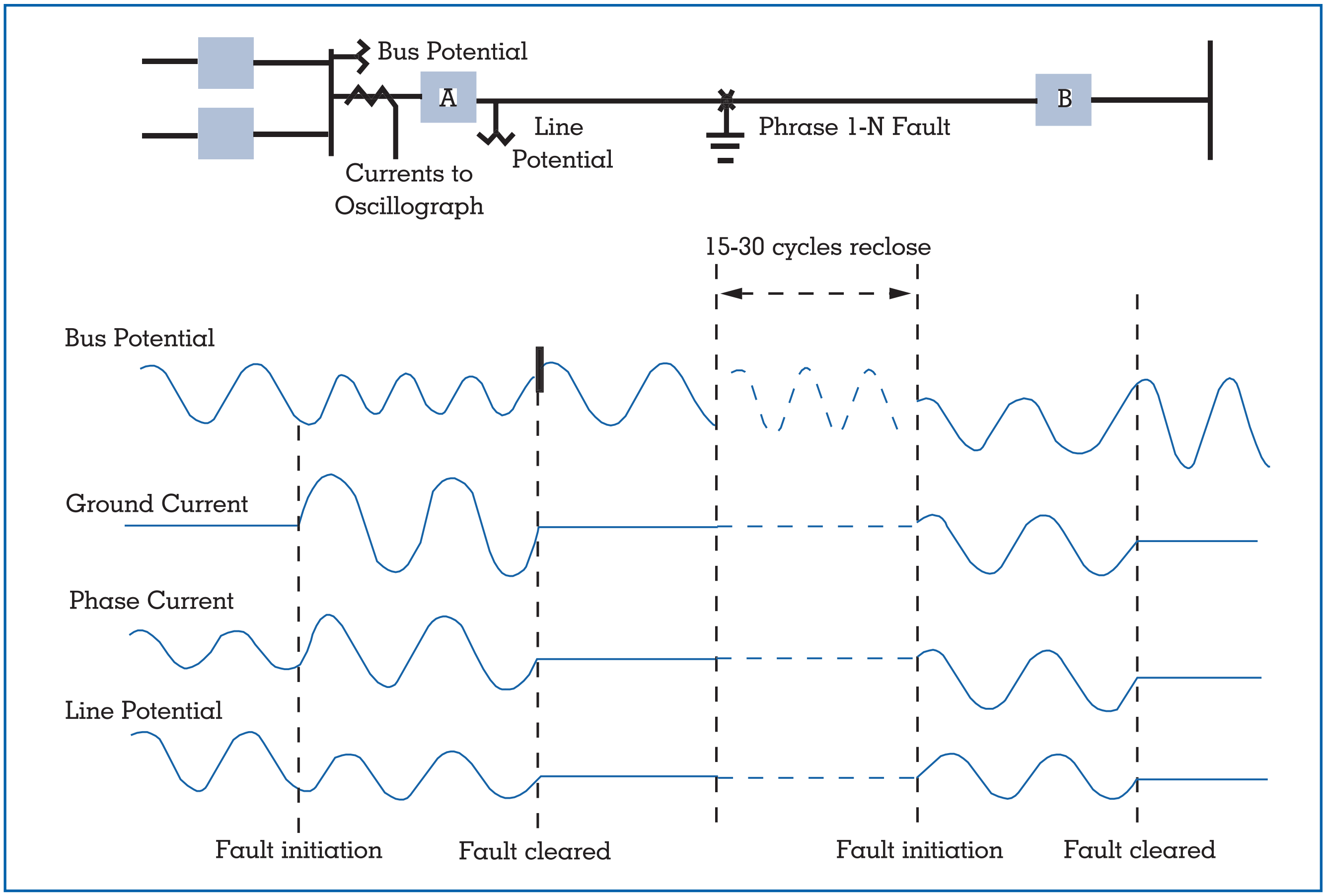


Figure 6.
Single phase-to-ground fault with unsuccessful high-speed reclose.

Fuse

The fuse is a level detector and is both the sensor and the interrupting device. It is installed in series with the equipment being protected, and it operates by melting a fusible element in response to the current flow.

Electromechanical Relays

The actuating forces are created by a combination of input signals, stored energy in springs, and dashpots. The plunger type relay consists of a moving plunger inside a stationary electromagnet. It is typically applied as an instantaneous level detector. The induction-type relay is similar to the operation of a single-phase ac motor in that it requires the interaction of two fluxes across a disc or cup. The fluxes can be produced by two separate inputs or by one input electrically separated into two components. Depending on the treatment of the inputs (i.e. one current separated into two fluxes, two currents, or a current and a voltage), this design can be used for a time-delay overcurrent relay, a directional relay, or a distance relay.

Solid-State Relays

All of the functions and characteristics of electromechanical relays can be performed by solid-state devices, either as discrete components or as integrated circuits. They use low-power components, either analog circuits for fault-sensing or measuring circuits as a digital logic circuit for operation. There are performance and economic advantages associated with the flexibility and reduced size of solid-state devices. Their settings are more repeatable and hold closer tolerances. Their characteristics can be shaped by adjusting logic elements as opposed to the fixed characteristics of induction discs or cups.

Computer Relays

The observation has often been made that a relay is an analog computer. It accepts inputs, processes them, electromechanically or electronically, to develop a torque or a logic output resulting in a contact closure or output signal. With the advent of rugged, high performance microprocessors, it is obvious that a digital computer can perform the same function. Since the usual inputs consist of power system voltages and currents, it is necessary to obtain a digital representation of these parameters. This is done by sampling the analog signals and using an appropriate computer algorithm to create suitable digital representations of the signals.

PROTECTION SCHEMES

Individual types of electrical apparatus, of course, require protective schemes that are specifically applicable to the problem at hand. There are, however, common detection principles, relaying designs and devices that apply to all.

Transmission Line Protection

Transmission lines utilize the widest variety of schemes and equipment. In ascending order of cost and complexity they are fuses, instantaneous overcurrent relays, time delay overcurrent relays, directional overcurrent relays, distance relays, and pilot protection. Fuses are used primarily on distribution systems. Instantaneous overcurrent relays provide a first zone protection on low-voltage systems. Time delay overcurrent relays provide a backup protection on low-voltage systems. Directional overcurrent relays are required in loop systems where fault current can flow in either direction. Distance relays provide a blocking and tripping function for pilot relaying and first, second, and third zone backup protection on high-voltage and extra-high-voltage systems. Pilot protection provides primary protection for 100 percent of the line segment by transmitting information at each terminal to all other terminals. It requires a communication channel such as power line carrier, fiber optics, microwave, or wire pilot.

Rotating Apparatus

The dominant protection scheme for generators and motors is the differential relay. Access to all entry points of the protected zone is usually readily available, no coordination with the protection of other connected apparatus is required, and the faulted zone is quickly identified. Motor protection also includes instantaneous and time delay overcurrent relays for backup.

Substation Equipment

Differential relaying is the universal bus and transformer protection scheme. The inrush current associated with power transformers requires a special differential relay utilizing filters to provide harmonic restraint to differentiate between energizing current and fault current.

Instantaneous and time delay overcurrent relays are the most common protective devices used on shunt reactors, capacitors and station service equipment.

CONTROL

Transmission line faults are predominantly temporary, and automatic reclosing is a necessary complement to the protective relaying function. The reclose time must be greater than the time required to dissipate the arc products associated with the fault. This varies with the system voltage and ranges from 15–20 cycles at 138 kV to 30 cycles for the 800 kV systems. Automatic reclosing requires that proper safety and operating interlocks are provided.

Rotating equipment, transformers, and cables do not, in general, have temporary faults, and automatic reclosing is not provided.

MONITORING

The importance of monitoring the performance of power system and equipment has steadily increased over the years.

Oscillographs and other fault recorders such as sequence of events are, by nature, automatic devices. The time frame involved in recognizing and recording system parameters during a fault precludes any operator intervention. The most common initiating values are currents and voltages associated with the fault itself. Phase currents increase, phase voltages decrease, and there is normally very little ground current, so all of these are natural candidates for trigger mechanisms. There are transient components superimposed on the 60 Hz waveform that accompany faults and other switching events. They are revealed in the oscillographic records and are an essential element in analyzing performance. Figure 6 is a typical record of a single phase to ground fault and unsuccessful high-speed reclose.

With the advent of digital relays, the situation changed dramatically. Not only could the relays record the fault current and voltage and calculate the fault location, they could also report this information to a central location for analysis. Some digital devices are used exclusively as fault recorders.

Stanley H. Horowitz

See also: Electric Power, Generation of; Electric Power, System Reliability and; Electric Power Transmission and Distribution Systems.

BIBLIOGRAPHY

Blackburn, J. L. (1952) *Ground Relay Polarization.* AIEE Trans., Part III, PAS, Vol. 71, December, pp. 1088–1093.

Horowitz, S. H., and Phadke, A. G. (1996). *Power System Relaying.* New York: John Wiley & Sons, Inc.

IEEE Power Engineering Society. (1980). *Protective Relaying for Power Systems*, ed. Stanley H. Horowitz. New York: IEEE Press.

IEEE Power Engineering Society. (1992). *Protective Relaying for Power Systems II*, ed. Stanley H. Horowitz. New York: IEEE Press.

IEEE Power System Relaying Comm. (1979). *Protection Aspects of Multi-Terminal Lines.* IEEE Special Publication No. 79 TH0056-2-PWR. New York: IEEE Press.

Lewis, W. A., and Tippett, L. S. (1947). *Fundamental Basis for Distance Relaying on 3-Phase Systems.* AIEE Trans., Vol. 66, pp. 694–708.

Mason, C. R. (1956). *The Art and Science of Protective Relaying.* New York: John Wiley & Sons.

Westinghouse Electric Corp., Relay Instrument Division. (1979). *Applied Protective Relaying.* Coral Springs, FL: Author.

ELECTRIC POWER, SYSTEM RELIABILITY AND

An electric power system involves the production and transportation of electrical energy from generating facilities to energy-consuming customers. This is accomplished through a complex network of transmission lines, switching and transformer stations, and other associated facilities. The primary objective in the design of the delivery system is the creation and operation of a safe, reliable, and economic link between the generating supply and the customer load points. Reliability has always been a primary concern in the design and operation of electric power systems. In most of the developed countries throughout the world, a very high level of reliability has been the norm for many years. This is particularly true in North America, where customers have come to expect electricity supply to be available virtually on demand.

The Canadian Electricity Association (CEA) has collected Canadian service continuity statistics for many years and publishes these data on an overall basis. Canadian electric utility reliability performance statistics for 1997 are shown in Table 1.

The SAIFI, SAIDI, and IOR indices are defined in Table 1 as follows:

- SAIFI is the average number of interruptions per customer served per year: SAIFI = (Total Customer – Interruptions) ÷ Total Customers Served.
- SAIDI is the system average interruption duration for customers served per year: SAIDI = (Total Customer – Hours of Interruption) ÷ Total Customers Served.
- IOR is the per unit of annual customer-hours that service is available: IOR = (8760 hours/year – SAIDI) ÷ 8760 hours/year

These data show that in 1997 Canadian electric power customers experienced an average of 2.35 power failures and on average were without electricity for 3.70 hours. Their demands were satisfied 99.9578 percent of the time. The SAIFI, SAIDI, and IOR values shown are averages and vary widely across the country and for different segments of any given area. In general, urban customers experienced much higher levels of reliability than did rural customers. The CEA statistics also show that approximately 80 percent of the interruptions seen by the average customer occurred due to failures within the low-voltage distribution system, where the supply configuration is basically radial in nature. The high-voltage network, which links the generating facilities with the major load points, is known as the bulk supply network and accounts for approximately 20 percent of the power failures experienced by an average customer. Distribution system failures are usually local in nature and tend to affect a relatively small

System Average Interruption Frequency (SAIFI)	=	2.35 int/system customer
System Average Interruption Duration (SAIDI)	=	3.70 hours/system customer
Index of Reliability (IOR)	=	99.9578%

Table 1.
1997 Canadian Electric Utility Reliability Performance
SOURCE: Canadian Electricity Association

number of customers, while failures in the bulk system can affect many customers. This is not always the case, as extreme weather conditions such as the 1998 ice storms in Ontario, Quebec, and the eastern United States affected both the bulk power and the distribution systems in these regions. The ability to generate sufficient electrical energy has traditionally not been a major concern in North America for some time, as vertically integrated utilities have planned and constructed their generating facilities to stringent reliability criteria. Regulatory authorities have viewed the establishment of adequate generation facilities as a mandatory requirement of reliable supply. This may change in the near future as the vertically integrated utility structure is discarded and replaced by independent, market-driven generation, transmission, and distribution companies. A major requirement in the new unbundled electric power system structure is open access to the bulk delivery system. This will permit competition among generating companies as well as customer choice in regard to supply. The overall integrity of the bulk delivery system is the vital link in this structure in order to retain the high level of reliability that customers have become accustomed to.

THE NORTH AMERICAN ELECTRIC RELIABILITY COUNCIL (NERC)

The bulk system has long been seen as a vital element in the maintenance of an economic and reliable supply of electrical energy in North America, and a high degree of redundancy has been incorporated to achieve this objective. The North American Electric Reliability Council (NERC) was formed in 1968, following the November 9–10, 1965, blackout that affected the northeastern United States and Ontario, Canada, in response to extreme concerns regarding the reliability of the rapidly developing interconnected power networks in North America. NERC is a not-for-profit corporation owned by ten regional councils, whose members come from virtually all segments of the electric power industry. Their parent companies account for virtually all the electrical energy supplied throughout the United States, Canada, and northern portions of Baja California, Mexico. The various regional councils are as follows:

- NPCC: Northeast Power Coordinating Council
- MAAC: Mid-Atlantic Area Council
- ECAR: East-Central Area Reliability Co-ordination Agreement
- MAIN: Mid-America Interconnected Network, Inc.
- MAPP: Mid-Continent Area Power Pool
- WSCC: Western Systems Co-ordinating Council
- SPP: Southwest Power Pool
- SERC: Southeastern Electric Reliability Council
- ERCOT: Electric Reliability Council of Texas
- FRCC: Florida Reliability Coordinating Council

The primary focus of NERC is on the bulk transmission systems spread throughout North America. The principal objective is to ensure that system disturbances is one area do not adversely affect other areas or regions. This is achieved through stringent and specific criteria; for example, a bulk power system should be able to withstand the loss of a large transmission component, a large generator, or a large area load. The reliability implications of customer supply in a specific or local area are not covered by NERC criteria unless that contingency event adversely affects a neighboring utility system. The 1965 blackout forcibly established recognition of the fact that reliable electric service is critical to the economic and social welfare of the millions of businesses and residents in the northeastern United States and eastern Canada. An immediate response to the blackout was the creation of the Northeast Power Coordinating Council (NPCC). The NPCC together with the nine other councils, form the present NERC:

- NPCC covers the northeastern United States and central and eastern Canada. NPCC's responsibility is to develop appropriate reliability criteria and guides, monitor the individual utility and participants' performance with these protocols, and thereby ensure that the individual bulk systems and therefore the

overall bulk system has been planned and is operating reliably.
- MACC was established in 1967 and encompasses nearly 50,000 square miles from Virginia to New York. MAAC contains the PJM centrally dispatched electric control area, which is the largest such area in North America and the third largest in the world. PJM Interconnection became the first operational independent system operator in the United States on January 1, 1998.
- ECAR was established in 1967, and its current membership includes 29 major electricity suppliers in nine east-central states serving more than 36 million people. The current full members of ECAR are those utilities whose generation and transmission have an impact on the reliability of the overall region.
- MAIN regular and associate members include investor-owned utilities, cooperative systems, municipal power agencies, independent power producers, marketers, and two municipal systems that together serve 19 million people living in a 120,000-square-mile area.
- MAPP was formed in the mid-1960s to perform regional planning of transmission and generation and has a wide variety of members. The MAPP organization performs three functions. It is a reliability council under NERC, a regional transmission group, and a power and energy market.
- WSCC covers the largest geographic area of the ten regional councils. Its 1.8-million-square-mile service area covers more than half the conterminous area of the United States. WSCC was formed in 1967 by 40 electric power systems and in 1999 had more than 100 member organizations, which provide electric service to over 65 million people. The region is divided into four major areas that reflect varying, and sometimes extreme, geographic, and climatic conditions.
- SPP consists of 55 member entities and serves customers in 8 southwestern states with a population of more than 25 million. Initially formed in 1941, SPP has 17 control areas with over 40,000 MW of generation.
- SERC was created in 1970 by 22 electric systems. The initial NERC agreement in 1968 was signed by 12 regional and area organizations, which included from the Southeast, the CARVA Pool, the Tennessee Valley Authority, the Southern Company, and the Florida Electric Power Coordinating Group. Subsequent discussions indicated that the overall reliability of the southeastern bulk power system would be better served by the creation of a regional reliability council with broader membership. This lead to the formation of SERC.
- As noted earlier, FRCC became the 10th Reliability Region of NERC in 1996.
- ERCOT provides the overall electric system coordination in the state of Texas and it consists of six different market groups with more than 80 voting and nonvoting members.
- The Florida peninsula, previously part of SERC, became a separate NERC Region in 1996 and was designated as the FRCC. It currently has 34 members consisting of investor-owned utilities, cooperatives, municipals, power marketers, and independent power producers.

Overall coordination of the ten councils is the responsibility of NERC. The advent of deregulation, competition, and the unbundling of traditional vertically integrated utilities has necessitated considerable reevaluation of the role that NERC will play in the future. The ten regional councils and NERC are continually evolving in regard to their scope and mandates. NERC assembled the Electric Reliability Panel in 1997 to recommend the best ways to set, oversee, and implement policies and standards to ensure the continued reliability of North America's interconnected bulk electric systems in a competitive and restructured industry. The resulting report of the panel stated that the introduction of competition within the electric power industry and open access to transmission systems require creation of a new organization with the technical competence, unquestioned impartiality, authority, and respect of participants necessary to enforce reliability standards on bulk electric power systems. Since that time, NERC has been aggressively working to create a new structure, designated as the North American Electric Reliability Organization (NAERO), to achieve these objectives. The importance of the bulk delivery system will not diminish in the new utility environment and will if anything became even more important in the maintenance of reliable electric energy supply. The events that lead to the Northeast blackout of 1965 and the creation of NERC and its constituent councils are important reminders of the need to continually appraise the security and adequacy of the highly interconnected bulk power network in North America.

THE NORTHEAST BLACKOUT OF 1965

The massive power failure of November 9, 1965, affected approximately thirty million customers and initiated the most intensive examination of power system planning and operating practices in the history of the electric power industry. The overall blackout was the result of a series of cascading events initiated by a major power surge. The network connections to the south and the east were too weak to withstand this massive power surge, and the overall network became unstable and ceased to function as an integrated whole. Network islands, in which the load exceeded the generation, were created and then subsequently collapsed. Many of these individual utility systems did not have adequate black-start procedures and equipment; therefore, service restoration was delayed for significant periods of time.

The Federal Power Commission (FPC) report on the blackouts, published in December 1965, stated that the initial cause of the blackout was the operation of a backup relay on one of the five main transmission lines taking power to Toronto from hydro facilities on the Niagara River. Due to an improper setting, the relay disconnected one line. The other lines were then overloaded and consequently tripped out successively. Approximately 1,500 MW of power generated at these hydro facilities, which power had been flowing to Toronto, then attempted to take an alternate route through the one remaining U.S.-Canadian interconnection, at Massena, New York. When this subsequently tripped, the now approximately 1,700 MW previously going to Canada surged into the U.S. network and initiated the breakup of systems in the northeastern United States.

The FPC report details the three basic stages of the overall blackout event. The first was the initial system shock due to the massive power surge. There was a widespread separation of systems throughout New York and New England in only seconds. If this had been the only system reaction, the resulting blackout would have impacted about one-third of the actual customers affected and only those in the northern and northwestern areas of the affected region. The second stage was the subsequent collapse of those utilities in eastern New York and New England that were separated from their normal interconnected grids. These islands were basically left with insufficient generating capacity to meet the connected load, resulting in declining frequency and voltage levels and individual system collapse. The third stage of the overall blackout was the inability of many utilities to restore power due to inadequate or nonexistent black-start capability at the fossil-fired steam-generating stations. Difficulties were also encountered in reenergizing high-voltage underground oil-filled transmission cables.

The initiating and subsequent events associated with the Northeast blackout revealed many deficiencies in utility planning, operating, and maintenance practices. The criteria and procedures now embedded in the NERC Planning and Operating Standards stem from the many studies, investigations, and reports initiated because of the Northeast blackout. The electric power industry has been in continuous evolution since its inception some one hundred years ago. It is quite possible, however, that the changes occurring now are the most dramatic in its history. Electric energy supply in the United States, Canada, and many other parts of the world is moving from the traditional regulated industry of the past to a more open competitive, market-based environment. These changes are being driven by both market and legislative forces and could create significant obstacles to the maintenance of a high level of bulk system reliability. The traditional vertically integrated utility “obligation to supply” will not exist in the new electricity market, which will contain many new players who, at best, will have only an indirect responsibility for reliability. The transmission system of the future will be used in many different ways than was envisaged when it was planned and constructed. Firm transfers within and across systems coupled with both short-term and long-term economic transactions will affect emergency support from interconnected systems. These factors will reduce system flexibility and therefore tend to reduce system reliability. Open access will increase the uncertainty associated with the timing and availability of new generation resources. These concerns have already become evident with service disruptions in California, New York, and Alberta Traditional opposition to the construction of high-voltage transmission due to land use, aesthetics, and electromagnetic fields, and the need for judicial reviews will not diminish in the new electric utility environment and will place severe constraints on the use of existing transmission systems.

The reliability of a modern electric power system depends on continuous real-time control of power and energy production, transmission line flows, system frequency, and voltage. This complex task will get more involved in the new environment with increased market participation on both the supply and the demand sides.

CUSTOMER CONSIDERATIONS

Customers will have a choice of their energy provider in the new utility environment. It is unlikely, however, that most customers have any real appreciation for those aspects that will affect reliability of supply or how a highly integrated electric utility system operates. An electric transmission system is not a conventional transportation system in which a product is dispatched from a source to a receiving point. An electric transmission system is a total energy system in which energy generation and consumption must remain continuously and instantaneously in balance. A customer cannot buy a particular unit of energy from a given operating unit, as the electric power system is a total energy system. Power flows in the system are dictated by the laws of physics, and flows instantaneously and automatically respond to changes in network conditions by following the paths of least impedance. Third-party transactions superimposed on an existing transmission network will change voltage profiles and transmission losses and will affect individual load point reliability levels.

Virtually all residential, commercial, and small industrial customers are served through low-voltage transmission networks. Only a relatively small number of large industrial consumers are served directly from the high-voltage transmission network. As previously noted, Canadian data indicate that approximately 80 percent of all the electric power system outages that an average Canadian consumer experiences are due to failures in the low-voltage distribution system. These events will not be affected by unbundling traditionally vertical utilities into separate functional zones involving different companies. It is going to be extremely difficult to improve customer reliability by unbundling the three functional zones of generation, transmission, and distribution, but there is considerable potential for seriously lowering customer reliability.

CLIMATE CHANGE

Electric power transmission and distribution systems are extremely vulnerable to adverse weather conditions. It appears to many people that extreme weather is becoming more frequent and also more violent. It has been suggested that this is due to increased heat and humidity in Earth's environment, leading to increases in both rain and snowfall. The effect of extremely adverse weather on both transmission and distribution facilities was clearly illustrated by "Ice Storm '98," which impacted eastern Ontario, southern Quebec, and parts of the northeastern United States. The weather condition was created by a combination of events that occurred over the North American continent. Moist, warm air from the Gulf of Mexico was pumped into southern Ontario and Quebec by a low-pressure system over the Texas panhandle. Coincident with this phenomenon, a large, stationary Arctic high-pressure area over Hudson Bay created a northeasterly circulation over central Quebec that moved very cold air into the Lawrence and Ottawa River valleys. The southerly warmer air current was unable to move the heavy, cold air and overrode the cold area at the contact surface, which resulted in considerable freezing rain. In the collision between the two air masses, the warm air was pushed upward and the cold air down. The snow that fell melted at the middle level but didn't have time to freeze again before hitting the ground. The ice that accumulated on the transmission and distribution facilities slowly increased as there were no periods of sunshine or thawing between the various periods of freezing rain.

The impact of "Ice Storm '98" was unprecedented in Canadian electric power system history. The Ontario Hydro Report provides a chronology of the twenty-three major storms that have impacted the system since 1942. The report also provides a detailed inventory of the effects, consequences, and mitigation measures taken due to the ice storm. The estimated direct total cost to Ontario due to the ice storm is $472 million (Canadian), which includes costs incurred by Ontario Hydro, local Ontario and federal governments, the Insurance Bureau of Canada and associated insurance companies, the Department of National Defence, and affected customers, including

System Average Interruption Frequency (SAIFI)	= 3.68 int/system customer
System Average Interruption Duration (SAIDI)	= 31.35 hours/system customer
Index of Reliability (IOR)	= 99.6422%

Table 2.
1998 Canadian Electric Utility Reliability Performance
SOURCE: Ontario Hydro, May 1, 1998.

businesses, farms, and residents. The monetary impacts in Quebec are expected to be considerably higher. Frequent questions asked immediately after the ice storm were: Why not put the transmission and distribution facilities underground and avoid the consequences of freezing rain? Why not increase the redundancy built into the network? Why not build bigger and stronger transmission facilities? All of these are technically possible and would have different impacts on the reliability of electric power system supply. All would have considerable economic consequences. It was estimated that the cost to replace the transmission and distribution systems in eastern Ontario would be about $11 billion, resulting in electricity rate increases of more than 11 percent. The question of balancing reliability and economics is an ongoing requirement in electric power system planning, operations and decision-making.

The impact of “Ice Storm ’98” on system reliability can be seen from the Canadian electric utility reliability performance statistics for 1998 in Table 2.

The impact of “Ice Storm ’98” can clearly be seen by realizing that the number of customer interruptions and the customer hours of interruption for all of Canada in 1997 were 24,280,244 and 38,130,783, respectively. The ice storm alone resulted in 12,332,950 customer interruptions and 282,576,829 customer hours of interruption in the utilities affected. Removing the ice storm incidents from the 1998 Canada-wide data results in a SAIFI of 2.46 and a SAIDI of 3.40. The IOR is 99.9612 percent. “Ice Storm ’98” had only a relatively moderate effect on SAIFI but a dominant effect on the customer hours of interruption and the SAIDI statistic due to the extremely long storm duration and the required restoration period. It is important to realize that the bulk transmission system retained its integrity according to NERC criteria, and the impact of the ice storm on Ontario Hydro facilities did not propagate into neighboring interconnected utilities.

CONCLUSION

Reliability of electric energy supply is an important requirement in modern society, and consumers in developed countries have grown to expect electricity to be available on demand. The electric utility industry in North America and throughout the rest of the developed world is undergoing considerable change as open transmission access and consumer choice are replacing traditional utility structures in which the obligation to meet customer requirements was a key component. One casualty in this move to the competitive marketplace may be the planning of adequate generating capacity to meet future load requirements. The bulk transmission system will become the focal point of power system operation and control in the new environment and the determining factor in retaining high levels of customer reliability and satisfaction. It is likely that judicial regulation will increase rather than decrease in the new environment, as regulatory authorities exercise increased vigilance to ensure that acceptable reliability levels are provided and competition is allowed to flourish. The utility industry has developed tremendously since the Northeast blackout of 1965 and has learned many lessons over the subsequent years. The requirement to exercise both flexibility and control in permitting open access and competition while maintaining an acceptable level of reliability will be the major challenge in the next decade.

Roy Billinton

See also: Climatic Effects; Consumption; Domestic Energy Use; Electric Motor Systems; Electric Power, Generation of; Electric Power, System Protection, Control, and Monitoring of; Electric Power Substations; Electric Power Transmission and Distribution Systems; Government and the Energy Marketplace; Regulation and Rates for Electricity.

BIBLIOGRAPHY

Canadian Electricity Association. (1997). *Service Continuity Report on Distribution System Performance in Canadian Electrical Utilities*. Montréal, Canada: Author.

“NERC Planning Standards.” (1997). Princeton, NJ: North American Electric Reliability Council.

Ontario Hydro. (1998). “Ice Storm 98: Electricity Supply Impacts of the January 1998 Ice Storm in Eastern Ontario.” Toronto: Author.

“Program Improvements in the CEA Service Continuity Report and Impact of the Ice Storm on CEA Data.” (1999). CEA Electricity ’99 Conference, Vancouver, B.C., March.

U.S. Federal Power Commission. (1965). *Northeast Power Failure, November 9 and 10, 1965*. Washington, DC: U.S. Government Printing Office.

U.S. Federal Power Commission. (1967). *Prevention of Power Failures: Volume I, Report of the Commission; Vol. II, Reliability of Electric Bulk Power Supply; Vol. III, Studies of the Task Groups on the Northeast Power Interruption*. Washington, DC: U.S. Government Printing Office.

ELECTRIC POWER MEASUREMENT

See: Units of Energy

ELECTRIC POWER SUBSTATIONS

An electric power substation is a facility that provides a junction between parts of the power grid. The substation's functions, critical for the proper operation of the power system, include the interconnection of power lines from different parts of the system; the monitoring and control of system operating conditions; and the protection of the power system equipment.

CLASSIFICATION AND GENERAL DESCRIPTION

Substations may be classified into one of several categories depending on their location and function within the system. Generator substations are located at the site of power generating stations and provide the connection to the transmission system. Bulk power substations link the transmission system to the subtransmission system, stepping the voltage down through a transformer (transformer substation), or linking high-voltage transmission lines from different parts of the system without changing the voltage (switching substation). A distribution substation provides the link between the subtransmission system and the much lower voltages of the distribution system. A converter station is a unique type of bulk power substation that provides a link between high-voltage alternating-current transmission lines and high-voltage direct-current transmission lines.

The siting of substations, electrical, geographic, economic, political, and aesthetic factors must be considered. The high voltages of the transmission system are utilized because the reduced currents result in more efficient power transmission. Therefore, substations are placed as close to the system loads as possible to minimize losses. This is constrained by the value and availability of real estate, as well as by the requirement that terrain be relatively level within the substation. Care is taken in substation placement, particularly in areas of dense population, that the location not obstruct scenic views or aesthetically depreciate commercial or residential developments. The physical size of substations can cover large areas because the high-voltage components are insulated from each other by air and thus must be separated by significant distances. Historically, these issues have limited the installation of large substations to areas of relatively sparse population. However, since the 1980s, substations have been insulated with pressurized sulfurhexafluoride gas (SF_6). Because of the highly insulating quality of SF_6, the size of these gas-insulated substations may be well under 25 percent of the size of an air-insulated substation with the same power-handling capability. In some applications, particularly those in proximity to population centers, the entire substation may be enclosed within buildings, reducing aesthetic concerns and deterioration by the environment. Nevertheless, air-insulated substations are still generally preferred because of the higher cost and environmental concerns regarding the release of SF_6 (which is being investigated as a greenhouse gas).

SYSTEM INTERCONNECTION

The primary function of substations is to provide an interconnection between transmission lines extending to other geographical areas and between parts of the system that may be operating at different voltages. A principal aspect of the substation design is the arrangement of connections through circuit breakers to common nodes called busses. Circuit breakers are large electrical switches that provide the ability to disconnect the transmission lines or transformers from the bus. Transformers provide a change in voltage.

Busses

Busses are typically made of aluminum or copper and are rigid bars in the substation, insulated from ground and other equipment through ample insulating material, typically air or sulfurhexafluoride. The arrangement of the busses in the substation may fall into a number of different categories; the most common are illustrated and explained in Table 1. The appropriate selection of configuration is made by carefully balancing cost, reliability, control, and space

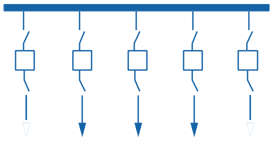
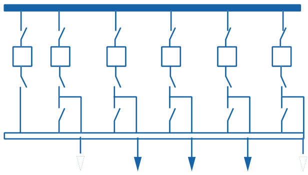
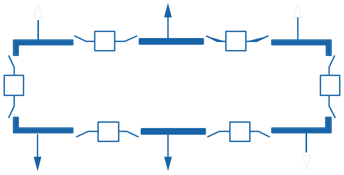
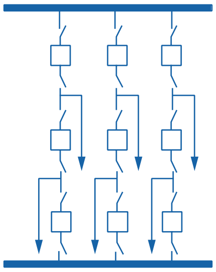
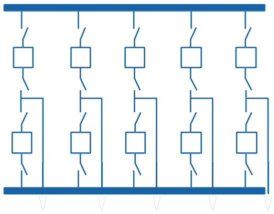
Single Bus		• All connections are tied to a single bus, with one circuit breaker for each bus. This arrangement is favored for its simplicity and low cost, although it is least desirable with regards to reliability. Maintenance to substation equipment requires that connections be removed from service.	• This type of bus is usually the configuration of choice in substations at or below 130 kV.
Main and Transfer Bus		• As with the single bus arrangement, each connection is linked to the main bus through a circuit breaker, but the breaker may be bypassed using disconnect switches through a transfer bus and another breaker to the main bus. This permits isolation of the circuit breaker for maintenance without loss of service to the connection.	• Used in more critical applications at or below 130 kV, and occasionally at higher voltages.
Ring Bus		• This scheme has all circuit breakers linked in a closed loop, with connections entering at the junction between breakers. This way, any connection may be isolated or any single circuit breaker removed without interrupting the other connections. This provides a higher level of redundancy than the systems mentioned above. Control and protective relaying issues are somewhat more complicated for this arrangement.	• Usually found in substations above 130 kV, in smaller substations. Often installed with the expectation of future expansion to a breaker-and-a-half scheme.
Breaker-and-a-half Scheme		• This scheme has two equal busses, with three breakers connected between them. Each connection may be linked to one of the busses through one breaker, and in the event that one breaker is out of service or in need of maintenance, the connection may still be served through the two breakers to the other bus. The name of this arrangement comes from the fact that two connections are served by three breakers, so that there is an average of one and a half breakers per connection. This scheme is less complicated than the ring bus, with higher reliability, but is more costly.	• Most common on systems above 130 kV.
Double Bus		• A double bus, double breaker arrangement provides a link to each bus through an independent breaker for each connection. This provides full redundancy in case of malfunction, or the need to perform maintenance on a circuit breaker or bus, but is the most expensive configuration.	• Usually found in most critical transmission substations and in generator substations.
Bus	Line, Transformer, or Load ↓	Disconnect Switch	Circuit Breaker ☐

Table 1.
Most Common Arrangements of the Busses in a Substation

constraints. If the substation is providing service to critical loads, the need for high reliability may warrant the higher cost of a more complex bus arrangement, while for less critical loads, space constraints may dictate a minimal bus arrangement.

Disconnect Switches

For every piece of equipment in a substation, manual switches—called disconnect switches—are provided to enforce complete electrical isolation from equipment before any service is performed. Disconnect switches are placed in clearly visible locations so maintenance personnel can continuously confirm that the equipment is isolated. The disconnect switch cannot interrupt current, so it is opened only when the current has already been interrupted by an automatic switch such as a circuit breaker.

Circuit Breakers

Circuit breakers are switches that are operated by a signal, from a relay or from an operator. The circuit breaker is designed to interrupt the very large currents that may occur when the system experiences a fault, such as a lightning strike or arc to ground (e.g., a tree falling on a line, or a line falling to the ground). Because these extremely large currents can cause severe damage to equipment such as transformers or generators, and because these faults can disrupt the proper operation of the entire power system, the circuit breakers are designed to operate rapidly enough to prevent damage to equipment, often in 100 milliseconds or less.

The circuit breaker contacts consist of two pieces of metal that are able to move with respect to each other. When the circuit breaker is closed, the contacts are touching and current flows freely between them. When the circuit breaker opens, the two contacts are separated, typically by a high-strength spring or a pneumatic operator. As the contacts separate, current continues to flow through them, and the material between them is ionized, forming a conducting plasma. To provide isolation, the plasma must be eliminated and the contacts be separated a sufficient distance to prevent the reinitiation of an arc. Several different technologies are implemented to give four common types of circuit breakers.

Air blast circuit breakers are insulated by air, and the plasma is extinguished as a blast of compressed air is blown between the contacts. These are less common than the other types and generally are no longer applied in new installations because of size, and problems with the maintenance of the compressors. Oil-filled circuit breakers have the contacts enclosed within a sealed tank of highly refined oil, with oil ducts designed to force oil between the contacts to quench the arc when the contacts open. These are common, but decreasing in popularity due to the environmental concerns associated with the risk of an oil spill. Although breaker failures occur only rarely, hundreds of gallons of oil may be spilled in a single failure, requiring very costly remedial procedures. The more popular breakers for high-voltage systems are gas-filled breakers that have the contacts enclosed within a sealed tank of pressurized SF_6. These have proved highly reliable, although there have been some environmental concerns about the release of the SF_6 when maintaining the device or when the tank ruptures. For lower-voltage applications (less than 34 kV), vacuum breakers are often used. These eliminate arcing by enclosing the contacts within an evacuated chamber. Because there is no fluid to be ionized, there can be no plasma formed. Their major benefit is a very fast response time and elimination of environmental concerns.

In addition to circuit breakers, there are other classes of automatic switches that can be controlled or operated remotely, but with current-interrupting capability. These include circuit switchers, reclosers, and sectionalizers.

Transformers

Power transformers perform the very important function of linking parts of the power system that are at different voltages. They are found exclusively in substations, except in the distribution system, where they may be mounted on poles or pads close to the loads they are serving.

SYSTEM MONITORING AND PROTECTION

The substation provides a monitoring point for system operating parameters. The power system is a highly complex and sensitive conglomeration of parts that must all be coordinated to function properly. For this reason, the operating conditions must be very closely observed and controlled. This is done by using specialized sensors to acquire the information and then communication systems to convey the information to a central point. For immediate response to system faults (such as damaged conduc-

Pend Oreille Utility District workers at a control box in a power substation. (Corbis-Bettmann)

tors, arcs to ground, or other undesirable operating conditions), a system of protective relaying (consisting of sensors and automated switches) is used to operate circuit breakers.

Instrument Transformers

The high voltages and currents seen in a substation exceed the voltage and current ratings of monitoring equipment, so instrument transformers are used to convert them to lower values for monitoring purposes. Instrument transformers may be categorized as current transformers (CTs) or voltage transformers (VTs), which are also sometimes designated as potential transformers. CTs typically consist of a toroidal core of magnetic material wrapped with a relatively high number of turns of fine wire, with the current to be measured passing through the middle of the toroid. These devices are often located in the bushings of circuit breakers and transformers so as to be able to measure the current in those devices. Bushings are the special insulated connections that allow the current to pass from the outside air into a sealed metal enclosure. VTs serve the function of stepping the voltage down to a measurable level. There is usually one connected to each of the substation busses. Most of the time VTs are constructed in essentially the same fashion as other transformers, although sometimes a capacitive coupling may enhance or replace the electromagnetics. Recent advances in technology have developed a new class of CTs and VTs that are optical devices that use specialized materials and advanced signal processing techniques to determine current based on the polarization of light as influenced by magnetic field strength, and voltage based on the polarization of light as influenced by electric field strength. While these devices are significantly more expensive than the traditional technologies, they provide higher accuracy and reliability and better electrical isolation.

Once the operating conditions have been measured, the information is conveyed to a central location using a system known as SCADA (Supervisory Control and Data Acquisition). The SCADA system data are displayed in the regional dispatch center to assist operators to know what actions must be taken for the best operation of the system.

Protective Relaying

Instrument transformers provide inputs to the automatic protection system. To provide a quick response to faults, a group of devices called relays accept the voltage and current signals, determine when abnormal conditions exist, and open the circuit breakers in response to fault conditions. The protection system design opens only the circuit breakers closest to the problem so that all of the rest of the system may resume normal operation after the fault is isolated from the system. Historically, determining which breakers to open has been done using various electromechanical devices that had the necessary comparisons and delays built into their design. These include overcurrent relays, directional relays, distance relays, differential relays, undervoltage relays, and others. These electromechanical devices have proven rugged and reliable since the early 1900s. In the late 1950s a new class of relays, solid-state relays, using analog circuits and logic gates, provided basically the same performance, but without any moving parts and hence reduced maintenance requirements. With the advent of low-cost-high level microprocessors, a new generation of relays has been born in which a single microprocessor-based relay performs all of the functions of several different electromechanical or solid-state relays. The microprocessor provides the benefits of higher accuracy, improved sensitivity to faults, better selectivity, flexibility, ease of use and testing, and self-diagnostic capabilities. They can be integrated into the SCADA system to communicate the cause of breaker opening, and can be operated, reset, and updated through remote access. These advantages are why microprocessor-based relays are found in most new installations and are also being retrofitted into many existing substations.

In addition to protection against excessive currents, equipment must be protected against excessive voltages that commonly result from lightning strikes or switching transients. Because of the high speed of these surges, relays and circuit breakers are unable to respond in time. Instead, this type of protection is provided by surge arrestors, which are passive devices that prevent overvoltages without moving parts. An air gap was the earliest type of surge arrestor, in which a special set of contacts are set a distance apart specified by the maximum tolerable voltage. When the voltage exceeds that threshold an arc forms, essentially shorting out the overvoltage. The newer surge arrester technology is the metal-oxide varistor (MOV). This is a device that behaves like a very large resistor at voltages below the specified threshold, but at voltages above the threshold, the resistance of the device drops precipitously, effectively drawing enough current to limit the voltage, but without shorting it to ground.

SYSTEM VOLTAGE CONTROL

Another of the principal functions of a substation is to provide the means to control and regulate voltages and power flow. These functions are provided either by feedback from an automated system or by remote instruction from the dispatch center using an array of devices and systems within the substation.

A load tap changer, an integral part of a power transformer, is a special switch that adjusts the voltage ratio of the transformer up or down to keep the load side voltage at the desired level despite changing voltages on the source side. Capacitor banks are used to raise the voltage in a substation when it has dropped too low, particularly in areas of large industrial loads. Shunt reactors are used to lower voltages that have risen too high due to the capacitance in the transmission or distribution line.

Another class of devices used to control the voltage is operated using powered electronic switches to continuously adjust the capacitance and/or inductance in a substation to keep the voltage at precisely the voltage desired. These devices are relatively new in deployment, having been developed with the advent of inexpensive and robust power semiconductor components. These devices are part of a group broadly known as FACTS (Flexible AC Transmission System) devices and include static var compensators, static synchronous compensators, and dynamic voltage restorers.

John A. Palmer

See also: Capacitors and Ultracapacitors, Electric Motor Systems; Electric Powers, Generation of; Electric Powers, System Protection, Control, and Monitoring of; Electric Power, System Reliability and; Electric Power Transmission and Distribution Systems; Insulation; Transformers.

BIBLIOGRAPHY

Asea Brown Boveri. (1988). *Selection and Application of Gas Insulated Switchgear*. North Brunswick, NJ: Author.
Bosela, T. R. (1997). *Introduction to Electrical Power System Technology*. Englewood Cliffs, NJ: Prentice-Hall.
Faulkenberry, L. M., and Coffer, W. (1996). *Electrical Power Distribution and Transmission*. Englewood Cliffs, NJ: Prentice-Hall.
Glover, J. D., and Sarma, M. (1994). *Power System Analysis & Design*, 2nd ed. Boston: PWS.

ELECTRIC POWER TRANSMISSION AND DISTRIBUTION SYSTEMS

The North American electric power transmission system has been described as the largest, most complex machine ever built by humanity. It is a massive network of generating stations, transmission lines, substations, distribution lines, motors, and other electrical loads all interdependently linked for the conversion, transportation, and control of electrical energy. Approximately 60 percent of all energy utilized in the United States passes through the interconnected electric power system. The major goal of the system is to most efficiently and reliably deliver electric power from generating stations to residential, commercial, and industrial consumers.

A small portion of the power system is depicted in Figure 1. The flow of energy is as follows: at generating stations, mechanical, chemical, or some other form of energy is converted into electricity, most often using a synchronous generator. The electrical output from the generator is converted, through a transformer, to a very high voltage, to be conveyed through transmission lines to transmission substations. Within the transmission substations, the voltage is stepped down to the subtransmission system, through which the power is conveyed to the distribution substations and into the distribution system. The distribution system delivers the power to residential, commercial, and industrial users, where the power is converted to light, heat, motion, or other desired forms of energy.

HISTORICAL CONTEXT AND TECHNOLOGICAL DEVELOPMENTS

The earliest commercial power system, believed to be Thomas Edison's Pearl Street station opened in 1882, consisted of a simple generator and a number of users. Within twenty years, more than 3,000 small electric generating stations were built in cities across the United States, each serving relatively local loads and with no interconnection. With the advent of the transformer, in about 1885, and the recognition that voltage drop and losses will be significantly reduced by stepping the voltage up and the current down, power transmission systems were created. The first demonstration of an ac power transmission line, in 1886, operated at 3,000 V over a distance of 4,000 ft (1,220 m). The first commercial transmission line in the United States was a 13-mile (21-km) transmission line operating at 3,300 V. As insulation systems improved and the technology of transformers was advanced, transmission voltage levels increased to 40 kV by 1907. This was a practical limit for the pin-type insulators (Figure 2) that were used to support the line on the towers, due to the structural stresses in the support. The voltage level was only able to increase further with the invention of the suspension insulator (Figure 2), which is in common use at the beginning of the twenty-first century. This increased the practical limit to about 150 kV, which was the limiting case because of corona. (Corona is a phenomenon in which the air in the vicinity of the energized surface is ionized because of the intensity of the electric field. It results in significant energy losses to the system.) The intensity of the electric field is reduced by having a larger-diameter conductor. The voltage was again increased in the 1960s after the realization that forming bundles of two, three, or four conductors could also mitigate the problem of corona. Power systems in the late 1990s operated at voltages as high as 765 kV in the United States, and as high as 1,100 kV in some parts of Europe.

SYSTEM INTERCONNECTIONS

Through the first several decades of commercial power systems, one generator or a small cluster of

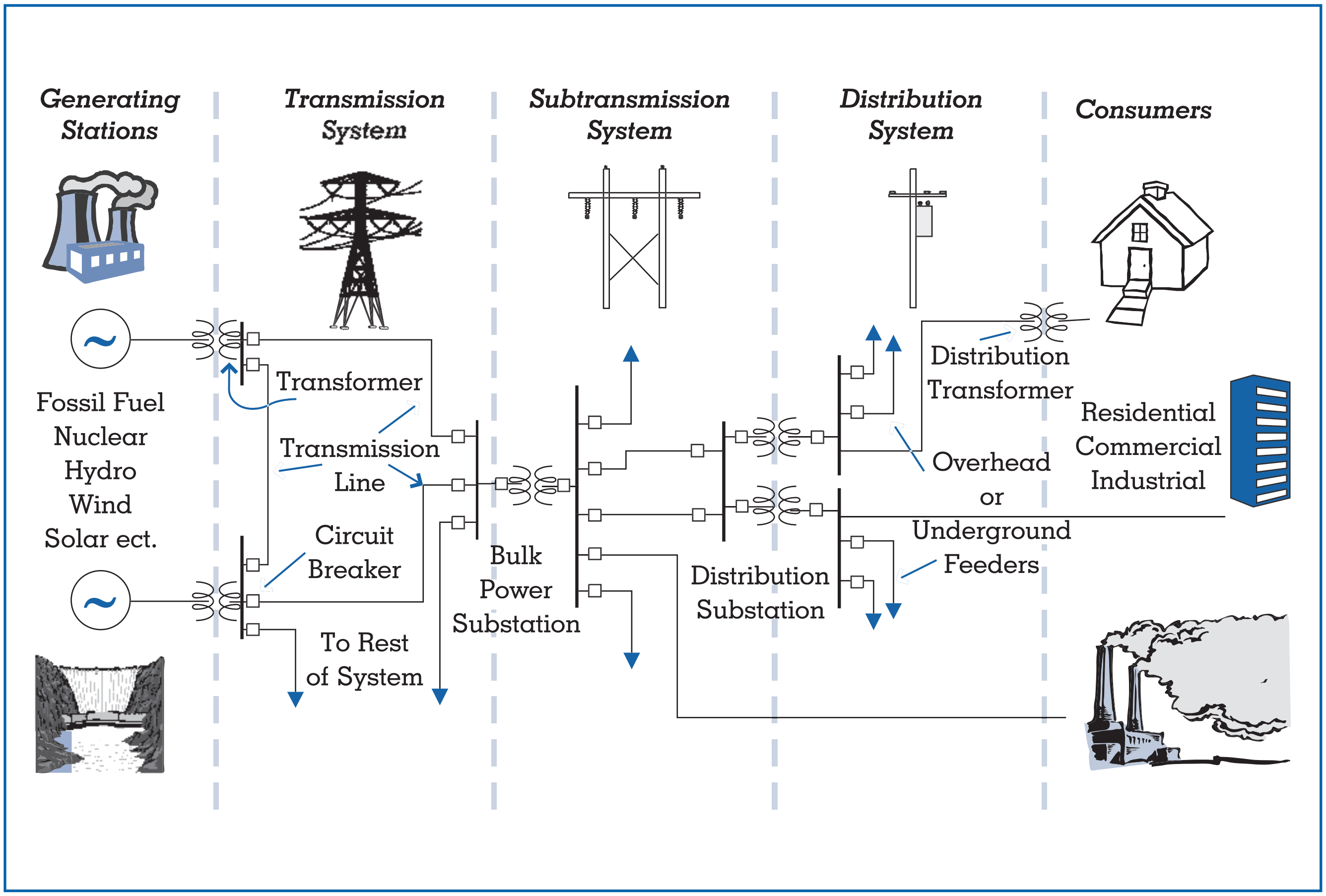


Figure 1.
A portion of the electric power transmission system.

generators would provide power to a group of users in a given region. In the 1930s, systems began to be interconnected for reliability and economic reasons. In the operation of any engineering system, equipment occasionally malfunctions due to degradation of the equipment itself or because of outside influences. When this happens in small systems, the load is no longer served and the region is subjected to blackout. Relatively minor changes to the system, such as the addition or removal of a single large load, also cause a significant impact on system operating frequency or other parameters. The interconnection of a large number of generators over a wide area avoids both types of problems. The loss of one generator is made up quickly by the controls on the other generators on the system. A change in load has a much smaller impact when the total load is many orders of magnitude larger. Additionally, the interconnection of several local systems permits economic transactions, so that a utility requiring more power may purchase from another generating company rather than utilizing its own generation facilities, which may not be available or may be more expensive to operate.

The North American power system (covering continental Canada, the conterminous United States, and parts of northern Mexico) is made up of four independent power systems, with special connections among them. The independent systems are the Eastern interconnection, the Western interconnection, Quebec, and Texas. Within any of the four systems, there is a high level of connectivity, and all generators have coordinated control systems to enable them to work together fully. Among the systems are one or more high-voltage direct current (HVDC) links, which permit the flow of power for commerce and reliability. Several other smaller power systems also operate in North America but do not have any interconnection to the larger systems. There are a large number of other interconnections internationally in Europe and elsewhere.

Despite the very strong advantages to interconnection, there is an inherent weakness in the interconnection of large systems covering vast distances. Under some operating conditions, the system becomes unstable, and a small change may have a large impact. This occurs most frequently when the system is heavily loaded, such as during the summer, under heavy air-conditioning loads. Interarea oscillations may occur as the internal control systems for the various generators respond at slightly different times, so that power flow through transmission lines may widely fluctuate and even change direction. Careful monitoring and control must be maintained to avoid unstable operating regimes, and extensive efforts have gone into the development of simulation techniques to predict potentially unstable operating conditions so they may be avoided. Computational analyses are typically done that assess the impact of single and double contingencies—that is, the widespread effect of the failure of a single piece of equipment or generating unit or the loss of a single transmission line, or some combination of events. Naturally, inasmuch as the system has so many elements all working together, it is impossible to predict every possible contingency. Because of this, occasionally unstable operating regimes are entered, which may result in local or even widespread outages.

The most infamous outage in U.S. history was the Northeast blackout of November 9–10, 1965, in which thirty million people lost power for as long as thirteen hours. In this case, large quantities of power were being transmitted over long lines to New York City. The initiating event was the tripping of a single transmission line on the Ontario–New York border. This resulted in several other transmission lines having to pick up the load that had previously been carried by that line, and those lines overheated and tripped, removing 1,800 MW of generation (at Niagara Falls) from the system. The entire Northeast power system became unstable and separated into a number of different isolated systems, none of which had a balance between generation and loads. This resulted in the remaining generation tripping off line, and a widespread outage covering much of New York, Ontario, New England, New Jersey, and Pennsylvania. The blackout was widespread, and the damages that resulted from looting and panic were extraordinary. In response to the outage, the National Electric Reliability Council (now the North American Electric Reliability Council—NERC) was created, with the responsibility to ensure system security and reliable operation. NERC is owned by ten regional coordinating councils consisting of various utilities, power producers, power marketers, and customers. While many other blackouts have occurred since the one in 1965, some affecting millions of customers, none has affected as many customers, and none has had an impact on the industry that was so widespread. This is largely due to the efforts of NERC to study and promote reliability and establish policies, guidelines, and standards conducive to reliable operation.

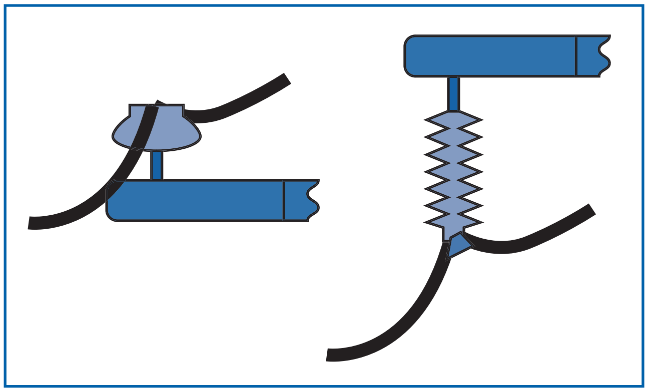


Figure 2.
Typical pin type insulator (left) and suspension insulator (right).

TRANSMISSION LINE CHARACTERISTICS AND THEIR IMPLICATIONS

The performance of a transmission line, and its limitations, are directly related to physical parameters that come from its design, construction, and even its location. Those parameters are common to many electrical circuits: resistance, inductance, and capacitance.

The series resistance of a transmission line is closely related to the losses that will be dissipated when current passes through the line (proportional to the square of the current magnitude). The resistance is proportional to the length of the line but inversely proportional to the cross-sectional area of the conductor.

The losses, and hence the effective resistance, are also increased by passing the line in close proximity to a noninsulating surface—for example, passing the line over seawater. The metal from which the conductor is made is also very important—for example, copper has a lower resistance, for the same geometry, than aluminum does. Also related to the losses of the transmission line is the shunt resistance. Under most circumstances, these losses are negligible because the conductors are so well insulated; however, the losses become much more significant as the insulators supporting the transmission line become contaminated, or as atmospheric and other conditions result in corona on the line.

The series inductance of the transmission line is a measure of the energy stored in the magnetic field of the conductor. High inductance is usually the limiting factor on the ability to transmit power over long distances, because the stability limit for power transfer is inversely proportional to the line inductance. Inductance increases for conductors that are farther apart, and decreases for conductors having a larger diameter.

The shunt capacitance of the transmission line is related to the energy stored in an electric field between conductors and/or earth. Capacitance negatively influences the operation of the transmission line by requiring higher currents from the generators (charging current), and by inducing a voltage rise (sometimes well in excess of safe operating limits) in a lightly loaded transmission line (the Ferranti effect). The capacitance increases when conductors are brought closer together, and it decreases for conductors having a smaller diameter.

The actual flow of power through the transmission system from one point to another is dictated by the operation of all of the generators and by the resistance, inductance, and capacitance of the system. Because those parameters are mainly characteristics of design, in the past very little flexibility in control was provided to change the path of power flow. In one well-known example, there is a lot of power generation that occurs in Niagara Falls, New York, and New York City consumes a lot of power. It has frequently happened that the flow of the majority of the power from Niagara Falls to New York City, rather than being along the most direct route, has been from Niagara Falls into Ontario, through Michigan, Ohio, Pennsylvania, and New Jersey before getting to New York City. Despite economic transactions and contractual agreements that would call for a direct route with minimized losses, the path of power flow is dictated by the laws of physics and the system parameters.

There are several techniques, however, that modify the system parameters to manipulate the power flow in localized regions. The one that has been in use the longest is modifying the parameters of generator operation, the quantity of power being produced, and the voltage at the terminals of the machine. This sometimes requires operation of less efficient and more costly generating stations, instead of optimizing efficiency and cost, and other limitations frequently come into play as well. The second approach is to install a very costly type of specialized transformer, a phase angle regulating transformer, at crucial points in the system. This device provides limited control, but can of modify power flows by imposing a change in phase angle in the voltage and current going through it. Switched series capacitor banks are sometimes used to reduce the effective inductance of a transmission line, permitting more power to flow through it, but these are not highly controllable devices either.

A fairly new technology that provides a high level of control for power flows and system stability is a class of devices known as flexible alternating current transmission systems (FACTS). A number of FACTS devices have been used at all different voltage levels. On the high-voltage system are devices such as static var compensators (SVC), thyristor controlled series capacitors (TCSC), static synchronous compensators (STATCOM), and universal power flow controllers (UPFC). These devices work by canceling out the inductance and/or capacitance of the transmission lines and the system loads. They are operated with high-power electronic devices and are fully controllable, either manually or automatically. Because of their dynamic nature and quick response time, they can respond to system disturbances and can provide an extraordinary increase to system stability.

Still another approach, which has been used since the 1960s to control power flow and transmit power over very long distances at high efficiencies, is high-voltage direct current transmission (HVDC). While they fill the same role as high-voltage ac transmission lines (bulk power transmission), HVDC lines are impervious to the effects of system inductance and capacitance, so power flow and system stability are not influenced by those parameters. Each end of the HVDC transmission line is located in a special converter station where high-power electronic devices convert alternating current to direct current at one end and reverse the process at the other end. HVDC power transmission is more efficient and more controllable, but the converter stations are very costly, so the planning and design of such a line includes a careful cost/benefit analysis in which typically the HVDC system is only chosen over a high-voltage ac line for very long distances or as an interconnection between independent systems.

TYPES OF TRANSMISSION LINES

For the ac or dc interconnections that span great distances, high-voltage transmission lines are almost exclusively overhead. These lines are typically constructed with aluminum wrapped around a reinforcing steel core. The conductors are connected to the support structures with suspension insulators that may be made of porcelain or polymers. The type of structure that supports the transmission line, the distance between structures, the arrangement of conductors, and so forth are functions of operating voltage, terrain, climate, right-of-way, aesthetic concerns, cost, etc. Common types of structures include treated wood, steel lattice, tubular steel, and concrete. Overhead power transmission lines are less expensive than underground transmission lines and are easier to install, inspect, and maintain. However, they are also more exposed to the hazards of environmental conditions such as severe winds, ice formation, and corrosion and contamination due to airborne pollutants.

Underground transmission lines are preferred in places where rights-of-way are severely limited because they can be placed much closer together than overhead lines. They are also favored for aesthetic reasons. They may be directly buried in the soil, buried in protective steel or plastic pipes, or placed in subterranean tunnels. The conductors are usually contained within plastic insulation encased in a thin metallic sheath. The conductors enclosed in steel pipes may be immersed in oil, which may be circulated for cooling purposes. For all types of underground lines, the capacitance is higher than for overhead lines, and the power transfer capability is usually limited by the resistive losses instead of the inductance. While not exposed to environmental hazards, underground cables are at risk of damage by rodents, construction, or geological instabilities.

In the 1990s, significant efforts have gone into the development of transmission lines made from superconductors, which are materials that have essentially no resistance when operating at extremely low temperatures. While a significant amount of energy is required to operate the cryogenic systems, the reduction in power loss and the increase in power transfer capability have made the technology appealing. Several prototype installations in various stages of design and preparation have provided encouraging preliminary results, but as of the end of the twentieth century, no full-scale applications have been deployed.

ENVIRONMENTAL IMPACT OF TRANSMISSION SYSTEMS

While many people are concerned about the aesthetic impact of transmission lines, several other aspects of the transmission line environmental impact must also be considered. There is a danger to wildlife imposed by the presence of high-voltage surfaces such as overhead transmission lines. For small birds that land on transmission lines there is very little hazard because there is no path for current to flow through them. However, birds with large wingspans or climbing rodents will often reach between surfaces energized at different voltages, killing them and damaging system equipment.

Overhead transmission lines require that the area beneath them be cleared of trees or tall shrubs, which may result in erosion. When the transmission line right-of-way is not kept clear, the transmission line may come into contact with vegetation, causing a fault on the system and possibly starting a fire. Chemical contamination of soil may result from some types of transmission structures, such as treated wood. Burial of underground cables also can impact the environment due to erosion.

Another environmental concern that has been raised is the fear that electromagnetic fields (EMF) may cause negative physiological effects. Various epidemiological studies have purported to find an association between the presence of transmission lines and different types of cancer, especially in children. These studies have found at most a very weak association, and were based on estimated and not measured electromagnetic fields. While the issue has been given significant media attention, inciting grave public concern and much new research, the medical and scientific communities have been unable to definitively confirm that EMF has any measurable physiological impact.

TRANSMISSION SYSTEM ORGANIZATIONAL STRUCTURE

Until recently, all functions related to the generation, transmission, and distribution of electric power in the United States were executed under the umbrella of monopolies regulated by state utility boards (under a variety of titles). Through the 1990s, legislation has resulted in the deregulation of that industry and many changes in the nature of commerce in the energy industry. The details of the new system are still in development, but the deregulated system as presently designed consists of a plan to separate the utilities into independent entities consisting of regulated businesses and unregulated businesses. The regulated businesses will include the transmission companies and the distribution companies. It is necessary to regulate these companies to utilize the assets that exist to provide power to individual users and to transfer bulk power without the proliferation of an excessive quantity of new infrastructure. The deregulated businesses will consist of generating companies responsible for the production of power, and retail companies responsible for sale, billing, and other energy-related services. Sale and marketing transactions will be conducted by power marketers, with transactions being conducted on the power exchange. The entity responsible for observing and operating the overall system, ensuring that all functions are executed within the safe and reliable operating limits, will be the independent system operator (ISO).

John A. Palmer

See also: Capacitors and Ultracapacitors; Electric Motor Systems; Electric Power, Generation of; Electric Power, System Protection, Control, and Monitoring of; Electric Power, System Reliability and; Electric Power Substations; Environmental Problems and Energy Use; Insulation; Transformers.

BIBLIOGRAPHY

Faulkenberry, L. M., and Coffer, W. (1996). *Electrical Power Distribution and Transmission*, Englewood Cliffs, NJ: Prentice-Hall.

Glover, J. D., and Sarma, M. (1994). *Power System Analysis & Design*, 2nd ed. Boston: PWS.

Gyugyi, L., et al. (1995). "The Unified Power Flow Controller: A New Approach to Power Transmission Control." *IEEE Transactions on Power Delivery* 10:1085–1093.

National Research Council. (1995). *EMF Research Activities Completed Under the Energy Policy Act of 1992: Interim Report, 1995.* Washington, DC: National Academy Press.

ELECTRIC VEHICLES

Electric vehicle (EV) is an abbreviated term for battery electric vehicle, which is a vehicle that uses energy from a battery to operate an electric motor that rotates the wheels. Throughout the twentieth century, a wide array of electric cars, trucks, buses, bikes, and scooters have been developed. However, except for the golf cart and special delivery vehicles, none has experienced any sustained commercial success.

THE BASICS

EVs use a variety of electric motors, dc, ac, and three-phase ac induction being the most common. No one electric motor is optimal for all vehicles. Motor selection criteria include power needs, variable speed needs, operating voltage, and physical space limitations. Researchers at the turn of the twenty-first century are looking into directly attaching electric motors to the wheels to eliminate the drive shaft and differential. This will improve efficiency by reducing weight and by minimizing the mechanical losses between the motor and wheel so that propulsion to the wheels is generated at the wheels.

EVs deliver superior efficiency because of the electric motor. Approximately 75 percent of the battery's chemical energy is used by the electric motor to generate the mechanical energy rotating the wheels. In addition, little electricity is wasted during recharging. The extremely efficient rechargeable batteries of electric vehicles result in somewhere between 70 to 90 percent of the electrical energy used to charge the batteries being available during discharging. In comparison, the internal combustion engine (ICE) vehicle converts only about 20 percent. This figure drops even further when taking into account the energy losses through the transmission and idling. Because the electric motor does not consume any energy while the vehicle is stationary, the EV operates even more efficiently than the ICE vehicle in urban settings, especially for EVs with regenerative braking technology, which recovers braking energy that normally would be dissipated as heat to charge the batteries.

The key component of EVs remains the energy source making propulsion possible: batteries. Advances in batteries are necessary to match the performance of ICE vehicles, particularly the high power-to-weight ratio for acceleration and the high energy-to-weight ratio for range. Since the heavier the battery, the lower the power output per pound of weight, designers face a seemingly no-win tradeoff: additional batteries added to achieve greater acceleration result in more weight that takes away from vehicle range and performance. However, some of the weight problem can be offset by teaming high power density/low energy density devices, such as ultracapacitors or flywheels, to handle peak loads with the low power density/high energy density battery.

Unfortunately, often the EV's efficiency and cost advantages are wiped out by the battery's abbreviated life expectancy. When battery life expectancy is around 2,000 miles (reported by testers of General Motors's EV, the Impact), then the recharging and replacement battery costs of an EV (over $1,500 for the GM Impact) alone can exceed the total operating costs for an ICE vehicle. Energy costs per mile for the GM Impact ($0.02 electricity plus $0.42 for batteries) were greater than the total operating costs for an ICE vehicle ($0.42).

The energy costs of building vehicles must also be considered. For ICE vehicles, more energy is usually used in construction of the vehicle than will be consumed in fuel for driving 100,000 miles. For the EV, the dynamics are even worse since the material and energy costs of batteries are considerable. Batteries are expensive since they entail a substantial amount of material (added weight) and often involve multiple complex construction. For example, the thirty-two advanced lead-acid batteries for the 1995 GM Impact weighed over 850 pounds.

EARLY DEVELOPMENT

EVs are not new. They actually originated in the 1880s. In fact, it may be difficult to imagine, but at one time EVs were more popular than ICE vehicles. In the early 1900s, of the approximately 8,000 motor vehicles in the United States, 40 percent were powered by steam, 38 percent by battery, and 22 percent by gasoline. Since steam-powered vehicles needed a constant supply of water and high steam pressure, they proved expensive and difficult to maintain, which lead to their demise by around 1910. The problems with internal combustion engine (ICE) vehicles at that time were difficulties in starting, noisiness, and unreliability. In contrast, early EV owners enjoyed immediate starts, quiet operation, minimum maintenance, and impressive performance for the times: speeds of 15 to 20 miles per hour and 30 to 40 miles between charges.

The popularity of EVs did not last. By the 1920s, the performance of ICE vehicles improved dramatically, and the earlier major drawbacks had been solved. Ironically, the replacement of the dangerous hand crank with a battery-powered electric starter was a major innovation accelerating ICE vehicle sales at the expense of EV sales. At the same time, there were no concurrent solutions to the limitations of battery technology for EVs that addressed the demand to drive faster and farther.

REEMERGENCE DURING THE OIL SHORTAGES

From 1920 until the oil shortage crisis of the 1970s, the only major uses of the EV were for golf carts and neighborhood delivery vehicles. In the United Kingdom, the lower operating costs and fewer repairs of the EV more than offset the disadvantages of its modest range, speed, and acceleration for these short-range delivery vehicles.

Several EVs came on the market in the 1970s, both in the United States and abroad, as a transportation alternative in answer to soaring gasoline prices and supply disruptions causing the long lines at gasoline stations. But as the oil shortages ended in the early 1980s, interest in EVs again faded. EV technology advanced, but not nearly enough to seriously compete with ICE vehicles.

THE AIR QUALITY CRISIS

Regulations imposed on auto makers to address a perceived air pollution crisis once again renewed EV interest in the early 1990s. Perhaps the most aggressive regulations were imposed by the California Air Resources Board (CARB) low emission vehicle (LEV) program mandating that zero emission vehicles

(ZEV) comprise 2 percent of manufacturers sales by 1998 and 10 percent by 2003. The belief among regulators was that mandates would spur innovation, create the economies of scale to lower costs (as a result of greater vehicle output), and incite the sales momentum that would begin a large-scale transition to EVs. Considering the state of EV technology at the time, the auto industry was very reluctant to spend the millions necessary to develop an electric car that their marketing research told them people did not want.

General Motors made the most noteworthy foray into the market in 1995 with the Impact (later renamed the EV1). Unlike the offerings of other auto makers, the Impact was an all-electric design from the ground up. The advanced lead-acid battery series delivered 320 volts, and with an inverter that converted direct current to alternating current for transfer to two induction motors (one for each front wheel), the car had impressive performance: quick and smooth acceleration from 0 to 60 mph in eight seconds, and a top speed of 110 mph. Other innovative features included the lowest-drag exterior design, high-pressure tires that had half the rolling resistance of ordinary tires, and a regenerative braking system. However, there were problems. Compared to ICE vehicles, the Impact was considerably more expensive; its lightweight body design was not as safe; and its range was limited to 55 to 95 miles between charges (depending on terrain, driving habits, and temperature). A full charge took 6 hours, and it turned out the batteries needed to be replaced after only 2,000 to 3,000 miles, not 20,000 miles as hoped. Attempting to address the battery shortcomings, in 1999 GM began offering more expensive nickel-metal hydride batteries as an option. These batteries extended the range to 75 to 130 miles, but also took slightly longer to recharge.

The much higher initial cost remains the most significant objection to EVs. Economies of scale could solve the cost problem, yet range, change time, and battery life expectancy will continue to be major disadvantages. Even if the range of EVs was extended to 300 miles between charges, battery life extended to 40,000 miles, and full recharge time reduced to 30 minutes, these are still limitations that ICE vehicle owners do not have to endure. Moreover, EVs offer limited functionality at a much higher cost (even with federal and state subsidies), while the public keeps looking for more functionality. For example, many sports utility vehicle purchases are driven by desire for their unmatchable multi-functionality. People use the vehicle for many things besides commuting: car pooling duties, hauling lumber, backroad touring, towing a boat, and plowing through the occasional snow storm, to name a few. The additional utility of these vehicles may cost the buyer a sizable premium, yet the buyer is happy to pay it for the security of having the added functionality, even though the advanced functionality may be terribly underutilized.

ICE innovations that resulted in significant emission reductions also hastened the demise of EVs. In fact, Toyota and Honda in 1999 announced the development of ICE vehicles that could meet California's Super Ultra Low Emission Standard proposed for 2004. It can be argued that these super ultra low emission vehicles (SULEVs) will generate levels of emissions comparable to EVs if the power plant emissions that generate the electricity to power the EV batteries are taken into account. However, figuring out EV emissions from the electric generation mix is a very imprecise science. Almost all of the electric energy will come from emission-producing coal and natural gas power plants, not emission-less hydroelectric and nuclear energy since these sources are "first through the meter." Thus EVs are not really zero emission vehicles. They are what Amory Lovins has called "elsewhere emission vehicles"—not occurring at the tailpipe but at a distant smokestack. Elsewhere emissions are still considered a worthwhile goal by many since the non-attainment of air quality standards is almost exclusively a problem for urban areas, and natural gas and coal power plants are usually all located outside of urban areas. Further, it is also easier to monitor and control emissions at a small number of stationary sources (power plants) than the tailpipe of every vehicle.

The combination of ICE innovation, the lack of EV advances, and the dismal acceptance of EVs by consumers forced California's ZEV program to suspend the vehicle sale mandates in 1995. The mandates were replaced by agreements with automakers to make a concerted effort and demonstrate technology. Nevertheless, declaring an inability to develop demand, Honda withdrew the EV Plus from the market in 1999, followed by General Motors withdrawing the EV1 (formerly Impact) in 2000. All automakers have shifted their efforts to developing the other more promising emerging electric vehicle options such as hybrid and fuel cell technologies.

Despite the repeated failures of EVs in the marketplace, there is widespread agreement that electric vehicles make sense for selected applications. For instance, city delivery vans have represented a good niche for EVs because they have sufficient space to be outfitted with large battery cases and do not need to have rapid acceleration or go long distances between charges.

As battery innovation continues, it is likely that the realm of applications for which EVs make sense will expand. A promising future market for electric vehicles may be electric scooters and electric motor-assisted bicycles. These types of vehicles will not only cut down on the urban air pollution problem, but also could help reduce congestion since the typical ICE vehicles being replaced are much larger. In particular, it is the hope in many highly populated urban areas of Asia that electric scooters and electric bicycles will become the personal transportation option of choice. In 2000, a nickel-metal hydride electric scooter developed by Ovonic Battery Company achieved a range of 73 miles with an efficiency equivalent to more than 300 miles per gallon of gasoline.

THE LARGE-SCALE RECHARGING CHALLENGE

By 1995 global passenger-car registrations surpassed 500 million, with the United States accounting for almost 160 million of the total. If significant performance improvements in battery technology occur that make EVs more attractive than ICE vehicles, major investments in the recharging infrastructure would be required to accommodate the millions of new EVs. This would present critical challenges for the electric utility industry, both at the bulk power system level and at the point of distribution or end use.

Those electric utility companies with low load growth and high generation capacity margins would benefit most from the increased power sales associated with having more electric vehicles. However, in the current transition to electric industry deregulation, generation capacity margins in many regions are declining to barely acceptable levels, since generating capacity is being added in response to market opportunities and not as part of a long range, comprehensive planning effort focusing on load growth. More importantly, a greatly increased number of EVs using charging systems located in residential areas could vastly change current load profiles. Electric power engineers are concerned that adding these large "single-phase loads" may adversely affect power quality, safety, and system reliability. (Single-phase loads are lower voltage loads placed on distribution systems that may unbalance three-phase systems and may require large current flows.)

Power system planners need to consider how the costs associated with electric vehicles should be passed along to consumers. Fortunately from a load balancing perspective, it is likely that most EV charging will occur in off-peak periods. Charging in off-peak periods would reduce utility costs and therefore should allow utilities to reduce customer rates, but this would require "time of day metering" that is not available in most service areas.

There are also public safety concerns associated with electric vehicle charging devices. Single-phase charging systems for home use could require current flows that far exceed 100 to 200 amp service installations that are typical in homes, and could result in overloads, breaker operations, and blown fuses on distribution systems. Older homes were not designed for these high current flows and it is likely that their electrical systems would have to be updated. In addition, charging systems would pose problems for people who live in apartments, coops, and condominiums who are accustomed to street parking. Three-phase charging systems located at service stations would be better for charging, since they deliver constant (nonpulsating) power at higher voltages and avoid the high current and unbalancing effects associated with single-phase systems. These systems could reduce some of the public safety concerns, produce more power, and charge batteries more quickly. Also, the service station charging approach would make it easier to achieve the goals set by the United States Advanced Battery Consortium: 3 to 6 hours for normal recharge time, and a fast recharge time of 50 percent of capacity in less than 30 minutes.

Both single-phase charging systems and three-phase charging systems would contain power electronic devices that as a side effect introduce waveform distortion and create "power quality" problems. Filtering devices used in conjunction with residential charging systems could be used to reduce harmonics and other power quality problems, but the cost of such filtering devices is currently quite high.

Some engineers have suggested that electric vehicle charging systems would have the least negative impact if integrated with emerging distributed resources technologies involving distributed generation and distributed storage. Distributed systems could help in addressing the new challenges of system operations in the open access, competitive, restructured electric power industry and relieve network congestion problems. In addition, the energy line losses could be significantly reduced by using distributed resources to meet electric vehicle charging requirements.

At present there is much speculation about the long-term future of EVs. Electric vehicles could dramatically lessen U.S. dependence on imported oil and help solve air quality problems, yet also will create new problems and challenges for electric utility system operators.

John Zumerchik
Fred I. Denny

See also: Batteries; Capacitors and Ultracapacitors; Electric Motor Systems; Emission Control, Vehicle; Environmental Problems and Energy Use; Flywheels; Fuel Cells; Fuel Cell Vehicles; Hybrid Vehicles; Materials; Transportation, Evolution of Energy Use and.

BIBLIOGRAPHY

Balzhiser, R. E., and Bryson, J. E. (1994). "The Strategic Role of Electric Vehicles." *Forum for Applied Research and Development* (Spring):31–34.

"Building the Electric Vehicle Future: EPRI's Vehicle Development Activities." (1987). Electric Power Research Institute Report EU.3017.11.87. Palo Alto, CA: Electric Power Research Institute.

Chachich, A. (1991). "Energy and Transportation: A Technological Update." In *The Energy Sourcebook*, eds. R. Howes and A. Fainberg. New York: American Institute of Physics.

"Charging Up Electric Vehicles." (1992). *EPRI Journal*, June.

D'Agostino. (1993). "The Electric Car." *IEEE Potentials* (February):28–32.

Fischetti, M. (1992). "Here Comes the Electric Car—It's Sporty, Aggressive and Clean." *Smithsonian* 23(1):34–43.

Flink, J. J. (1970). *America Adopts the Automobile, 1895–1910*. Cambridge, MA: MIT Press.

Greene, D. (1996). *Transportation and Energy*. Lansdowne, VA: Eno Transportation Foundation Inc.

Miller, K. (1996). "The Future Is Electric." *Life* 19(2):110–116.

Riezenmann, M. J. (1992). "Electric Vehicles." *IEEE Spectrum* 29(11):18–21.

Sperling, D. (1995). *Future Drive: Electric Vehicles and Sustainable Transportation*. Washington, DC: Island Press.

U.S. Department of Energy. (1995). "DOE'S Energy Partnerships for a Strong Economy: Electric Utilities Meet the Challenge." February/March, p. 3.

EMISSION CONTROL, POWER PLANT

Power plant emissions result from the combustion of fossil fuels such as coal, gas, and oil. These emissions include sulfur dioxide (SO_2), nitrogen oxides (NO_x), particulate matter, and hazardous air pollutants, all of which are subject to environmental regulations. Another emission is carbon dioxide (CO_2), suspected of being responsible for global warming.

Historically, under both federal and state regulations, the demand for gas to heat homes and to meet needs of business and industry took priority over utility use to generate electricity. These restrictions have been eased by amendments to the Fuel Use Act in 1987, and, as a result, new gas-fired generation units are being constructed. However, coal-fired units continue to provide over 50 percent of the total utility generation of electricity.

Until the late 1960s, a typical electric utility scenario was one of steadily growing electricity demand, lower costs of new power plants through technological advances, and declining electricity prices. Utility companies appeared before Public Utility Commissions (PUCs) to request approval for rate reductions for customers. In the 1970s, multiple factors caused costs to increase dramatically: fuel costs escalated, primarily because of oil embargoes; legislators passed more stringent environmental laws requiring huge investments in emissions control technology; and these costs escalated as inflation and interest rates soared. When utility companies requested rate increases to cover these higher costs, PUC hearings were no longer sedate and routine. Cost recovery pressures continued throughout the 1980s as environmentalists succeeded in lowering emissions limits.

FORMATION OF POLLUTANTS IN UTILITY BOILERS

Sulfur is found in coal in organic forms as well as inorganic forms such as pyrites, sulfate, and elemental. Organic sulfur is the most difficult to remove, and reliable analytical methods are required to support coal-cleaning technologies designed to remove the sulfur prior to burning the coal. At the present time, most utilities burn “uncleaned” coal, and, upon combustion, most of the sulfur is converted to SO_2, with a small amount further oxidized to form sulfur trioxide (SO_3)

$$\mathrm{S(coal)} + O_2 \rightarrow SO_2$$
$$SO_2 + \tfrac{1}{2}\,O_2 \rightarrow SO_3$$

The sulfur content of coals available to utilities ranges from about 4 percent in high-sulfur coals to less than 1 percent in some Western coals. Although transportation costs may be higher for Western coals, many Eastern utilities elect to burn Western coals to comply with increasingly stringent SO_2 regulations.

NO_x emissions are less dependent on the type of coal burned, and two oxidation mechanisms are associated with the release of NO_x into the atmosphere during the combustion process. Thermal NO_x results from the reaction of nitrogen in the combustion air with excess oxygen at elevated temperatures, and fuel NO_x is a product of the oxidation of nitrogen chemically bound in the coal.

Hazardous air pollutants (HAPs) are substances that may cause immediate or long-term adverse effects on human health. HAPs can be gases, particulates, trace metals such as mercury, and vapors such as benzene. For coal-fired power plants, the HAPs of most concern are metals such as mercury, arsenic, and vanadium.

All combustion processes produce particulate matter. Amounts and size distribution of the particulates emitted depend on a number of factors, including fuel burned, type of boiler, and effectiveness of collection devices.

A wide variety of control technologies have been installed by utilities throughout the United States to reduce the emissions of these pollutants. At the same time, research on new technologies is being conducted to ensure compliance with future environmental standards.

ENVIRONMENTAL ISSUES

The legislation most responsible for addressing power plant emissions is the Clean Air Act. Initially established in 1970 with major amendments in 1977 and 1990, it provides for federal authorities to control impacts on human health and the environment resulting from air emissions from industry, transportation, and space heating and cooling. In the original 1970 programs, National Ambient Air Quality Standards (NAAQS) were established for six "criteria" air pollutants—SO_2, NO_x, particulate matter, ozone, lead, and carbon monoxide—at a level to protect human health and welfare and the environment with a "margin of safety." New Source Performance Standards (NSPS) were set for major new facilities projected to emit any pollutant in significant amounts. To receive an operating permit, a new unit must meet or exceed control standards established by the Environmental Protection Agency (EPA). In the 1977 Amendments, permits required control levels for new plants that were not only as stringent as NSPS but also reflected the best available technologies.

The reduction of atmospheric concentrations of the sulfur and nitrogen oxides blamed for acid rain was a major issue in the debate that led to the 1990 Clean Air Act Amendments (CAAA). The final legislative action is one of the most complex and comprehensive pieces of environmental legislation ever written.

The 1990 CAAA contain the following sections:

Title I: Provisions for Attainment and Maintenance of National Ambient Air Quality Standards
Title II: Provisions Relating to Mobile Sources
Title III: Hazardous Air Pollutants
Title IV: Acid Deposition Control
Title V: Permits
Title VI: Stratospheric Ozone Protection
Title VII: Provisions Relating to Enforcement.

Titles I and IV are most relevant to SO_2 and NO_x control. Title I establishes a 24-hour average ambient air standard for SO_2 of 0.14 ppm. The NO_x provisions require existing major stationary sources to apply reasonably available control technologies and new or modified major stationary sources to offset their new emissions and install controls representing the lowest achievable emissions rate. Each state with an ozone nonattainment region must develop a State Implementation Plan (SIP) that includes stationary NO_x emissions reductions.

Title IV, the Acid Rain Program, addresses controls for specific types of boilers, including those found in coal-fired power plants. A two-phase control strategy was established. Phase I began in 1995 and originally affected 263 units at 110 coal-burning utility plants in twenty-one eastern and midwestern states. The total of affected units increased to 445 when substitution or compensating units were added. In Phase II, which began January 1, 2000, the EPA has established lower emissions limits and also has set restrictions on smaller plants fired by coal, oil, and gas. For example, the Phase I SO_2 emissions limit is 2.5 lb/million Btu of heat input to the boiler whereas the Phase II limit is 1.2 lb/million Btu. In both phases, affected sources will be required to install systems that continuously monitor emissions to trace progress and assure compliance.

One feature of the new law is an SO_2 trading allowance program that encourages the use of market-based principles to reduce pollution. Utilities may trade allowances within their system and/or buy or sell allowances to and from other affected sources. For example, plants that emit SO_2 at a rate below 1.2 lb/million Btu will be able to increase emissions by 20 percent between a baseline year and the year 2000. Also, bonus allowances will be distributed to accommodate growth by units in states with a statewide average below 0.8 lb/million Btu.

The Clean Air Act of 1970 and the Amendments of 1977 failed to adequately control emissions of hazardous air pollutants, that are typically carcinogens, mutagens, and reproductive toxins. Title III of the 1990 Amendments offers a comprehensive plan for achieving significant reductions in emissions of hazardous air pollutants from major sources by defining a new program to control 189 pollutants.

Although the petrochemical and metals industries were the primary focus of the toxic air pollutants legislation, approximately forty of these substances have been detected in fossil power plant flue gas. Mercury, which is found in trace amounts in fossil fuels such as coal and oil, is liberated during the combustion process and these emissions may be regulated in the future. EPA issued an Information Collection Request (ICR) that required all coal-fired plants to analyze their feed coal for mercury and chlorine. Since these data will be used in making a regulatory decision on mercury near the end of the year 2000, it is critical that the power industry provide the most accurate data possible.

In 1987, health- and welfare-based standards for particulate matter (measured as PM_{10}, particles 10 micrometers in diameter or smaller) were established. A 10 micrometer (micron) particle is quite small; about 100 PM_{10} particles will fit across the one millimeter diameter of a typical ballpoint pen. For PM_{10} particles, an annual standard was set at 50 micrograms per cubic meter (50 $\mu g/m^3$) and a 24-hour standard was set at 150 $\mu g/m^3$.

Since these PM_{10} standards were established, the EPA has reviewed peer-reviewed scientific studies that suggest that significant health effects occur at concentrations below the 1987 standards. In addition, some studies attributed adverse health effects to particles smaller than 10 microns. In July 1997, the EPA, under the National Ambient Air Quality Standards (NAAQS), added standards for particulate matter with a diameter of 2.5 microns or less ($PM_{2.5}$). The annual $PM_{2.5}$ standard was set at 15 $\mu g/m^3$ and the 24-hour $PM_{2.5}$ standard was set at 65 $\mu g/m^3$.

Through implementing new technologies and modifying unit operating conditions, the electric utility industry has significantly reduced the emissions of SO_2, NO_x, and particulates since passage of the 1970 Clean Air Act and its subsequent amendments. With full implementation of Title IV of the 1990 CAAA, the 1990 baseline level of more than 14.5 million tons of SO_2 will be reduced to 8.9 million tons per year. NO_x emissions during Phase I will be reduced by 400,000 tons per year, and Phase II will result in a further reduction of 1.2 million tons per year. Particulate control devices, installed on nearly all coal-fired units, have reduced particulate emissions from more than three million tons per year in 1970 to less than 430,000 tons per year in 1990.

Over the years, utilities have funded research to develop technologies that not only will meet existing standards but also will meet future emissions reductions based on continuing concerns about acid rain, ozone, fine particulates, and other environmental issues. To remain competitive in the global economy, utilities must seek technologies that balance the conflicting drivers of productivity demands, environmental concerns, and cost considerations. Engineering designs should minimize costs and environmental impact, and, at the same time, maximize factors such as reliability and performance.

CLEAN COAL TECHNOLOGY PROGRAM

The Clean Coal Technology (CCT) Program, a government and industry cofunded effort, began in 1986 with a joint investment of nearly \$6.7 billion. The recommendation for this multibillion dollar program came from the United States and Canadian Special Envoys on Acid Rain. The overall goal of the CCT Program is to demonstrate the commercial readiness of new, innovative environmental technologies. The program is being conducted through a multiphased effort consisting of five separate solicitations administered by the U.S. Department of Energy. Industry proposes and manages the selected demonstration ventures. Many of the projects funded in the first stages of the program are generating data or have finished their testing program. Within the next few years, the United States will have in operation a number of prototype demonstration projects that promise to meet the most rigorous environmental standards.

The CCT projects, in general, are categorized as follows:

1. Advanced electric power generation
    a. Fluidized bed combustion
    b. Integrated gasification combined cycle
    c. Advanced combustion systems
2. Environmental control technologies
    a. Sulfur dioxide control technologies
    b. NO_x control technologies
    c. Combined SO_2/NO_x control technologies
3. Coal processing for clean fuels
    a. Coal preparation technologies

   b. Mild gasification
4. Indirect liquefaction
5. Industrial applications.

The following sections describe the various compliance options for controlling emissions from utility power plants.

SULFUR DIOXIDE CONTROL TECHNOLOGIES

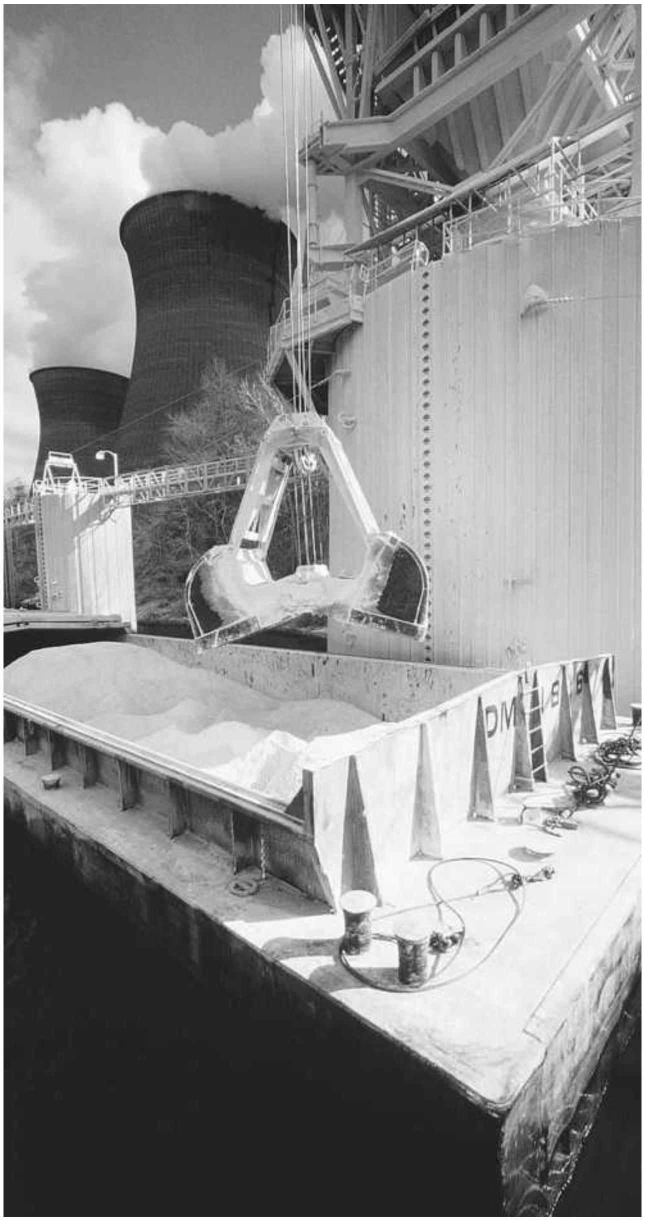
Three major compliance options for SO_2 emissions available to utilities using coal-fired boilers are to switch fuels, purchase/sell SO_2 allowances, or install flue gas desulfurization (FGD) technologies. Costs, availability, and impact on boiler operation must be considered when evaluating switching to low-sulfur coal or natural gas. As more utilities enter the free market to purchase SO_2 allowances, prices will rise. Therefore, to minimize costs and, at the same time, meet environmental standards, power producers should continuously monitor the tradeoffs among these three options.

Although FGD processes, originally referred to as scrubbing SO_2 from flue gas, have been available for many years, installations in the United States were quite limited until passage of the Clean Air Act of 1970. Even then, installations were usually limited to new facilities because existing plants were exempt under the law.

Projects in the CCT program demonstrated innovative applications for both wet and dry or semidry FGD systems. The wet FGD systems, which use limestone as an absorber, have met or exceeded the 90 percent SO_2 removal efficiency required to meet air quality standards when burning high-sulfur coal. The dry or semidry systems use lime and recycled fly ash as a sorbent to achieve the required removal.

In wet FGD systems, flue gas exiting from the particulate collector flows to an absorber. In the absorber, the flue gas comes into contact with the sorbent slurry. The innovative scrubbers in the CCT program featured a variety of technologies to maximize SO_2 absorption and to minimize the waste disposal problems (sludge).

A number of chemical reactions occur in the absorber beginning with the reaction of limestone ($CaCO_3$) with the SO_2 to form calcium sulfite ($CaSO_3$). The calcium sulfite oxidizes to calcium sulfate ($CaSO_4$) which crystallizes to gypsum ($CaSO_4 \bullet 2H_2O$). The gypsum crystals are either stored in onsite waste disposal landfills or shipped to a facility where the gypsum is used in manufacture of wallboard or cement. The scrubbed gas then passes through mist eliminators that remove entrained slurry droplets. Recovered process water is recycled to the absorption and reagent preparation systems.


Lime being unloaded at a power plant. The material will be used to reduce sulfur dioxide emissions as part of its flue gas desulfurization system. (Corbis Corporation)

In the dry or semidry FGD system, the sorbent, usually lime, is injected into the flue gas stream before the particulate collection device. The lime, fed as a slurry, quickly dries in the hot gas stream and the particles react with the SO_2. The resulting particles are then removed, along with the fly ash, by the particulate collection device.

NO_x CONTROL TECHNOLOGIES

Combustion modifications and postcombustion processes are the two major compliance options for NO_x emissions available to utilities using coal-fired boilers. Combustion modifications include low-NO_x burners (LNBs), overfire air (OFA), reburning, flue gas recirculation (FGR), and operational modifications. Postcombustion processes include selective catalytic reduction (SCR) and selective noncatalytic reduction (SNCR). The CCT program has demonstrated innovative technologies in both of these major categories. Combustion modifications offer a less-expensive approach.

Because NO_x formation is a function of the temperature, fuel-air mixture, and fluid dynamics in the furnace, the goal of a combustion modification is to mix fuel and air more gradually to reduce the flame temperature (lower thermal NO_x production), and to stage combustion, initially using a richer fuel-air mixture, thus reducing oxidation of the nitrogen in the fuel. LNBs serve the role of staged combustion.

Overfire air (OFA) is often used in conjunction with LNBs. As the name implies, OFA is injected into the furnace above the normal combustion zone. It is added to ensure complete combustion when the burners are operated at an air-to-fuel ratio that is lower than normal.

Reburning is a process involving staged addition of fuel into two combustion zones. Coal is fired under normal conditions in the primary combustion zone and additional fuel, often gas, is added in a reburn zone, resulting in a fuel rich, oxygen deficient condition that converts the NO_x produced in the primary combustion zone to molecular nitrogen and water. In a burnout zone above the reburn zone, OFA is added to complete combustion.

By recirculating a part of the flue gas to the furnace, the combustion zone turbulence is increased, the temperature is lowered and the oxygen concentration is reduced. All of these factors lead to a reduction of NO_x formation.

Boilers can be operated over a wide range of conditions, and a number of operational changes have been implemented to reduce NO_x production. Two promising technologies for staged combustion are taking burners out-of-service (BOOS) and biased firing (BF). With BOOS, fuel flow is stopped but air flow is maintained in selected burners. BF involves injecting more fuel to some burners while reducing fuel to others. Another operational modification is low excess air (LEA) which involves operating at the lowest excess air while maintaining good combustion. Depending on the type of boiler and the orientation of the burners, operators have viable choices to reduce NO_x production. Advances in boiler control systems enable operators to minimize NO_x and maximize performance.

Postcombustion processes are designed to capture NO_x after it has been produced. In a selective catalytic reduction (SCR) system, ammonia is mixed with flue gas in the presence of a catalyst to transform the NO_x into molecular nitrogen and water. In a selective noncatalytic reduction (SNCR) system, a reducing agent, such as ammonia or urea, is injected into the furnace above the combustion zone where it reacts with the NO_x to form nitrogen gas and water vapor. Existing postcombustion processes are costly and each has drawbacks. SCR relies on expensive catalysts and experiences problems with ammonia adsorption on the fly ash. SNCR systems have not been proven for boilers larger than 300 MW.

COMBINED SO_2/NO_x/PARTICULATE CONTROL TECHNOLOGIES

The CCT program involves a number of projects that achieve reduction of SO_2, NO_x, and particulate emissions in a single processing unit. The technologies described are uniquely combined to achieve project goals and, at the same time, to provide commercial-scale validation of technologies for utilities to consider in order to meet environmental standards.

PARTICULATE CONTROL TECHNOLOGIES

The two major compliance options for particulate control are electrostatic precipitators and fabric filters (baghouses). Dust-laden flue gas enters a precipitator and high voltage electrodes impart a negative charge to the particles entrained in the gas. These negatively charged particles are then attracted to and collected on positively charged plates. The plates are rapped at a preset intensity and at preset intervals, causing the collected material to fall into hoppers. Electrostatic precipitators can remove over 99.9 percent of the particulate matter.

Fabric filters (baghouses) represent a second accepted method for separating particles from a flue gas stream. In a baghouse, the dusty gas flows into and through a number of filter bags, and the particles are retained on the fabric. Different types are available to collect various kinds of dust with high efficiency.

FUTURE INTEGRATED FACILITY

As the global economy expands and worldwide population increases, the demand for additional electric power will grow. The Utility Data Institute (UDI), a Washington, D.C.-based trade organization, estimates that new generating plants totalling 629,000 MW in capacity will be built worldwide by 2003. UDI projects that 317,000 MW, or more than half of this new capacity, will be installed in Asia. Estimates for other regions are: North America, 81,000 MW; European Union, 78,000 MW; Latin America, 55,000 MW; the Middle East, 34,000 MW; Russia and former Soviet Union, 34,000 MW; and other, 30,000 MW. UDI forecasts that fossil fuels will account for 57 percent of the new generating plants, with coal taking 31 percent, gas 19 percent, and oil 7 percent.

The opportunities and threats of the 1990s and for the foreseeable future are related to competition and deregulation. The challenge for utilities will be to produce electric power as cheaply as possible while still complying with environmental regulations.

The Electric Power Research Institute (EPRI), founded by the electric utility industry to manage technology programs, envisions the evolution of a fully integrated facility that produces numerous products in addition to electricity. The first step is to remove mineral impurities from coal. Some of the clean coal could be gasified to provide not only fuel for fuel cells but also products such as elemental sulfur and chemical feedstocks. The remainder of the clean coal can be used in conventional boilers or fluidized bed combustion systems to generate process steam and electricity. The ash resulting from the combustion process can be mined for valuable trace metals before it is used in applications such as road construction. By employing advances from the CCT program, the integrated facility will allow utilities to provide their customers with reliable electrical service and to meet present and future environmental standards.

Charles E. Hickman

See also: Air Pollution; Climatic Effects; Coal, Consumption of; Energy Management Control Systems.

BIBLIOGRAPHY

Air Pollutants. (2000). Vancouver: Greater Vancouver Regional District. <http://www.gvrd.bc.ca/services/air/pollution/pollution.html>.

Clean Coal Technology Compendium. (1999). Los Alamos, NM: Los Alamos National Laboratory. <http://www.laml.gov/projects/cctc>.

Clean Coal Technology Program. (2000). Washington, DC: U.S. Department of Energy, Office of Fossil Energy. <http://apollo.osti.gov/fe/cct.html>.

Electric Power Research Institute. (1987). "Electrostatic Precipitator Guidelines. Volume 1: Design Specifications; Volume 2: Operations and Maintenance; Volume 3: Troubleshooting." *EPRI Journal CS-5198*. Final Report (June).

Electric Power Research Institute. (1999). "Coal Technologies for a New Age." *EPRI Journal* (Summer):6.

Electric Power Research Institute. (1999). "Helping Coal Plants Respond to Mercury Directive." *EPRI Journal* (Summer):6.

Fabric Filters. (1998–2000). Pittsburgh: Wheelabrator Air Pollution Control, Inc. <http://www.wapc.com/ff.htm>.

"Monitoring Effort Initiated Under PM$_{2.5}$ Program." (1998). *Clean Coal Today* (30):8–10.

Nitrogen Oxides Reduction Program: Final Rule for Phase II (Group 1 and Group 2 Boilers). (1996). Washington, DC: U.S. Environmental Protection Agency. <http://www.epa.gov/acidrain/nox/noxfs3.html>.

Overview: The Clean Air Act Amendments of 1990. (1990). Washington, DC: U.S. Environmental Protection Agency Office of Air and Radiation. <http://www.epa.gov/oar/caa/overview.txt>.

U.S. Department of Energy, Clean Coal Technology Topical Reports. (1995). "SO_2 Removal Using Gas Suspension Absorption Technology." Report No. 4 (April). Washington, DC: U.S. Government Printing Office.

U.S. Department of Energy, Clean Coal Technology Topical Reports. (1999). "Advanced Technologies for the Control of Sulfur Dioxide Emissions from Coal-Fired Boilers." Report No. 12 (June). Washington, DC: U.S. Government Printing Office.

U.S. Department of Energy, Clean Coal Technology Topical Reports. (1999). "Reburning Technologies for the Control of Nitrogen Oxides Emissions from Coal-Fired Boilers." Report No. 14 (May). Washington, DC: U.S. Government Printing Office.

U.S. Department of Energy, Clean Coal Technology Topical Reports. (1999). “Technologies for the Combined Control of Sulfur Dioxide and Nitrogen Oxides Emissions from Coal-Fired Boilers.” Report No. 13 (May). Washington, DC: U.S. Government Printing Office.

U.S. Environmental Protection Agency. (1997). “EPA’s Revised Particulate Matter Standards.” Fact Sheet (July). Washington, DC: U.S. Government Printing Office.

Utility Data Institute. (1995). “UDI Sees Electricity Demand Surging with Global Economy.” *Wind Energy Weekly* No. 648 (May 29).

EMISSION CONTROL, VEHICLE

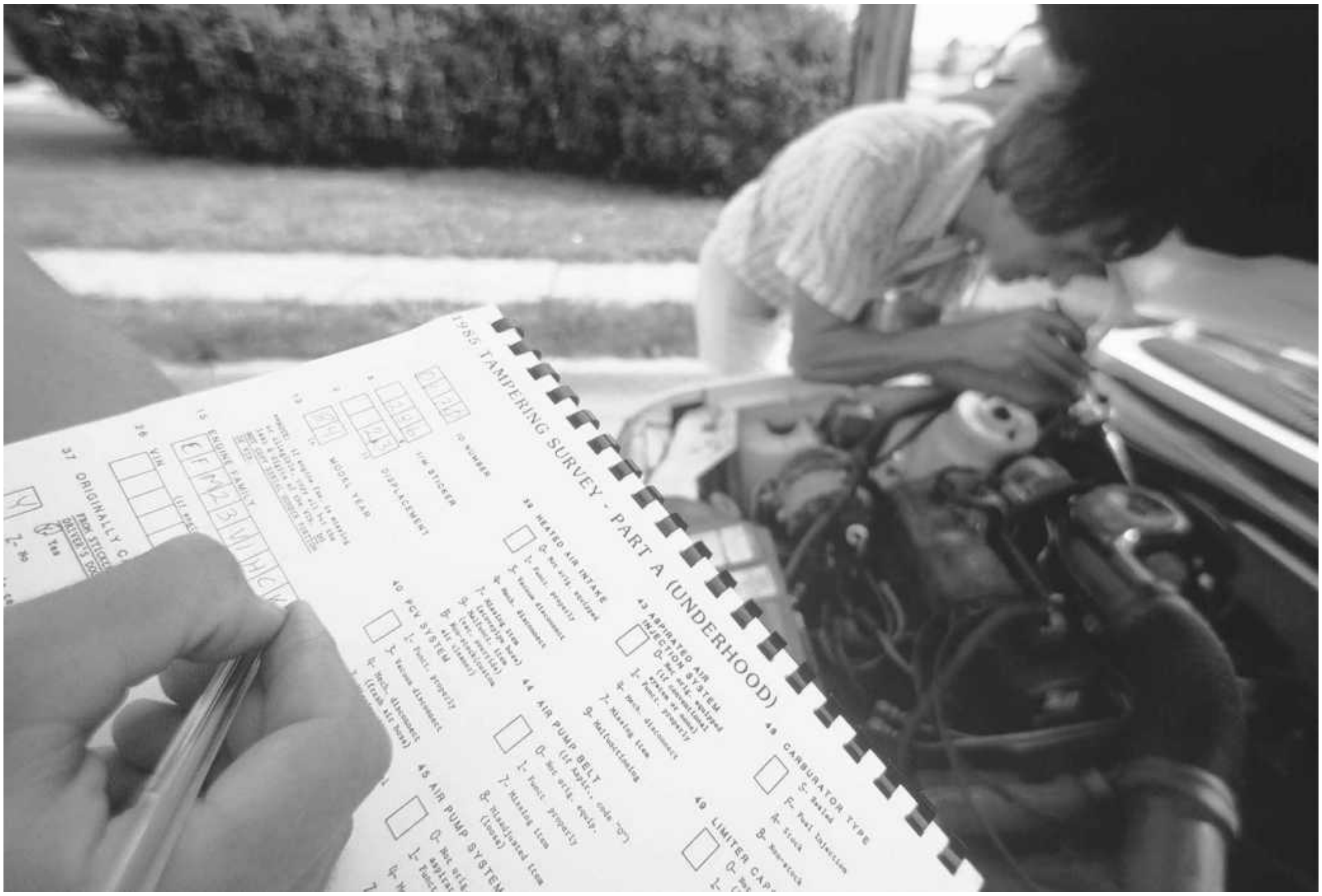
Researchers linked automobile use to air pollution in the early 1950s when A. Haagen-Smit of the California Institute of Technology and fellow researchers began to unravel the complex atmospheric chemistry that leads to the formation of photochemical smog (ozone). Ozone is a strong lung and throat irritant that decreases lung function, increases respiratory problems, and complicates heart disease. Moderate ozone concentrations also damage materials and crops, increasing the cost of living. Ozone forms in the atmosphere when oxides of nitrogen (NO_x) and hydrocarbons (HC) mix and react in the presence of sunlight. Because onroad automobiles, trucks, passenger vans, and sport utility vehicles are typically responsible for about 30 percent of a region’s HC and NO_x emissions, transportation is a significant contributor to smog problems in urban areas.

Onroad transportation sources are also responsible for more than 60 percent of regional carbon monoxide emissions. Carbon monoxide is a colorless, odorless gas that interferes with oxygen transfer in the bloodstream. With a higher affinity than oxygen to bind with red blood cell hemoglobin, with continual exposure, CO gradually displaces oxygen in the bloodstream. Because CO disperses well, it tends to be a hotspot pollutant, with troubling concentrations occuring in areas of high vehicle activity and poor air circulation, such as urban street canyons. When transportation facilities are constructed, engineers model the microscale air quality impacts to ensure that the highway design will not result in unhealthy CO levels downwind from the facility.

Carbon monoxide (CO) and hydrocarbon (HC) emissions arise from incomplete combustion, when petroleum hydrocarbons (gasoline or diesel fuel) do not completely oxidize to carbon dioxide and water. Hydrocarbon emissions also result from gasoline evaporation (liquid leaks, daily heating of fuel and tank vapors, seepage from fuel lines and other components, and displacement of fuel vapors during refueling). A number of toxic air contaminants, such as benzene and 1,3-butadiene are also associated with unburned and partially burned fuel hydrocarbons. Oxides of nitrogen (NO_x) form in the high temperature and pressure environment of the engine cylinder when the elemental nitrogen in air reacts with oxygen. Higher levels of NO_x form at the higher engine temperatures, which unfortunately correspond to peak fuel efficiency.

Given the health impacts that arise from exposure to vehicle emissions and their byproducts, regulatory agencies have focused on motor vehicle emissions control in their efforts to clear the air. Five basic strategies are employed for reducing onroad vehicle emissions: (1) reducing the emissions from new vehicles that displace the older high-emitting vehicles that are scrapped, (2) accelerating vehicle fleet turnover to get new vehicles into the fleet more quickly, (3) reducing emissions from in-use vehicles, (4) reducing travel demand to reduce vehicle activity, and (5) improving traffic flow to reduce emission rates.

The primary focus of federal environmental policy over the past 30 years has been on reducing the emissions from new vehicles. Strategies to enhance vehicle fleet turnover have not prove cost effective over the long term, but such strategies can provide useful short-term emissions reductions. Strategies aimed at limiting in-use vehicle emissions began around 1983 and continue today. Strategies designed to reduce travel demand and improve traffic flow, generically classified together as transportation control measures, have achieved mixed results to date. While demand management measures do work for individual large employers, regional demand management programs have failed to garner public support and have not provided cost-effective emissions reductions. On the other hand, technology-based traffic flow improvement programs, such as traffic signal optimization, can still provide significant emissions reductions at the regional level.



Surveyors from the Environmental Protection Agency inspect the engines and exhaust systems of volunteers' cars for emission control tampering and fuel switching. (Corbis Corporation)

VEHICLE STANDARDS AND EMISSION CONTROLS TECHNOLOGY

In 1963, new cars emitted nearly eleven grams per mile of hydrocarbons (HC), four grams per mile of oxides of nitrogen (NO_x), and eighty-four grams per mile of carbon monoxide (CO). Public pressure to reduce vehicle emissions began to mount in the early 1960s. Manufacturers responded to the general public pressure (and a few specific emissions control regulations) by adding positive crankcase ventilation (PCV) systems to new vehicles. A PCV valve prevents the release of unburned fuel from the crankcase by sending these vapors back to the intake air for combustion. With the passage of the Clean Air Act of 1970, the U.S. Environmental Protection Agency (EPA) began implementing a series of comprehensive regulations limiting the gram-per-mile emissions from new motor vehicles. Manufacturers must produce vehicles that comply with the EPA standards, as measured in the laboratory on EPA's federal test procedure, but manufacturers are free to develop and implement any combinations of control systems they choose. Assembly-line testing and in-use surveillance testing and recall ensure that manufacturers comply with the certification standards.

Emissions standards in 1970 and 1972 were designed to reduce HC and CO emissions by 60 percent and 70 percent, respectively. In response to evaporative emissions standards, manufacturers developed onboard charcoal canisters in the early 1970s to capture gasoline vapors driven from the gas tank (and from the carburetor fuel bowl) by the daily rise and fall of ambient temperature. Exhaust emissions control strategies generally focused on de-tuning the engine to increase exhaust manifold temperature and adding a smog pump that delivered fresh air into the exhaust manifold to oxidize unburned CO and HC.

Manufacturers added exhaust gas recirculation (EGR) systems to counter the increased in-cylinder NO_x formation associated with higher operating temperatures. The EGR recycles a portion of the exhaust stream back into the engine intake air. The relatively inert exhaust gas, containing carbon dioxide and water but little oxygen, serves as a combustion buffer, reducing peak combustion temperatures.

By 1975, new cars were required to meet a 1.5 gram-per-mile HC standard, a 15 gram-per-mile CO standard, and a 3.1 gram-per-mile NO_x emissions standard. In response to the regulatory requirements, the automotive industry added new innovative emissions control technologies. Vehicles came equipped with oxidation catalysts (platinum and palladium on an alumina honeycomb or pellet substrate) designed to convert the CO and partially burned HC in the exhaust stream to CO_2 and water. The catalytic converter allows oxidation to occur at temperatures as low as 300°C, so that oxidation of the exhaust stream can continue downstream of the exhaust manifold. In 1975, lead was eliminated from the gasoline supply for these new vehicles, not because of lead's known harmful health effects, but because lead would foul the new catalytic converters of the new vehicles. A single tank full of leaded gasoline is enough to significantly and permanently reduce the efficiency of the catalytic converter. To reduce contamination of catalysts on new vehicles, the size of the opening to the gasoline tank fill neck was narrowed so that only the nozzles from unleaded gasoline pumps could be inserted into the fill neck during refueling.

Reduction catalysts that convert NO_x back to nitrogen and oxygen under conditions of low oxygen concentration began to appear in the late 1970s. At this time, vehicles began to employ dual-bed catalyst systems. These dual-bed systems employed a reduction catalyst followed by an oxidation catalyst, with fresh air (and thus additional oxygen) injected between the two catalyst beds. Dual-bed systems were capable of controlling NO_x, CO, and HC in a sequential mode.

The emissions reductions provided by the catalytic converter also allowed engineers to re-tune their engine designs and add an improved proportional EGR system. By modifying the EGR that recycles exhaust in proportion to intake air (as a function of engine speed) rather than at a constant rate, emissions could be better controlled over a wider range of operating conditions. The new EGR systems provided the added bonus of improved vehicle performance, balancing some of the efficiency losses associated with the use of catalytic converters.

Probably the most significant control technology breakthrough came in 1977, when Volvo released a computer-controlled, fuel-injected vehicle equipped with a three-way catalyst. The new catalytic converters employed platinum, palladium, and rhodium to simultaneously reduce NO and oxidize CO and HC emissions under carefully controlled oxygen conditions. The new Bosch fuel injection system on the vehicle provided the precise air/fuel control necessary for the new catalyst to perform effectively. The combined fuel control and three-way catalyst system served as the foundation for emissions control on the next generation of vehicles.

By 1981, exhaust emissions standards had tightened to 0.41 grams-per-mile HC, 3.4 grams per mile CO, and 1.0 gram-per-mile NO_x. Manufacturers turned from carburetors, to single-point throttle-body injection, and then to multi point fuel injection systems. With each shift in technology, better control over the air and fuel mixture and combustion was achieved. Better control over the delivery and mixing of air and fuel provided significant emissions and performance benefits. New computer-controlled variable EGR significantly reduced NO_x formation. Smog pumps had also given way to lightweight, inexpensive, pulse air injection systems, significantly improving engine performance. Using the natural pressure variations in the exhaust manifold, fresh air flows in to the manifold through a one-way reed valve and helps to oxidize CO and HC. Finally, many of the new vehicles now came equipped with the improved three-way catalytic converters (TWC) that debuted in 1977.

The new three-way catalytic converters required precise control of fuel/air ratio, so onboard computers became necessary to monitor the combustion process and rapidly adjust the air/fuel mixture through closed loop control. The goal of the computer program is to keep combustion at stoichiometric proportions, where there is just enough air (and therefore oxygen) delivered to completely oxidize the fuel. The stoichiometric ratio for an average fuel is roughly 14.7 kilograms of air per kilogram of fuel. An oxygen sensor in the exhaust manifold monitors the oxygen concentration of the exhaust gases to determine if the combustion mixture contained sufficient oxygen. A reading of zero oxygen in the exhaust gas probably indicates that too much fuel was mixed with the intake air (consuming all of the oxygen before combustion was completed) while a high oxygen concentration in the exhaust gas indicates that too little fuel was mixed with the intake air. The computer processes and evaluates multiple readings each second making minute adjustments to the amount of fuel delivered to the intake air (closing the loop between computer action, sensor reading, and computer response). The computer never achieves a perfect stoichiometric mixture; the air and fuel mix instead alternates between slightly rich and slightly lean. However, the extremely rapid measurement and computer response minimizes emissions formation by responding rapidly to changes in engine operation.

The efficiency of the three-way catalytic converter is also a function of air/fuel ratio. At the stoichiometric air/fuel ratio of 14.7 kilograms of air per kilogram of fuel, the relative air/fuel ratio known as λ equals 1.0. Figure 1 illustrates catalytic converter efficiency for each pollutant as a function of relative air/fuel ratio λ (where a positive λ indicates a lean mixture and a negative λ indicates a rich mixture). The closer the mixture stays to stoichiometric, the more efficient the catalyst at reducing the combined emissions of the three pollutants.

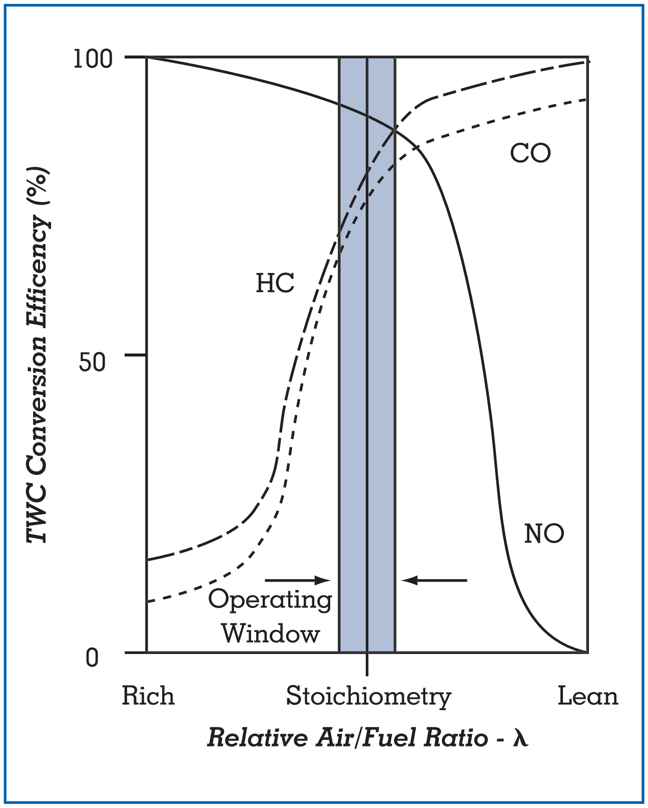


Figure 1.

SOURCE: Chowanietz, 1995.

By 1994, EPA had further tightened the standards to 0.25 grams-per-mile for HC and 0.4 grams per mile for NO_x. Hence, new vehicle HC emissions had now dropped nearly 98 percent and NO_x emissions had dropped 90 percent compared to the level of the 1960s. Manufacturers were also required to ensure that the emissions control systems would endure for at least 100,000 miles. To meet the stringent 1994 standards, manufacturers relied on improved technology and materials, and more advanced computer systems to monitor combustion and rapidly adjust a variety of operating parameters (fuel metering, spark timing, and EGR) to optimize vehicle performance and minimize emissions. Advanced exhaust treatment systems (such as electrically heated or close-coupled catalysts), that manufacturers originally believed in the 1980s would be necessary to comply with these standards, have not been needed.

In 1999, the EPA proposed stringent standards applicable to model year 2004 vehicles. Thus, the EPA continues to implement technology-forcing regulations, in which EPA tasks manufacturers with an emissions standard, and industry must develop technologies to enable the vehicles to comply. When these standards are in place, new vehicles will emit less than 1 percent of the HC and NO_x emissions of their 1960s counterparts (see Figure 2). Advanced computer controls, variable valve timing, and improved catalysts will continue to provide significant reductions. New control systems are also likely to focus on reducing emissions immediately following the engine start. Advances in diesel emissions control technologies may yield viable light-duty diesel vehicles.

Vehicles powered by alternative fuels have yet to make significant inroads into public ownership. Battery technology has not advanced sufficiently to deliver low-cost electric vehicles capable of providing comparable vehicle performance and more than 100 miles between recharging. However, new hybrid electric vehicles that perform on a par with current vehicles and never require recharging began entering the marketplace in 2000. These hybrid electric vehicles are achieving emissions levels as low as 10 percent of 1999 emissions levels, qualifying well below the

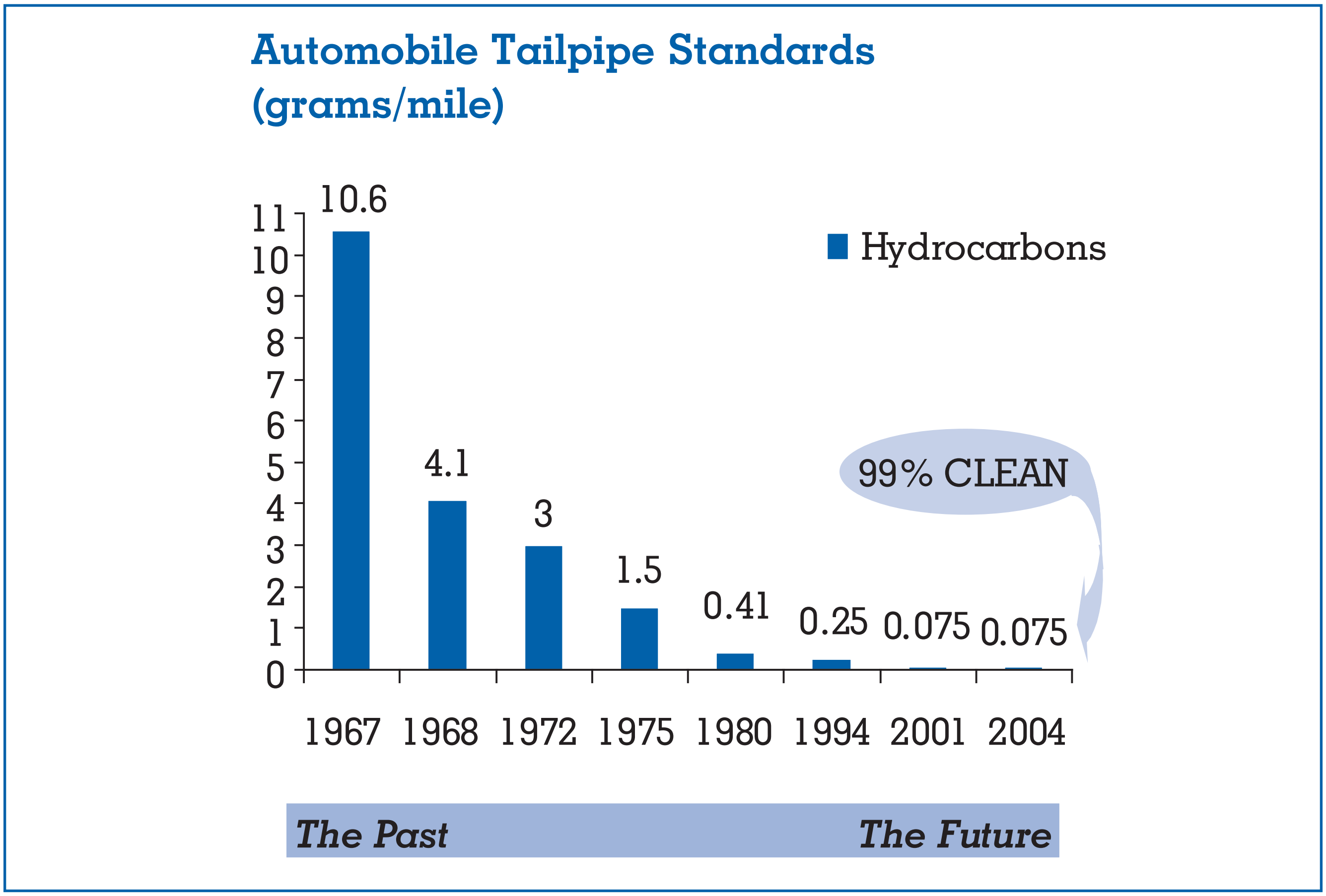


Figure 2.

SOURCE: Alliance of Automobile Manufacturers (2000)

2004 standards. Fuel cell vehicles, which convert the chemical energy of a fuel into electricity, are expected to provide near-zero emissions within the decade.

LIGHT-DUTY TRUCKS AND SPORT UTILITY VEHICLES

Sales of sport utility vehicles and light-duty trucks have topped 45 percent of the new vehicle sales market in the late 1990s, and shares are still climbing. In some months, light-truck and SUV sales exceed those of automobiles. Current emissions regulations are currently less stringent for the vast majority of these light-duty trucks (LDTs) and sport utility vehicles (SUVs) than they are for automobiles. These vehicles are heavier than automobiles, employ larger engines, and the drivetrain (engine, transmission, rear differential, and tire diameter) is designed to handle heavier loads. As such, the emissions from these vehicles are naturally higher. New and in-use LDTs and SUVs exhibit much higher emissions levels than do automobiles (roughly 30 percent more HC and CO, and 85 percent more NO_x per mile traveled). Light-duty trucks also have a significantly longer lifespan in the fleet than do automobiles, compounding the emissions impact over time.

The environmental community has argued, for some time, that the actual onroad duties of the vast majority of LDTs and SUVs are not significantly different from the duties performed by automobiles. That is, most of these vehicles are simply making commute, shopping, and recreation trips that do not require enhanced performance. This issue, combined with the availability of emissions control systems to significantly lower emissions from LDTs and SUVs, led the EPA to harmonize the two standards under the new Tier 2 program. Beginning in 2004, through a phase-in schedule, LDTs and SUVs are required to meet the same emissions standards as automobiles.

Although the lightest of current onroad trucks already meet the same standards as automobiles, the new certification requirement will bring the remaining 80 percent of light-duty truck sales into alignment with the more stringent emissions standards.

HEAVY-DUTY VEHICLES

Onroad and off-road heavy-duty vehicles (greater than 8,500 pounds gross vehicle weight rating) contribute significantly to emissions of NO_x, which in turn participate in ozone formation. As one would expect, heavy-duty engines are large, and the engine load for a vehicle carrying a 60,000-pound payload is extremely high. Most heavy-duty trucks operate on the diesel cycle, an engine cycle that produces much higher temperature and pressure conditions, leading to the formation of significantly greater NO_x levels per mile traveled. Given the stringent controls implemented for light-duty vehicles, it is not surprising that the heavy-duty vehicle contribution as a percentage of regional emissions has been increasing. Projections for the Los Angeles basin in 2010 indicate that without further controls, heavy-duty vehicles will contribute more than 55 percent of onroad NO_x emissions.

According to industry experts, the state of emissions control for heavy-duty engines in the 1990s was at the level of technical advancement that we were achieving for light-duty vehicles in the 1970s. In the last few years, new, highly effective diesel particulate trap and catalyst systems have been developed for heavy-duty diesel vehicles. Many of these new system designs are currently on the road undergoing performance and durability testing. All of the new technologies that are forthcoming were developed in response to new EPA heavy-duty vehicle certification standards that are affective as of 2004. Significant reductions in heavy-duty vehicle emissions are on the horizon.

ENGINE START EMISSIONS

Exhaust emissions are high during the first one to three minutes of engine operation, until combustion stabilizes and the catalytic converter reaches approximately 300°C (known as light-off temperature, when the catalyst begins controlling emissions). Peak catalyst efficiency occurs between 400°C and 800°C. A vehicle that sits more than an hour is usually considered to be starting in a cold-start mode, because the temperature of the catalytic converter has dropped significantly since the vehicle was last used. The average vehicle on the road in 2000 emitted 2 to 4 grams of HC, 1 to 3 grams of NO_x, and 30 to 50 grams of CO for each cold engine start. Vehicles starting in warm-start mode (less than one hour of parking time) produce significantly lower emissions than a cold start, but still contribute significantly to overall trip emissions. Hot starts after 10 minutes or parking time still produce nearly 0.5 gram of HC, 0.5 gram of NO_x, and 20 grams of CO per start for the average vehicle. New emissions models are forthcoming that estimate engine start emissions as a continuous function of park-time distributions.

Because gram/mile emissions from modern vehicles are so low, engine start emissions have become a large fraction of the emissions associated with a vehicle trip. For a typical twenty-mile commute trip in a 1994 vehicle, roughly 30 percent of the CO and 10 percent of the NO_x and HC can be attributed to the cold start. For a ten-mile trip, the overall emissions are about 35 percent lower, but the cold start contributions rise to approximately 50 percent of the CO and 20 percent of the NO_x and HC. On very short trips, total trip emissions are lower still, but the cold start contribution dominates the total. For a half-mile trip, most vehicles never achieve catalyst light-off, and more than 95 percent of the CO and 80 percent of the HC and NO_x trip emissions can be attributed to cold-start operation (as compared to the same trip made by a fully-warmed-up vehicle). It is important to note that a single trip of twenty miles will result in significantly lower emissions than ten trips of two miles each. Trip chaining, where the end of one trip serves as the beginning of the next trip after a short parking period, also results in significantly lower emissions than if each trip results in a cold engine start.

Given the importance of engine start emissions in urban areas, new emissions control systems are likely to focus on achieving instant catalyst light-off. Catalyst manufacturers will increase catalyst surface area and use materials and designs that are resistant to damage from high-temperature exhaust gas. Such designs will allow placement of catalysts closer to the exhaust manifold where higher temperatures will help the catalyst reach light-off much more quickly.

ENRICHMENT EMISSIONS

In recent years, research has demonstrated that real-world vehicle emissions under typical onroad operating conditions can differ significantly from the emissions observed in the laboratory under standard federal test procedures. The occurrence of enrichment, when the air/fuel mixture becomes rich for a few moments, results in orders of magnitude increases in CO and HC emissions rates for short periods. N. Kelly and P. Groblicki (1993) first reported indications that enrichment conditions were likely to be causing a significant portion of vehicle emissions not captured during standard laboratory certification tests. Numerous studies since then have identified enrichment as a widespread concern.

Carbon monoxide emissions rates (grams/second) under enrichment conditions for the very cleanest of vehicles can soar as high as 2,500 times the emissions rate noted for stoichiometric conditions. Although most vehicles spend less than 2 percent of their total driving time in severe enrichment, this can account for up to 40 percent of the total CO emissions (LeBlanc et al., 1995). Hydrocarbon emissions rates can rise by as much as a factor of a hundred under enrichment conditions. Enrichment activity is usually associated with high power demand and engine load conditions, such as high-speed activity, hard accelerations, or moderate accelerations under moderate to high speeds. However, enrichment also occurs during hard deceleration events. When the throttle plate snaps shut during a rapid deceleration event, the rapid decrease in intake manifold pressure vaporizes liquid fuel deposits, causing the fuel mixture to become rich.

All vehicles undergo some enrichment. Fuel enrichment sometimes results from malfunctions of vehicle sensors and control systems. When engine and exhaust gas sensors fail to provide appropriate data to the onboard computer under certain operating conditions, the computer sends inappropriate control commands to fuel injectors and spark advance units. Depending upon the type and extent of component failure, such malfunctions can result in a super-emitter, with significantly elevated emissions rates under all operating conditions. It is interesting to note that engine manufacturers have engineered occurrences of enrichment through the onboard vehicle computer software. Because peak engine torque develops when the air/fuel mixture is slightly rich, manufacturers sometimes use enrichment to improve vehicle performance. Enrichment can increase acceleration rates, improve engine performance while hill-climbing or running accessories such as air conditioning, and can be used to control cylinder detonation. In addition, the cooling properties associated with vaporizing and partially combusting excess fuel lowers peak combustion temperatures, protecting cylinders, valves, and catalysts from high-temperature damage during high RPM activity.

When enrichment episodes occur in the real world, but not in the laboratory under federal certification tests, real-world emissions are significantly higher than predicted. Further complicating emissions prediction is that aggressive driver behavior and complex traffic flow characteristics play a large role in enrichment occurrence. Current vehicle activity simulation models can predict average speeds and traffic volumes very well, but poorly predict the hard-acceleration events that lead to enrichment.

The federal test procedure for new vehicle certification is limited to a maximum acceleration rate of 3.3 mph/second and a maximum speed of 57 mph (and even that speed is for a very short duration). Based upon extensive data collected in Baltimore, Spokane, and Atlanta, more than 8.5 percent of all speeds exceeded 57 mph, and more than 88 percent of trips contained acceleration activity exceeding 4 mph/second. In fact, more than one-third of the trips monitored included an acceleration rate at some point during the trip of more than 7 mph/second. Similarly, more than 15 percent of the deceleration activity exceeded -3.5 mph/second. Hence, enrichment events are significant in real-world emissions inventories.

To counter the elevated emissions associated with enrichment, the EPA has adopted supplemental federal test procedures. The new laboratory test procedures contain higher speeds, higher acceleration and deceleration rates, rapid speed changes, and a test that requires the air conditioning to be in operation. These tests increase the probability that vehicles will go into enrichment under laboratory test conditions. Hence, manufacturers have an incentive to reduce the frequency of enrichment occurrence in the real world. Future catalytic converters and emissions control systems will be resistant to the high-temperature conditions associated with engine load, and will be less likely to require enrichment for protection. Thus, enrichment contributions to emissions will continue to decline.

IN-USE VEHICLE EMISSIONS

New vehicle emissions standards have served as the primary means for reducing vehicle emissions over the last thirty years. However, urban areas must wait for years before the purchase of new vehicles significantly reduces onroad emissions. Meanwhile, daily motor vehicle emissions remain dominated by the small fraction of very high-emitting vehicles. The average car on the road emits three to four times more pollution than new standards allow, and minor control system malfunctions greatly increase emissions. Numerous research studies conclude that a small fraction of onroad vehicles contribute a large fraction of fleet emissions. Some researchers argue that as few as 5 percent of the vehicles are causing 40-50 percent of onroad emissions, but published estimates of super-emitter contribution estimates vary widely. These research studies rely upon laboratory data collected on certification tests, field data collected using portable testing systems, data from remote sensing devices that estimate pollutant concentrations in vehicle tailpipe exhaust plumes, or laboratory or roadside inspection and maintenance data. The controversy surrounding the wide range of super-emitter contribution estimates stems from significant differences in the vehicles sampled, data collected, and the analytical methods and assumptions employed in the various analyses. Although the contribution percentage is uncertain, it is clear that a small fraction of super-emitting onroad vehicles contribute disproportionately to emissions.

To reduce emissions from onroad vehicles, urban areas have turned to inspection and maintenance (I/M) programs. By 1983, sixty-four cities nationwide had established I/M programs, requiring passenger vehicles to undergo a visual inspection and a two-speed idle test to detect severely malfunctioning emissions control systems. Many areas are now adopting advanced I/M programs, which require vehicle testing on a garage treadmill to better identify problem vehicles. Enhanced I/M programs achieve greater emissions reductions than standard I/M programs. However, other states are beginning to restructure and sometimes eliminate statewide inspection and maintenance programs, because the annual fees and testing hassle are not popular with the public. Furthermore, some studies indicate that the emission reduction benefits of I/M programs, while still significant, may be achieving only half of their current modeled emissions reductions. When a state eliminates or scales back an I/M program, the state is responsible for identifying other sources of emissions reductions.

New onboard diagnostics (OBD) systems bridge the gap between new vehicle certification and the in-use compliance verification of I/M. Onboard diagnostics systems detect failures of the engine sensors and the control actuators used by the onboard computer to optimize combustion and minimize emissions. Federal and California OBD programs introduced in 1994 detect component failures (such as an oxygen sensor) by continuously monitoring and evaluating the network of sensor readings to detect erroneous or illogical sensor outputs. Such OBD systems employ detailed computer programs that can change the control logic, discard the inputs from bad sensors, and ensure that emissions remain low even when failures do occur. A malfunction indicator lamp (MIL), or Check Engine light, illuminates on the dashboard when the OBD system identifies problems. Engine computers facilitate repair by reporting trouble codes to mechanics through handheld diagnostic tools that interface with the engine computer. Under I/M programs, vehicles with OBD-reported malfunctions cannot be re-registered until the problem is diagnosed and repaired. The new OBD systems are designed to improve the effectiveness of I/M and minimize lifetime emissions from the vehicle.

Super-emitters behave differently than their normal-emitter counterparts. Whereas normal-emitting vehicles may exhibit high emissions under a hard acceleration or high speeds, vehicles classified as super-emitters tend to exhibit elevated emissions under almost every operating condition. New emissions models will likely track the activity of high-emitting vehicles separately, applying different emission rate algorithms to this activity. Similarly, these new emissions models will also model the effect of I/M programs as decreasing the fraction of onroad high-emitting vehicles.

CLEANER FUELS

Numerous fuel properties affect evaporative and exhaust emissions. Refiners can modify fuel vapor pressure, distillation properties, olefin content, oxygen content, sulfur content, and other factors to reduce emissions. In 1989, the EPA set fuel volatility limits aimed at reducing evaporative emissions. In 1992, manufacturers introduced oxygenated gasoline into cities with high wintertime CO levels. By 1995, the EPA's reformulated gasoline (RFG) program required the sale of special gasoline in nine metropolitan areas that do not meet national clean air standards for ozone. RFG yielded a 15 percent reduction in HC emissions without increasing NO_x emissions, at a cost of somewhere between four and seven cents per gallon. Fuels had to include 2 percent oxygenate by weight (ethanol or MTBE), but manufacturers could adjust a variety of other gasoline properties to achieve the mandated emissions reduction.

Proposed fuel regulations associated with EPA's 2004 vehicle standards program (Tier 2) will substantially reduce the allowable sulfur content of fuel, significantly enhancing the effectiveness of advanced catalytic converters. Sulfur in gasoline temporarily deactivates the catalyst surface, thereby reducing catalyst efficiency. The sulfur reductions are critical for enabling the vehicle emissions control technology to meet Tier 2 standards. By 2006, Tier 2 regulations will require an average fuel sulfur level of 30 ppm, with an 80 ppm cap. This is a substantial decrease from current average sulfur levels of 340 ppm. Vehicle manufacturers estimate that the 90 percent reduction in fuel sulfur will reduce NO_x emissions from the new, low-emitting vehicles by 50 percent, at a marginal cost of between two and four cents per gallon. Vehicle manufacturers argue that further reducing sulfur levels from 30 ppm to 5 ppm will provide additional emissions benefits at a cost somewhere between two and three additional cents per gallon.

FUEL ECONOMY IMPROVEMENTS AND EMISSIONS

In the 1970s, manufacturers requested and received some delays in the implementation of new vehicle certification standards. EPA granted these delays to help manufacturers balance emissions reduction efforts with their efforts to increase corporate average fuel economy. At that time, almost every control system (smog pumps, EGR, and catalytic converters) resulted in a fuel economy penalty. The direct relationship between increased emissions control and decreased fuel economy was broken in the late 1980s with the widespread adoption of advanced computer-controlled fuel injection and spark timing systems. Smog pumps were removed and other devices that reduced fuel economy were improved with computer control. Today, the same technologies that reduced motor vehicle emissions (electronic fuel injection, spark timing, and computer control) have improved fuel economy (or provided improved engine power output in lieu of fuel economy improvements). In general, reduced fuel consumption results in emissions reductions from the vehicle, and reduced vehicle refueling minimizes evaporative emissions. Improving fuel economy is sometimes referred to as the forgotten emissions control strategy.

CONCLUSION

Despite the emissions rate reductions achieved during the last thirty years from new and in-use vehicles, rapid growth in vehicle use has offset a good portion of the total potential reductions. Population growth continues at a rate between 1 percent and 2 percent per year, but the number of trips per day and vehicle miles of travel are increasing at double or triple that rate in many areas. More people are making more trips and driving farther each day. As vehicle miles of travel continue to increase, so do congestion levels. The net effect is that more people are making more trips and driving farther under conditions that increase emission rates. Manufacturers sell nearly 16 million vehicles per year in the United States. More importantly however, the average vehicle lifespan of nearly fourteen years continues to increase. Given the tremendous growth in vehicle use and the emissions rate increases that come with congestion and an aging onroad fleet, reducing onroad vehicle emissions remains extremely important for air quality. Without the previous 30 years of transportation emissions controls, urban air quality would have continued to degrade. Instead, the most polluted areas of the United States have experienced significant air quality improvements.

Randall Guensler

See also: Traffic Flow Management.

BIBLIOGRAPHY

Bosch, R. (1996). *Automotive Handbook, 4th Edition.* Warrendale, PA: Society of Automotive Engineers.

Cadle, S. H.; Gorse, R. A.; and Lawson, D. R. (1993). "Real-World Vehicle Emissions: A Summary of the Third Annual CRC-APRAC On-Road Vehicle Emissions Workshop." *Journal of the Air and Waste Management Association* 43(8): 1084–1090.

Chatterjee, A.; Wholley, T.; Guensler, R.; Hartgen, D.; Margiotta, R.; Miller, T.; Philpot, J.; and Stopher, P. (1997). *Improving Transportation Data for Mobile Source Emissions Estimates* (NCHRP Report 394). Washington, DC: National Research Council, Transportation Research Board.

Chowanietz, E. (1995). *Automobile Electronics.* Warrendale, PA: Society of Automotive Engineers.

Chrysler Corporation. (1998). *Emissions and Fuel Economy Regulations.* Auburn Hills, MI: Chrysler Environmental and Energy Planning, Environmental Team.

Degobert, P. (1995). *Automobiles and Pollution.* Warrendale, PA: Society of Automotive Engineers.

Faiz, A.; Weaver, C.; Walsh, M. (1996). *Air Pollution from Motor Vehicles.* New York: The World Bank.

Guensler, R.; Bachman W.; and Washington, S. (1998). "An Overview of the MEASURE GIS-Based Modal Emissions Model." In *Transportation Planning and Air Quality III,* ed. T. Wholley. New York: American Society of Civil Engineers.

Heck, R. M., and Farrauto. R. J. (1995). *Catalytic Air Pollution Control.* New York: Van Nostrand Reinhold.

Heywood, J. (1988). *Internal Combustion Engine Fundamentals.* New York: McGraw-Hill.

Kelly, N., and Groblicki, P. (1993). "Real-World Emissions from a Modern Production Vehicle Driven in Los Angeles." *Journal of the Air and Waste Management Association* 43(10)1351–1357.

LeBlanc, D.; Saunders, F. M.; and Guensler; R. (1995). "Driving Pattern Variability and Potential Impacts on Vehicle CO Emissions." *Transportation Research Record* 1472:45–52.

National Research Council. (1991). *Rethinking the Ozone Problem in Urban and Regional Air Pollution.* Washington, DC: National Academy Press.

Seinfeld, J. H. (1986). *Atmospheric Chemistry and Physics of Air Pollution.* New York: John Wiley and Sons.

Society of Automotive Engineers. (1998). *Advanced Converter Concepts for Emission Control (SP-1352).* Warrendale, PA: Author.

South Coast Air Quality Management District. (1997). *Air Quality Management Plan, Chapter 2: Air Quality and Health Effects.* Diamond Bar, CA: South Coast AQMD.

St. Denis, M. J.; Cicero-Fernandez, P.; Winer, A. M.; Butler, J. W.; and Jesion, G. (1995). "Effects of In-Use Driving Conditions and Vehicle/Engine Operating Parameters on "Off-Cycle" Events: Comparison with FTP Conditions." *Journal of the Air and Waste Management Association.* 44(1)31–38.

U.S. Department of Transportation, Bureau of Transportation Statistics. (1996). *Transportation Statistics Annual Report 1996: Transportation and the Environment.* Washington DC: U.S. Department of Transportation.

U.S. Environmental Protection Agency. (1995). *Air Quality Trends (EPA-454/F-95-003).* Research Triangle Park, NC: EPA Office of Air Quality Planning and Standards.

U.S. Environmental Protection Agency. (1996). "Final Regulations for Revisions to the Federal Test Procedure for Emissions from Motor Vehicles." *Federal Register.* 61:54852–54906.

U.S. Environmental Protection Agency. (1999). "Fleet Characterization Data for MOBILE6 (EPA420-P-00-011; M6.FLT.007)." Ann Arbor, MI: Office of Mobile Sources.

U.S. Environmental Protection Agency. (2000). "Control of Air Pollution from New Motor Vehicles: Tier 2 Motor Vehicle Emissions Standards and Gasoline Sulfur Control Requirements." *Federal Register* 65:6698–6870.

Wark, K.; Warner, C. F.; and Davis, W. T. (1998). *Air Pollution: Its Origin and Control.* Menlo Park, CA: Addison-Wesley.

ENERGY CONSUMPTION

See: Supply and Demand and Energy Prices

ENERGY ECONOMICS

Energy economics is the application of economics to energy issues. Central concerns in energy economics include the supply and demand for each of the main fuels in widespread use, competition among those fuels, the role of public policy, and environmental impacts. Given its worldwide importance as a fuel and the upheavals in its markets, oil economics is a particularly critical element of energy economics. Other efforts have treated natural gas, coal, and uranium. Energy transforming and distributing industries, notably electric power, also receive great attention. Energy economics addresses, simultaneously as well as separately, both the underlying market forces and public policies affecting the markets.

Economic concerns differ sharply from those of natural scientists and engineers. The most critical difference is in the outlook towards supply development. Many economists argue that market forces allow smooth adjustments to whatever happens to the physical stock of resources. Potentially, these market forces can produce resources cheaper than other methods presently employed to cause adjustments. At worst, the cost rises will be gradual and manageable. In contrast, these economists stress the harmful effects of governments on energy.

A major influence on this optimism in market forces is observation that, historically, technical progress has promoted energy market development. However, the processes by which this progress emerges have not been conducive to formal economic analysis. Rather than wait for a satisfactory model, economists must treat new technology as an unexplained but vital element of energy markets. A related consequence is that established energy sources get almost all the attention. Economics can say little about products that have not emerged.

RESEARCH IN ENERGY ECONOMICS

Scholarly study of the issues in energy economics ranges from massive tomes to short articles. Government and international agencies, consulting firms, industrial corporations, and trade associations also produce a vast body of research in energy economics.

THE PRACTITIONERS

The practitioners of energy economics variously identify themselves as energy economists, mineral economists, natural resource economists, and industrial organization economists. Separate professional societies exist to represent each of three specialties: resource and environmental economics, mineral economics, and energy. These associations do not interact with one another In addition, academic programs exist in each of these areas.

SPECIALTIES AND SUBDISCIPLINES

In each case, quite different bases created the specialties. Energy economics as a separate field emerged with the energy turmoil of the 1970s. Many people were suddenly drawn into dealing with energy issues and felt a strong need for organizations exclusively dealing with their concerns. In contrast, mineral economics emerged in the 1930s from interactions among mineral engineers, geologists, and economists on how the insights of their individual fields could be combined to deal with the problems of all forms of mineral extraction. Mineral economists took a broad view of minerals that gave a prominent role to energy, but many energy economists viewed mineral economists as concerned only with rocks. A more critical problem with mineral economics was that its reach was and remains limited to specialized academic programs and long-established mineral agencies such as those in the U.S. Department of the Interior, and that it was dominantly North American. The energy economists attained greater breadth in the identity and nationality of participants.

Resource economics has at least two bases. Many natural scientists raise widely accepted broad concerns over natural resource availability. A massive federal government study of the problem was instituted during the Truman administration. One result of the effort was the Ford Foundation endowment that established Resources for the Future, a Washington, D.C., research institute, devoted to the study of all natural resource problems. The institute has attracted leading figures in all areas of resource economics, including the environment. Academic programs in resource economics grew mainly from efforts to broaden the scope of agricultural economics, a discipline long practiced principally at the land-grant colleges of major U.S. agricultural states. The scope went on to encompass consideration of environmental problems.

WHY SUBFIELDS EXIST

As is standard in the development of subdisciplines, these fields arose from an intuitive perceived need. Nothing more than the ability to maintain a critical mass of participants adequately justifies the separations. Although the areas deal with different bodies of fact, this distinction is insufficient to justify a field. If it were, we might also have automotive economics or baseball economics. All involve application of economic principles to particular problems. Finding unifying analytic bases for separation is problematic.

Similar arguments, however, can be made about the formal justifications made for even the largest, longest-established branches of economics. The line between pure theory and industrial economics is fuzzy; international trade economics is largely a demonstration that nationality should be economically irrelevant. Given the vast expansion of modern economics, it is necessary to subdivide into convenient sectors. The energy and related realms reflect, on a considerably smaller scale, the value of such segmentation in uniting people with similar interests. Thus, the argument that special analytic issues arise does not justify the existence of special fields; a community of interests is what dominates.

MINERAL DEPLETION AND ITS POLICY IMPLICATIONS

A perennial subject of concern is the long-term availability of energy resources. Natural scientists and even many economists not specializing in natural resources stress the importance of the physical limits to the amount of available energy and to the thermodynamic forces through which initial use precludes economic reuse. Natural resource economists counter that this approach provides an incomplete and misleading vision of the "natural" resource sector. While the physical endowment is limited by definition, its usability is not. Existence does not guarantee economic value. First, a use for the "resources" must be found. Then techniques must be designed and profitably implemented for finding ("exploration"), converting the finds to producing properties ("development"), processing the output to a useable form, and moving the production to customers.

Resource economics stresses that, to date, experience is that the conversion of unutilized resources to profitable ones has moved ahead of demand growth and resource commodity prices have fallen over time. Moreover, M. A. Adelman's classic studies of petroleum supply development have shown that exploration is an ongoing effort to expand the potential for supply expansion. A backlog of developable prospects always exists. When profitable, the much more expensive development stage is undertaken.

Resource pessimists counter that this process cannot proceed forever because the eternal persistence of demand for any given commodity that is destroyed by use must inevitably lead to its depletion. However, the eternal persistence assumption is not necessarily correct. The life of a solar system apparently is long but finite. Energy sources such as nuclear fusion and solar energy in time could replace more limited resources such as oil and natural gas. Already, oil, gas, nuclear power, and coal from better sources have displaced traditional sources of coal in, for example, Britain, Germany, Japan, and France.

Alarms about depletion arose long before massive energy consumption emerged. Experience suggests that the fear is premature. Acting politically to save energy resources may prove more wasteful than allowing consumption. Indeed, the exhaustion problem already is extensively abused to justify undesirable policies. Every local energy producer that reaches the end of its economic life argues that it should be preserved as a hedge against exhaustion (or whatever other evil that can be thought up).

Markets themselves are structured to be very responsive to whatever problems arise with resource supply. Depletion is an economic problem, and markets are capable of reacting to it. Economic limits necessarily are greater than physical ones. That is, much that is physically present will never be economical to employ. Total depletion, if it occurs, will repeat on a large scale what has already occurred for specific suppliers. Costs will become prohibitive, output will start declining, and a steady, recognizable move to extinction will follow.

Elaborate economic theories of exhaustible resources shows that the cost and price pressures associated with such resource depletion provide incentives for slowing depletion. Prudent investors are rewarded doubly. The impending decline in supply pushes up prices and thus the gross income from sales. The restraint slows the depletion of low cost resources, lowers production costs, and thus the net payoff to delayed sales. The theory also indicates that the distribution of payoffs between the two sources can differ greatly. Thus, while we can expect to observe steady upward pressures on energy prices, many different patterns are possible. Price rises will be persistent but not necessarily at a constant rate. The absence of evidence of such price pressures is strong evidence that exhaustion is not an immediate threat. Whatever pattern proves optimal can emerge in the marketplace, and the skepticism among many economists about government foresight are amply proved by prior energy experience. Appraisals of prospects will change with improved knowledge and whatever else permanently alters the situation. Competition among specialized private investors is more adaptable than public policy.

To complicate matters, a government program to assist investment in minerals will not necessarily slow depletion. Investment aid, to be sure, encourages depletion-reducing investment in delaying production. However, the aid also stimulates depletion-increasing investments in production facilities. Which effect predominates depends on the circumstances. Depletion retardation is more likely for producers with presently large excesses of price over cost. Depletion stimulation is more likely when prices are close to costs.

ECONOMIC EFFICIENCY VERSUS SUSTAINABLE DEVELOPMENT

Since the 1980s, the persistence of concerns over both the maintenance of natural resource commodity supply and environmental quality has become restated as a search for sustainable development. The concern is that unregulated markets will produce a pattern of natural resource commodity production that unduly favors the present and near-term future over the longer-term, and generates too much pollution.

Some resource economists fervently support the concept of sustainability. Others argue that the principle is less coherent, comprehensible, and compelling than prior concepts, particularly the core economics principle of efficiency. For economists, the choice of terminology is secondary. The primary concern is resolving the underlying problems of possible market inefficiencies and the ability of governments to cure them.

Richard L. Gordon

BIBLIOGRAPHY

Adelman, M. A. (1995). *The Genie out of the Bottle: World Oil since 1970.* Cambridge, MA: MIT Press.

Barnett, H. J., and Morse, C. (1963). *Scarcity and Growth: The Economics of Natural Resource Availability.* Baltimore: Johns Hopkins Press for Resources for the Future.

Brennan, T. J.; Palmer, K. L.; Koop, R. J.; Krupnick, A. J.; Stagliano, V.; and Burtraw, D. (1996). *A Shock to the System: Restructuring America's Electricity Industry.* Washington, DC: Resources for the Future.

Gilbert, R. J., and Kahn, E. P., eds. (1996). *International Comparisons of Electricity Regulation.* Cambridge, Eng.: Cambridge University Press.

Simon, J. L. (1996). *The Ultimate Resource 2.* Princeton, NJ: Princeton University Press.

ENERGY INTENSITY TRENDS

Energy intensity is defined as the ratio of energy used to some measure of demand for energy services. There is no one measure of energy intensity. The measure depends on the universe of interest being measured. If the universe of interest is an entire country, then the measure needs to reflect the energy used and the demand for energy services for that country. A country's demand for energy services is usually measured as the dollar value of all goods or services produced during a given time period, and is referred to as Gross Domestic Product (GDP). If the universe of interest is households, energy intensity might be measured as energy used per household or size of the housing unit. If the universe of interest is the iron and steel industry, the measure for energy intensity could be energy used per ton of raw iron produced or energy per dollar value of the raw iron.

Energy-intensity measures are often used to measure energy efficiency and its change over time. However, energy-intensity measures are at best a rough substitute for energy efficiency. Energy intensity may mask structural and behavioral changes that do not represent "true" efficiency improvements. A shift away from producing products that use energy-intensive processes to products using less-intensive processes is one example of a structural change that might be masked in an energy-intensity measure. It is impossible to equate one energy-intensity measure to some "pure" energy efficiency. Therefore a set of energy-intensity measures should be developed—keeping in mind the caveats underlying the measures.

ENERGY-INTENSITY TRENDS IN THE UNITED STATES, 1972 TO 1986

Before the 1970s the United States experienced a time of falling energy prices and ample supplies of petroleum. In 1973, crude petroleum prices shot up by 400 percent. In the early 1980s, the growth of economic activity outpaced the demand for energy. Between 1972 and 1986, energy consumption per dollar of GDP declined at an average annual rate of 2.1 percent. Energy per dollar of GDP is a useful measure. It is important, however, to understand the factors that lie behind any changes in the measure.

At first, short-term behavioral changes, such as lowered thermostats and reduced driving, were common in reducing energy demand, but their overall effects on demand were small. While these transient changes were taking place, other, more fundamental changes were working their way into energy-using processes. Examples were the introduction of automobile fuel economy standards, appliance efficiency standards, and the movement away from energy-intensive processes in the manufacturing sector.

In 1975, Congress responded to the oil crisis of 1973 by passing the Energy Policy and Conservation Act. This legislation established Corporate Average Fuel Economy (CAFE) standards. By 1985, CAFE standards required that all new passenger cars had to have an

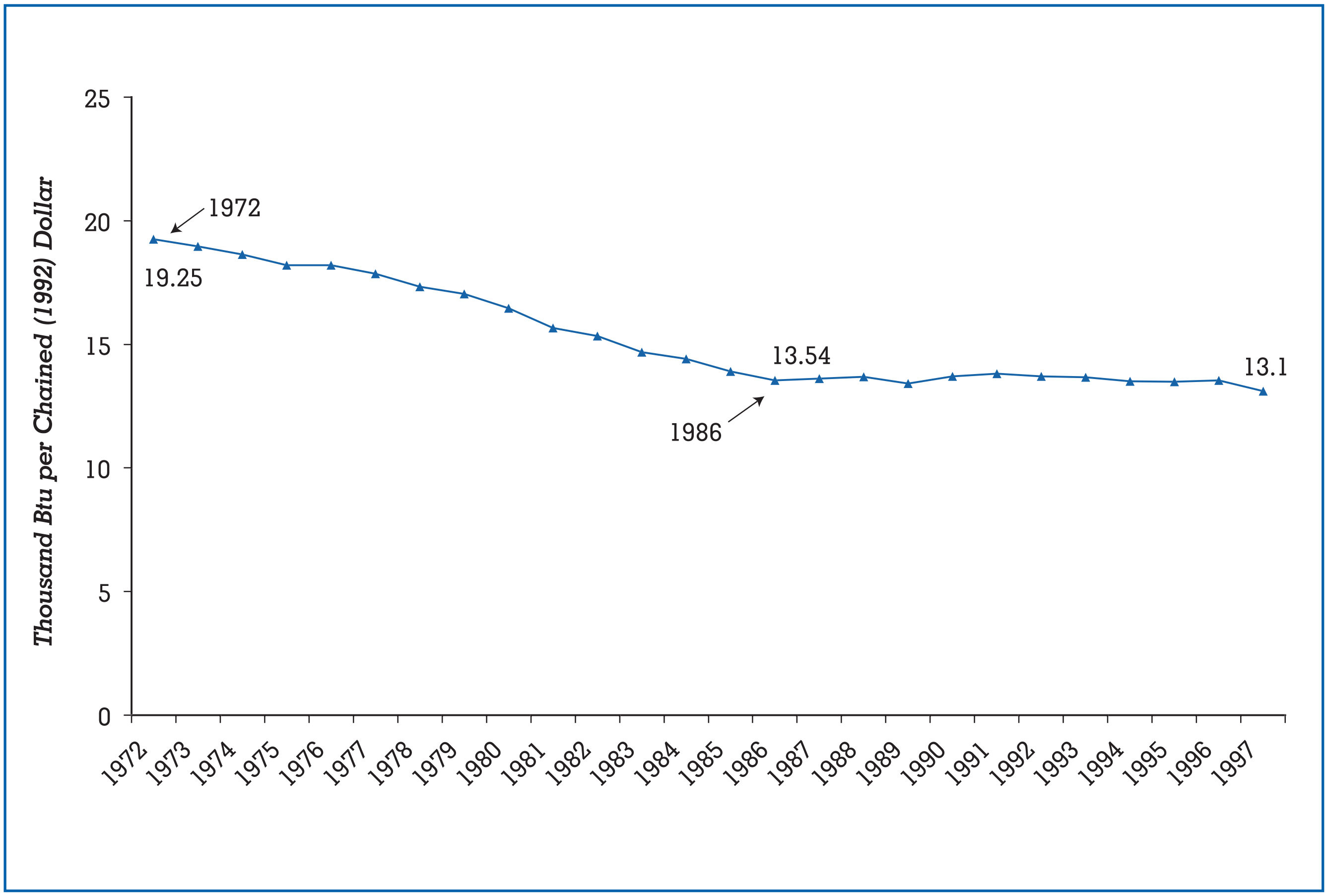


Figure 1.
U.S. Energy Consumption per Dollar of GDP, 1972–1997
source: Energy Information Administration, Annual Energy Review 1997.

average gas mileage of 27.5 miles per gallon (mpg) of gas, nearly double the mileage of typical cars made in the mid 1970s. Light trucks had to average 20.6 mpg.

Additionally, as a response to rising energy prices and uncertainty of supply, several states adopted appliance efficiency standards. At the federal level, the National Appliance Energy Conservation Act of 1987 established the first national standards for refrigerators and freezers, furnaces, air conditioners, and other appliances. The Energy Policy Act of 1992 added national standards for incandescent and fluorescent lights, small electric motors, office equipment, and plumbing products.

The movement away from energy-intensive processes in the manufacturing sector was an important force in the reductions in energy intensity during these years. One of the most noticeable shifts was in the primary metal industry. In 1976, 128 million tons of raw iron were produced by 1986 production had fallen to 81.6 million tons. Domestic primary production of aluminum was 3.7 million metric tons in 1972, but fell to 3 million metric tons in 1986.

ENERGY-INTENSITY TRENDS IN THE UNITED STATES, 1986 TO 1997

After 1986, the CAFE and appliance standards in place resulted in stock turnovers to more efficient automobiles and appliances. However, the decline in energy consumption per dollar of GDP slowed appreciably and, between 1986 and 1997, the energy intensity trend remained rather flat. Other forces in the U.S. economy were pushing energy consumption higher, resulting in increases in the energy-intensity measure.

Between 1987 and 1997, twelve million households were added to the country's housing stock, representing a 13 percent increase. Since most of this increase took place in the West and the South, the demand for electricity for central air conditioning increased as well. In 1987, 52 percent of all households in the South had central air conditioning. In 1997, this percentage was 70 percent—a 35 percent increase. Additionally, new households have been getting larger, resulting in an increased demand for heating, air conditioning, lighting, and appliances. New energy-consuming devices such as VCRs, microwaves ovens, and home computers were also purchased on a wide scale.

The energy intensity measure, miles per gallon, for the stock of passenger cars increased from 17.4 in 1986 to 21.3 in 1996. However, average miles driven per car per year increased from 9,464 to 11,314. There has also been a change in the mix of new vehicles purchased. In 1980, 38 percent of all new vehicles were subcompacts, falling to 18 percent in 1997. The market share for small vans was less than 1 percent in 1980, but grew to 19 percent in 1997. Small utility vehicles had a market share of 3.4 percent in 1980, growing to 26 percent in 1997. Although the larger vehicles became more efficient, more were being purchased and were being driven more.

Although the manufacturing sector continued its decline in the production of energy-intensive products, between 1986 and 1997 the service sector continued to grow. Not only did the commercial building stock increase, but the use of office equipment—from computers to copy machines—has grown rapidly. In just three years, between 1992 and 1995, the number of personal computers and computer terminals in commercial buildings increased from 29.8 million to 43 million (45%).

Clearly the United States is doing more with less energy. Although total energy has grown as demand for goods and services has climbed, energy use per person has hardly changed, 348 million Btu per person in 1972 and 352 million Btu in 1997. During this same time period, energy per GDP has declined 32 percent. State and federal energy-efficiency standards, consumer behavior, and structural shifts, have fueled this decline.

Stephanie J. Battles

See also: Appliances.

BIBLIOGRAPHY

The Aluminum Association. (1997). *Aluminum Statistical Review for 1996.* Washington, DC: Department of Economics and Statistics.

American Iron and Steel Institute (1998). *Annual Statistical Report 1988.* Washington, DC: Office of Energy Markets and End Use.

Energy Information Administration. (1995*). Measuring Energy Efficiency in the United States' Economy: A Beginning*, DOE/EIA-0555 (95)/2. Washington, DC: Office of Energy Markets and End Use.

Energy Information Administration. (1998). *Annual Energy Review 1997*, DOE/EIA-0384 (97). Washington, DC: Office of Energy Markets and End Use.

Heal G., and Chichilnisky, G. (1991). *Oil and the International Economy.* Oxford, England: Clarendon Press.

International Energy Agency. (1997). *Indicators of Energy Use and Efficiency.* Paris, France: Organization for Economic Cooperation and Development.

U.S. Bureau of the Census. (1997). *Statistical Abstract of the United States,* 117th ed. Washington, DC: U.S. Government Printing Office.

U.S. Department of Energy. (1998). *Transportation Energy Data Book: Edition 18.* ORNL-6941. Oak Ridge, TN. Oak Ridge National Laboratory

ENERGY MANAGEMENT CONTROL SYSTEMS

EVOLUTION PERSPECTIVE OF EMCS

The primary purpose of energy management control systems (EMCS) is to provide healthy and safe operating conditions for building occupants, while minimizing the energy and operating costs of the given building. Aided by technological developments in the areas of electronics, digital computers, and advanced communications, EMCS have been developed to improve indoor quality while saving more energy.

Electromechanical Timers

The earlier devices used in the first half of the twentieth century to control building loads (such as lighting and space conditioning) were electromechanical timers, in which a small motor coupled to a gearbox was able to switch electrical contacts according to a predefined time schedule. Normally the output shaft of the gearbox causes one or more pairs of electrical contacts to open or close as it rotates. These electromechanical devices were simple and reliable, and they are still used to control lights and ventilation in some buildings. However, for lighting and space conditioning, the main drawbacks of this type of timer was inflexibility. Manual intervention is required to change settings, and the operation mode is essentially without feedback (open-loop controllers), since the schedule of operation is not easily influenced by the variables in the controlled process.

Electronic Analog Controllers

The development of electronic circuitry capable of processing sensor signals made possible the appearance of electronic controllers able to respond to variable conditions. For example, to control street lighting a light sensor coupled to a simple electronic amplifier and switch can turn off the lights during the day, avoiding energy waste. In some cases the analog electronic controllers were coupled with electromechanical timers, providing the possibility of controlling the output as a function of time and/or operation conditions.

Although electronic circuitry has been available since the beginning of the twentieth century, the invention of the transistor in 1948 was a key milestone, bringing a decrease in costs, an improvement in reliability, and a reduction in the size of control circuits. During the 1950s and 1960s, electronic analog controllers became widely available for controlling lighting, heating, ventilation, and air conditioning (HVAC). Although these controllers made possible improved control based on information from sensor feedback (closed-loop control), they still suffer from lack of flexibility, since they are able to implement only simple control strategies and require manual change of settings.

Digital Controllers

The invention of the microprocessor in 1970 signaled a radical change in the area of building controls, allowing the development of increasingly powerful EMCS. From the outset, microprocessor-based EMCS could be easily programmed for changeable and variable time schedules (e.g., workday versus weekend operation, public holiday operation). In addition to this flexibility, increasingly powerful microprocessors, along with new energy management hardware and software, allowed for the progressive implementation of sophisticated control algorithms, optimization of building operation, and integration of more functions into building control and supervision.

Most new hardware and software development in the EMCS market now aims at utilizing the full potential of EMCSs and at making the information obtained more usable and accessible. This accessibility has created a secondary benefit with a greater potential for customer/utility communication. In many ways the EMCS evolution and market penetration in the industrial, commercial, and to a certain extent, residential markets, has equaled that of the personal computer.

During the 1970s EMCS digital controllers were mostly electronic time clocks that could be programmed according to a variable schedule. However, increasingly powerful, low-cost microcomputers have rapidly improved computing power and programmability and hence increased the application of EMCS. This rapid improvement has in turn introduced benefits such as the following:

- flexible software control, allowing simple modification;
- built-in energy management algorithms, such as optimal start of space conditioning, limiting the electricity peak demand; also, control of HVAC and lighting loads in practically all EMCS applications;
- integrated process, security, and fire functions;
- monitoring capability allowing identification of faulty equipment and analysis of energy performance;
- communication with the operator and other EMCS that can be remotely located.

These possibilities make EMCS increasingly attractive for energy monitoring, control, and utility/customer interface. The integration of functions associated with safety and security (e.g., intrusion and fire) as well as those associated with building diagnostics and maintenance are also expanding the potential market of EMCS. The application of advanced EMCS with these features is leading to the appearance of “smart buildings,” which can adjust to a wide range of environmental conditions and which offer comfort and security with minimal use of energy.

CURRENT EMCS CHARACTERISTICS

An EMCS is usually a network of microprocessor-based programmable controllers with varying levels of

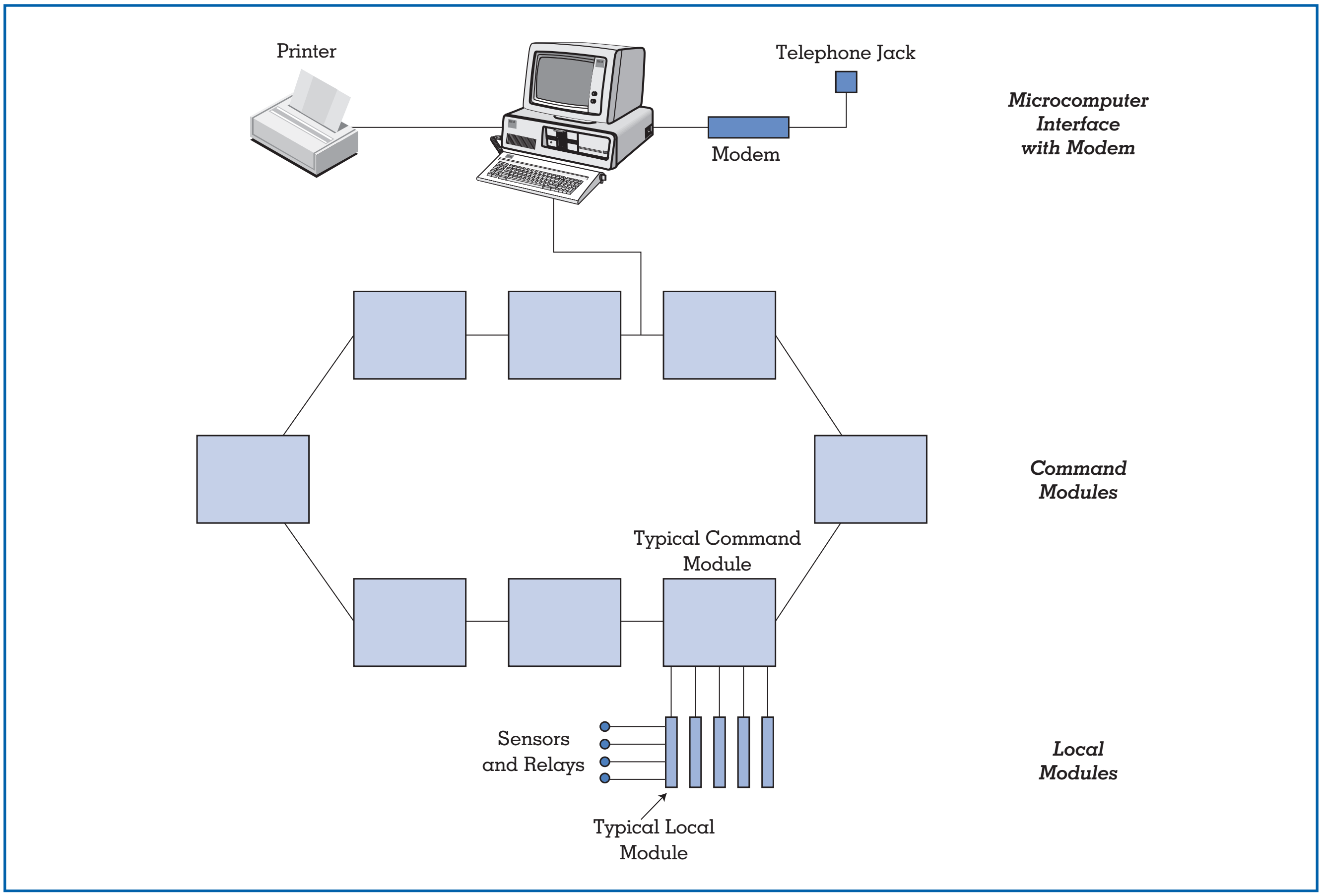


Figure 1.
Schematic of a typical EMCS with local control modules, command control modules, and personal computer interface.

intelligence and networking capabilities. This "typical" EMCS has three components (see Figure 1):

- Local control module: directly wired to sensors and actuators (e.g., equipment to control space conditioning, such as pumps and fans);
- Command control module: makes control decisions;
- Personal computer interface: simplifies operator control of the system through a user-friendly interface.

An important feature of the typical system is its modularity. The most powerful EMCS installations have all three components, but often only a single module is necessary for simple applications, such as controlling a single air-conditioning unit, or for most applications in the residential sector. Thus local and control modules are capable of stand-alone operation without higher-level components. The functions, hardware, software, and communication characteristics of these three components are as follows.

Local Control Modules

A local control module performs the following basic functions in an EMCS:

- receives information from sensors;
- controls actuators, through relays, to switch equipment on and off or to change its variable output;
- converts analog sensor data to a digital form;
- performs direct digital control;
- communicates with the command module.

Modern EMCS use a variety of sensors, including temperature, humidity, occupancy, light, pressure, air flow, indoor air quality, and electric power (normally pulses from power meters). The actuators are the units that can influence the state of the system, including chillers, pumps, fans, valves, and dampers.

Command Control Module

The command control module, or control module, is the real intelligence of the EMCS; all programming and control software resides here. Command control module software can create reports (e.g., for recording the historical status of variables) and perform various demand-limiting schemes. The common energy management strategies offered at this level include proportional-integral-derivative (PID) control loops, duty cycling, and optimal start/stop of HVAC units. Also available are economizer control (use of free cooling with outside air when the outside temperature is below the target temperature) and programmed start/stop with demand limiting of selected loads.

The command module can be programmed either through a keypad or by downloading programs from the PC host. The information presented in command module reports is easy to change, but it must follow a prearranged format. These reports usually show temperatures, peak demand, whole-building energy, equipment status, maintenance records, and alarm records. Automatic reporting of a system alarm, such as machine failure, by phone or by electronic mail is another typical software feature.

Command modules communicate with other modules through a local area network (LAN). Through this LAN, command modules receive information from the local control modules and store data. These data can be stored from a week to two years, depending on the recording interval and the number of points to be monitored. Unlike host-based systems, which use a central computer to interrogate each command module individually, the computer interface can tap into the network like any other command module.

A recent data communications protocol for Building Automation and Control Networks (BACnet), ASHRAE Standard 135-1995, is an important step to ensure that controllers made by different manufacturers can communicate with each other in a simple way, avoiding the expense of additional interface hardware and communication software.

Personal Computer Interface

The personal computer interface allows for easy operation of an EMCS, but all of the system's control functions can be performed in its absence. This user interface serves three purposes:

- storage of a backup of the command module's programs to be used in case of power or system failure;
- archiving of trend data for extended periods of time;
- simplification of programming and operation of the system through a user-friendly interface.

The software at the user-interface level usually involves block programming, whereby the operator can define control loops for different sensors as well as other specific control strategies in a BASIC-type, easily understood programming format. This block structure allows modification or enabling of different program blocks, avoiding the necessity of rewriting the main program. Programs at this level are menu-driven, with separate sections for trend reports, programming, graphics, etc. Visual displays, such as the schematic of HVAC equipment interconnections, are used to make information easier to comprehend. Building floor plans can be incorporated to display information such as room temperatures.

INTEGRATION OF FUNCTIONS

Because of increasing computation power, EMCSs can integrate other functions and also be coupled to other processes. By helping coordinate plant operation, the integration of functions also makes the acquisition of EMCSs more attractive, since a significant portion of the hardware can be shared by different applications. For tracking the performance of the building operation and carrying out the required maintenance, monitoring and data-logging functions are essential in an integrated system, since operators need to be aware of the operating status of each part of the plant and to have access to reports of previous performance. Also, alarm display and analysis is a very convenient function for carrying out diagnostics that can be performed by an integrated system. Safety and security functions can be easily and economically integrated, although some installations prefer separate systems due to potential litigation problems.

The use of peak electricity demand-reducing technologies, such as thermal storage and peak-shaving involving the use of a standby generator, also can benefit the control and monitoring capabilities of EMCSs. In thermal storage (ice or chilled water for cooling), the EMCS can schedule the charging of the system during off-peak hours to optimize savings in both peak demand and energy costs. Temperatures in the storage system are monitored to minimize energy and demand costs without adversely affecting plant operations or products. Thermal storage, already an attractive option for space cooling in commercial and industrial buildings and in several food industries (dairy, processed meat, fish) and for space cooling, has become an even more effective energy saver thanks to EMCSs. An EMCS also can determine when the use of an existing standby generator during periods of peak demand can reduce costs, taking into account the load profile, demand and energy costs, hour of the day, fuel costs, etc. The standby generator also can be put on-line on request from the utility in times of severe peak demand or loss of generating capacity reserve margins.

EMCS FOR ENERGY SAVINGS

Typical savings in energy and peak demand are in the range of 10 to 15 percent. These savings are achieved by reducing waste (e.g., switching off or reducing the lights and space conditioning in nonoccupied spaces), by optimizing the operation of the lighting (e.g., dimming the lights and integration with natural lighting), and space conditioning (e.g., control of thermal storage).

THE EMCS FOR REAL-TIME PRICING

State-of-the-art systems with flexible software allow for utility interface. These systems are all currently capable of data monitoring and of responding to real-time pricing, depending only on the software installed on the user-interface computer and the number of sensors the customer has installed for end-use monitoring.

To reduce use of electricity when electricity costs are highest, EMCS use a network of sensors to obtain real-time data on building-operating and environmental conditions. Some large electricity consumers are connected with the utility through a phone line for the communication of requests to reduce the peak electricity load and for present and forecasted demand. The building operator traditionally closes the link, instructing the EMCS to respond to the utility's signals. However, this current manual load shedding/shifting response to utility prices is too labor-intensive and operationally inefficient for large-scale implementation.

Energy management systems have been developed that can control loads automatically in response to real-time prices. Real-time prices are sent to the customer, whose EMCS can modulate some of the loads (e.g., air conditioning, ventilation, nonessential lighting). Thus peak demand can be reduced at times of high electricity price, maintaining all essential services. An EMCS could be used to modulate HVAC load by controlling temperature, humidity, volatile organic compounds (VOCs), and carbon dioxide levels within a window of acceptance, the limits of which may be adjusted as a function of the real-time prices. In theory this strategy can save energy and substantially decrease peak demand. The most attractive candidates for ventilation control performed by EMCSs are: large commercial buildings with long thermal time constants (or with thermal storage), buildings with low pollutant emission from building materials, buildings that house furnishings and consumer products, and buildings that require a large volume of ventilation (or circulation) air per occupant.

THE EMCS FOR DATA MONITORING

Parallel to the increase in performance of microcomputers in the late twentieth century been an exponential decrease in the price of semiconductors and memory. This means that data-logging functions can be added at a small extra cost, considerably increasing the usefulness of a given EMCS.

Because monitoring and data-logging facilities add to initial cost, some customers may be hesitant to choose between a simple configuration without those capabilities and a more complex model including them. The benefits associated with monitoring often outweigh the price premium, as they allow the user to:

- tune the performance of a system and check energy savings;
- record the load profile of the different operations or major pieces of equipment within the plant;
- find the potential for improvements;
- allocate energy charges to each operation or product line;
- submeter electricity end use, allowing more accurate charging of costs to specific processes or divisions;
- check worn or faulty equipment (e.g., declining fan pressure in an air-handling unit may mean a clogged filter);
- check utility meter readings (potentially a particularly useful feature in plants using power electronics con-

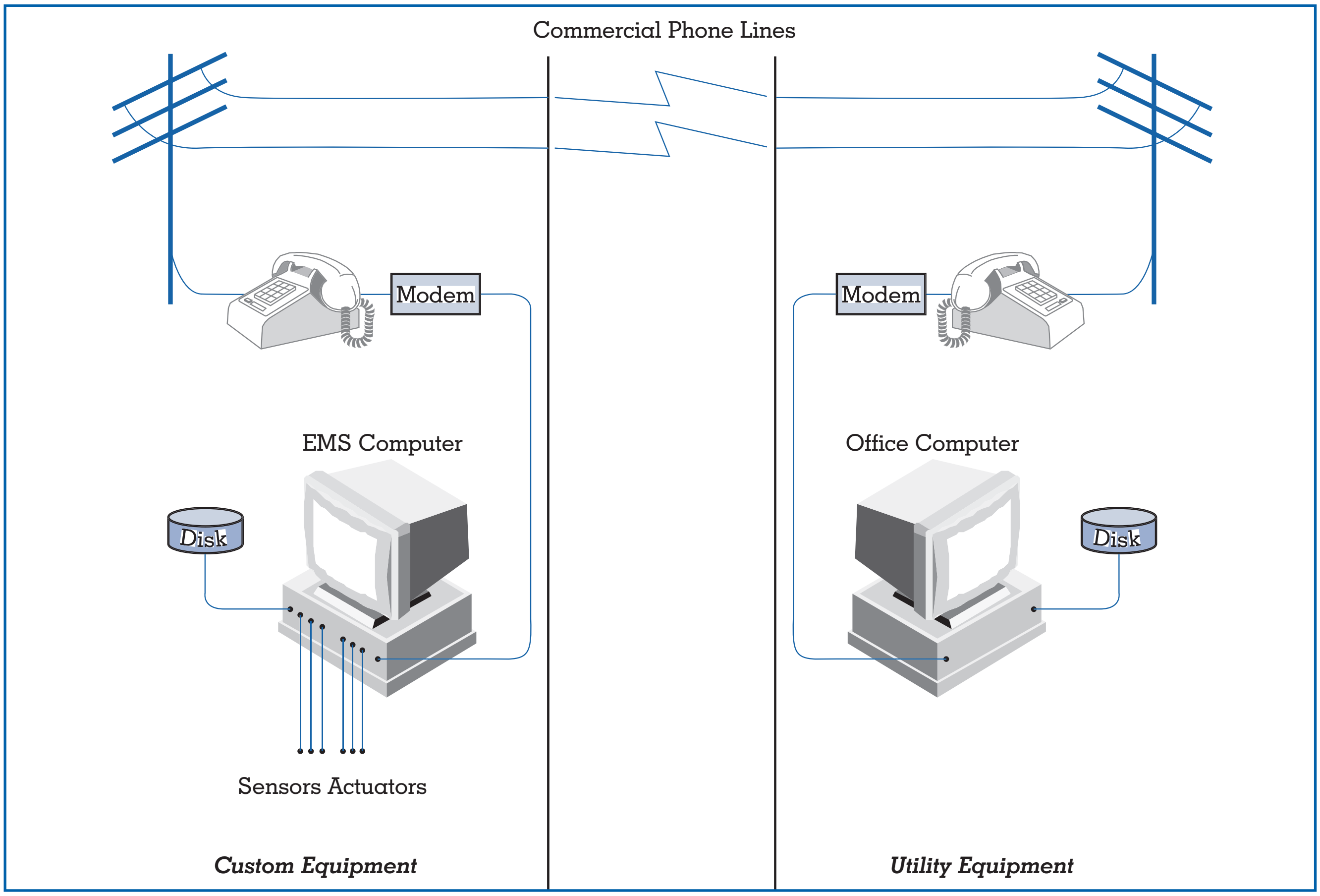


Figure 2.
EMCS-based monitoring of building and end uses. A utility computer can interface the customer EMCS and obtain building and energy-use information.

trol devices, which can generate errors in conventional power meters).

Figure 2 shows a diagram of the layout of EMCS-based end-use monitoring in which the data collected by the EMCS can be remotely monitored by the utility or by the maintenance staff in the company's central office.

The load data can be disaggregated into main uses by using suitable algorithms that take into consideration sensor information, plant equipment, and plant schedule. Figure 3 shows hourly electricity use of an office building, broken down by major end uses. This type of data analysis can be useful for monitoring building performance and for identifying opportunities to save energy at peak demand.

TRENDS IN THE EVOLUTION OF ENERGY MANAGEMENT AND CONTROL SYSTEMS

Future trends for the evolution of EMCS are likely to involve improvements in user interface, easier access, better controls, and advances in integration, namely including the following:

- better access to system information, with more remote diagnosis and maintenance capabilities;
- easier installation and programming, with advanced graphics user interface;
- smaller, more distributed controllers and unitary control for more advanced energy management optimization;
- easier integration of products from different manufacturers, and continuing effort toward communication standardization;

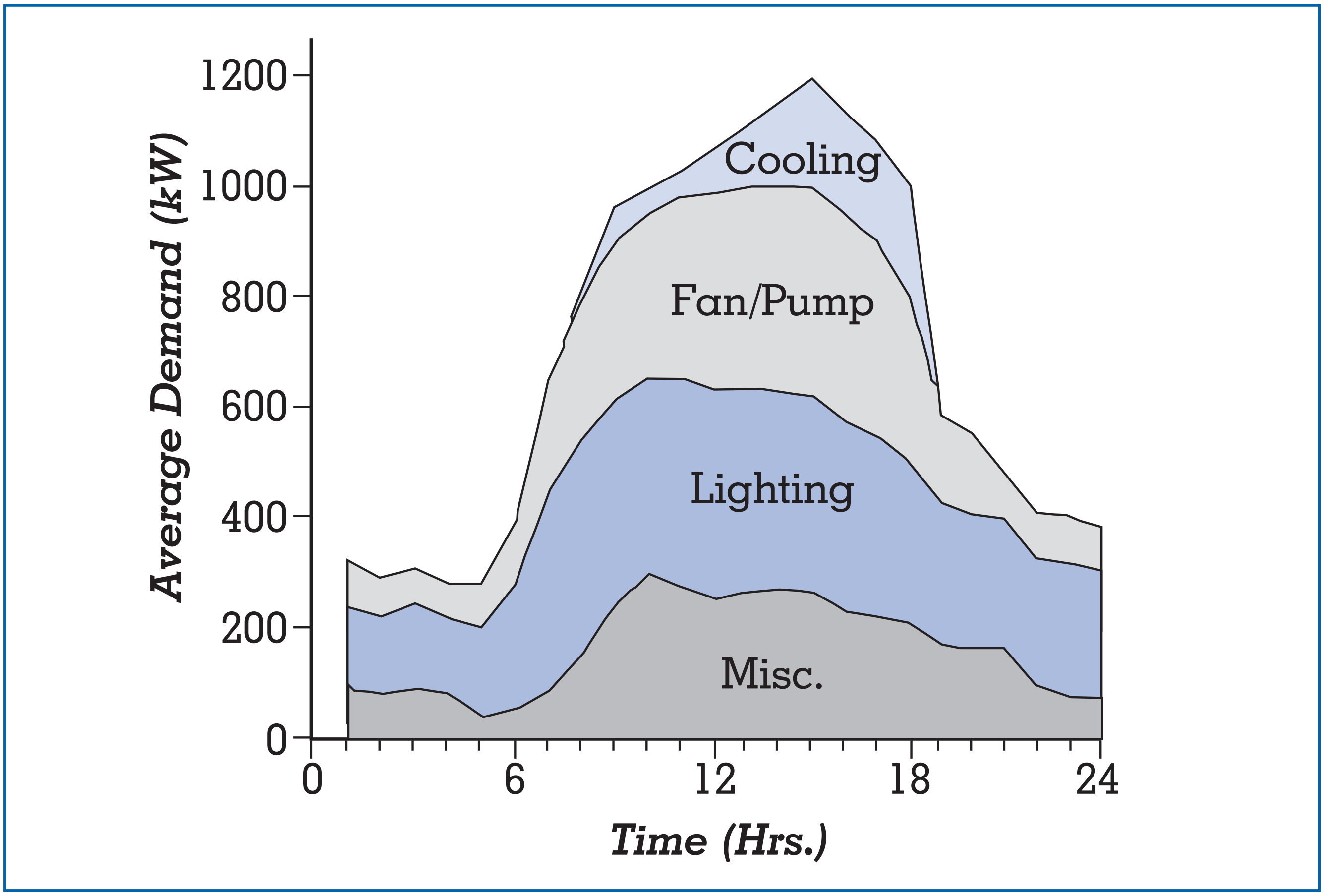


Figure 3.
An example of disaggregated hourly end-use data obtained from whole-building load.

- less integration of fire and security functions.

One of the most significant EMCS trends is toward better access to the information gathered by the EMCS. The EMCS industry is developing smaller, distributed units with increased programmability and control capabilities. These smarter local units offer increased system versatility and reliability and allow smaller units to perform functions that would previously have required larger, more expensive systems. Another trend in EMCS hardware development has been the integration of sensors and controllers from various manufacturers. The work performed toward the development of a standard communications protocol (BACnet) is of the greatest importance in ensuring communication compatibility between and among equipment made by various manufacturers.

One system feature of particular relevance is remote troubleshooting. For example, in case of an alarm signal in the building, the user can trace back through the system to find the cause of the alarm. In fact, manufacturer representatives can perform much of the routine troubleshooting over the phone. Whenever an operator has a software problem, the representatives can call up the system to correct programming problems or help develop new applications. Both preventive maintenance (carried out at programmed intervals of operation time) and predictive maintenance management (carried out when the plant sensors detect a deterioration in the equipment performance) can be incorporated in powerful EMCS, including databases containing details (even images) of the spare parts required for maintenance.

The creation of graphics can be menu-driven, often utilizing a building floor plan or system schematic to display the collected data. The floor plan is first drawn by the customer, and then variables, such as current room temperature, are superimposed.

Optical scanners also can allow easier graphics creation, and user-selected video frames can be incorporated into the software displays.

Voice communication capabilities, including voice recognition and speech synthesis, are also being increasingly used to provide a simpler user interface. Thus, for example, verbal instructions can be given for resetting set points (temperature, etc.), or to request other actions from the system. Fire and security monitoring is one EMCS feature, which may require its own dedicated system, although this leads to higher costs. This trend is conditioned by insurance and liability issues.

Anibal T. de Almeida
Hashem Akbari

See also: Efficiency of Energy Use; Electric Power, System Protection, Control, and Monitoring of; Energy Economics; Industry and Business, Productivity and Energy Efficiency in; Risk Assessment and Management.

BIBLIOGRAPHY

Akbari, H.; de Almeida, A. T.; Connell, D.; Harris, J.; and Warren, M. (1988). "Use of Energy Management Systems for Performance Monitoring of Industrial Load-Shaping Measures." *Energy, The International Journal* 13(3):253–263.

Akbari, H.; Heinemeier, K.; Le Coniac, P.; and Flora, D. (1988). "An Algorithm to Disaggregate Commercial Whole-Building Electric Hourly Load into End Uses." *Proceeding of the ACEEE 1988 Summer Study on Energy Efficiency in Buildings* 10:13–26.

American Society of Heating, Refrigeration, Air-Conditioning Engineers (ASHRAE). (1995). *ASHRAE Standard 135-1995: BACnet—Data Communication Protocol for Building Automation and Control Networks.* Atlanta: Author.

Electric Power Research Institute (EPRI). (1985). *Control Strategies for Load Management.* Palo Alto, CA: Author.

Electric Power Research Institute (EPRI). (1986). *Energy Management Systems for Commercial Buildings.* Palo Alto, CA: Author.

Fisk, B., and de Almeida, A. T. (1998). "Sensor-Based Demand Controlled Ventilation." *Energy and Buildings* 29(1):35–44.

Heinemeier, K., and Akbari, H. (1992). "Evaluation of the Use of Energy Management and Control Systems for Remote Building Monitoring Performance Monitoring." *Proceedings of the ASME International Solar Energy Conference.*

ENGINES

An engine is a machine that converts energy into force and motion. Possible sources of energy include heat, chemical energy in a fuel, nuclear energy, and solar radiation. The force and motion usually take the form of output torque delivered to a rotating shaft. (Torque is the twisting effort developed around a center of rotation. In an engine it is conceptually quantified by the product of the radius from the center of the output shaft to a second point at which a tangential force is applied, and the magnitude of that applied tangential force.) In contrast, the output of a jet or a rocket engine is simply the thrust force derived from its high-velocity exhaust jet.

Most often the input energy to an engine is derived through combustion of a fuel. The result is a combustion engine. Combustion engines can be classified according to the nature of their combustion and its initiation. Possible combinations are depicted in Figure 1. First, combustion may occur either continuously or intermittently. Second, that combustion may occur either external to the engine or internally, within the engine.

In the continuous external-combustion engine, a fuel is burned outside the confines of the machine responsible for the conversion of energy into useful work. The heat energy generated through combustion is then transferred into a working medium that undergoes a repetitive cycle of thermodynamic processes. The portion of this engine that converts the heat energy into work is appropriately termed a heat engine because the input heat energy need not necessarily come from combustion. The second law of thermodynamics states that not all of the heat energy transferred into the working medium of the heat-engine cycle can be converted into output work. The first law of thermodynamics states that the difference between the heat transferred into the engine cycle and the work produced as a result of that cycle must be rejected to the surroundings.

The steam engine used by utilities to generate electricity is an example of a continuous external-combustion engine. A fuel—usually coal, oil, or natural gas—is burned to generate heat. That heat is transferred into the working medium of the engine cycle—namely, water—through a heat exchanger known as a boiler or a steam generator. The engine cycle produces work in the form of torque on a rotating shaft, usually by means of a steam turbine. The difference between the heat added to the cycle in the steam generator and the net work produced by the turbine is rejected to the environment in a condenser that returns the steam to water at its original pressure and temperature for a repetition of the cycle. Other examples of the continuous external-combustion engine include the Stirling engine and the closed-cycle gas turbine. In these engines the working medium remains in the gaseous phase throughout the cycle and is often hydrogen, helium, or some gas other than air.

Because the steam engine described is a heat engine of the external-combustion type, the cycle experienced by the working medium can be executed without combustion. In some steam engines, for example, the required input heat is supplied by a nuclear reactor. Stirling engines have been operated on radiant energy supplied by the sun.

The dominant continuous internal-combustion engine is the gas turbine, which is used in both torque-producing and thrust-producing applications. It involves continuous compression in an aerodynamic compressor; continuous combustion in a burner that in principle resembles a household oil burner; and continuous work extraction in a turbine that both drives the compressor and, in torque-producing configurations of the engine, delivers engine output. In applications where fuel economy is of great importance, a heat exchanger may be added that transfers heat from the turbine exhaust gas to the burner inlet air to preheat the combustion air. The torque-producing gas turbine is used for electric power generation, ship propulsion, in military tanks, and as an aircraft turboprop (propjet) engine. In on-road automotive applications, the gas turbine has never progressed beyond the demonstration stage because it has not been commercially competitive with existing automotive piston engines.

In thrust-producing gas turbines, the turbine extracts only enough energy to drive the compressor and engine accessories. The remaining available energy is converted to a high-velocity exhaust jet that provides a forward thrust. The aircraft turbojet and its cousin the turbofan (fanjet) embody this concept. The turbofan differs from the turbojet in that the turbine also drives a low-pressure-rise compressor, whose airflow bypasses the burner and the turbine and joins the turbine exhaust in the jetstream. This results in the production of thrust by a jet that has a higher flow rate but a lower velocity than the turbojet. As a result, the turbofan offers higher propulsion efficiency at a lower aircraft speed than the turbojet. The turbojet and turbofan have driven the large piston engine once used in passenger airliners and military aircraft into obscurity.

Despite the existence in Figure 1 of an intermittent external-combustion engine as a possibility, no such engine is known to have been placed in service. However, the intermittent internal-combustion engine is another matter. It is applied to everything from chain saws and lawn mowers to large trucks, locomotives, and huge oceangoing vessels. The intermittent internal-combustion engine can be further subdivided as to the method used for initiating combustion and the nature of the air-fuel charge that is prepared for that ignition.

The most common internal-combustion engine in use is the homogeneous-charge spark-ignition (HCSI) engine. The ubiquitous reciprocating-piston gasoline engine serves as an example. In it, fuel and air are premixed upstream of the cylinder or cylinders with the objective of supplying to the engine a homogeneous mixture of the two. Combustion is initiated at the appropriate time in the engine cylinder by discharging an electric spark in the entrapped air-fuel mixture. The hot products of combustion expand, generating useful work in the form of force on the moving piston. That force is transformed into torque on a rotating shaft by the engine mechanism. The spent combustion products are exhausted to the environment as heat energy and replaced by a fresh air-fuel charge for repetition of the engine cycle. The internal-combustion engine operates on a repetitive mechanical cycle that consecutively repeats the events comprising it. However, it does not follow a true thermodynamic cycle, as is done in the heat engine, because the working medium is never returned to its original state. Instead, the working medium is exhausted from the engine as products of combustion, to be replaced by a fresh charge.

Although this description of the reciprocating-piston engine fits the vast majority of gasoline engines in service, the same cycle of events can be executed in a variety of other kinematic arrangements. One such alternative approach is the rotary engine, which avoids the oscillatory force production of the recipro-

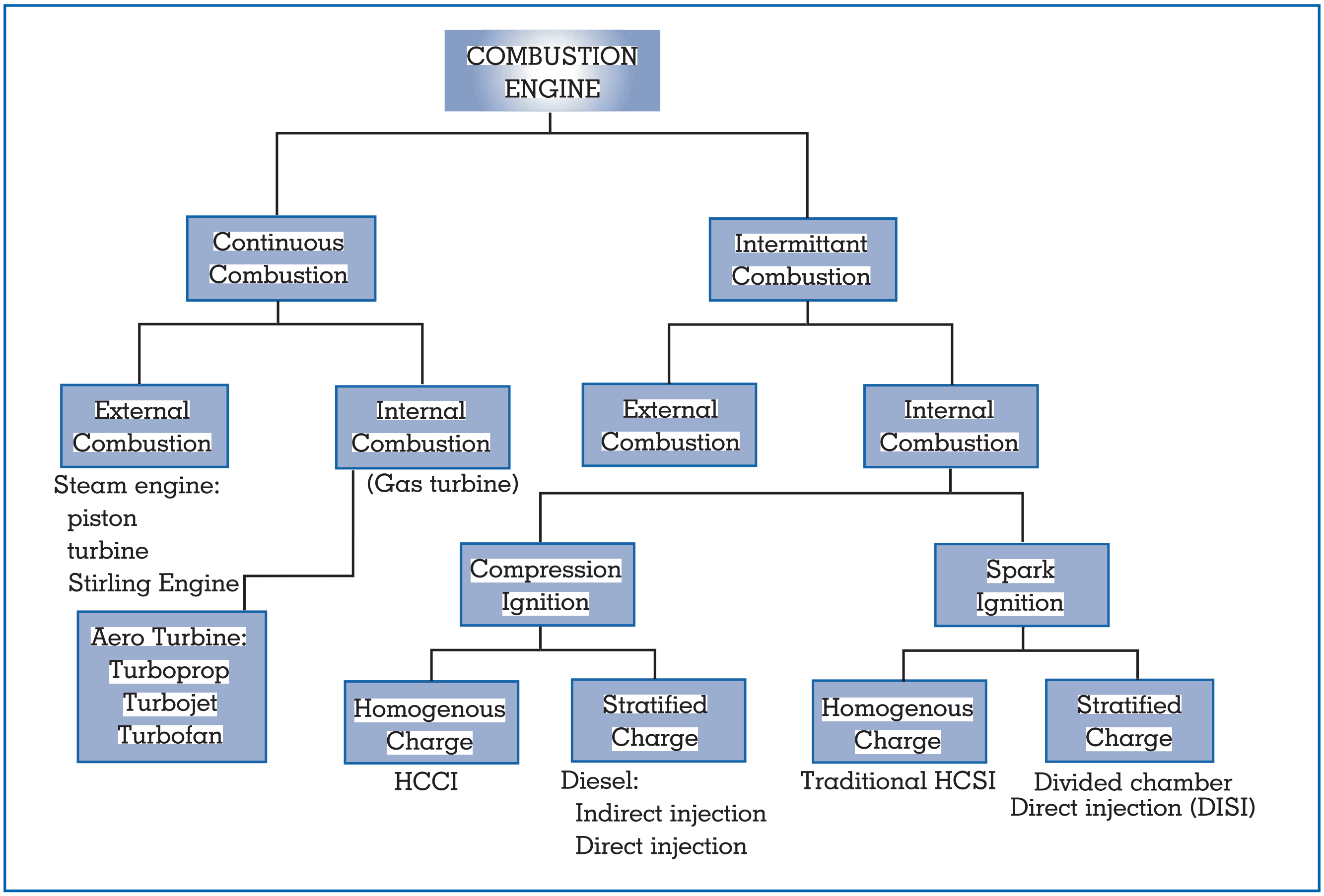


Figure 1.
Combustion engine tree.

cating-piston. The best known rotary engine is the Wankel engine. Just like the reciprocating-piston engine, it consecutively effects an intake of fresh mixture, combustion of that mixture, expansion of the resulting hot gasses to produce output work, and expulsion of the hot exhaust products to prepare for repetition of the cycle. The varying cylinder volume of the reciprocating-piston configuration is mimicked in three chambers, each enclosed by one flank of a three-sided rotor and the inner wall of a specially shaped stationary housing within which that rotor turns.

In the stratified-charge spark-ignition engine, homogeneous mixing of the air and fuel is avoided in favor of creating a mixture in the region surrounding the spark plug that is richer in fuel than the cylinder average. One approach to charge stratification once experiencing limited application in automobiles was the divided-chamber engine, in which a richer-than-average mixture was inducted into a prechamber connected to the main chamber above the piston by a restrictive passageway. A spark plug in the prechamber ignited the segregated rich mixture, which then sent a jet of burning gas into the main chamber to serve as a powerful ignition source for a leaner-than-average mixture inducted into that chamber.

The developing trend is to eliminate the prechamber and stratify the charge by injecting the fuel directly into the cylinder. This approach, known by several different names but adequately described as the DISI (Direct-Injection-Stratified-Charge) engine, involves careful control of both injection timing and in-cylinder air motion. Charge stratification of this nature facilitates using a cylinder-average air-fuel ratio that is higher (leaner) than could normally be expected to burn satisfactorily if the air and fuel were mixed homogeneously. Burning such an overall-lean mixture has the potential for better fuel economy than the traditional homogeneous-charge engine provides.

The compression-ignition stratified-charge engine is commonly known as the diesel engine. It has also been implemented both with and without a divided chamber. In the divided-chamber, or indirect-injection (IDI) version, fuel is injected only into the prechamber, where combustion begins. The direct-injection (DI) configuration has no prechamber, with fuel being injected directly into the volume above the piston. In either case, the fresh charge of air is compressed to a sufficiently high pressure and temperature that the injected fuel autoignites. Such autoignition sites are generally found on the periphery of the injector spray plume and eliminate the need for a spark plug.

On the century-old time scale of the internal-combustion engine, the homogeneous-charge compression-ignition (HCCI) engine is a comparative newcomer. It involves preparation of a homogeneous charge upstream of the engine cylinder or cylinders. That charge is then autoignited in the cylinder by compression, as in the diesel, with combustion beginning at many distributed ignition sites essentially concurrently. If the mixture is too fuel-rich, combustion occurs too quickly and generates excessive noise. If the mixture is too fuel-lean, combustion becomes incomplete and erratic. Typically, exhaust gas is recirculated into the inlet mixture to help control combustion. Compression ignition of the homogeneous charge cannot be used over the complete operating range of the engine. Currently a popular subject of research study, the homogeneous-charge compression-ignition concept has seen very limited commercial application to date.

Charles A. Amann

See also: Aircraft; Combustion; Diesel Cycle Engines; Gasoline Engines; Spacecraft Energy Systems; Steam Engines; Stirling Engines; Thermodynamics; Turbines, Gas; Turbines, Steam.

BIBLIOGRAPHY

Amann, C. A. (1980). "Why not a New Engine?" *SAE Transactions* 89:4561–4593.

Heywood, J. B. (1988). *Internal Combustion Engine Fundamentals.* New York: McGraw-Hill.

ENTHALPY

See: Thermodynamics

ENTROPY

See: Thermodynamics

ENVIRONMENTAL ECONOMICS

In dealing with environmental questions, economists emphasize efficiency, social welfare, and the need for cost accountability. A basic principle for efficiency is that all costs be borne by the entity who generates them in production or consumption. For example, production and consumption of diesel fuel will be socially inefficient if significant resulting costs are shifted to others who happen to be downwind or downstream from the refinery that makes the fuel or the truck that burns it. The benefits of making and using the fuel should exceed the cost—society at large—or else the process reduces total social welfare.

Information is the key to such internalizing "external costs." If the generator of pollution damage is known, along with the victim and the size of the damage, then the polluter can be held accountable. Historically, if wrongful damage is done, courts in the United Kingdom, the United States, Canada, and nations with similar legal systems have been willing to force compensation by polluters or, when the damage is great enough, to order cessation of the pollution. Small damage is ignored. But if no one knows whether the damage is serious, or who caused it, then regulation may be instituted to cope with it. However, since the regulator may not know more than courts could learn at trial, the results of regulation vary from increasing efficiency to reducing it. An analysis of the problems facing a regulator provides context for examining the key issues in environmental economics.

Cost accountability will be most prevalent when property rights to land and resources are clearly defined. Clear property rights make the owners of a resource responsible, in most countries, for the way that resource is used and for the harms it may cause others. An owner's right to the use of property does not include the right to use it in ways that impose a cost on others. In such cases, courts have historically held owners of a polluting plant or business responsible for harm they may cause other parties. Clear property rights make owners face the cost of inefficient use of a resource and thus encourage owners to ensure that their property or equipment is put to the most highly valued use. Property rights provide what economists consider the *incentives* to ensure that resources are used efficiently (maximizing net value) and in a way that constrains negative impacts on other individuals.

When ownership rights are less well defined, or not easily defended, the incentives for resource owners to efficiently use resources in a safe, non–polluting way, are decreased or removed. Individuals are less likely to take expensive, time-consuming action to protect a resource that they do not own, and by which they are not directly affected financially. For instance, most landowners would be quick to take action to prevent garbage generated by a local business from piling up in their own backyard. However, they would be less likely to take actions to prevent the same business from polluting a nearby lake or river. The reason is that any one individual has less direct, or at least less obvious, interest in the lake or river, than they do in their own property—and usually less ability to affect the outcome.

This incentive problem is increased as the number of polluters and the number of land owners increase so that it is difficult to pinpoint specific incidents of pollution and their effect on individuals. The case of air pollution from cars and multiple factories is a classic example of this information problem. In a large metropolitan area, there are millions of automobiles and many factories that could contribute to air pollution. There are also millions of individuals who could be harmed by that pollution. But it is generally difficult, if not impossible, for one individual to identify a specific problem they have experienced due to air pollution and then to pinpoint the source of that problem. Such situations, where property rights are not well defined, as is the case with air, and where it is difficult to identify a particular source of pollution, often lead to calls for government regulations to prevent a certain activity, or to reduce a certain activity such as exhaust from automobiles, in an effort to prevent harm to others.

Just as in the decision-making process of individual land owners, incentives are important in the government decision-making process. Economists have identified a characteristic of the government decision-making process that can allow the concerns of special interest groups can take precedence over the interest of the general public. The principle of rational voter ignorance states that since the cost of obtaining information about political issues is high, and any individual voter is likely to pay a small portion of the cost as well as reap a small portion of the benefit of any government action, individual voters are not likely to take the time to become well informed on specific issues. In contrast, politically organized special interest groups, such as firms in a polluting industry, will pay a heavy price for any new regulations that might be directed toward them. Therefore, they have a financial incentive not only to be well informed on the issues affecting them, but also to spend time and money trying to influence the government to ensure that they do not bear the cost of regulations.

With strong incentives for businesses and other special interest groups, and weak incentives for individual voters, it is not surprising that many environmental regulations have often been less successful at preventing harm than the more traditional property rights-based approaches. Thus, while privately owned lands and resources are generally healthy and well preserved, many resources that are not owned, such as air or many waterways, are polluted.

Many economists have therefore become disappointed in the effectiveness of traditional regulatory solutions to environmental problems. They look to market incentives such as those provided by private property rights, and market-like mechanism, where polluters must bid for or trade for the right to release potentially harmful emissions, as policy alternatives.

Richard L. Stroup

See also: Acid Rain; Air Pollution; Atmosphere.

BIBLIOGRAPHY

Brubaker, E. (1995). *Property Rights in Defence of Nature.* Toronto: Earthscan Publications.

Field, B. C. (1997). *Environmental Economics: An Introduction*. Boston: Irwin McGraw-Hill.

Wills, I. (1997). *Economics and the Environment: A Signalling and Incentives Approach*. St. Leonards, NSW, Aust.: Allen & Unwin.

ENVIRONMENTAL PROBLEMS AND ENERGY USE

For millions of years humans existed in harmony with nature. But the Industrial Revolution, and the exploration and development of energy to fuel that revolution, began a period of ever-growing fossil fuel combustion that resulted in greater water pollution, air pollution, deforestation, and growing atmospheric carbon dioxide concentrations. Nuclear energy, which was supposed to be the solution to these problems associated with fossil fuel production and combustion, turned out to present an equally great threat to the environment in terms of safety and waste disposal. Because almost no energy source is totally benign, and all the major energy sources have unwanted drawbacks, billions of dollars are being spent each year on scientific research to find ways to lessen the environmental impact of the sources in use, to make the more benign sources more cost-competitive, to improve the energy efficiency of technology, and to determine if the high energy-consuming habits of humans are putting the planet in peril by irreversibly and harmfully altering the atmosphere.

THE ENERGY-ENVIRONMENT LINK

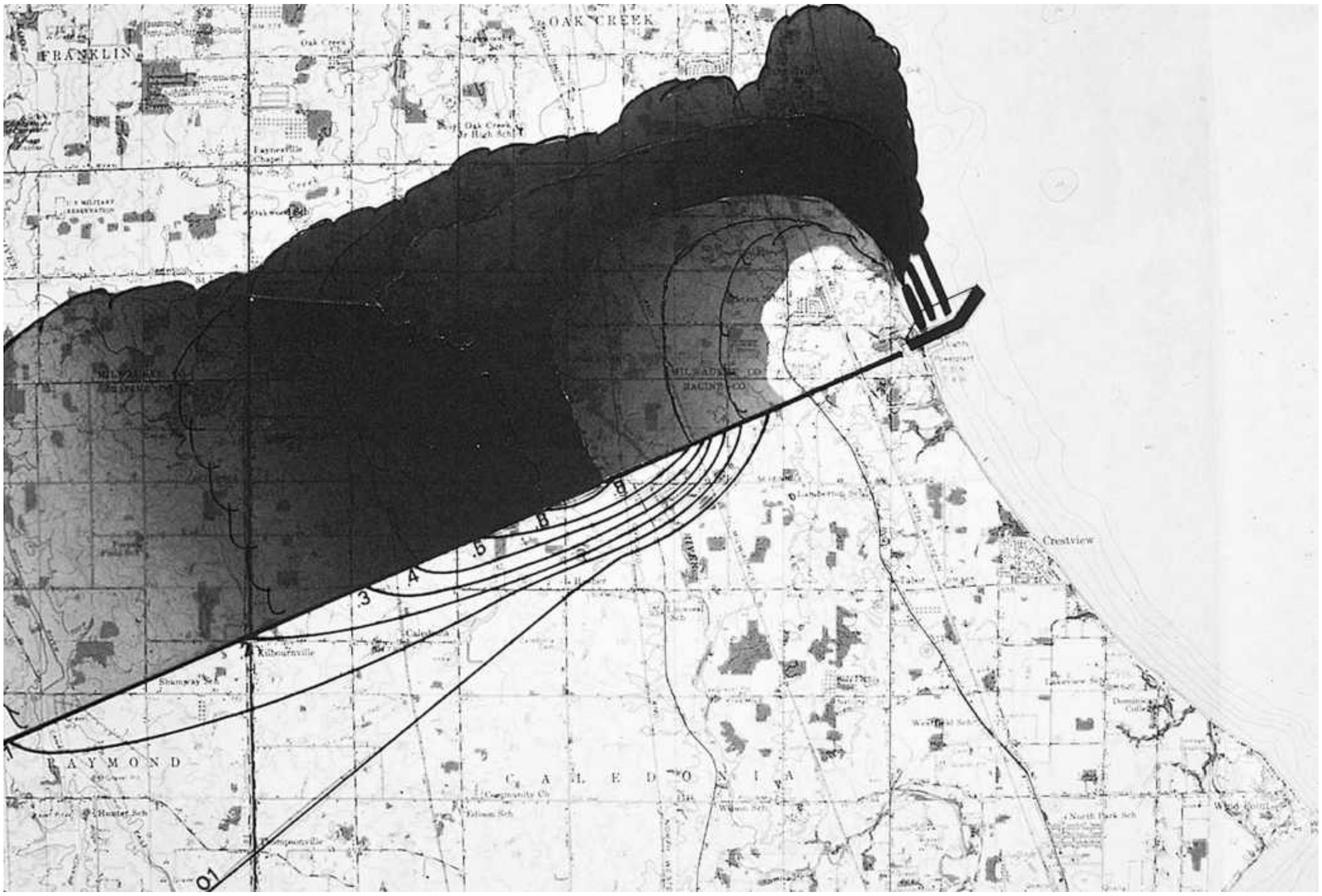
Environmental protection and resource use have to be considered in a comprehensive framework, and all of the relevant economic and natural scientific aspects have to be taken into consideration. The concepts of "entropy" and "sustainability" are useful in this regard. The entropy concept says that every system will tend toward maximum disorder if left to itself. In other words, in the absence of sound environmental policy, Earth's energy sources will be converted to heat and pollutants that must be received by Earth. The concept of sustainability has to do with the ability of a population to engage in economic activity and energy development without creating future irreversible problems for the environment and therefore for the economy.

The energy issues associated with environmental protection are complex ones involving trade-offs. Energy is an integral aspect of civilization. People in the industrialized nations use energy for heating, cooling, lighting, cooking, entertainment, transportation, communication, and for a variety of other applications, including home security systems and fire protection systems. Energy powered the Industrial Revolution and is now necessary for the age of information technology.

The developing world is rapidly exploiting energy supplies so that they, too, can benefit from industrial growth and economic development to enjoy the kind of comforts and conveniences available in the highly industrialized nations. Half of the new electric power generation facilities to be installed in the first ten years of the twenty-first century will be in China, largely because it is one of the fastest-growing economies in the developing world, and more than one-fifth of the world's population resides within its borders.

Some environmentalists believe that the root cause of environmental problems associated with energy is population growth: The more people there are, the more the demand for energy, and the greater the adverse environmental impact. Unfortunately, there is no easy solution. Most people living in the developing world live in dire poverty. Developing nations trying to solve the poverty problem have historically pushed for economic growth, but over the last quarter of the twentieth century more nations made greater efforts to control population as well. It is widely believed that the sustainable solution to the energy and environment problem requires an effort on both fronts. In 1900 the world's population totaled about 1.6 billion people. The total went over 5 billion by 1990, reached nearly 6 billion in 2000, and the United Nations projects 9.35 billion by 2050. Virtually all of the projected increase will come from the developing world (from 4.75 billion to 8.20 billion).

As the world's population increases, economic affluence is also increasing in many regions. Affluence usually means more business providing more goods and services; more homes and business with more climate control; more use of appliances and electronic technology in homes and businesses; and more miles logged by more automobiles, trucks, trains, and air-



Computers can be used to generate models of power plant pollution such as the one seen here. (FMA Production)

planes. Despite impressive gains in energy efficiency, the demand for energy has continued to grow throughout the world. Total energy consumption, including energy use in all sectors and non-energy uses of fossil fuels, increased from 66 quadrillion Btus in 1970 to more than 95 quadrillion Btus by 2000 in the United States. Over the same period, total world energy consumption had increased to more than 400 quadrillion Btus, from 207 quadrillion Btus in 1970. In the United States, about 64 percent of the 95 quadrillion Btus was consumed directly in end uses, and about 36 percent of the total was consumed for electric power production.

The fear associated with the growing populations and affluence in the developing world is that there will be far more people demanding far more energy to power technology. If economic growth and affluence accelerate in the developing world and if its population assumes the high energy use behavior of the United States—where energy use per capita is ten to twenty times that of the developing world—environmental problems will significantly worsen. For example, China's economy grew more than tenfold from 1953 to 1989, and at the same time its energy consumption grew 18-fold. Because China has vast reserves of coal, which is by far the dirtiest-burning fossil fuel and the one that emits the most carbon dioxide, it is highly likely it will double the amount of coal burned between 1998 and 2015 just to maintain its present economic growth rate.

The greatest growth in energy consumption has come in the form of electrical energy. From 1970 to 1990, the amount of electricity consumed worldwide more than doubled. This increase in electrical energy demand required a corresponding increase in generating plants, substations, and transmission lines. However, because of improvements in the efficiency of equipment—for both new sites and upgrades—this has not resulted in a doubling of either coal and natural gas production, or of generating and distribution equipment.

The fuels being used today for electric power production are coal, oil, natural gas, uranium (for nuclear power), and various other materials, including refuse-derived fuels. Most of the electric power in the United States is produced using coal-fired generating plants. Coal met 52 percent of the total requirements for electricity generation in 1997. In the same year, nuclear power plants provided 18 percent, natural-gas-fired power plants 14 percent, and hydroelectric plants about 10 percent. Oil, or petroleum, makes a relatively small contribution, only about 3 percent. In the future non-hydro renewable-energy technologies, which accounted for less than 2 percent of the total in 1997, may make a larger contribution if policies are implemented to curtail carbon emissions to combat global warming.

With the unprecedented levels of energy consumption taking place in all sectors of the economy, it begs the question of sustainability and the capacity of Earth's environment to withstand this consumption. Most environmentalists feel that the heavy reliance on fossil fuels, and the emissions from combustion, are already beyond Earth's sustainable capacity. Others are more optimistic, believing that the greater energy demands of humans can be fulfilled with only a minimal negative impact on the environment, and that science and technology can solve all the environmental problems associated with energy production and consumption.

ENVIRONMENTAL ACTION

Public concern with controlling population growth, managing economic growth, and ensuring that growth does not adversely affect the environment has found a voice through the many environmental groups that have come into existence, such as the Sierra Club, the Environmental Defense Fund, Friends of the Earth, the Bullitt Foundation, the Wilderness Society, and Greenpeace. Environmental advocacy organizations collectively have an annual budget of more than $50 million, and push for greater spending for environmental mitigation and cleanup efforts, which already run into the billions of dollars annually. In 1997 the Department of Energy (DOE) spent more than $6 billion, about one-third of its budget, for nuclear waste management and clean up at federal facilities.

In 1997 the Department of Energy spent about one-third of its budget for nuclear waste management and clean up at federal facilities. (Field Mark Publications)

The first Earth Day, April 22, 1970, was supported, directly or indirectly, by more than 20 million Americans. It is often cited as the beginning of the environmental awareness movement in the United States. However, prior to that day several individuals spoke to the need for environmental protection. Notable are the writings of Henry David Thoreau and Rachel Carson and the public policies of Presidents Theodore Roosevelt and Franklin D. Roosevelt. At Walden Pond, Thoreau saw an unspoiled and undeveloped forest as the means to preserve wild animals. Rachel Carson's book *Silent Spring* described the dangers associated with using DDT and other pesticides. Theodore Roosevelt and Franklin D. Roosevelt expanded the national park system to provide a habitat for wildlife and to prevent the destruction of natural lands through development and exploitation.

During the last third of the twentieth century, many federal legislative proposals have addressed environmental protection and resource conservation that has had a profound impact on the energy industry. At the federal level, environmental regulations have been managed by the U.S. Environmental Protection

Agency (EPA), which was established in 1970, the same year the first Clean Air Act was passed into law. In 1972 the Clean Water Act became law, and in 1973 the Endangered Species Act became law. Other important federal environmental legislation includes the Resource Conservation and Recovery Act, passed in 1976; the Response, Compensation, and Liability Act of 1980; the Nuclear Waste Policy Acts of 1982 and 1987; and the Low-Level Radioactive Waste Policy Acts of 1980 and 1985. From 1980 to 2000 these environmental regulations, and the enforcement efforts of the EPA, have had a much greater impact on decisions made in the energy industry than all the policy initiatives implemented by the DOE.

Some of the environmental issues associated with energy production and utilization that are being given the most attention by maker of public policy and by environmental groups include clean air, global climate change, nuclear power, electric and magnetic fields, oil spills, and energy efficiency.

Clean Air

Public concerns about air quality led to the passage of the Clean Air Act in 1970 to amendments to that act in 1977 and 1990. The 1990 amendments contained seven separate titles covering different regulatory programs and include requirements to install more advanced pollution control equipment and make other changes in industrial operations to reduce emissions of air pollutants. The 1990 amendments address sulfur dioxide emissions and acid rain deposition, nitrous oxide emissions, ground-level ozone, carbon monoxide emissions, particulate emissions, tail pipe emissions, evaporative emissions, reformulated gasoline, clean-fueled vehicles and fleets, hazardous air pollutants, solid waste incineration, and accidental chemical releases.

Energy use is responsible for about 95 percent of NO_x (precursor of smog—i.e., urban ozone) and 95 percent of SO_x (acid rain). The choice for the energy industry was to switch to "cleaner" energy sources, or to develop the technology that will make using the "dirtier" energy sources less harmful to the environment. For utilities making electricity generation decisions, this meant going with natural gas instead of coal for new generation, switching from high-sulfur eastern U.S. coal to low-sulfur western U.S. coal, and installing flue gas desulfurization equipment to comply with the sulfur dioxide provisions of the Clean Air Act amendments of 1990. In the transportation industry, most of the dramatic reductions in vehicle emissions have come about from advances in engine technology, not from improvements in fuels.

Global Climate Change

There is growing concern that certain human activities will alter the Earth's climate. The global climate change issue is perhaps better termed "the greenhouse issue" because the concern is that certain "greenhouse gases," including carbon dioxide, CFCs, HFCs, PFCs, SF6, nitrous oxide, and methane will accumulate in Earth's upper atmosphere and cause a heating effect similar to that in a greenhouse.

The global climate change issue is of particular concern to those in the energy field because energy production and consumption involve combustion, which has been a major factor in increasing atmospheric carbon dioxide concentrations since the beginning of the Industrial Revolution. Energy use in the United States is responsible for about 98 percent of human-generated carbon dioxide emission.

Concerns about global climate change have led to extensive research and high-level international debates about the need for targets and timetables to reduce carbon dioxide emissions. Some policymakers believe that current uncertainties in how to approach the issue do not justify an all-out effort to reduce carbon dioxide emission, while others feel that this is a crisis needing immediate attention.

There are many uncertainties about global climate change, the workings of global greenhouse gas sources and sinks, and techniques to reduce or sequester emissions. It has been established that the CFCs used for airconditioning have higher global warming potential than carbon dioxide. CFC-reduction methods have been implemented. Research is under way to better understand how agriculture and forestry produce and absorb gases such as carbon dioxide, nitrous oxide, and methane. Methane is produced by rice cultivation, animal waste, and biomass burning. Nitrous oxide is produced from cultivation, fossil fuel/biomass burning, and fertilizer use.

The Intergovernmental Panel on Climate Change predicted in 1993 that a doubling of carbon dioxide concentrations by 2100 will occur under a "business as usual" scenario. However, technologies now exist where greenhouse gas growth rate can be reduced. The electric utility industry advocates an evolution toward highly efficient electrotechnologies, with more of the electricity produced by natural gas, as a way to mitigate carbon dioxide emissions.

Transportation

The transportation sector is a major polluter of the environment. In 1994 there were 156.8 million noncommercial vehicles on the road, averaging 19.8 miles per gallon, traveling an average of 11,400 miles, and burning an average of 578 gallons of fuel annually per vehicle. The trend has been one of more vehicles more miles driven, and more roads demanded by drivers.

Transportation accounts for about one-fourth of the primary energy consumption in the United States. And unlike other sectors of the economy that can easily switch to cleaner natural gas or electricity, automobiles, trucks, nonroad vehicles, and buses are powered by internal-combustion engines burning petroleum products that produce carbon dioxide, carbon monoxide, nitrogen oxides, and hydrocarbons. Efforts are under way to accelerate the introduction of electric, fuel-cell, and hybrid (electric and fuel) vehicles to replace some of these vehicles in both the retail marketplace and in commercial, government, public transit, and private fleets. These vehicles dramatically reduce harmful pollutants and reduce carbon dioxide emissions by as much as 50 percent or more compared to gasoline-powered vehicles.

Technology is making possible more fuel-efficient and cleaner-running automobiles, but it cannot do anything to reduce the number of automobiles or the ever-increasing number of highways. As long as the law requires that all the revenue from the federal fuel tax be used to build and maintain highways, the number of miles of highway in the United States will continue to increase at an explosive rate. The amount of road-building that has taken place in America since the early 1950s have produced an environment well suited for the automobile but not for plants, wildlife, or even people. In fact, physicist Albert Bartlett showed that the number of miles of highway will approach infinity under extended projections.

Besides all the gaseous and liquid wastes of transportation that result from energy use, and the loss of natural environment to roadways, there is also the solid-waste problem of disposal—vehicles and components such as tires and batteries. Responding to the growing disposal problem, many manufacturers are building automobiles that contain far more recyclable parts.

Water Pollution

The major energy-related sources of water pollution are from thermal pollution, surface water pollution from oil spills, polychlorinated biphenyls, and groundwater contamination.

Thermal Pollution. Thermal pollution occurs as a result of hot-water emissions from electric power plants. High-temperature steam, which causes the turbine blades to rotate, passes by the blades, cools and condenses to liquid water. This condensation liberates energy that must be removed by circulating water in pipes in contact with the condenser. Each second the water must remove more than a billion joules of heat. When this heated water is discharged into a body of water, it can alter aquatic life in two ways: Significantly warmer water retains less oxygen, making it more difficult for many species to survive, and warmer water favors different species than cooler water. The altered ecosystem may harm all species; attract less desirable organisms; or could even improve survivability of the most desirable organisms, especially during the winter months.

Oil Spills. Oil spills occur from oil pipeline leaks, oil tanker accidents, or submarine oil drilling operations. The two major ocean drilling accidents—oil wells blowing out—were the 1969 Santa Barbara Channel spill and the 1979 Yucatan Peninsula spill, in Mexico. The Yucatan spill spewed out more than three million barrels before being capped in 1980. Both caused damage to beaches and marine life, but the smaller Santa Barbara spill was far more devastating because of unfavorable winds following the accident.

The largest oil spill in U.S. history occurred in March 1989, when the tanker *Exxon Valdez* ran aground in the Prince William Sound inlet in the Gulf of Alaska, spilling 11 million gallons (42 million liters) of crude oil. The resulting slick covered more than 1,000 miles (1,600 kilometers) of the Alaska coastline and caused an estimated $3 billion to $15 billion in environmental damages. The spill killed hundreds of thousands of fish and seabirds and thousands of otters. The tanker's captain, Joseph J. Hazelwood, had reportedly been drinking before the accident and had turned control of the ship over to the third mate. The state of Alaska in March 1989 brought a criminal indictment against Hazelwood for his role in the disaster, but in March 1990 he was acquitted of the most serious criminal charges against him. Prosecutors were unable to convince jurors that

Figure 1.
Although the pipeline from the North Slope to the port of Valdez was the route chosen and completed in 1977, the scientific community preferred the trans-Alaska-Canada route; its use would have prevented the *Exxon Valdez* tanker accident of 1989.

Hazelwood was legally drunk at the time of the accident. In September 1994, a federal court jury ordered Exxon Corp. to pay $5 billion in punitive damages to Alaskan fishermen, local residents, and property owners. The fine was reported to be the highest punitive award ever levied against a corporation and also the largest ever in an environmental pollution case. Since the *Exxon Valdez* spill, a number of safeguards have been instituted to help prevent future oil spills, such as the mandating of double-walled tankers and better pipelines.

Some environmentalists anticipated a major accident like that of the *Exxon Valdez*. When planning for the Trans-Alaska Pipeline took place during the energy crisis of the early 1970s, it was controversial because of the conflicts of balancing the needs of oil resource development and the known and predicted environmental problems. The tremendous political pressure to quickly deliver new oil supplies for national security reasons hastened a decision. Many feel that the Prudhoe Bay to Valdez Trans-Alaska Pipeline was chosen because it meant the most rapid resource development. An alternative trans-Alaska-Canada route—which would have avoided the marine environment tanker risk and earthquake zone pipeline risk—was favored by environmentalists, but was rejected during the political process (see Figure 1). It would have taken longer to construct, yet this route would have made it possible to include an oil and gas pipeline in one corridor.

Polychlorinated Biphenyls. Polychlorinated biphenyls (PCBs) are carcinogenic and adversely affect the liver, nervous system, blood, and immune response system. Moreover, PCBs persist a long time in the natural environment and become concentrated in the higher parts of the food chain. A major source of PCBs occurred during the production of electrical capacitors and transformers. It is believed that the Hudson River has more than 295,000 kilograms of PCBs that were discharged into the river by General Electric from 1950 to 1977. The controversy is whether to spend millions removing and treating the river sediment, or to allow the river to clean itself by the natural process of sediment transport to the ocean. Removal and treatment, has been advocated by the EPA but has not yet been initiated.

Groundwater. The major threat to groundwater is from leaking underground fuel storage tanks. It is unknown how many underground storage tanks leak, how much gets into the groundwater, and what impact the leaks have on human health. The primary worry is not with the fuels themselves but with the "cleaner burning" additive methyl-t-butyl ether (MTBE). Ironically, MTBE, whose use the EPA mandated in 1990 as a way to improve air quality by lowering harmful emissions from vehicles, may turn out to be the most serious threat ever to drinking water. It is a suspected carcinogen, it migrates quickly into groundwater, and there is no known way to treat groundwater that is polluted with MTBE. The federal government was looking at banning MTBE as of 2000, but because of its prevalent use during the 1990s and the extent of underground fuel tank leakage, MTBE in groundwater will be an environmental problem for years to come. There is also a problem in finding a gasoline additive substitute for MTBE. The leading replacement contender, ethanol, has a lower octane rating (106 compared to 116 for MTBE) and is considerably more expensive.

Nuclear Power

Public opposition to commercial nuclear power plants began with the misperception that the plants could explode like nuclear weapons. The nuclear industry made progress in dispelling this misperception, but suffered major setbacks when an accident occurred at the Three-Mile Island nuclear power plant in Pennsylvania and at the Chernobyl nuclear power plant in the USSR.

March 28, 1979, an accident at Three-Mile Island resulted in a small release of radiation when a pressure relief valve became stuck open. According to a report by the U.S. Nuclear Regulatory Commission (NRC), the dose of radiation received by the people in the immediate area of Three-Mile Island was much less than the radiation dose that the average member of the U.S. population receives annually from naturally occurring radiation, medical use of radiation, and consumer products. In subsequent years law suits have been filed by plaintiffs contending that high radiation exposure levels at Three-Mile Island caused them to develop cancer, but the courts have ruled that the plaintiffs failed to present any evidence that they were exposed to enough radiation to cause their cancers.

A much more serious nuclear accident occurred at Chernobyl in the USSR on April 26, 1986, when one of the Chernobyl units experienced a full-core meltdown. The Chernobyl accident has been called the worse disaster of the industrial age. An area comprising more than 60,000 square miles in the Ukraine and Belarus was contaminated, and more than 160,000 people were evacuated. However, wind and water have spread the contamination, and many radiation-related illnesses, birth defects, and miscarriages have been attributed to the Chernobyl disaster.

The fear of accidents like Chernobyl, and the high cost of nuclear waste disposal, halted nuclear power plant construction in the United States in the 1980s, and in most of the rest of the world by the 1990s. Because nuclear fusion does not present the waste disposal problem of fission reactors, there is hope that fusion will be the primary energy source late in the twenty-first century as the supplies of natural gas and petroleum dwindle.

Electric and Magnetic Fields

Questions are being raised about the possible health effects of electric and magnetic fields from electric power transmission, distribution, and end-use devices. Electric and magnetic fields exist in homes, in workplaces and near power lines. Electric fields exist whenever equipment is plugged in, but magnetic fields exist only when equipment is turned on. Both electric and magnetic fields become weaker with distance from their source. Science has been unable to prove that electric and magnetic fields cause adverse health effects, but further investigation is under way. Research is focusing on the possible association between magnetic field exposure and certain types of childhood cancer. While laboratory experiments have shown that magnetic fields can cause changes in living cells, it is unclear whether these experiments suggest any risk to human health.

Energy Efficiency and Renewable Energy

Energy efficiency and renewable energy are being promoted as sustainable solutions to the world's energy needs. Through energy efficiency less energy is used, generally resulting in reduced demands on natural resources and less pollution. And if less energy is being consumed, a far greater fraction can come from renewable sources.

Some energy efficiency strategies include moving toward more fuel-efficient cars, electricity-stingy appliances, and "tighter" homes with better windows to reduce heating and air conditioning requirements. Because of dramatic improvements in the energy efficiency of appliances, particularly refrigerators, far fewer coal-fired power plants needed to be built in the 1980s and 1990s. However, the price of energy is relatively low in the United States, so businesses and citizens do not feel an urgency to conserve energy and purchase only energy-efficient products. In Japan and Europe, where citizens pay two to three times as much as Americans for gasoline and electricity, energy consumption per capita is half what it is in the United States. Many environmentalists feel the U.S. government needs to institute a carbon tax, mandate that manufacturers provide more energy-efficient appliances, and develop other incentives to encourage purchase of energy-efficient products.

During the 1970s and early 1980s, many in the environmental movement suggested a low-energy or soft-path-technology future, one that would involve greater conservation, increased efficiency, cogeneration, and more use of decentralized, renewable energy sources. Moreover, the claim was that this could be done without a reduction in the quality of life if future development included more energy-efficient settlement patterns that maximized accessibility of services and minimized transportation needs, if agricultural practices involved less energy in the production of food and emphasized locally grown and consumed foods, and if industry followed guidelines to promote conservation and minimize production of consumer waste. By 2000, many of these gains in energy efficiency were realized—energy is being used far more efficiently than ever before—yet the ever-growing U.S. population and new uses of technologies necessitated more fossil fuel energy production, which made low-energy-future scenarios in the United States impossible (see Figure 2). Nevertheless, if not for energy efficiency improvements, many more electricity generating facilities would have been built.

THE ENERGY ENVIRONMENT FUTURE

Since the 1970s, technological advances have solved many environmental problems associated with energy production and consumption, and proven that more energy consumption does not necessarily mean more pollution. The fossil fuel industries are producing and distributing more energy less expensive-

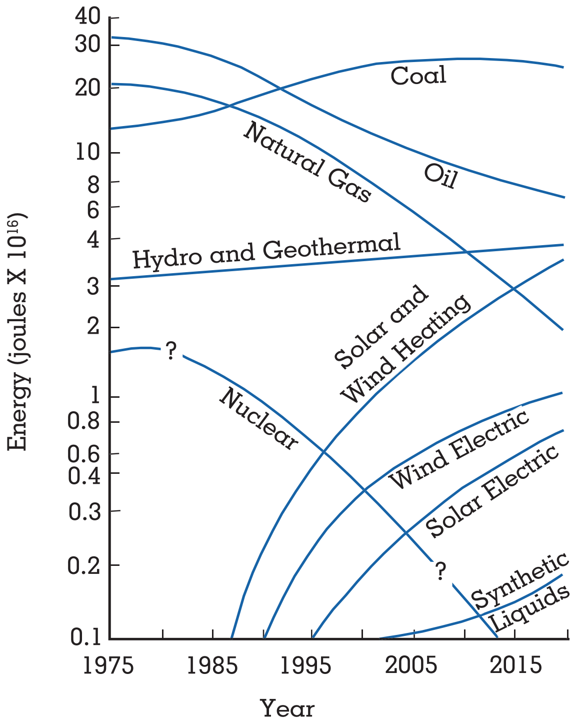


Figure 2.

A low-energy-future scenario envisioned in the late 1970s. By 2000, the energy being consumed from petroleum, natural gas, coal, and nuclear power were all on the rise, and renewable energy supplied a smaller fraction of total energy.

SOURCE: Steinhart et. al, 1978.

ly than ever before while creating less of an impact on the environment; the electricity-producing industry is generating more electricity with less fossil fuel and less harmful pollutants; and the automobile industry's best 2000 model year cars produce one-twentieth the harmful emissions of 1975 automobile models and at the same time offer far better performance and fuel economy. Certainly there are still major environmental problems. Nuclear waste is piling up, awaiting political decisions for its long-term storage. Underground fuel tanks, whose numbers are in the hundreds of thousands, are prone to leaks and migration into the drinking water supply. Deforestation continues in many parts of the world where people desire the wood for fuel and the land clear-cut for agriculture.

Hopefully, as the ongoing energy and environment debate continues, policymakers will choose a middle ground between the proposals of a few extreme environmentalists who would greatly cut back on energy development, and the proposals of a few of those in the energy industries who would unnecessarily exploit the environment to increase energy supplies. Reasonable strategies can ensure that there is sufficient energy for economic development to continue, but with minimal adverse environmental effects.

Fred I. Denny

See also: Acid Rain; Air Pollution; Atmosphere; Carson, Rachel; Climatic Effects; Disasters; Environmental Economics; Fossil Fuels; Gasoline and Additives; Gasoline Engines; Government and the Energy Marketplace; Nuclear Fission; Nuclear Fusion; Nuclear Waste.

BIBLIOGRAPHY

Bartlett, A. A. (1993). "The Arithmetic of Growth: Methods of Calculation." *Population and the Environment* (14):359–387.

Bartlett, A. A. (1969). "The Highway Explosion." *Civil Engineering* (December):71–72.

Buell, P., and Girard J. (1999). *Chemistry: An Environmental Perspective*. Englewood Cliffs, NJ: Prentice-Hall.

Botkin, D. B., and Keller, E. A. (1999). *Environmental Science: Earth as a Living Planet*, 3rd ed. New York: John Wiley & Sons.

Steinhart, J. S.; Hanson, M. E.; Gates, R. W.; Dewinkel, C. C.; Brody, K.; Thornsjo, M.; and Kambala, S. (1978). "A Low-Energy Scenario for the United States: 1975–2000." In *Perspectives on Energy*, 2nd ed., ed. L. C. Ruedisli and M. W. Firebaugh. New York: Oxford University Press.

ENVIRONMENTAL PROTECTION AGENCY

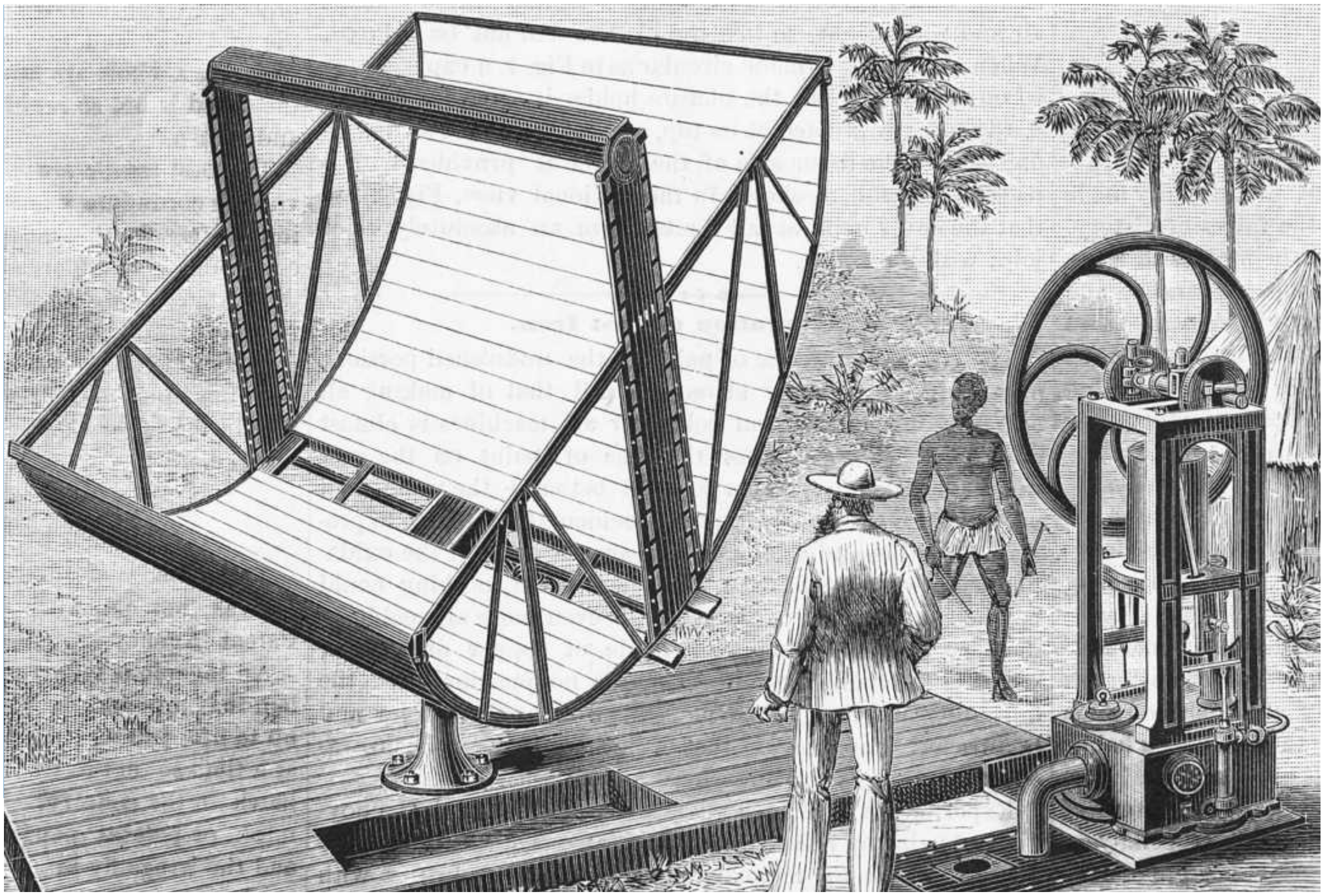
See: Government Agencies

ERICSSON, JOHN (1803–1889)

John Ericsson, born at Langbanshyttan, in the province of Varmland, Sweden, on July 31, 1803, was the youngest of three children of a mine owner. Ericsson acquired his skills as a mechanical draftsman at the age of thirteen, working on the construction of the Göta Canal. Against all advice Ericsson left the canal company to embark on a military career in 1820. At an early age Ericsson had experimented with mechanics and continued to do so when in the military, constructing what he termed a "flame engine" to challenge the steam engine. This engine worked by internal combustion, and the success of a small model prompted Ericsson to travel to England in 1826 to demonstrate and patent his invention.

Although the engine was not a success, Ericsson's trip to London allowed him to meet John Braithwaite, a machine manufacturer, who had the expertise to put Ericsson's ideas into practice. In 1828 Ericsson, with Braithwaite, patented the principle of artificial draft in steam boilers. The principle of forced draft was applied to a fire engine and a locomotive entered for the Rainhill locomotive trials of 1829.

Ericsson continued his search for a substitute for steam, and in 1833 patented his caloric engine. Fitted with what Ericsson termed a regenerator, this engine allowed heat to be re-used resulting in savings in fuel. These savings, Ericsson believed, made the engine suitable for marine use. The tubular heat exchanger was an invention of Robert Stirling, but Ericsson was not always inclined to admit priority or give credit to those whose ideas he put into practical form. Ericsson unsuccessfully opposed Stirling's second patent application in 1827. Ericsson demonstrated a five horsepower engine, generating much fierce



Undated illustration of John Ericsson's solar engine. (Corbis Corporation)

debate and being dismissed by many as an unworkable attempt at perpetual motion.

Ericsson lived an expensive lifestyle. The costs of his experiments plunged him into debt, confined him twice to a debtors' prison, and eventually led to his declaration of bankruptcy in 1835. In 1836 Ericsson patented a rotary propeller for ships. With the financial backing of the U.S. Consul in Liverpool, Ericsson built in 1837 a 45-foot boat that successfully demonstrated the superiority of screw over paddlewheel, and gained him orders for two iron ships from the United States. Although Ericsson built another caloric engine in 1838, he failed to convince the British establishment of the benefits of either his propeller or engine. He was persuaded to travel to New York the following year, never to return to England or Sweden.

Once in New York, Ericsson met with Harry DeLameter of the DeLameter Iron Works and immediately sought out backers for his inventions. Work on the caloric engine with a wire gauze regenerator continued. Surprisingly, Ericsson adopted an open cycle, although in his patent he had stated the benefits of using air under pressure. Such was Ericsson's faith in his engine that by 1852 he had raised sufficient capital to build a caloric-powered ship to challenge the dominance of the marine steam engine. The 250-foot ship, built of wood, was fitted with sails and paddlewheels, although Ericsson had successfully applied his propeller to the U.S. warship *Princeton*. The trial voyage of the caloric-ship in January 1853 was a stage-managed affair in which Ericsson avoided giving precise details of the ship's machinery.

A subsequent trip was made to Washington, a round trip of some five hundred miles. Unfortunately, the ship sank in April in freak weather conditions and, although successfully salvaged, her engines were replaced with steam engines. Ericsson's detractors claimed the ship was underpowered and the giant fourteen-foot diameter pistons impractical. By adopting an open cycle, Ericsson was unable to prove convincingly that his giant caloric engines fitted with regenerators could successfully overthrow the marine steam engine. Ericsson was to subsequently destroy all details and drawings of the ship's engines. The concept that "caloric" was a fluid that could be used over and over was not sustained in the eyes of Ericsson's critics.

The failure of the caloric ship and the financial losses incurred by his backers did not diminish Ericsson's enthusiasm for the air engine. The stationary steam engine had weaknesses: It required skilled operators and incurred heavy insurance costs. Ericsson again successfully raised funds to develop a small open-cycle stationary caloric engine with only a rudimentary regenerator. For Ericsson and his backers it was a financial success. Marketed as requiring no water, it could be operated by unskilled labor and, perhaps most important of all, would not explode.

During the American Civil War, Ericsson designed a steam-driven iron-clad ship, a concept initially rejected by the Navy Board as being too novel. The *Monitor,* built through the intervention of President Lincoln, but at Ericsson's expense, brought to a sudden end the era of wooden warships.

Ericsson devoted much time to the study of tidal action, although schemes for tidal power were abandoned, as he later commented as "not being able to compete with the vast energy stored in lumps of coal. But the time will come when such lumps will be as scarce as diamonds." A life-long interest in solar power proved more fruitful. Ericsson experimented with both condensing steam engines and his caloric engine, using silvered glass reflectors. Ericsson built in 1873 a closed-cycle solar motor along the lines of Stirling's engine, developed with a view of irrigating the parched lands of western states; however, little interest was shown in the sun-motor by the agricultural community. Ericsson patented the design in 1880 as a stationary engine that could be heated by liquid, gas, or solid fuels. It was sold in large numbers.

Shortly after joining the army, Ericsson met and later became betrothed to Carolina Lillieskold, the daughter of an army captain. From this relationship, Ericsson's only child, a son Hjalmer, was born in 1824; however, Captain Lillieskold's hostility prevented a marriage. Ericsson left Sweden two years later without knowing he had fathered a son. It was Ericsson's brother Nils who eventually learned of the birth and arranged for the child to be raised by their mother, although for nearly fifty years Ericsson had no contact with his son. In London, in 1836, Ericsson married Amelia Byam, fourteen years his junior. Amelia joined her husband in New York but found him absorbed in his work. The marriage was not a happy one, and Amelia returned to England. Ericsson continued to support her financially until her death in 1867, but they never met again.

For all Ericsson's years in America, he rarely travelled outside of New York. Because of failing health he withdrew from public life, but resolved to die in harness and continued inventing. John Ericsson passed away, in seclusion, on the morning of March 8, 1889.

Congress, which had had an uneasy relationship with the living Ericsson, and unable to decide how to honor the inventor of the *Monitor*, acquiesced to a suggestion from the Swedish government that Ericsson's body be returned to Sweden. In 1926 Ericsson was finally honored with a statue unveiled in Washington, D.C. by President Calvin Coolidge and Gustaf Adolf, Crown Prince of Sweden.

Robert Sier

See also: Stirling, Robert.

BIBLIOGRAPHY

Church, W. C. (1890). *The Life of John Ericsson.* New York: Scribner.

Ericsson, J. (1876). *Contributions to the Centennial Exhibition.* Philadelphia: John Ericsson.

Miller, E. M. (1978). *U.S.S. Monitor: The Ship That Launched a Modern Navy.* Annapolis: Leeward Publications.

Norton, W. A. (1853). "On Ericsson's Hot Air or Caloric Engine." *American Journal of Science and Arts,* 45.

Sier, R. (1987). *A History of Hot Air and Caloric Engines.* London: Argus Books.

ETHANOL

See: Biofuels

ETHICAL AND MORAL ASPECTS OF ENERGY USE

The production and use of energy gives rise to a wide range of ethical and moral issues. Worldwide there are four general energy options available, each of which can raise significant ethical questions. We can continue to rely primarily on fossil fuels, presently estimated to account for more than 80 percent of worldwide energy use. Second, we could shift to greater reliance on nuclear energy. Third, we could develop alternative energy sources such as wind, solar, and geothermal power. A fourth alternative would focus on conservation and energy efficiency and seek to decrease the overall demand for energy.

Continuing dependence on fossil fuels raises several major ethical issues. Ethical questions concerning our responsibilities to future generations are raised by the fact that fossil fuels are a nonrenewable energy source, so that every barrel of oil or ton of coal burned today is forever lost to future generations. Further, the by-products of fossil fuel combustion pose hazards to both present and future generations.

Nuclear energy also faces major ethical challenges. Nuclear power generates toxic wastes that remain hazardous for thousands of generations. Even assuming that the operation of nuclear power plants can be made safe, disposal of nuclear wastes can jeopardize the health and safety of countless future people. Further, the proliferation of nuclear technology that is necessary for generating nuclear energy also raises ethical concerns of international peace and security.

Energy sources such as solar, wind, and geothermal power are often proposed as renewable and non-polluting alternatives to fossil and nuclear fuels. But here, too, ethical challenges must be faced. Over the short term, alternative energy sources will likely be more expensive relative to fossil and nuclear fuels. Such price differentials mean that safer and cleaner energy sources will more likely be available to wealthy countries and individuals while the poor will continue relying on more dangerous and polluting energy sources. As a result, questions need to be raised concerning equality and fairness in the distribution of alternative energy sources. Further, development of these alternative technologies may require government subsidies and incentives, which can raise additional questions of freedom, fairness, and equality.

Finally, conservation and energy efficiency also raise ethical challenges. A significant decrease in energy consumption is possible in only two ways: we significantly decrease demand or we significantly decrease the population of people demanding energy. Either option raises major questions concerning values such as individual freedom of choice, property rights, fairness, and equal opportunity, as well as ethical issues regarding population policy, standards of living, and quality of life.

INDIVIDUAL CHOICE AND ENERGY POLICY

We can begin to focus on the specific value issues involved in energy by reflecting on the nature of ethics itself. At the start of a dialogue on ethics, Socrates once said that “we are dealing with no small thing but with how we ought to live.” There is no more fundamental ethical question than this: How ought we to live? But as Socrates understood, this question can be interpreted in two ways. “We” can mean each one of us individually, or it may mean all of us collectively. Taken in the first sense, ethics is concerned with how I should live my life, with the choices I ought to make, with the type of person I should be. We can refer to this sense of ethics as morality. Taken in the second sense, this question refers to how we ought to live together in society. Ethics in this sense raises questions of public policy, law, institutions, social justice. To distinguish this sense of ethics from morality, we can refer to these as questions of social ethics.

Although important questions of individual morality can be involved with energy issues, the production and use of energy primarily raises questions of social ethics and public policy. This emphasis can be explained simply by the magnitude of energy issues. Such questions as resource conservation, global warming, nuclear waste disposal, and pollution will not be resolved through individual action alone. However, before turning to the public policy perspective, it will be helpful to consider some aspects of individual energy choices.

Some would argue that energy policy ought to be left to individual choice. In such a situation, some individuals may choose a frugal and conservative lifestyle that demands relatively little energy resources. One thinks here of the common environmental adage "live simply so that others may simply live." Other individuals may choose a more energy-intensive lifestyle. Either way, one could argue that the choice ought to be left to the moral values of individuals.

This option is favored by those who defend economic markets as the most ethically appropriate approach to public policy. This view argues that individual consumers, relying on their own preferences, should be free to choose from a variety of energy options. The working of a competitive market would then guarantee the optimal distribution of both the benefits and burdens of energy production. Consumers who value safe and clean energy sources would be free to choose wind or solar power and, presumably, would be willing to pay more for these benefits. Assuming a social system in which government subsidies were eliminated and external costs such as pollution were fully internalized, this economic arrangement would most efficiently satisfy the greatest number of individual desires while also respecting individual freedom of choice.

Defenders of this approach could point to the deregulation of the electric utility industry within the United States during the late 1990s as an example. As the industry becomes deregulated, a number of new firms stepped into the market to offer consumers a wider range of energy choices. In many areas, consumers who are willing to pay higher prices are able to purchase energy generated from environmentally friendly, "green" sources.

There are major problems with this individualistic approach to energy policy, however. The ideal market of economic theory exists nowhere in reality. Further, even market defenders acknowledge cases in which markets fail. Significantly, some paradigmatic examples of market failure, such as the externality of pollution and monopolistic control of production, are associated with the production of energy. More importantly, perhaps, crucial ethical questions can be missed if we only consider the perspective of individual values and choice.

Consider how a single individual might deliberate about the consumption of fossil fuels. Burning fossil fuels increases the amount of greenhouse gases released into the atmosphere and strong evidence suggests that this can lead to global warming. Given this scenario, does morality require that individuals refrain from driving automobiles?

To answer this question an individual might well weigh the benefits of driving a private car against the costs of increased greenhouse gas emissions. An average car might burn between two and three gallons of gasoline each hour. Given that an average driver might drive only a few hours each day, the amount of fuel burned in this activity will make little difference in the amount of atmospheric carbon dioxide. Weighed against the convenience and freedom of driving one's own car and the market-established price of gasoline, it may well be reasonable for an individual to decide that there is no significant moral issue involved in driving.

However, if we extend this line of reasoning across a large population, a decision that seems minor to an individual can turn out to have enormous social implications. If millions of people make the same seemingly reasonable decision and burn millions of gallons of gasoline each day, the atmospheric consequences are significant. We are here faced with ethical and policy questions that would never arise from an individual's point of view. For example, we might consider increasing taxes on gasoline, requiring automobile manufacturers to improve mileage efficiency, subsidizing mass transit, providing tax incentives for alternative fuel transportation, or even prohibiting private automobiles in urban areas. The crucial point is that none of these questions would ever arise from the perspective of an individual facing a single choice. Recognizing that these are important ethical questions that deserve consideration, we recognize the need for treating public policy questions as distinct from individual moral questions.

A second inadequacy of the individualistic approach is that it can underestimate the influence which social practices have upon individual choice. As individuals, we pursue goals based on our interests and desires. But a complete ethical analysis should include an examination of the source of those interests and desires.

If we take consumer demand as a given, then the major task for energy policy is to produce enough energy to satisfy that demand. Alternative policies will then be judged in terms of how well they accomplish that task. But when we recognize that the demand for an energy-intensive lifestyle is a product of social and cultural factors, and that these factors themselves can be influenced by public policy, then we see the need to ask questions that might ordinarily be ignored by individuals.

For example, should we pursue policies that would discourage energy use and encourage conservation? Are all uses of energy equally valid on ethical grounds? Should energy producers be discouraged from advertising, or should they be required to promote conservation? Should poor, less-developed countries receive subsidized energy resources from the developed, industrialized countries? Again, these are questions that are raised only from a public policy perspective. Clearly, an adequate ethics of energy must move beyond moral questions and focus on social and public policy perspectives.

SOCIAL ETHICS AND ENERGY POLICY: PRESENT GENERATIONS

Turning to social ethics, we can distinguish two general types of ethical questions that pertain to energy policy: questions of justice in the present, and questions concerning our responsibilities to future generations. Issues concerning social justice for present generations can be categorized in terms of debates between utilitarian (maximizing beneficial consequences) and deontological (acting in accord with moral principles and duties) approaches to ethics. Ethical questions concerning future generations involve both the content and the very existence of such duties.

Utilitarian ethics holds that energy policy ought to be set according to the general ethical rule of maximizing the overall good. For example, if oil exploration in an Arctic wilderness area would produce greater overall social happiness than would preservation of the wilderness area, utilitarian ethics would support exploration. Utilitarianism is a consequentialist ethics in which good and bad policy is a function of the consequences that follow from that policy. Policies that increase net social benefits are right, those that decrease net social benefits are wrong. Thus, utilitarianism employs what can be thought of as an ethical cost-benefit methodology, weighing the benefits and harms of various alternatives and promoting that option which proves most useful in maximizing benefits over harms. Because energy is something valued only for its usefulness, utilitarian ethics seems well suited for establishing energy policy.

Explained in such general terms as maximizing the good, utilitarianism is an intuitively plausible ethical theory. Disagreements occur when defenders attempt to specify the content and meaning of the good (or "happiness"). An important contemporary version of utilitarianism identifies happiness with the satisfaction of individual desires or, simply, getting what one wants. Sometimes identified as preference utilitarianism, this view equates the good with the satisfaction of individual preferences and is closely associated with the goal of microeconomic efficiency. This particular version of utilitarianism has had a profound impact on energy policy, especially energy policy as it is found in liberal democratic societies. From this perspective, the goal of energy policy is to optimally satisfy the demand for energy while minimizing any potential harms that might result.

Two trends within this general utilitarian approach dominate energy policy. One holds that there are experts who can predict policy outcomes, determine relative risks and benefits, and administer policies to attain the goal of maximum overall happiness. These experts, trained in the sciences, engineering, and the social sciences, are best situated to predict the likely consequences of alternative policies. Scientific understanding of how the world works enables these experts to determine which policies will increase the net aggregate happiness. This version of utilitarian thinking typically supports government regulation of energy policy and, as a result, is often criticized on ethical grounds as involving paternalistic interference with individual decision-making and property rights.

A second trend within the utilitarian tradition argues that efficient markets are the best means for attaining the goal of maximum overall happiness. This version would promote policies that deregulate energy industries, encourage competition, protect property rights, and allow for free exchanges. In theory, such policies would direct rationally self-interested individuals, as if led by an "invisible hand" in famed economist Adam Smith's terms, to the optimal realization of overall happiness. As with the approach that relies on energy experts, the market approach agrees that the goal of energy policy ought to be the optimal satisfaction of consumer demand.

Both approaches share two fundamental utilitarian assumptions. First, utilitarianism is a consequentialist ethics that determines right and wrong by looking to the results of various policies. Second, they hold that ethics ought to be concerned with the overall, or aggregate, welfare. Deontological ethics (the word is derived from the Greek word for duty) rejects both of these assumptions.

Committed to the ethical maxim that the ends don't justify the means, deontological ethics rejects the consequentialism of utilitarianism for an ethics based on principles or duties. There are many cases in which ethics demands that we do something *even if* doing otherwise would produce greater overall happiness. On this view, right and wrong policy is a matter of acting on principle and fulfilling one's duties. Respect for individual rights to life and liberty or acting on the demands of justice are common examples of ethical principles that ought not be sacrificed simply for a net increase in the overall happiness.

An especially troubling aspect of utilitarianism is the emphasis on collective or aggregate happiness. This seems to violate the ethical principle that individuals possess some central interests (to be distinguished from mere preferences) that ought to be protected from policies aimed simply at making others happier. Most of us would argue that individuals have rights that ought not to be sacrificed to obtain marginal increases in the aggregate overall happiness. Our duty to respect the rights of individuals, to treat individuals as ends in themselves, and not as mere means to the end of collective happiness, is the hallmark of deontological ethics.

Nowhere is this concern with individual rights more crucial than in questions concerning the justice of energy policy. Distributive justice demands that the benefits and burdens of energy policy be distributed in ways that respect the equal dignity and worth of every individual. A *prima facie* violation of justice occurs when social benefits and burdens are distributed unequally. Particularly troubling inequalities occur when the benefits of policy go to the powerful and wealthy, while the burdens are distributed primarily among the poor and less powerful.

Consider, as an example, the logic of a policy decision to build and locate an electric generating plant or oil refinery. Economic considerations such as the availability of ample and inexpensive land, and social considerations such as zoning regulations and political influence, would play a major role in such a decision. In practice, this makes it more likely that plants and refineries, as well as waste sites and other locally undesirable land uses, will be located in poorer communities whose population is often largely people of color.

Evidence suggests that this is exactly what has happened. Beginning in the 1970s, sociologist Robert Bullard studied the location of hazardous and polluting industries within the United States. He found that many such industries, including many electric generating plants and oil refineries, are disproportionately located in minority communities. This is not to claim that there has been an intentional social policy to unfairly burden minority communities. But it does suggest that the economic and political factors that give rise to such decisions have much the same practical effect. Upper income levels disproportionately benefit from inexpensive energy to fuel consumerist lifestyles, while lower income minority communities carry a disproportionate burden of energy production.

Much the same has been said on the international level. Debates concerning international energy justice occurred frequently during the Earth Summit in Rio de Janeiro in 1992 and the Kyoto conference on global warming in 1997. Representatives of the less developed countries argue that industrialized countries have long benefited from readily accessible and inexpensive energy resources, which have been primarily responsible for the proliferation of greenhouse gases and nuclear wastes. However, after having attained the benefits of this lifestyle, the industrialized countries now demand a decrease in greenhouse emissions, conservation of resources, and a reduction in the use of nuclear energy. These policies effectively guarantee that the non–industrialized world will remain at an economic and political disadvantage. Many argue that justice would demand industrialized countries carry a heavier burden for decreasing energy demands, reducing greenhouse emissions, storing nuclear wastes, and for conserving non–renewable resources. Influenced by such reasoning, the majority of industrialized countries accepted greater responsibility at the Kyoto conference for reducing greenhouse gases.

From a strictly utilitarian perspective, unequal distribution of the benefits and burdens of energy production and use might be justified. Utilitarians have no in-principle objection to unequal distribution. If an unequal distribution would create a net increase in the total aggregate amount of happiness, utilitarians would support inequality. Deontologists would argue that these practices treat vulnerable individuals as mere means to the end of collective happiness and, thus, are unjust and unfair. Such central interests as health and safety ought not be sacrificed for a net increase in overall happiness. Moral and legal rights

The Hanford Nuclear Reservation in eastern Washington stores class A low-level radioactive waste. (Corbis)

function to protect these interests from being sacrificed for the happiness of others.

Legal philosopher Ronald Dworkin suggests that individual rights can be thought of as “trumps” which override the non-central interests of others. Public policy issues that do not violate rights can be appropriately decided on utilitarian grounds and properly belong to legislative bodies. However, when policies threaten central interests, courts are called upon to protect individual rights. The judiciary functions to determine if rights are being violated and, if so, to enforce those rights by overruling, or “trumping,” utilitarian policy.

Critics raise two major challenges to deontological approaches. Many charge that deontologists are unable to provide a meaningful distinction between central and non-central interests. Lacking this, public policy is unable to distinguish between rights and mere preferences. From this perspective, the language of rights functions as a smoke-screen raised whenever an individual’s desires are frustrated by public policy. The inability to distinguish central from non–central interests has given rise to a proliferation of rights claims that often obstructs effective public policy.

Critics might cite the NIMBY (not in my back yard) phenomenon as a case in point. A small minority is sometimes able to thwart the common good by claiming that their rights are being violated when, in fact, this minority simply does not want to bear its share of the burden. The cessation of nuclear power plant construction within the United States might provide an example of this. By claiming that nuclear plants threaten such rights as health and safety, opponents to nuclear power have been able to block further construction. If, however, there is little evidence of any actual harm caused by nuclear plants, these opponents may have succeeded in obstructing a beneficial public policy by disguising their preferences as rights. A similar claim could be made concerning debates about locating such locally undesirable but socially beneficial projects as oil refineries and electric generating plants.

A second challenge argues that deontologists are unable to provide a determinate procedure for deciding between conflicting rights claims. Even if we can distinguish rights from mere preferences, effective policy needs a procedure for resolving conflicts. Returning to the analogy of trump cards, deontologists are challenged to distinguish between the ace and deuce of trumps.

Consider, for example, one possible scenario that could follow from the Kyoto Protocol on carbon reduction. Developing countries claim, as a matter of right, that the United States should bear a greater responsibility to reduce carbon emissions. Failing to do so would violate their rights to equal opportunity and fairness. One means by which the United States could meet the Kyoto targets would involve significantly scaling back its energy-intensive agriculture and military sectors. However, because the United States is a major exporter of food products and because its military protects many democracies throughout the world, these options might well threaten the basic rights to food, health, and security for many people in the developing world. In turn, that could be avoided if the United States scaled back its industrial rather than agricultural or military sectors. But this would threaten the freedom, property rights, and economic security of many U.S. citizens. Deontologists are challenged to provide a decision procedure for resolving conflicts between such rights as equal opportunity, fairness, food, health, security, property, and freedom. According to critics, no plausible procedure is forthcoming from the deontological perspective.

SOCIAL ETHICS AND ENERGY POLICY: FUTURE GENERATIONS

Many energy polices also raise important ethical questions concerning justice across generations. What, if any, responsibilities does the present generation have to posterity? This question can be raised at many points as we consider alternative energy policies.

Fossil fuels are a nonrenewable resource. Whatever fossil fuel we use in the present will be forever lost to posterity. Is this fair? The harmful effects of global warming are unlikely to occur for many years. Should we care? Is it ethical to take risks with the welfare of future generations? Nuclear wastes will remain deadly for thousands of generations. Does this require us to change our behavior now? Do we have a responsibility to invest in alternative energy sources now, even if the benefits of this investment go only to people not yet born? Given the energy demands made by increasing populations, what is an ethically responsible population policy?

In many ways, debates surrounding our ethical responsibility to future generations parallel the debates described previously. Utilitarians are concerned with the consequences that various policies might have for the distant future. Committed to the *overall* good, utilitarians must factor the well-being of future people into their calculations. Some argue that future people must count equally to present people. Others, borrowing from the economic practice of discounting present value of future payments, argue that the interests of present people count for more than the interests of future people. Counting future people as equals threatens to prevent any realistic calculation from being made since the further into the future one calculates, the less one knows about the consequences. Discounting the interests of future people, on the other hand, threatens to ignore these interests since, eventually, any discount rate will mean that people living in the near future count for nothing.

In contrast to this utilitarian approach, some argue that future people have rights that entail duties on our part. For example, in 1987 the UN-sponsored Brundtland Commission advocated a vision of sustainable development as development "which meets the needs of the present without sacrificing the ability of the future to meets its needs." This suggests that future people have an equal right to the energy necessary to meet their needs. However, if future generations have a right to energy resources in an equal amount to what is available to us, we would be prohibited from using any resources, because whatever we use today is denied forever to the future. On the other hand, if future generations have a right to use energy resources at some point in the future, why do we not have an equal right to use them today?

As can be seen from these examples, even talking about ethical responsibilities to future people can raise conundrums. Some critics claim that talk of ethical responsibilities to distant people is nonsense and that present energy policy should be governed solely by a concern for people living in the present and immediate future.

Two challenges are raised against claims that present generations have ethical responsibilities to future generations. The first is called the problem of "disappearing beneficiaries." Consider the claim that present generations ought to decrease our reliance on fossil fuels to ensure a future world protected from the harmful effects of global warming. The defense of this view argues that future people would be made better-off by this decision. But we need to ask "better-off" than what? Intuitively, we would say that they will be better-off than they would have been otherwise. However, this argument assumes that the people who benefit will be the same people as those who would exist if we adopted the alternative policy of continued reliance on fossil fuels. Yet alternative policy decisions as momentous as international energy policy, population controls, or significant conservation measures, would surely result in different future people being born. When we consider alternative policy decisions, we actually are considering the effects on two, not one, sets of future people. The future population that would be harmed by our decision to continue heavy use of fossil fuels is a different population than the one that would benefit from major conservation programs. Thus, it makes little sense to speak about one future generation being made better or worse-off by either decision. The potential beneficiaries of one policy disappear when we choose the alternative policy.

The second challenge is called the argument from ignorance. Any discussion of future people and their happiness, their needs and preferences, their rights and interests, forces us to make assumptions about who those people might be and what they will be like. But realistically, we know nothing about who they will be, what they will want or need, or even *if* they will exist at all. Since we are ignorant of future people, we have little basis to speak about our responsibilities to them.

The implication of these arguments is that energy policy ought to be set with due consideration given only to present generations. While we might have responsibilities which *regard* future people (present duties to our children affect the life prospects of future generations), we have no direct responsibilities *to* future people. We can have no responsibility to that which does not exist.

Plausible answers can be offered to these challenges. While we may not know who future people will be, if they will be, or what their specific interests and needs might be, we do know enough about future people to establish present ethical responsibilities to them. Just as in cases of legal negligence where we hold individuals liable for unintended but foreseeable and avoidable harms that occur in the future, it can be meaningful to talk about foreseeable but unspecific harms to unknown future people. Surely we have good reasons for thinking that there will be people living in the future, and that they will have needs and interests similar to ours. To the degree that we can reasonably foresee present actions causing predictable harms to future people, we can acknowledge the meaningfulness of ethical responsibilities to future generations. While the present may not have specific responsibilities to identifiable future people, it does make sense to say that we have responsibilities to future people, no matter who they turn out to be.

What might our responsibilities to the future be? How do we balance our responsibilities to future people against the interests, needs, and rights of the present? Perhaps the most reasonable answer begins with the recognition that we value energy not in itself but only as a means to other ends. Thus, the requirement of equal treatment demanded by social justice need not require that future people have an equal right to present energy resources but, rather, that they have a right to an equal opportunity to the ends attained by those resources. Our use of fossil fuels, for example, denies to them the opportunity to use that fuel. We cannot compensate them by returning that lost energy to them, but we can compensate them by providing the future with an equal opportunity to achieve the ends provided by those energy resources. Justice requires that we compensate them for the lost opportunities and increased risks that our present decisions create.

While there are practical difficulties in trying to specify the precise responsibilities entailed by this approach, we can suggest several general obligations of compensatory justice. First, our responsibility to the future should include a serious effort to develop alternative energy sources. Continued heavy reliance on fossil fuels and nuclear power places future people at risk. Justice demands that we minimize that risk, and investment in alternative energy sources would be a good faith step in that direction. Arguments of this sort could justify government expenditures on research into fusion and renewable energy sources. Second, we have a responsibility to conserve nonrenewable resources. Wasting resources that future people will need, especially when we presently have the technology to significantly increase energy efficiency, makes it more difficult for future people to obtain a lifestyle equal to ours. Finally, it would seem we have a responsibility to adopt population policies and modify consumption patterns to moderate worldwide energy demand over the long term.

Joseph R. DesJardins

See also: Conservation of Energy; Culture and Energy Usage; Government and the Energy Marketplace.

BIBLIOGRAPHY

Brundtland, G. (1987). *Our Common Future.* New York: Oxford University Press.

Bullard, R. (1993). *Confronting Environmental Racism.* Boston: South End Press.

Bullard, R. (1994). *Dumping in Dixie.* Boulder, CO: Westview Press.

Daly, H., and Cobb, J. (1989). *For the Common Good.* Boston: Beacon Press.

DesJardins, J. (1997). *Environmental Ethics: An Introduction to Environmental Philosophy.* Belmont, CA: Wadsworth Publishing

DesJardins, J. (1999). *Environmental Ethics: Concepts, Policy, Theory.* Mountain View, CA: Mayfield Publishing.

Dworkin, R. (1977) *Taking Rights Seriously.* Cambridge, MA: Harvard University Press.

MacLean, D., and Brown, P., eds. (1983). *Energy and the Future.* Lanham, MD: Rowman and Littlefield.

Partridge, E., ed. (1980). *Responsibilities to Future Generations.* Buffalo: Prometheus Press.

Rachels, J. (1999). *The Elements of Moral Philosophy.* New York: McGraw-Hill.

Schneider, S. (1989). *Global Warming: Are We Entering the Greenhouse Century?* San Francisco: Sierra Club Books.

Shrader-Frechette, K. S. (1989). *Nuclear Power and Public Policy.* Dordrecht, Holland: Reidel.

Shrader-Frechette, K. S. (1991). *Risk and Rationality.* Berkeley: University of California Press.

Simon, J. (1981). *The Ultimate Resource.* Princeton: Princeton University Press.

Sikora, R. I., and Barry, B., eds. (1978). *Obligations to Future Generations.* Philadelphia: Temple University Press.

Wenz, P. (1988). *Environmental Justice.* Albany: State University of New York Press.

Westra, L., and Wenz, P., eds. (1995). *Faces of Environmental Racism.* Lanham, MD: Rowman and Littlefield Publishers.

EXPLOSIVES AND PROPELLANTS

Explosions occur when gases in a confined space expand with a pressure and velocity that cause stresses greater than the confining structure can withstand. The gas expansion can be caused by rapid generation of gaseous molecules from a solid or liquid (e.g., an explosive) and/or rapid heating (as in a steam "explosion"). An explosion can be low-level, yet still dangerous, as in the deflagration (rapid burning) of a flour dust and air mixture in a grain elevator, or very intensive, as in the detonation of a vial of nitroglycerin.

Explosives and propellants are mixtures of fuel and oxidizer. The intensity of combustion is determined by the heat of combustion per pound of material, the material's density, the gas volume generated per volume of material, and the rate of deflagration or detonation. The latter, the most important variable, is determined by the speed at which fuel and oxidizer molecules combine.

The first explosive was black gunpowder, invented by the Chinese in the Middle Ages. In gunpowder, the fuel is powdered sulfur and charcoal, and the oxidizer is saltpeter (potassium nitrate). When heated, the oxidizer molecule decomposes to form potassium oxide (a solid), nitrogen and nitrous oxides (gases), and excess pure oxygen that burns the fuel to form more gases (carbon oxides, sulfur oxides). The gases generated by this rapidly burning mixture can explode (rupture its container), as in a firecracker, or propel a projectile, as in a rocket or a gun. Because it takes time for oxygen to diffuse to the fuel molecules the explosion of gunpowder is a rapid burning, or deflagration, not the high-rate detonation characteristic of a high explosive.

While propellants are formulated to burn rapidly and in a controlled manner, they can go from deflagration to detonation if mishandled. High explosives, on the other hand, are designed to detonate when activated. Here oxidizer and fuel are always situated in the same molecule, and in the right proportions, as determined by the desired end-products. Once initiated, gases are formed too fast to diffuse away in an orderly manner, and a shock wave is generated that passes through the explosive, detonating it almost instantaneously. This shock wave and the resultant

Explosive Substances

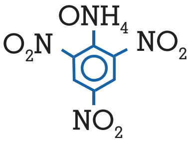
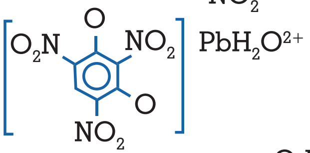
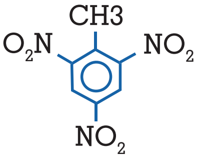
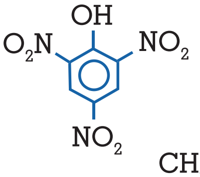
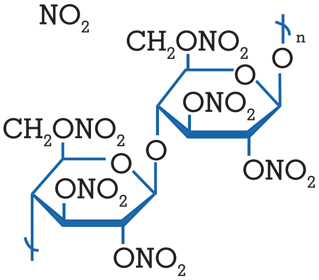
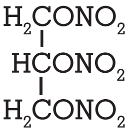
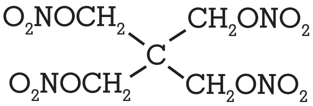
Name	*Formula*	*Use*
ammonium nitrate	NH_4NO_3	solid oxidizer
ammonium perchlorate	NH_4ClO_4	solid oxidizer
lead azide	$Pb(N_3)_2$	primary explosive
ammonium picrate		secondary high explosive
lead styphnate		primary explosive
2,4,6-trinitrotoluene		secondary high explosive
picric acid		secondary high explosive
nitrocellulose		secondary explosive used in propellants
nitroglycerin	 $\begin{array}{c} H_2CONO_2 \\ \vert \\ HCONO_2 \\ \vert \\ H_2CONO_2 \end{array}$	liquid secondary explosive ingredient in commercial explosives and propellents
nitromethane	CH_3NO_2	liquid secondary explosive
pentaerythritol tetranitrate	 $C(CH_2ONO_2)_4$	secondary high explosive used as booster

Table 1.
Explosive substances.

high-velocity expansion of gases can cause great damage, even if there is no confining container to rupture.

A propellant typically burns at the rate of about 1 cm/sec, some 10,000 times the rate coal burns in air. However high explosives "burn" at a rate some 100,000 to 1,000,000 times faster than propellants because the reaction rate is controlled by shock transfer rather than heat transfer. Thus, although there is more energy released in burning a pound of coal than a pound of dynamite, the power output, which is the rate of doing work, can be 100 times larger for a propellant, and 100 million times larger for a high explosive.

Black gunpowder, was the only explosive (actually a propellant) known until 1847, when the Italian chemist, Ascani Sobrero, discovered nitroglycerin. This rather unstable and shock-sensitive liquid was the first high explosive. It was too dangerous to use until Alfred Nobel developed dynamite in 1866, a 50 percent mixture of nitroglycerin stabilized by absorption in inert diatomaceous earth. The much safer trinitrotoluene (TNT) was synthesized soon after, and military needs in the two World Wars led to a number of new high explosives. Almost all of these are made by nitration of organic substrates, and they are formulated to be simultaneously safer and more energetic (see Table 1). For commercial uses, low cost is of paramount importance, and it has been found that a mixture of ammonium nitrate, a fertilizer ingredient, and fuel oil (ANFO) gives the most "bang for the buck."

By designing an explosive charge to focus its blast on a small area, a so-called shaped or armor-piercing charge, gas velocities as high as 20,000 mph and at pressures of 3 million psi can be achieved. Such a force pushes steel aside through plastic flow, much as a knife cuts through butter. Under the high pressure, the metallic steel behaves like a viscous liquid, the same way plastics flow when extruded through dies to make various shapes.

The thrust of recent explosives research continues to emphasize increased safety and control. These properties are achieved by the use of primary and secondary explosives. The secondary explosive is the main ingredient in a charge, and it is formulated to be stable in storage and difficult to initiate. Some secondary explosives just burn slowly without detonating if accidentally ignited. The initiator or primary explosive is a small quantity of a more sensitive material fashioned into a detonator or blasting cap. It is attached to the main charge just before use and is activated by a fuse, by percussion (as in a gun), or by an electrical current.

The key to safety in explosives manufacturing is to use isolated high-velocity nitric acid reactors that have only a very small hold up at any one time (that is, only a small amount of dangerous material is "held up" inside the reactor at any time). Units are widely spaced, so any accident involves only small amounts of explosive and does not propagate through the plant. Fire and electrical spark hazards are rigorously controlled, and manpower reduced to the absolute minimum through automation.

The recent rise in the use of expolosives in terrorist activity poses new challenges to industry and law enforcement. This challenge is being met by the use of sophisticated chemical detection devices to screen for bombs and more rigorous explosive inventory safeguards and controls. Plans have also been proposed to tag explosives with isotopes to make them easier to trace if misused.

Herman Bieber

See also: Nuclear Energy; Nuclear Fission; Nuclear Fusion.

BIBLIOGRAPHY

Harding, G., ed. (1993). *Substance Detection Systems, Vol. 2092.* Bellingham, WA: Proceedings of the International Society for Optical Engineering.

Squire, D. R. (1991). *Chemistry of Energetic Materials.* San Diego, CA: Academic Press.

Urbanski, T. (1985). *Chemistry and Technology of Explosives* New York: Pergamon Press.

U.S. Bureau of Alcohol, Tobacco and Firearms. (1986). *Firearms and Explosives Tracing.* Washington, DC: Author.

U.S. Senate, Committee on the Judiciary. (1993). *Terrorism in America.* Washington, DC: U.S. Government Printing Office.

EXTERNAL COMBUSTION ENGINES

See: Engines

F

FARADAY, MICHAEL (1791–1867)

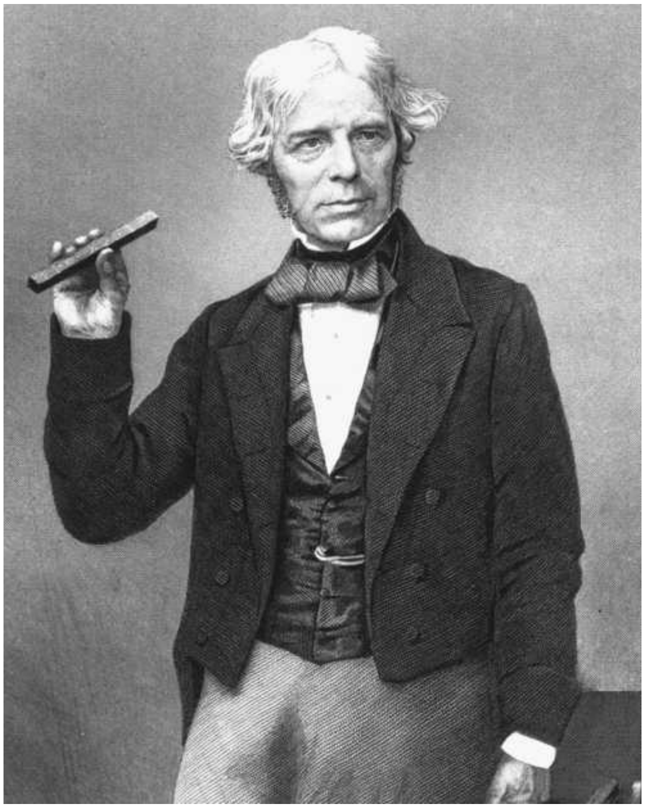
Michael Faraday has been called the "patron saint of electrical engineering." He produced the first electric motor and the first electric generator and is considered the greatest experimental scientist of the nineteenth century. Faraday came from humble beginnings. He was born in a village that is now part of London, and his father was a migrant blacksmith who was often ill and unable to support his family. Faraday often went hungry as a child and his only formal education was at a Sandemanian Church Sunday school. (The Sandemanians were a small fundamentalist Christian sect, and Faraday later became an elder of the church.) At age thirteen, he was apprenticed to a bookbinder for seven years. In addition to binding the books, he read them voraciously. Although he completed the apprenticeship, he subsequently sought a way out of a trade that he considered selfish and vicious.

Faraday's great opportunity came when a friend offered him a ticket to attend the lectures on chemistry given by Sir Humphrey Day, the director of the Royal Institution in London. After attending the lectures, Faraday sent Davy a neatly bound copy of his notes and asked for employment. In 1812 Davy did require a new assistant and, remembering the notes, hired Faraday. Davy was a leading scientist of his time and discovered several chemical elements, but it has been said that Faraday was his greatest discovery. Faraday was given quarters at the Royal Institution where he was to remain for forty-five years (staying on even after his marriage). Davy and his wife took Faraday with them as secretary to Europe on a grand tour in 1813. Despite the hostilities between France and England, they received Napoleon's permission to meet with French scientists in Paris. During this time Faraday's talent began to be recognized internationally.

In 1820 Faraday finished his apprenticeship under Davy and in the following year married and settled into the Royal Institution. Faraday's early reputation as a chemist was so great that in 1824 he was elected to the Royal Society. In 1825 Davy recommended that Faraday succeed him as director of the Royal Institution. The appointment paid only a hundred pounds a year, but Faraday soon received some adjunct academic appointments that enabled him to give up all other professional work and devote himself full-time to research. Faraday's scientific output was enormous, and at the end of his career, his laboratory notebooks, which covered most of his years at the Royal Institution, contained more than sixteen thousand neatly inscribed entries, bound in volumes by Faraday himself.

In his early work, Faraday was primarily a chemist. He liquefied several gases previously considered incapable of liquefaction, discovered benzene, prepared new compounds of carbon and chlorine, worked on new alloys of steel, and discovered the laws of electrolysis that bear his name. The latter discovery became the basis for the electroplating industry that developed in England during the early nineteenth century. In 1821, however, Hans Christian Oersted's discovery that an electric current could produce a magnetic field led Faraday away from chemistry for a while. With a better understanding of electric and magnetic fields, Faraday succeeded in building the first elementary electric motor.

Faraday returned to chemistry, but after 1830, his investigations again concentrated on electric and magnetic phenomena. He had become convinced that the reverse action to the phenomenon discovered by Oersted was also possible, that a magnetic



Michael Faraday. (Corbis Corporation)

field could produce an electric current. He also believed that he could induce a current in a circuit using an electromagnet like the one invented by Sturgeon in 1824. Wrapping two wires many times around opposite sides of an iron ring, Faraday discovered that when an electric current in one wire was turned on or off a current appeared in the other wire. Thus Faraday discovered the law of electromagnetic induction that bears his name: The electromotive force (voltage) induced in a circuit is equal to the (negative) time rate-of-change of the magnetic flux through the circuit. This law was not only the basis for the first elementary electric generator that Faraday produced, but also for all subsequent electric power dynamos that employ coils rotating in a magnetic field to produce electric power. Although Faraday is credited with the discovery because he was the first to publish his work, it was later learned that induction had been discovered shortly before by Joseph Henry, then an instructor in an obscure school in Albany, New York.

In the decade that followed this discovery, Faraday continued to make fundamental discoveries about electricity and magnetism. He showed that the electricity obtained from various sources was the same, analyzed the effects of dielectrics on electrostatics, and studied electric discharges in gases. At the end of 1839, however, his health broke down. There is reason to believe that Faraday may have suffered from mercury poisoning, a common affliction in that period. He did not return to work completely until 1845, when he discovered the phenomena of magnetically induced birefringence in glass and of diamagnetism. In 1846 he published a paper in which he suggested that space was a medium that bore electric and magnetic strains and that these strains were associated with the propagation of light. Later scientists recognized that these ideas were the forerunners of the modern theory of electromagnetic propagation and optical fields.

Faraday worked alone; he had no students, just ordinary assistants. Although devoted to laboratory work, he was also a brilliant public lecturer. Personally, he was invariably described as a gentle and modest person. He never forgot his humble beginnings, and he had a clear view of his own worth and a disdain of the class system. In 1858, when Queen Victoria, in view of his lifetime of great achievement, offered him a knighthood and the use of a house, Faraday accepted the cottage but refused the knighthood, stating that he preferred to remain "plain Mr. Faraday."

Faraday's activity slowed after 1850, and, in 1865, a progressive loss of memory forced his complete retirement. He died in 1867 and he was buried, not in Westminster Abbey, but perhaps more befitting his egalitarian ideals, in Highgate Cemetery, London.

Leonard S. Taylor

See also: Electric Motor Systems; Electric Power, Generation of; Magnetism and Magnets; Oersted, Hans Christian.

BIBLIOGRAPHY

Atherton, W. A. (1984). *From Compass to Computer.* San Francisco: San Francisco Press.

Segrè, E. (1984). *From Falling Bodies to Radio Waves.* New York: W. H. Freeman.

Meyer, H. W. (1971). *A History of Electricity and Magnetism.* Cambridge, MA: MIT Press.

FEDERAL ENERGY REGULATORY COMMISSION

See: Government Agencies

FERMI, ENRICO

(1901–1954)

Enrico Fermi was both a brilliant theorist and an unusually gifted experimentalist — a combination of talents seldom found among twentieth-century physicists. Born in 1901 in Rome, Fermi obtained his doctor's degree in physics magna cum laude from the University of Pisa at the age of 21, with a dissertation on x-rays.

After two years of post-doctoral research at Max Born's Institute in Göttingen, and then with Paul Ehrenfest in Leiden, Fermi taught for two years at the University of Florence, where he soon established his reputation by developing what are now known as the Fermi-Dirac statistics. In 1926 he was appointed to a full professorship in physics at the University of Rome, where he quickly gathered around him a group of talented young faculty members and students, who helped him make a name for Rome in the fields of nuclear physics and quantum mechanics. His theoretical work culminated in a 1933 theory of nuclear beta decay that caused a great stir in world physics circles, and is still of major importance today.

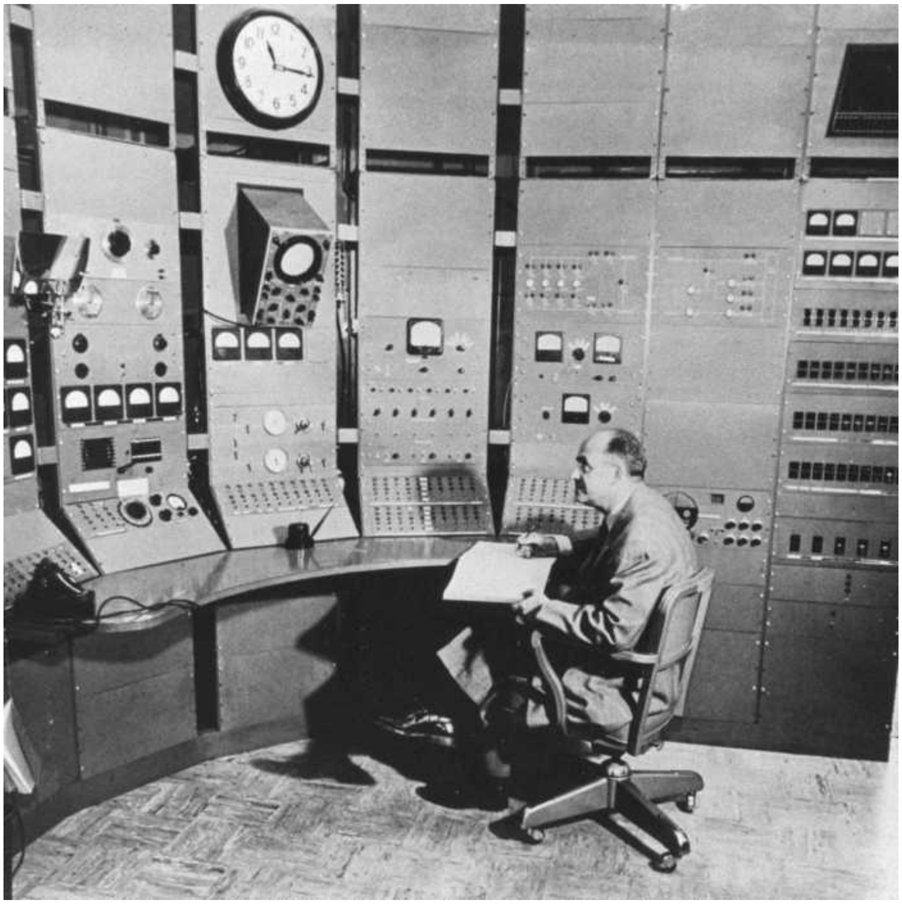
Fermi had been fascinated by the discovery of the neutron by James Chadwick in 1932. He gradually switched his research interests to the use of neutrons to produce new types of nuclear reactions, in the hope of discovering new chemical elements or new isotopes of known elements. He had seen at once that the uncharged neutron would not be repelled by the positively-charged atomic nucleus. For that reason the uncharged neutron could penetrate much closer to a nucleus without the need for high-energy particle accelerators. He discovered that slow neutrons could be produced by passing a neutron beam through water or paraffin, since the neutron mass was almost equal to that of a hydrogen atom, and the consequent large energy loss in collisions with hydrogen slowed the neutrons down very quickly. Hence these "slow" or "thermal" neutrons would stay near a nucleus a longer fraction of a second and would therefore be more easily absorbed by the nucleus under investigation. Using this technique, Fermi discovered forty new artificial radioactive isotopes.

In 1934 Fermi decided to bombard uranium with neutrons in an attempt to produce "transuranic" elements, that is, elements beyond uranium, which is number 92 in the periodic table. He thought for a while that he had succeeded, since unstable atoms were produced that did not seem to correspond to any known radioactive isotope. He was wrong in this conjecture, but the research itself would eventually turn out to be of momentous importance both for physics and for world history, and worthy of the 1938 Nobel Prize in Physics.

Fermi's wife, Laura, was Jewish, and as Hitler's influence over Mussolini intensified, anti-Jewish laws were passed that made Laura's remaining in Italy precarious. After accepting his Nobel Prize in Stockholm, Fermi and his wife took a ship directly to the United States, where they would spend the rest of their lives. Enrico taught at Columbia University in New York City from 1939 to 1942, and at the University of Chicago from 1942 until his death in 1954.

In 1938 Niels Bohr had brought the astounding news from Europe that the radiochemists Otto Hahn and Fritz Strassmann in Berlin had conclusively demonstrated that one of the products of the bombardment of uranium by neutrons was barium, with atomic number 56, in the middle of the periodic table of elements. He also announced that in Stockholm Lise Meitner and her nephew Otto Frisch had proposed a theory to explain what they called "nuclear fission," the splitting of a uranium nucleus under neutron bombardment into two pieces, each with a mass roughly equal to half the mass of the uranium nucleus. The products of Fermi's neutron bombardment of uranium back in Rome had therefore not been transuranic elements, but radioactive isotopes of known elements from the middle of the periodic table.

Fermi and another European refugee, Leo Szilard, discussed the impact nuclear fission would have on physics and on the very unstable state of the world



Enrico Fermi. (Library of Congress)

itself in 1938. The efforts of Szilard in 1939 persuaded Albert Einstein to send his famous letter to President Franklin D. Roosevelt, which resulted in the creation of the Manhattan Project to contruct a nuclear bomb. Fermi was put in charge of the first attempt to construct a self-sustaining chain reaction, in which neutrons emitted by a fissioning nucleus would, in turn, produce one or more fission reactions in other uranium nuclei. The number of fissions produced, if controlled, might lead to a useful new source of energy; if uncontrolled, the result might be a nuclear bomb of incredible destructive power.

Fermi began to assemble a "nuclear pile" in a squash court under the football stands at the University of Chicago. This was really the first nuclear power reactor, in which a controlled, self-sustaining series of fission processes occurred. The controls consisted of cadmium rods inserted to absorb neutrons and keep the reactor from going "critical." Gradually the rods were pulled out one by one, until the multiplication ratio of neutrons produced to neutrons absorbed was exactly one. Then the chain reaction was self-sustaining. To proceed further would run the risk of a major explosion. Fermi had the reactor shut down at exactly 3:45 P.M. on December 2, 1942, the day that is known in history as the beginning of nuclear energy and nuclear bombs.

Fermi lived only a little more than a decade after his hour of triumph. He spent most of this time at the University of Chicago, where, as in Rome, he surrounded himself with a group of outstanding graduate students, many of whom also later received Nobel Prizes. Fermi died of stomach cancer in 1954, but his name remains attached to many of the important contributions he made to physics. For example, element 100 is now called Fermium.

Fermi's overall impact on physics is well summarized by the nuclear physicist Otto Frisch (1979,

p. 22): “But occasionally one gets a man like Enrico Fermi, the Italian genius who rose to fame in 1927 as a theoretician and then surprised us all by the breathtaking results of his experiments with neutrons and finally by engineering the first nuclear reactor. On December the second, 1942, he started the first self-sustaining nuclear chain reaction initiated by man and thus became the Prometheus of the Atomic Age.”

Joseph F. Mulligan

BIBLIOGRAPHY

Allison, S. K., et al. (1955). “Memorial Symposium in Honor of Enrico Fermi held on April 29, 1955 at meeting of the American Physical Society in Washington, D.C.” *Reviews of Modern Physics* 27:249–275.

Fermi, E. (1962–1965). *Collected Papers*, 3 vols., ed. E. Segrè, et al. Chicago: University of Chicago Press.

Fermi, L. (1954). *Atoms in the Family: My Life with Enrico Fermi.* Chicago: University of Chicago Press.

Frisch, O. (1979). *What Little I Remember.* Cambridge: Cambridge University Press.

Segrè, E. (1970). *Enrico Fermi, Physicist.* Chicago: University of Chicago Press.

Segrè, E. (1971). “Fermi, Enrico.” In *Dictionary of Scientific Biography,* ed. Charles Coulston Gillispie, Vol. 4, pp. 576–583. New York: Scribner.

Smyth, H. D. (1945). *Atomic Energy for Military Purposes.* Washington, DC: U.S. Government Printing Office.

Wattenberg, A. (1993). “The Birth of the Nuclear Age.” *Physics Today* 46(January):44–51.

FERMI NATIONAL ACCELERATOR LABORATORY

See: Government Agencies

FIELD ENERGY

See: Waves

FISSION, NUCLEAR

See: Nuclear Fission

FISSION BOMB

See: Nuclear Fission

FLYWHEELS

Flywheels store kinetic energy (energy of motion) by mechanically confining motion of a mass to a circular trajectory. The functional elements of the flywheel are the mass storing the energy, the mechanism supporting the rotating assembly, and the means through which energy is deposited in the flywheel or retrieved from it.

Energy can be stored in rings, disks, or discrete weights, with spokes or hubs connecting the storage elements to shafts, and bearings supporting the assembly and allowing it to rotate. Energy may be transferred into or out of the wheel mechanically, hydraulically, aerodynamically, or electrically.

HISTORY

Ubiquitous in rotating machinery, flywheels have been used as a component of manufacturing equipment since their application in potters’ wheels before 2000 B.C.E.

Flywheels attained broad use during the Industrial Revolution. In the embodiment of this era, flywheels used heavy rims built from cast iron to damp pulsations in engines, punches, shears, and presses. Often the pulsations to be damped arose from reciprocating motive forces or reciprocating end processes. The conversion of reciprocation into rotation enabled formatting of the flow of this energy. The most important types of formatting were transportation of energy by shafts, conversion of torque and speed by gears, and damping by flywheels.

A 23-ton flywheel in the Palace of Engineering at the 1924 British Empire Exhibition. (Corbis Corporation)

Flywheels are found in internal-combustion engines, where they damp out torque pulses caused by the periodic firing of cylinders. In this application, energy is stored very briefly before it is used— for less than one revolution of the wheel itself.

The evolution of flywheel materials and components and a systemic approach to design have led to the development of stand-alone flywheel energy storage systems. In these systems the rotating element of the flywheel transfers energy to the application electrically and is not directly connected to the load through shafting. These systems typically store energy that is released over many revolutions of the wheel, and as a system may be used in place of electrochemical energy storage in many applications. By being separate and distinct from the process it supports, the stand-alone flywheel system may use materials and components optimally. Of the various flywheel types, stand-alone systems will typically have the highest energy and power density as well as the highest rim speed and rotation rate.

TECHNOLOGY

The kinetic energy stored in the flywheel rotor is proportional to the mass of the rotor and the square of its linear velocity. Transformed into a cylindrical system, the stored kinetic energy, KE, is

$$KE = \tfrac{1}{2} J \omega^2 = \tfrac{1}{2} mr^2\omega^2$$

where ω is the rate of rotation in radians per second and J is the moment of inertia about the axis of rotation in kilogram-meter^2. For the special case of a radially thin ring, the moment of inertia is equal to its mass, m, multiplied by the square of its radius, r. This radius is also known as the radius of gyration.

Stress in the rim is proportional to the square of linear velocity at the tip. When rotor speed is dictated

Applications	*Types*	*Unique Attributes*
Stationary engines (historic), damp pulsations	Spoked hub, steel rim	Massive, low-speed rotors, belt or shaft mechanical connection to application
Automobile engine, damp out torque for	Solid metal rotor	Mounted on engine shaft, very low cost
Satellite stabilization and energy storage	Control moment gyro/reaction wheel	Lightweight, extremely long life, and high reliability
Stationary UPS system	Steel or composite rotor in vacuum	Electrical connection to application; high power density (relatively high power generator)
Energy storage for hybrid propulsion	Composite rotor in vacuum	Electrical connection to application; high energy density (lightweight rotor)

Table 1.
Flywheel Applications, Types, and Unique Attributes

by material considerations, the linear velocity of the tip is set, and rotation rate becomes a dependent parameter inversely proportional to rotor diameter. For example, a rotor with a tip speed of 1,000 meters per second and a diameter of 0.3 meter would have a rotation rate of about 63,700 rpm. If the diameter were made 0.6 meters instead, for this material the rotation rate would be about 31,850 rpm.

To maximize stored energy, the designer seeks to spin the rotor at the highest speed allowed by the strength of the materials used. There is a trade-off between heavier, lower-strength materials and stronger, lighter materials. For a thin rim, the relationship between rim stress and specific energy or energy stored per unit mass of rim is given by

$$KE/m = \sigma_h/2\rho$$

where σ_h is the hoop stress experienced by the ring in newtons per square meter, and ρ is the density of the ring material in kilograms per meter^3. Thus high specific energy corresponds directly to high specific strength: $\sigma sh/\rho$ and rotors made from carbon composite may be expected to store more energy per unit weight than metal rotors.

Since energy is proportional to the square of speed, high performance will be attained at high tip speed. Rims produced from carbon fiber have attained top speeds in excess of 1,400 meters per second and must be housed in an evacuated chamber to avoid severe aerodynamic heating.

The flywheel rim is connected to a shaft by spokes or a hub. The rotor experiences high centrifugal force and will tend to grow in size while the shaft will not. The spoke or hub assembly must span the gap between the shaft and the rim, allowing this differential growth while supporting the rim securely. High-speed composite rims may change dimension by more than 1 percent in normal operation. This large strain or relative growth makes hub design especially challenging for composite flywheels.

Bearings support the shaft and allow the flywheel assembly to rotate freely in applications where the flywheel is a component in a more extensive rotating machine, the rim and hub are supported by the shafting of the machine, and no dedicated bearings are used. In stand-alone systems, flywheel rotors are typically supported with hydrodynamic or rolling element bearings, although magnetic bearings are sometimes used either to support part of the weight of the rotor or to levitate it entirely.

Compact, high-rim-speed, stand-alone flywheel systems may require that the bearings run continuously for years at many tens of thousands of revolutions per minute. Smaller rotors will operate at higher rotation rates.

Historically, flywheels have stored and discharged energy through direct mechanical connection to the load. The flywheel may be affixed to the load or may communicate with the load through gears, belts, or shafts. Stand-alone flywheel systems may convert electrical energy to kinetic energy through a motor, or convert kinetic energy to electricity through a generator. In these systems the flow of energy into and out of the flywheel may be regulated electronically using active inverters controlled by microprocessors or digital signal processors.

FUTURE ADVANCES

The flywheel has become an integral energy storage element in a broad range of applications. New designs and innovations in current designs and diverse opportunities for bettering flywheel performance continue to emerge.

The modern stand-alone flywheel embodies a number of sophisticated technologies, and advances in these technologies will yield further improvement to this class of flywheel. Flywheel progress will track development in power electronics, bearings, and composite materials. Power electronics such as active inverters and digital signal processors will continue to become power dense, reliable, and inexpensive. Rolling element and magnetic bearings will mature to a point where decades-long operating life is considered routine.

The energy stored in a flywheel depends on the strength of the rotor material. Carbon fiber tensile strength remains well below theoretical limits. Expected increases in strength along with reduction in cost as the use of this material expands will translate into more energy dense, less expensive rotors.

Progress, which is likely to be incremental, is certain to improve the performance and energy efficiency of all applications.

Donald A. Bender

See also: Bearings; Conservation of Energy; Heat and Heating; Kinetic Energy; Materials; Potential Energy; Storage; Storage Technology; Tribology.

BIBLIOGRAPHY

Genta, G. (1985). *Kinetic Energy Storage*. London: Butterworths.

Post, R. F.; Fowler, T. K.; and Post, S. F. (1993). “A High-Efficiency Electromechanical Battery.” *Proceedings of the IEEE* 81(3).

Post, R. F.; and Post, S. F. (1973). “Flywheels.” *Scientific American* 229(Dec.):17.

U.S. Department of Energy. (1995). *Flywheel Energy Storage Workshop*. Springfield, VA: U.S. Department of Commerce.

FOSSIL FUELS

Fossil fuel is a general term for any hydrocarbon or carbonaceous rock that may be used for fuel: chiefly petroleum, natural gas, and coal. These energy sources are considered to be the lifeblood of the world economy. Nearly all fossil fuels are derived from organic matter, commonly buried plant or animal fossil remains, although a small amount of natural gas is inorganic in origin. Organic matter that has long been deeply buried is converted by increasing heat and pressure from peat into coal or from kerogen to petroleum (oil) or natural gas or liquids associated with natural gas (called natural gas liquids). Considerable time, commonly millions of years, is required to generate fossil fuels, and although there continues to be generation of coal, oil and natural gas today, they are being consumed at much greater rates than they are being generated. Fossil fuels are thus considered nonrenewable resources.

This article provides a brief historical perspective on fossil energy, focusing on the past several decades, and discusses significant energy shifts in a complex world of constantly changing energy supply, demand, policies and regulations. United States and world energy supplies are closely intertwined (Figure 1). World supplies of oil, gas and coal, are less extensively developed than those of the United States and the extent of remaining resources is extensively debated. Fossil fuels such as coal, natural gas, crude oil and natural gas liquids currently account for 81 percent of the energy use of the United States, and in 1997 were worth about $108 billion. For equivalency discussions and to allow comparisons among energy commodities, values are given in quadrillion British thermal units (Btus). For an idea of the magnitude of this unit, it is worth knowing that 153 quadrillion Btus of energy contains about a cubic mile of oil. The world consumes the equivalent of about 2.5 cubic miles of oil energy per year, of which 1 cubic mile is oil, 0.6 cubic mile is coal, 0.4 cubic mile is gas, and 0.5 cubic

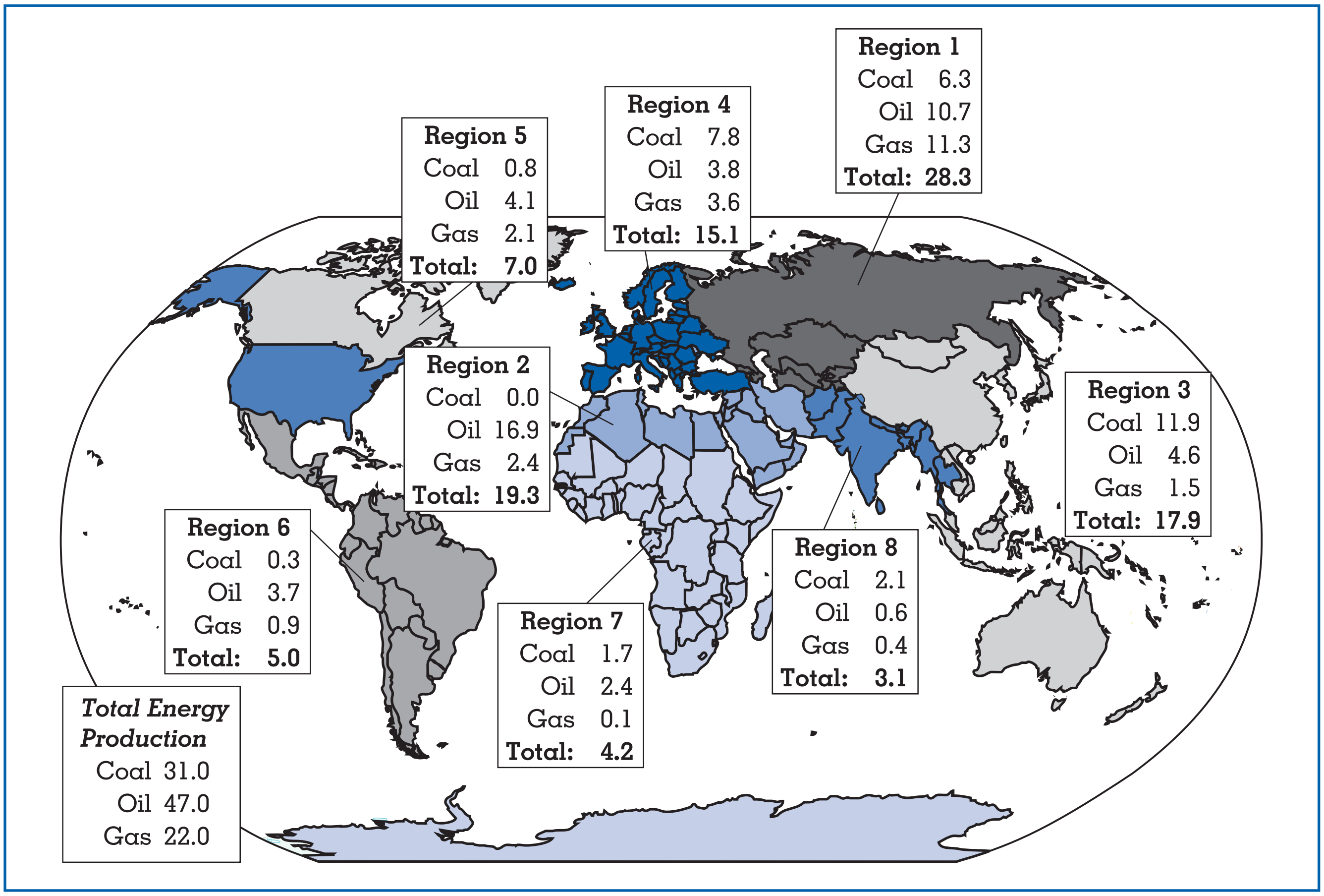


Figure 1.
Percentage comparisons of fossil fuel production for coal, oil, and natural gas for the world exclusive of the United States.

mile is all other forms of energy. Total oil reserves are about 32 cubic miles of oil and total energy reserves are equivalent to about 90 cubic miles of oil. Quadrillion Btus can also be converted to billion barrel oil equivalents (BBOE) at the ratio of about 5:1 (i.e., 5 quadrillion Btu to 1 BBOE).

Both energy consumption and production more than doubled between 1960 and 2000, reflecting the United States's increasing need for energy resources (Figure 1). In 2000 the United States was at its historically highest level of both fossil energy consumption and production. However, significant shifts in domestic energy consumption and production occurred between 1980 and 2000. Production of oil and natural gas in the United States did not meet consumption between 1970 and 2000.

Surprisingly, coal is the largest energy source in the United States and the world (Figure 1), despite perceptions that it has been replaced by other sources. In 1997 production of both coal (23.2 quadrillion Btus, or about 4.6 billion barrels of oil) and natural gas (19.5 quadrillion Btus, or about 3.9 billion barrels of oil) on an energy equivalent basis exceeded U.S. domestic oil production (13.6 quadrillion Btus, equivalent to about 2.7 billion barrels, or 3.1 billion barrels of oil if natural gas liquids are included). Coal production in the United States nearly doubled from 1970 to 2000 (from about 600 million tons to about 1 billion tons produced annually). Meanwhile, petroleum consumption at 18.6 million barrels of oil per day is near the all-time high of 18.8 million barrels of oil per day in 1978. Net U.S. petroleum imports (8.9 million barrels of oil per day) in 1997 were worth $67 billion and exceeded U.S. petroleum production (8.3 million barrels of oil per day). Concerns about petroleum supplies, specifically oil shortages, caused the United States to build a strategic petroleum reserve in 1977 that currently holds 563 million barrels of oil or about sixty-three days worth of net imported petroleum. This strategic supply is about half of the 1985 high when the reserve would have provided 115 days worth of net imported petroleum. Although U.S. oil and gas production generally declined from 1970 to 1993, natural gas production has been increasing since the mid-1980s and energy equivalent production from natural gas exceeded U.S. oil production in the late 1980s. Natural gas is expected to play an increasing role in the United States in response to both environmental concerns and anticipated major contributions from unconventional (or continuous) natural gas sources requiring reductions in carbon emissions, particularly if the Kyoto Protocol is passed by the United States.

Due to factors such as extreme fluctuations in commodity prices, particularly oil prices, wasteful oil and gas field practices, and perceived national needs, U.S. government involvement in energy markets has been part of U.S. history from the turn of the century. The U.S. began importing oil in 1948. The 1973 oil embargo by the Organization of Petroleum Exporting Countries (OPEC) made a strong impact on U.S. policy makers who responded by developing regulations designed to encourage new domestic oil production. A two-tiered oil pricing system was introduced that changed less for old oil and more for new oil. The prospect of higher prices for new oil produced record high drilling levels, focusing on oil development, in the late 1970s and early 1980s. Domestic production fell dramatically in the early 1980s following oil price deregulation that permitted world market forces to control oil prices. Since that time, the United States has returned to a period of reliance on oil supplied by the OPEC comparable to early 1970s levels.

Regulations also have a strong impact on natural gas supply, demand and prices. Exploration for and development of natural gas historically have been secondary to oil because of the high costs of transportation, as well as a complex transportation and marketing system that allowed for U.S.-federally regulated interstate gas pipelines but essentially unregulated intrastate pipelines. The natural gas supply shortfalls in 1977 and 1978 resulted in The Natural Gas Policy Act of 1978, which was designed to deregulate natural gas on a 10-year schedule. The Act also extended U.S. federal regulation to all pipelines and gave incentives to explore for and develop certain classes of resources such as unconventional gas. Section 29 of the 1980 Windfall Profits Tax provided tax credits for coal-bed methane, deep-basin gas, and tight-gas reservoirs. These tax credits were graduated and substantial; for example, the tax credit on coal-bed methane was 90 cents per million Btus and 52 cents per million Btus for tight-gas reservoirs by the end of 1992 when the credit was terminated. The price of coal-bed methane during the years the credit was in effect ranged from $1 to $3 per million Btus. Drilling for Section 29 gas wells increased during the late 1980s and early 1990s, allowing drillers to establish production prior to 1993 and thus take advantage of the tax incentive, which remains in effect for gas produced from these wells through 2003.

By contrast, U.S. coal resources are not restricted by supply; however, the environmental consequences of coal use have had a major impact on coal development. The Power Plant and Industrial Fuel Use Act of 1978 was developed in response to perceived natural gas shortages; it prohibited not only the switching from oil to gas in power generation plants, but also the use of oil and gas as primary fuel in newly built large plants. However, coal remains the least expensive source of energy; consequently, coal has soared from 1980 to 2000. The most significant change in the use of coal reflects compliance with the Clean Air Act Amendments of 1990 (CAAA 90), which have stringent sulfur dioxide emission restrictions. Production of coal that complies with this Act has caused a shift from production east of the Mississippi to west of the Mississippi. Many western states have substantial coal resources, particularly low-sulfur coal resources, such as those in the Powder River Basin of Wyoming. In fact, Wyoming has been the largest coal-producing state since 1988. CAAA 90 allows utilities to bring coal-fired generating units into compliance, for example, by replacing coal-fired units with natural gas, or by using renewable or low-sulfur coals. Conversion to natural gas from coal in power plants, made feasible by the relatively low costs of natural gas generation in the late 1980s and early 1990s has eroded coal's 1990 share of 53 percent of domestic electricity generation. Clearly policy and regulations such as CAAA

90 impact U.S. coal quality issues and promote low-sulfur coal and natural gas usage.

Many of the resource additions to natural gas will come from unconventional accumulations, also called continuous deposits, such as tight-gas reservoirs and coal-bed methane. Since deregulation of the oil and gas industry in the 1980s, policy decisions have increasingly affected the energy industry (particularly of natural gas) by providing tax incentives to produce unconventional or continuous hydrocarbon accumulations. However, there are associated costs. For example, coal-bed methane is an important and growing natural gas resource, but the disposal cost of waste water generated by coal-bed methane production is thirty-eight times greater than the cost of disposing waste water generated by an onshore conventional gas well. Similarly, the electricity generated by alternative energy sources, such as windmills near San Francisco, is three times more expensive than electricity produced by conventional electric generators.

Increased energy use both for coal and natural gas likely will have the greatest impact on western U.S. federal lands because major unconventional or continuous natural gas deposits are known to exist in Wyoming, Utah, Montana, and New Mexico. The U.S. Geological Survey, in cooperation with the Department of Energy, estimates that an in-place natural gas resource in the Green River Basin in Wyoming is more than ten times larger than a recent estimate of the entire recoverable conventional natural gas resources of the United States. However, recoverable resources make up only a small percentage of this large in-place resource. These continuous deposits are distributed over a wide area as opposed to conventional resources that are localized in fields. The majority of conventional undiscovered oil and natural gas resources will likely be found on U.S. federally-managed lands, and thus the federal government and policies will continue to play an increasing role in energy development.

Implementation of the 1998 Kyoto Protocol, which is designed to reduce global carbon emissions, will have dramatic effects on fossil fuel usage worldwide. The Kyoto Protocol mostly affects delivered prices for coal and conversion of plants to natural gas, nuclear and/or renewable resources. However, as pointed out by the International Energy Agency, increased natural gas consumption in the United States may likely have the effect of increased reliance on imported oil in the short-term to fulfill transportation sector needs. New technologies, such as gas to liquid conversion, have the potential to create revolutionary industrywide change. This change is contingent upon sufficient commodity prices to support the necessary infrastructure to develop underutilized gas resources worldwide.

Some economists argue that civilization moves toward increasingly efficient energy resources, moving from sources such as wood to coal to oil to natural gas and ultimately to non-carbon based energy sources in 100 to 200 year cycles. Others argue for a far more complicated process involving demand, supply and regulations. Fossil fuels will remain critical resources well into the next century. In the meantime, their abundance and potential shortages are debated.

Thomas S. Ahlbrandt

See also: Geography and Energy Use; Reserves and Resources.

BIBLIOGRAPHY

Ahlbrandt, T. S. (1997). "The World Energy Program." *U.S. Geological Survey Fact Sheet.* FS 007–97.

Ahlbrandt, T. S., and Taylor, D.J. (1993). "Domestic Conventional Natural Gas Reserves—Can They Be Increased by the Year 2010?" In *The Future of Energy Gases*, ed. David B. Howell. U.S. Geological Survey Professional Paper 1570, 527–546.

Campbell, C. J. (1997). *The Coming Oil Crisis.* Brentwood, United.Kingdom: Multi-Science Publishing Company.

Edwards, J. D. (1997). "Crude Oil and Alternate Energy Production Forecasts for the Twenty-first Century: The End of the Hydrocarbon Era." *American Association of Petroleum Geologists Bulletin* 81:1292–1305.

Gautier, D. L.; Dolton, G. L.; Attanasi, E. D. (1998). *1995 National Oil and Gas Assessment of Onshore Federal Lands.* U.S. Geological Survey Open-File Report 95–75.

International Energy Agency. (1998). *World Energy Outlook.* Paris, France: International Energy Agency/OECD.

Jackson, J., ed. (1997). *Glossary of Geology*, 4th ed. Alexandria, Virginia: American Geological Institute.

Klett, T. R.; Ahlbrandt, T. S.; Schmoker, J. W.; and Dolton, G. L. (1997). *Ranking of the World's Oil and Gas Provinces by Known Petroleum Volumes.* U.S. Geological Survey Open File Report 97–463, CD-ROM.

Law, B. E., and Spencer, C. W. (1993). "Gas in Tight Reservoirs—An Emerging Major Source of Energy." In *The Future of Energy Gases*, ed. David B. Howell. U.S. Geological Survey Professional Paper 1570, 233–252.

Masters, C. D.; Attanasi, E. D; and Root, D. H. (1994). "World Petroleum Assessment and Analysis." *Proceedings* *of the 14th World Petroleum Congress*. London: John Wiley and Sons.

Munk, N. (1994). "Mandate Power." *Forbes*, August, 41–42.

Nakicenovic, N. (1993). "The Methane Age and Beyond." In *The Future of Energy Gases*, ed. David B. Howell. U.S. Geological Survey Professional Paper 1570, 661–676.

National Petroleum Council. (1992). *The Potential for Natural Gas in the United States*. Washington, DC, National Petroleum Council.

U.S. Department of Energy. (1998). *Comprehensive National Energy Strategy*. Washington, DC: Author.

U.S. Energy Information Administration. (1998). *Annual Energy Review 1997*. Washington, DC: Department of Energy/Energy Information Administration.

U.S. Energy Information Administration. (1998). *International Energy Outlook 1998*. Washington, DC: Department of Energy/Energy Information Administration.

U.S. Energy Information Administration. (1998). *What Does the Kyoto Protocal Mean to U.S. Energy Markets and the U.S. Economy?*. Washington, DC: Department of Energy/Energy Information Administration.

U.S. Geological Survey. (1993). *The Future of Energy Gases*. U.S. Geological Survey Circular 1115.

FOURIER, JEAN BAPTISTE JOSEPH

(1768–1830)

Every physical scientist knows the name Fourier; the Series, Integral, and Transform that bear his name are essential mathematical tools. Joseph Fourier's great achievement was to state the equation for the diffusion of heat. The Fourier Series is a series of origonometric terms which converges to a periodic function over one period. The Fourier Integral is the limiting form of a Fourier Series when the period of the periodic function tends to infinity. The Fourier Transform is an analytic tool derived from the coefficients of the integral expansion of a Fourier Series. He pioneered the use of Fourier Series and Integrals because he needed them to solve a range of problems related to the flow of heat. Fourier was not the first person to realize that certain trigonometric expressions could be used to represent certain functions but he did develop the systematic use of such expressions to represent *arbitrary* functions.

BIOGRAPHY

The son of a tailor, Joseph Fourier was a member of a large family. Both of his parents died by the time he was nine. His education began at a local, church-run, military school, where he quickly showed talent in his studies and especially in mathematics. His school persuaded him to train as a priest. While preparing to take holy orders he taught his fellow novices mathematics. Fourier may well have entered the priesthood, but due to the French Revolution new priests were banned from taking holy orders. Instead he returned to his home town of Auxerre and taught at the military school. His friend and mathematics teacher, Bonard, encouraged him to develop his mathematical research, and at the end of 1789 Fourier travelled to Paris to report on this research to the Academie des Sciences.

Inspired by what he had experienced in Paris, Fourier joined the local Popular party on his return to Auxerre. In later years Fourier's involvement in local politics would lead to his arrest. He was arrested twice but each time was granted clemency.

France lost many of its teachers during the first years of the Revolution. One of the solutions to the shortage of teachers was the establishment of the Ecole Normale in Paris. Fourier, as a teacher and an active member of the Popular Society in Auxerre, was invited to attend in 1795. His attendance at the short-lived Ecole gave him the opportunity to meet and study with the brightest French scientists. Fourier's own talent gained him a position as assistant to the lecturers at the Ecole Normale.

The next phase of his career was sparked by his association with one of the lecturers, Gaspard Monge. When the French ruling council, the Directorate, ordered a campaign in Egypt, Monge was invited to participate. Fourier was included in Monge's Legion of Culture, which was to accompany the troops of the young general Napoleon Bonaparte (even then a national hero due to his successful campaigns in Italy).

The Egyptian campaign failed, but Fourier along with his fellow scientists managed to return to France. The general Fourier had accompanied to Egypt was now First Consul. Fourier had intended to return to Paris but Napoleon appointed Fourier as prefect of Isère. The prefecture gave Fourier the resources he needed to begin research into heat propagation but thwarted his ambition to be near the capital.

Jean Baptiste Joseph Fourier. (Library of Congress)

THEORIES OF HEAT PROPAGATION

The prevailing theory of heat, popularized by Simeon-Denis Poisson, Antoine Lavoisier and others, was a theory of heat as a substance, "caloric." Different materials were said to contain different quantities of caloric. Fourier had been interested in the phenomenon of heat from as early as 1802. Fourier's approach was pragmatic; he studied only the *flow* of heat and did not trouble himself with the vexing question of what the heat actually was.

The results from 1802–1803 were not satisfactory, since his model did not include any terms that described why heat was conducted at all. It was only in 1804, when Jean-Baptiste Biot, a friend of Poisson, visited Fourier in Grenoble that progress was made. Fourier realized that Biot's approach to heat propagation could be generalized and renewed his efforts. By December of 1807 Fourier was reading a long memoir on "the propagation of heat in solids" before the Class of the Institut de France.

The last section of the 1807 memoir was a description of the various experiments which Fourier had undertaken. It concentrated on heat diffusion between discrete masses and certain special cases of continuous bodies (bar, ring, sphere, cylinder, rectangular prism, and cube). The memoir was never published, since one of the examiners, Lagrange, denounced his use of the Fourier Series to express the initial temperature distribution. Fourier was not able to persuade the examiners that it was acceptable to use the Fourier Series to express a function which had a completely different form.

In 1810, the Institut de France announced that the Grand Prize in Mathematics for the following year was to be on "the propagation of heat in solid bodies." Fourier's essay reiterated the derivations from his earlier works, while correcting many of the errors. In 1812, he was awarded the prize and the sizeable honorarium that came with it.

Though he won the prize he did not win the outright acclaim of his referees. They accepted that Fourier had formulated heat flow correctly but felt that his methods were not without their difficulties. The use of the Fourier Series was still controversial. It was only when he had returned to Paris for good (around 1818) that he could get his work published in his seminal book, *The Analytical Theory of Heat*.

THE FOURIER LEGACY

Fourier was not without rivals, notably Biot and Poisson, but his work and the resulting book greatly influenced the later generations of mathematicians and physicists.

Fourier formulated the theory of heat flow in such a way that it could be solved and then went on to thoroughly investigate the necessary analytical tools for solving the problem. So thorough was his research that he left few problems in the analysis of heat flow for later physicists to investigate and little controversy once the case for the rigour of the mathematics was resolved. To the physical sciences Fourier left a practical theory of heat flow which agreed with experiment. He invented and demonstrated the usefulness of the Fourier Series and the Fourier integral—major tools of every physical scientist. His book may be seen both as a record of his pioneering work on heat propagation and as a mathematical primer for physicists. Lord Kelvin described the *Theory of Heat* as "a great mathematical poem."

Fourier's own attitude to his work is illustrated by the "Preliminary Discourse" he wrote to introduce his book. As a confirmed positivist, he stated that whatever the causes of physical phenomena, the laws governing them are simple and fixed—and so could be discovered by observation and experiment. He was at pains to point out that his work had application to subjects outside the physical sciences, especially to the economy and to the arts. He had intended to write a companion volume to his *Theory of Heat* that would cover his experimental work, problems of terrestrial heat, and practical matters (such as the efficient heating of houses); but it was never completed.

His talents were many: an intuitive grasp of mathematics, a remarkable memory and an original approach. Fourier was a man of great common sense, a utilitarian, and a positivist.

David A. Keston

See also: Heat and Heating.

BIBLIOGRAPHY

Bracewell, R. N. (1989). "The Fourier Transform." *Scientific American* 6:62.

Grattam-Guiness, I. (1972). *Joseph Fourier (1768–1830): A Survey of His Life and Work*. Cambridge, MA: MIT Press.

Herivel, J. (1975). *Joseph Fourier: The Man and the Physicist*. New York, NY: Clarendon Press.

FREIGHT MOVEMENT

The history of freight movement is the history of continuing improvements in speed and efficiency. Customers demand speed to meet a schedule; shippers demand efficiency to lower costs and improve profitability. Efficiency for the shipper can be broken down into three major components: equipment utilization, labor productivity, and energy efficiency. Dramatic strides in improving labor productivity, equipment innovations, and better utilization of that equipment have been made since the 1970s. It has been estimated that freight industry energy efficiency gains have saved the United States more than 15 percent (3.8 quadrillion Btus) to 20 percent between 1972 and 1992. All sectors of the freight transportation industry—air, water, rail, and truck—have been able to pass these savings on to their customers through lower prices and faster and more dependable service.

THE NATURE OF FREIGHT AND ENERGY

Movement of freight accounts for one-third of all U.S. transportation energy consumption. But if the U.S. share of the export/import cargo shipping market was included, the figure would be even higher.

The U.S. economy has changed drastically since the early 1970s. There has been a great change in shipping as well as a vast change in what is being shipped. Though the need to ship raw materials such as coal and minerals and basic commodities such as farm products has grown, the shipment of components and finished products has expanded much more rapidly. Awareness of shifting trends is important because each cargo industry segment has unique needs (see Table 1).

For the majority of shipped goods, the scope and influence of freight on an economy are often underappreciated because much of freight movement is "hidden." By the time a typical automobile is fully assembled and delivered, it has gone through thousands of transportation steps. It consists of numerous basic materials and thousands of components, all of which need to be fabricated, assembled, and sent to the automobile manufacturer for final assembly.

From the raw materials to the shipment of the finished product—including all the costs associated with shipments of various components to manufacture a product—the total cost for shipping is approximately only 2 percent of the average cost of the product in the United States. The shipping of the finished product, which usually adds another 1 to 3 percent to the selling price, is highly variable because of the tremendous variations in the mass and volume of the freight being shipped.

Of course, these costs can change with fluctuations in the price of energy. When the energy crisis of 1973 drove up oil prices, transportation costs were more affected than manufacturing costs. And the bigger the product and the greater the number of components needed for it, the greater was the run-up in freight costs. The transportation industry passed along these higher energy costs to the customers, making the industry a prime mover in pushing up inflation. Likewise, as energy prices stabilized or fell

Freight	*Mass to Volume Ratio*	*Delivery Speed and Why*	*Usual Mode and Why*
Commodities (coal, grain, raw materials)	Great	Slowest. Speed is not necessary. Because customers use large quantities of a commodity, customers want a reliable steady supply.	Waterways and Railways. This is the least expensive means of delivery.
Perishable Food (meat, fish and produce)	Moderate	Fast. Refrigerated transport is very expensive.	Refrigerated Truck/Container or air for Fresh Fish.
Flowers	Low	Fast. Flowers are perishables with a very short shelf life.	Air from South and Central America to the United States.
Manufactured Components	Varies	Fast. The just in time requirements of manufactures demand quick dependable deliveries.	Truck.
Finished Products	Varies	Fairly Fast. Delivery as quick as possible for payment as soon as possible.	Truck and Rail.
Spare Parts	Varies	Usually fast. It is extremely important to keep equipment fully utilized.	Air and Truck.
High Value Goods (gold, diamonds, jewels)	Moderate	Speed and the prevention of pilferage is foremost.	Air. Low volumes and greater security precautions.
Documents and Packages for Service Economy	Low	Very fast. The service economy requires quick communication.	Air.

Table 1.
Cargo and How It Is Moved
The wide variety of freight and the wide variety of needs favors a wide variety of different transportation modes. Although the railways and waterways move greater tonnage, trucking commands more than 80 percent of freight revenue.

through most of the 1980s and 1990s, the cost of freight transportation stabilized or fell as well.

The railways and waterways should continue to secure a larger share of the freight transportation market due to the energy efficiency advantages they command. Studies have found that railways can move competitive traffic at a fuel saving typically in the range of 65 to 70 percent compared with trucks (Blevins and Gibson, 1991) (see Table 2).

Although freight modes differ in propulsion sys-

Mode	*Typical Capacity*	*Propulsion System*	*Customer Price/Speed for Shipping (40 x 8 x 8 container size)*	*Typical Speed in United States*	*Mileage (ton-miles /gallon)*		*Energy Consumption (Btu/ton-mile)*	
					1975	*2000*	*1975*	*2000*
Container Ship	82,000 tons	Large Diesel	$500 Port-to-Port charge for container transported 2000 miles in 4 days	24 knots	182	230 est.	550	400
Container Train	7,000 tons	Large Diesel	$700 Door-to-Door Intermodal Container, Los Angeles to Chicago in 50 hours	60 to 70 mph (a)	185	380	685	370
Truck	35 tons	Diesel	$1,500 Door-to-Door, Los Angeles to Chicago in 40 hours	65 mph	44	44	2,800	2,800
747-400 Cargo Aircraft	100 tons	Jet Engine	$66 for 20 lb. package overnight	500 mph	3	6 est.	42,000	19,150

Table 2.
Energy Requirements for Freight Transportation
Much higher energy costs for the air and trucking industries are passed along in the form of higher rates, rates customers willingly pay for faster service. However, when crude oil prices more than doubled between 1998 and 1999 (fuel costs increased by about $200 million from 1998–1999 to 1999–2000 for Federal Express), both air freight and trucking operations resisted raising rates fearing competition from waterborne shipping and railroads, which have been closing the faster-services advantage.
Note: (a) The less vibration resulting from welded rail makes higher speed travel possible.

tems, resistance encountered, and speeds traveled, waterborne shipping and railroads will always have an inherent energy efficiency advantage over trucking for several reasons. First, propulsion becomes more efficient the longer the ship, train, or truck. When a truck adds a second or a third trailer (where allowed by law), it can double or triple the tonnage transported for only an incremental increase in fuel consumption. Thus, on a ton-miles-per-gallon basis, a large truck will be more energy-efficient than a small truck, a large ship more than a small ship, and a large plane more than a small plane. Second, trucks must possess the reserve power to overcome gravity; the need to negotiate grades that are usually limited to 4 percent, but that can reach as high as 12 percent. This “backup” reserve power is more easily met by locomotives and is not needed for the gradeless waterways. Third, weight reduction for trains and trucks is more important because it lowers the need for the reserve power needed to change speeds and climb grades. However, much more energy goes into propelling the nonfreight component of the truckload than the trainload. And finally, trucks are at a distinct disadvantage in terms of air and rolling resistance. Locomotives and container ships move at slower speeds and present a greater frontal area in relation to the volume and mass of the load, so the amount of additional energy needed to overcome air resistance is only a fraction of what a truck encounters; moreover, rolling resistance is far less for a steel wheel rolling on a steel rail than for a truck tire on a paved road.

Some of trucking’s disadvantage relative to shipment by water or rail can be explained by the type of product shipped. As more heavier freight is moved by rail, low-weight voluminous shipments continue to command a larger share of trucking freight. So although the volume of freight moved by trucks has increased, the energy intensity (ton-miles per gallon) of trucking has stayed the same. There have been technical improvements in trucking (e.g., better engines, less resistance), but they have been offset by shipment of less dense goods that fill up trailer space well before reaching the truck's weight limit; increased highway speeds, resulting in greater aerodynamic drag; and slow turnover among trucking fleets. Taking the differences of freight in mind, one study found that rail shipment achieved 1.4 to 9 times more ton-miles per gallon than competing truckload service (Department of Transportation, 1991). This estimate includes the fuel used in rail switching, terminal operations (e.g., loading and unloading a 55-foot trailer weighing approximately 65,000 pounds), and container drayage (e.g., trucking freight across town from one rail terminal to another).

Trucking also lacks the flexibility of rail shipment. If a railroad has to ship many lightly loaded containers, the railroad can simply add more train cars. Moreover, railroads can add train cars with only an incremental increase in fuel consumption. Constrained by regulations limiting trailer number, length, and weight, the same flexibility is not available to trucks.

THE ERA OF TRUCKING

Of all transportation options, trucking commands a lower share of ton-miles per gallon shipped than by rail or waterway, yet if considering only the volume of goods shipped, the percentage moved by truck would be much greater than by railway or waterway. In 1997 alone, trucks logged more than 190 billion miles, a 135 percent increase from 1975.

The ability of the trucking industry to thoroughly dominate transportation from 1960 until 1980 began to become apparent shortly after the conclusion of World War II because of four very favorable market conditions: (1) the shifting of freight away from military to commercial; (2) the building and completion of an extensive interstate highway system speeding delivery; (3) the decision of companies to expand and relocate facilities away from city centers and near interstate highway loops; and (4) the inability or unwillingness of railroads to adapt to these changes.

Trucking grew unabatedly until the energy crises of the 1970s. In an era of cheap energy and extensive interstate highways, trucking could offer speedy dependable service at a very competitive price. The only area in which trucking could not compete effectively with rail was for heavy commodities such as coal and grains. Railroads had a huge energy advantage, yet for many reasons—some beyond their control—could not adapt to the more competitive environment.

By the 1980s it was apparent that the real threat to trucking dominance was not the railroads but congestion. Highway congestion started to significantly worsen during the late 1970s and early 1980s, eroding much of the speed advantage and the accentuating the energy and labor productivity disadvantages of the railroads. Congestion continued to increase through the 1980s and 1990s, and it promises to worsen in the future. It has been estimated that growing traffic congestion might lower the on-road fuel economy of trucking by up to 15 percent by 2010 (compared to 1990), and the labor productivity of trucking will drop as well, as more time is spent stuck in traffic rather than moving freight.

Air quality suffers from this increasing congestion as well. Motor vehicles create the majority of air quality problems in urban areas, so for cities to comply with stringent ambient air quality standards, they will have to reduce motor vehicle emissions. Trucking accounts for only 4 percent of the U.S. motor vehicle fleet, yet can easily be responsible for 30 to 40 percent of the air quality problems because, in comparison to automobiles, the fleet is far older, is driven far more miles each year, and the emissions per vehicle are far greater.

Another congestion-related problem facing trucking is safety. Although crashes per mile decreased almost 50 percent from the 1970s to 2000, each year more than 5,000 people die in accidents involving trucks. Much of the reason for these accidents has been attributed to unsafe trucks, high speed limits, and driver fatigue. In 1997, trucks were involved in for 22 percent of the motor vehicle fatalities (excluding single-car accidents) in the United States.

For the economy as a whole, traffic jams in the ten most congested U.S. cities cost more than $34 billion a year, according to Texas Transportation Institute. In recognition of the part that trucking plays in this congestion problem, Congress instituted the Congestion Mitigation and Air Quality Improvement Program as part of the Internodal

Regulation or Subsidy	***Distortion***
Federal regulation of railroad rates to ensure that railroads did not unfairly use their monopoly power.	Allowed trucking companies to cherry-pick the most desirable traffic, leaving more volume of the less desirable traffic to the railroads. Repealed 1981.
Property tax abatements encouraging companies to locate and expand in suburban and rural areas.	If they did not make the siting decision for property tax saving reasons, efficient transportation would have been a higher priority. Started in 1960s and more prevalent now than ever.
Public funding for road and highway construction and reconstruction. Federal highway funding alone was $217 billion in 1998.	Although truckers pay road and fuel taxes, it pays only a small fraction of the cost of safety inspections and road damage from trucking. The taxes paid/benefits received imbalance has only grown through the years.
City and state subsidies in the bidding war for the next generation of ports.	The ship lines will make redevelopment decisions based more on who offers the bigge subsidy rather than the best rail links.
Anti-trust law that discourages railroad communication and cooperation necessary to make intermodal transportation work.	Without any coast-to-coast railroads, the railroads must cooperate to deliver containers across country.

Table 3.
Regulations and Subsidies That Have Distorted Energy Use in the Freight Transportation Market

Surface Transportation Efficiency Act of 1991 (ISTEA). A major goal of this program is to move more of the truck traffic to rail as a means of reducing congestion, improving air quality, and saving energy.

The trucking industry has been trying to make up for the growing labor and energy-efficiency advantages of the railroads by pushing for favorable legislation to raise the maximum workweek from sixty to seventy-two hours, gain higher weight limits, and allow more multi-trailers. These dynamics are moving the industry into what economist Michael Belzer calls "sweatshops on wheels."

Yet despite all the problems associated with the 190 billion miles logged by truckers each year, trucks are vital because they offer an unmatchable service: door-to-door delivery. Companies are located everywhere from city centers to distant suburban locations and rural areas, so the door-to-door railway delivery is not even an option for most businesses. Even if the United States made massive investments in railway infrastructure and raised the taxes of the trucking industry to fully account for its contribution to road and environmental damage, the mismatch of railroads to customer needs ensures that there always will be a great need for trucking. Except for the commodities and basic-materials industries, few other companies since the 1950s gave rail transportation a high enough priority in expansion and relocation plans. The corporate exodus from the cities long ago cemented many freight transportation and energy use patterns that are nearly impossible to change.

INTERMODALISM

It is a tedious, difficult debate to compare the advantages of one freight mode over another. Each mode has its distinct advantages. The important thing for most nations is to try to make each mode as efficient as possible. If each transportation mode can become equally efficient, and if government policy does not distort the market, the market will fairly dictate the most efficient way to move freight (see Table 3). All else being equal, it usually will be the most energy efficient choice too.

As the freight transportation industry moves into the twenty-first century, the greatest potential for improving efficiency is found along the railways because rail is the centerpiece in the evolution toward intermodalism. The intermodal industry consists of all modes of transportation working together to ship goods, whether it be via rail carrier, sea vessel, air transporter, or truck. Each mode has one thing in common: the delivery of a consignee's goods directly to the final destination, regardless of how the goods were originally shipped.

The first piggyback move, now called intermodal, came about when the Barnum & Bailey Circus went from state to state and town to town on a special train of flat railroad cars in 1872. Today this is referred to as a dedicated train. The cars were loaded with the tents, animals were put in caged containers, and passenger cars were hooked behind them carrying all of the circus personnel and wares. When the "circus train" arrived at the desired destination, the circus personnel would then untie the cables that secured the containers to the cars, back up a large portable ramp that was ten feet wide and approximately sixty feet long with a twenty degree angle, and attach it to the train where the locomotive was disconnected. The container cage would have wheels mounted under the container so that a tractor could back up to the ramp and attach to the container, pulling the container down for unloading. The same operation would also be used for loading.

Intermodal Evolution

To capitalize on the energy-efficiency advantages of intermodal freight movement, and also to come closer to matching the speed and reliability of trucking, the railroads had to overcome the equipment and infrastructure problems plaguing the industry, the aversion to cooperate in developing intraindustry and interindustry standards to achieve coast-to-coast delivery, price inflexibility due to regulation, and the reluctance to cooperate and customize equipment and processes to meet the needs of customers.

Prior to the containerization movement, work at ports and railyards was very slow and labor-intensive because much of the freight being moved entailed unprotected palleted cargo from ships and less-than-carload freight moved in boxcars. It took a lot of human energy to load and unload cargo in this manner and the upstart trucking industry had huge labor productivity advantages over the railroads.

The most important step in making intermodal shipment possible was the development of container standards and lifting equipment to meet the needs of ship lines and railroads. When Mi-Jack and the Santa Fe Railroad introduced a reliable overhead rubber-tired crane in 1963 to the piggyback (intermodal) industry, it started to take shipping out of the dark ages as loading or unloading times went from 45 minutes per load (circus ramping) to 2.5 to 3 minutes (rail terminal operators in 2000 guarantee the railroads 1 to 1.5 minutes per lift). All the associated operating costs and personnel requirements of circus ramping were eliminated as the new lift equipment, whether it be a sideloader (forklift truck) or a mobile gantry crane, greatly reduced energy requirements.

Although the containerization of cargo dates to the 1920s and earlier, the movement really accelerated during World War II, when the U.S. Armed Forces started to use containers to ship top-priority supplies because the cargo could be kept secret, handled less, and loaded and unloaded faster from truck to rail to sea vessel. Prior to containerization, pilferage was high because most freight was exposed and stacked on a 4-foot-square pallet for shipment.

After World War II, Malcolm McLean recognized the huge potential of containerization for the com-

Mode	*Cost*
Double-Stack Container	30 cents per mile
Piggyback Trailers on Spines	40 cents per mile
Truckload Highway Carrier	75 cents per mile

Table 4.
An example of costs for 1,000 mile haul
Although rail costs are route specific and vary tremendously, shipping by double-stack container is usually the most economical option, largely reflecting the energy savings. The energy-related advantage of the double-stack over the piggyback trailer comes from more cargo weight per rail car ("leaving the wheels behind," stacking, and lighter cars). As a result, a train can carry more cargo without increasing length (more cargo per 6,000 ft. of train). Trains also suffer less air resistance since double-stack sits lower. However, when drayage (local trucking) is expensive (up to $4.00 per mile in some areas), much of the economic advantage of intermodal shipping using double-stack containers is negated.

mercial market. McLean, who was in the trucking industry at that time, purchased two surplus sea vessels and started his own shipping company, SeaLand, first shipping fifty-eight containers from New York to Houston on April 26, 1956. Shipping would never be the same. SeaLand became very successful quickly, and eventually other shipping lines, such as Matson, also began to ship containers, in 1958. By the early 1960s every shipping line in the world was handling containers to ship cargo.

By using railroads to transport their containers, the shipping lines discovered in the 1960s that they could bypass the costly Panama Canal in going from Japan to Europe. The container could be transported from the Pacific coast port to the Atlantic coast port and placed on an awaiting ship, improving in-transit time and vastly reducing the cost of shipping and fuel. This is called the landbridge, which differs from the microbridge, wherein the container is loaded on a ship in Japan, unloaded onto a Santa Fe container car in San Francisco for delivery in Chicago, and then unloaded to a chassis for trucking to its final destination.

However, the containerization movement still had a major problem: standardization of equipment. It was very difficult to get the various shipping lines to agree on a standard container corner casting that any spreader on the crane could handle to facilitate the loading and unloading of the container. Matson and SeaLand were reluctant to change their design to the 1965 standards of the International Standard Organization and the American Standard Association (ISO-ASA). It took a major effort of the lifting equipment manufacturers Drott and Mi-Jack, with the help of railroads, to finally start converting all containers to the new standards a few years later. During the interim, Mi-Jack designed a special cobra-head latch that could accommodate all three corner castings (each required a different picking point to lift).

Another major innovation was the introduction of the double-stack car by Southern Pacific in the late 1970s. The shipping lines realized the economic advantages of shipping two or three containers on a double-stack car. Because it was in essence a skeleton frame, it weighed much less than a conventional container on flat car (COFC). In the process, significant weight reduction of the container itself was achieved. Lighter containers carried on lighter double-stack cars, with a better aerodynamic profile than piggyback trailers, dramatically improved railroad energy efficiency.

Charlie Kaye, the President of XTRA, wisely took advantage of railroads shortage of capital by having XTRA lease containers and trailers to the railroad on a per-diem basis in the early 1960s. For the railroads, it was a good arrangement. By using a leasing company for trailers and container equipment, it freed up capital to invest in new locomotives and track improvements.

In the early 1980s, Don Oris of American President Lines followed Kaye's lead. This shipper purchased its own double-stack railroad cars and did not rely on railroads to furnish equipment for shipping containers from the West Coast. Because it was now involved in both sea and rail transport, American President Lines saw the benefit of modifying the double-stack car design to fit all lift equipment at port and rail terminals.

Leasing became so popular that by the 1990s, railroads could bid among many leasing companies for the various types of equipment required. This leasing movement, and the start of the boom in imports, marked the beginning of railroads shipping more containers than trailers. The 1980 balance of 40 percent containers and 60 percent trailers began to change, shifting to 60 percent containers and 40 percent piggyback trailers by 1998.

Largely due to stifling regulations, the railroads had no other choice but to lease equipment. In the early 1970s the rate of return for the rail industry was in the 1 to 2 percent range, and bankruptcy was commonplace. By the mid-1970s, 25 percent of the nation's rail miles had to be operated at reduced speed because of dangerous conditions resulting from lack of investment in infrastructure.

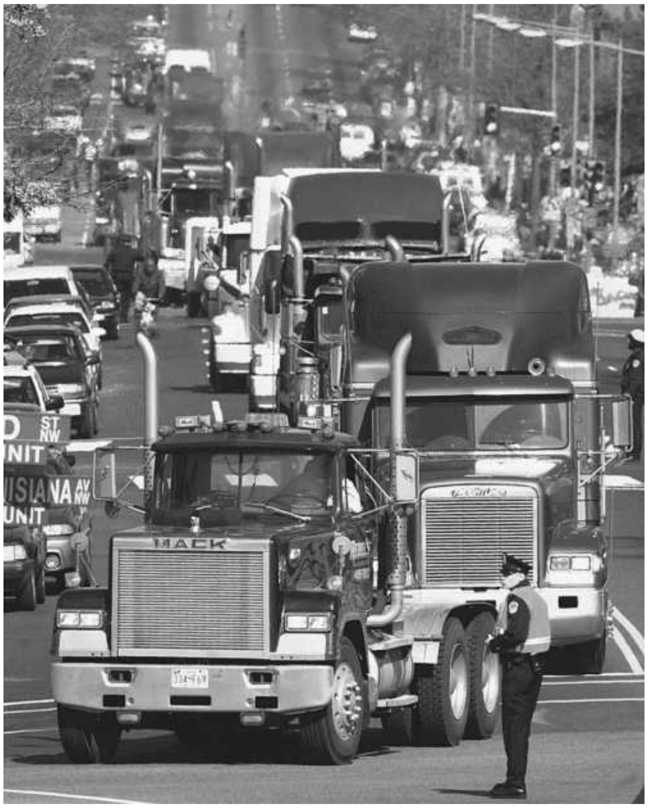
It took the combination of the Staggers Act of 1980 deregulating the railroads and the boom in imports starting in the 1980s for the railroads to be confident that they could recoup their investments in terminals, lifting equipment, new lightweight trailers on flat cars (TOFC) and double-stack cars, and energy-efficient locomotives. New terminals were designed for handling TOFC-COFC as well as very sophisticated lift equipment for loading trailers and containers weighing 40,000 to 60,000 pounds. This equipment had to operate every 1.5 to 2 minutes, five to seven days a week, working eighteen to twenty hours per day depending on the volume at the terminal.

Equally impressive was the introduction of new locomotive technology. For approximately the same fuel economy, a top-of-the-line locomotive in 1997 could generate twice the horsepower, pull more than twice the load, and reach a top speed 5 mph faster (75 mph) than a top-of-the-line locomotive in 1972. The better locomotives and lighter double-stack cars of the 1990s could replace as many as 280 trucks from the roadways for every 6,000 ft. train.

As intermodal shipping increased in volume, each mode of transportation realized that it needed to cooperate with competitors to increase the volume of freight. Mike Haverty, president of the Santa Fe Railroad in 1989, convinced two major truck carriers, J. B. Hunt and Schneider International, which previously were major competitors of the railroads, to become their shipping partners. Both agreed to put piggyback trailers on TOFC cars for trips greater than 500 miles for customers who required delivery within ten to fourteen hours. By 1992 Hunt realized the benefits of leaving the wheels behind, and began a transition to container and double-stack operations.

Railroads also entered into contracts with shipping lines to deliver containers after they were unloaded at ports. Again, truckers were working with shipping lines and airlines, as well as railroads, to deliver containers or trailers to their final destination. Each of the modes of transportation needed the other to complete the seamless delivery of freight from point of origin to final destination.

All these favorable dynamics were responsible for intermodal volume growing from 3 million containers and trailers in 1980 to nearly 9 million in 1998, as railroad customer rates fell from approximately 3.2 cents to 2.5 cents per ton-mile (1.5 cents if adjusted for inflation).



Trucks line North Capitol Street in Washington, D.C., as they convoy toward the U.S. Capitol on February 22, 2000. Hundreds of trucks entered the district to protest a steep rise in diesel fuel prices. (Corbis Corporation)

Future of Freight: Reducing Bottlenecks

If energy prices remain stable, customer rates will continue to fall as the freight transportation industry gets bigger and faster. On the equipment front, the trend is for bigger container ships, more powerful locomotives, and more multi-trailered trucks, regulations permitting. On the logistics front, there is certain to be more consolidation and cooperative efforts of railroad, shipping lines and truck carriers to reduce energy requirements and expedite cargo delivery. However, expansion and consolidation are not going to solve all freight transportation problems. To handle ever greater and faster-moving volume, the freight industry will have to find solutions to bottleneck problems.

The benefits of eliminating bottlenecks are twofold: energy efficiency and speedier deliveries. In the air freight area, hub-and-spoke systems are very efficient because large carriers such as Federal Express

A tugboat escorts a cargo ship in the Panama Canal. (Corbis Corporation)

have the volume and an efficient hub facility to quickly move freight through the terminal. The same is true for some of the major interstate trucking firms. Intermodal rail and shipping industries are a bit different: it takes a great deal of cooperation for different companies to share facilities and coordinate activities.

Going big is certain to help ton miles per gallon fall for sea vessels. The new container sea vessel *March Regima* can ship more than 8,000 TEUs (one TEU is equivalent to one 20-foot container), and even larger vessels are planned to handle 10,000 TEUs. These new vessels will be more than twice the size of the average medium-size to large-size container sea vessel of the 1990s, such as the Ckass C10, which can ship 4,500 TEUs, with 45,000 hp to travel at 22 knots to 25 knots, or the Class 11 container sea vessel, which can ship 4,800 TEUs requiring approximately 66,000 hp to travel at 25 knots to 30 knots.

To realize even greater efficiency with these large vessels, new port terminals are being designed to eliminate storage areas there and to provide faster loading and unloading sequences that will enable a ship to depart as quickly as possible. Depending on the size of the vessel, the cost to be in port can be as low as $5,000 per hour or as high as $15,000 per hour. These new ports will do away with the need to have ship-to-shore cranes unload containers on chassis, to be driven to a storage area at the port for pickup. Instead, loading and unloading of containers directly from vessels and railcars will eliminate extra steps in the process, dramatically cutting the time a larger-container vessel needs to be in port. Ports will be handling much more traffic in much less space.

Joint ventures such as the revitalization of the Panama Canal Railway—a 47-mile railroad running parallel to the Panama Canal and connecting ports on the Atlantic and Pacific Oceans—should be more common. Kansas City Southern Railroad and Mi-

Jack Products, a U.S. Class 1 railroad and North America's leading independent intermodal terminal operator, respectively, have invested $60 million for start-up expenses to revive and modernize the 143-year old first transcontinental railroad of the Americas. Fully operational in 2000, the revitalized railroad provides an efficient intermodal link for world commerce and complements the existing transportation provided by the canal, the Colón Free Trade Zone, and the port terminals.

This landbridge for the Americas reduces truck traffic on the major highway between Balboa and Colón by operating continuously, with a capacity for ten trains running in each direction every twenty-four hours. The new line accommodates locomotives and rolling stock capable of traveling at speeds of up to 65 to 95 km/hr, reducing energy requirements as well as emissions. The fine-tuned coordination of this port-to-port rail transport system, which allows railcars to pull up alongside vessels, helps the two coastal ports work like one enormous hub. Since the Panama Canal cannot handle the growing fleet of larger-container ships, this giant port-to-port terminal could eventually handle the majority of container traffic.

In the United States, the most important step being taken by railroads to relieve congestion bottlenecks is a new major hub terminal that would eliminate crosstown drayage. The through-port terminal will have a rail-mounted crane straddling ten or more tracks with the capacity to immediately select any group of containers to transfer from one train to another. For example, a westbound and an eastbound train arrive at the through-port terminal. Designated containers going east are on the westbound train. They will be transferred to the eastbound train for final destination. The same procedure will be performed for north and south, northeast and southeast, etc. The interchange, which normally would have taken one to three hours to unload a container or trailer and deliver it across town to an eastbound terminal from a westbound terminal, will now take only minutes using the through-port interchange operation.

Another alternative being considered is the in-line terminal design. This design consists of a trackside operation and storage area along trackside, in-line with the ramping and deramping operation of handling containers and trailers. It will also include one-way traffic on each side of the storage area. Truck drivers will be in and out of the terminal with their cargo in fewer than ten minutes, whether they are picking up or delivering containers or trailers. The gate dispatcher will now act as a traffic controller, similar to those used in the airline industry, and will be responsible for guiding the truck driver and freight to the delivery site of the trailer or container as well as the yard section or storage area location. With present terminal design, it takes approximately thirty minutes to two hours to pick up or deliver a trailer or container, from the arrival at the entrance gate to the delivery of the container and the driver's exit of the terminal. With an in-line terminal design, an intermodal truck driver's time will be more productive than that of the nonintermodal counterpart, who usually spends approximately 25 percent of the time waiting to load and unload.

THE WAYS OF MAJOR EXPORTERS

Unlike the United States, countries such as Japan and China are far greater exporters than importers, and they site manufacturing facilities with easy access to waterways. Many of the ports operate as effectively as U.S. ports, but the intermodal operations needed for fast, dependable overland service lag well behind U.S. operations.

Railroads also will need to do a better job of catering to customer needs so that customers choose intermodal transport for all shipping greater than 500 miles and not just for longer distances. United Parcel Service (UPS) was an early intermodal pioneer using piggyback to ship freight in the 1960s, and has continually encouraged railroads to invest in lifting equipment, railroad cars, and terminal infrastructure to improve ramping and deramping. To this day, it is still one of the largest customers shipping trailers in the intermodal industry. However, UPS is in the transportation business. For intermodal to expand significantly, the industry has to convince major industry of intermodal's benefits. "How close is the nearest intermodal terminal?" will need to become a top industry priority for expansion plans. More companies need to adopt a variation on the mine-mouth-to-boiler thinking of utilities for constructing new coal-fired power

NOVEL ENERGY SAVINGS: QUILTING INSTEAD OF REFRIGERATING

"Don't cool it, cover it" is the latest trend in transportation. Shipping temperature-sensitive refrigerated cargo has always been one of most expensive means to move freight. Just to run a compressor to refrigerate an 800 cubic foot trailer for three days can take more than 70 gallons of diesel fuel. Aside from fuel, refrigerated units must have regularly scheduled maintenance performed as well, which is why the shipping price to move a refrigerated trailer cross-country can be three or more times that of using a dry trailer. For companies in competitive marketplaces, this is a burdensome cost to absorb.

The Cargo Quilt was developed to offer an alternative, to drastically lower the cost of moving temperature-sensitive cargo by avoiding refrigeration. After the freight is loaded on a dry trailer or container, the Cargo Quilt, which works similarly to a thermos bottle, is draped over the cargo. Depending on the bulk of the shipment, it can maintain the temperature—whether hot or cold—from five to thirty days. And because there are no mechanical parts that can malfunction (e.g., thermostat or compressor), there is no chance of spoilage.

Some ultra-temperature-sensitive cargo will always need refrigeration, yet the manufacturer of the Cargo Quilt estimates that up to 50 percent of temperature-sensitive freight currently serviced by refrigerated trailers and containers could use a Cargo Quilt in a dry trailer at substantial savings. And as each mode of freight transportation lowers its guaranteed in-transit time, this percentage will continue to rise.

Quilts also reap savings by allowing the intermixing of chilled, frozen and freezables within the same mechanical trailer. Oscar Meyer Company uses a chilled unit for the majority of its meat products, yet needs to keep four to six pallets of turkey nuggets frozen as well. It achieves a dual-temperature zone container by covering the frozen turkey nuggets in a PalletQuilt. All the perishables arrive as loaded: some chilled, some frozen. Coors, Nestlé and Anheuser-Busch are just a few of the other major corporations that have switched to the Cargo Quilt and dry trailers.

Who would have ever thought that the Cargo Quilt, which has nothing to do with improving propulsion or reducing resistance, would turn out to be one of the most promising energy-conserving transportation innovations of the 1990s?

plants. The closer the intermodal terminal, the less of a concern is trucking and traffic congestion.

Railroads are expected to become more aggressive in moving away from their middleman status as intermediary to trucking and shipping. However, the mountainous debt taken on due to railroad mergers and acquisitions in the late 1990s brings into question railroads' ability to make future investments in infrastructure and marketing channels.

Intermodal service must improve for corporations that adopted just-in-time inventory in the 1980s and 1990s. Many companies feel that the benefits of keeping minimal stock are worth the premium price paid for faster door-to-door delivery by truck, which is a major reason why truck freight revenues remain at more than 80 percent of the country's total freight revenues.

There is a grave transportation risk in relying on just-in-time production methods. Any freight transportation system breakdown can be catastrophic for companies and nations. If key parts get stuck at the border, or if a natural disaster destroys a highway or a rail line, some factories might have to close down, others write off permanent losses. And if it is a component for an essential product or service, the ripple effect could cost the economy billions of dollars.

Automobile companies are major users of just-in-time methods, coordinating the delivery of thousands of parts from many different suppliers. A transportation delay of a few hours for one part could shut down an assembly line for half a day.

For the intermodal equipment and infrastructure investments to fully pay off, and to grab a larger share of just-in-time freight, will require further computerization improvements in tracking and processing equipment for the industry, as well as by the U.S. Customs Service. Although the 1980s transition from pen and paper to computerization by Customs has helped speed intermodal traffic, the tripling of imports from 1985 to 2000 put tremendous strain on the computer system used to process container shipments. This system, which is being upgraded, has created costly shipment delays for companies relying on just-in-time production methods. And there remains the fear of a major computer system breakdown that might take weeks to resolve and that would cost many companies billions of dollars as goods are not transported and as assembly plants are idled waiting for parts.

The intermodal industry seems to be addressing all these challenges and is highly likely to continue to grow and shift more freight toward rail because of the large advantages over roadways in safety, congestion, pollution, noise, land use, and energy consumption. Nothing is more energy-efficient than moving goods intermodally, and nobody does it better than the United States. Government and business people from around the world come to America to watch, learn, and marvel at the way the different modes cooperate to deliver cargo so efficiently.

John Zumerchik
Jack Lanigan, Sr.

See also: Air Travel; Government Approaches to Energy Regulation; Government and the Energy Marketplace; Locomotive Technology; Propulsion; Ships; Tires; Traffic Flow Management; Transportation, Evolution of Energy Use and.

BIBLIOGRAPHY

Abacus Technology Corp. (1991). *Rail vs.Truck Fuel Efficiency: The Relative Fuel Efficiency of Truck Competitive Rail Freight and Truck Operations Compared in A Range of Corridors.* Washington, DC: U.S. Department of Transportation, Federal Railroad Administration, Office of Policy.

Belzer, M. H. (2000). *Sweatshops on Wheels : Winners and Losers in Trucking Deregulation.* New York: Oxford University Press.

Berk, G. (1994). *Alternative Tracks: The Constitution of American Industrial Order, 1865–1917.* Baltimore: John Hopkins University Press.

Blevins, W. G., and Gibson, A. W. (1991). “Comparison of Emissions and Energy Use for Truck and Rail.” *Transportation Association of Canada Conference, Vol. 4, Conference Proceedings.*

Davis, S. C. (1997). *Transportation Energy Databook 17,* ORNL-6919. Oak Ridge, TN: Oak Ridge National Laboratory.

DeBoer, D. J. (1992). *Piggyback and Containers: A History of Rail Intermodal on America’s Steel Highways.* San Marino, CA: Golden West Books.

Greene, D. L. (1996). *Transportation and Energy.* Washington, DC: Eno Transportation Foundation, Inc.

Kahn, A. M. (1991). *Energy and Environmental Factors in Freight Transportation.* Ottawa: A. K. SocioTechnical Consultants for Transport Canada, Economic Research Branch.

FRICTION

See: Tribology

FUEL ADDITIVE

See: Gasoline and Additives

FUEL CELLS

A fuel cell is equivalent to a generator: it converts a fuel’s chemical energy directly into electricity. The main difference between these energy conversion devices is that the fuel cell acccomplishes this directly, without the two additional intermediate steps, heat release and mechanical motion.

A fuel cell has two basic elements: a fuel delivery system and an electro-chemical cell that converts the delivered fuel into useful electricity. It is this unique combination that enables fuel cells to potentially offer the best features of both heat engines and batteries. Like batteries, the cell generates a dc electric output and is quiet, clean, and shape-flexible, and may be manufactured using similar plate and film-rolling processes. By contrast, the fuel delivery system ensures that fuel cells, like heat engines, can be

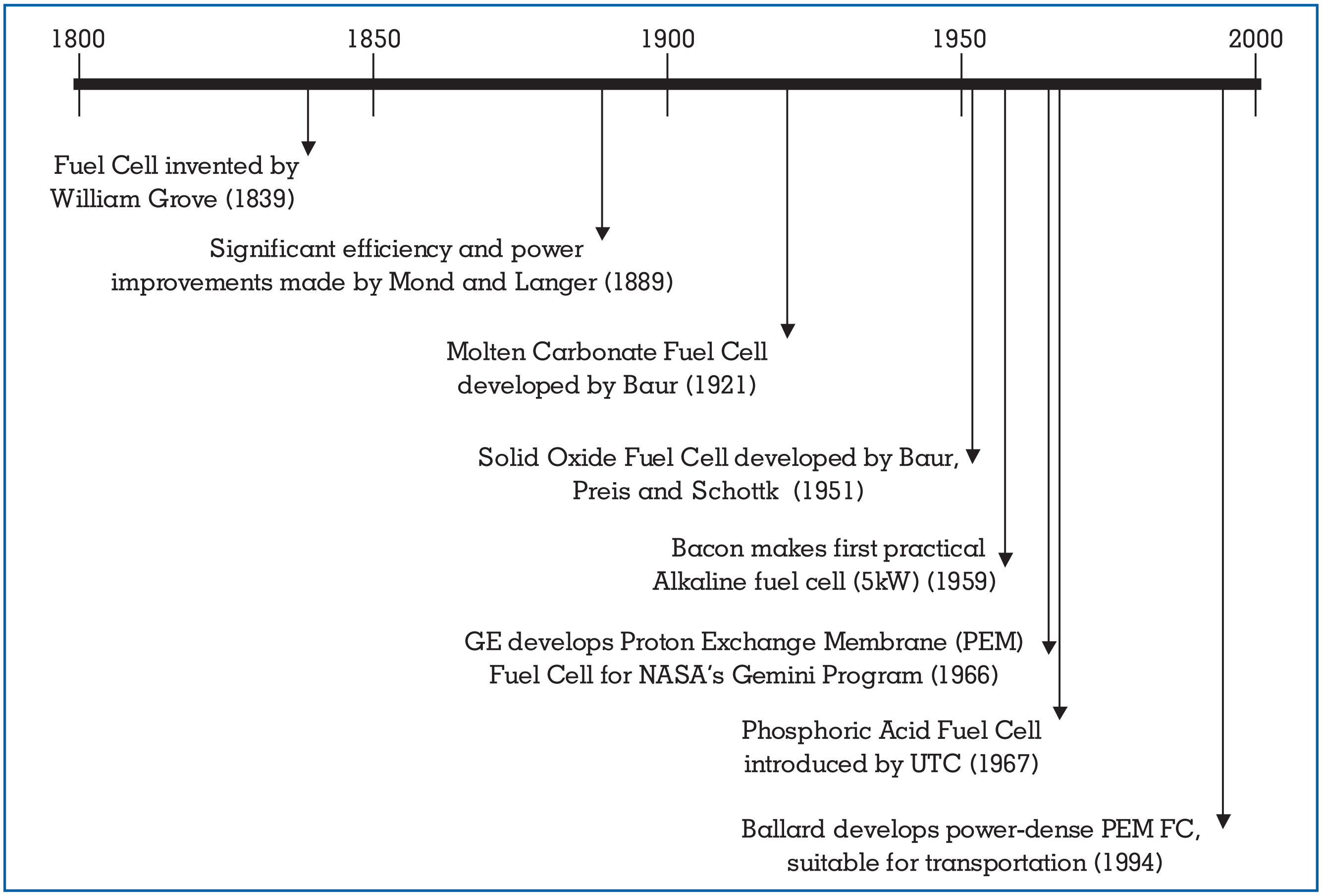


Figure 1.
Timeline of major fuel cell milestones.

quickly refueled and operate for long periods between stoppages.

"Fuel cell" is an ambiguous term because, although the conversion occurs inside a fuel cell, these cells need to be stacked together, in a fuel cell *stack*, to produce useful output. In addition, various ancillary devices are required to operate the stack properly, and these components make up the rest of the fuel cell *system*. In this article, fuel cell will be taken to mean *fuel cell system* (i.e., a complete standalone device that generates net power).

HISTORICAL INTEREST IN FUEL CELLS

Although fuel cells were invented over 150 years ago, Figure 1 reveals that there have been only a few key milestones in fuel cell development. For this reason they have only recently attracted significant and widespread interest from governments, research laboratories and major corporations. Two developments are behind this shift: impressive recent technology advances, and growing concern over the state of the environment.

For many years after their invention in 1839 by an English lawyer, Sir William Grove, fuel cells were little more than a laboratory curiosity because their performance was unreliable and few uses could be found for them. With rapid developments in electricity during the late 1800s it is surprising that fuel cells did not manage to compete with electrochemical batteries or generators as a source of electricity. In retrospect, the invention of the automobile in 1885 could have stimulated fuel cell development because a battery's limited energy storage makes it unsatisfactory for transportation. However, the internal combustion engine was introduced soon after the fuel cell, managed to improve at a faster rate than all alternatives, and has remained the prime mover of choice.

It was not until the 1960s that fuel cells successfully filled a niche that the battery or heat engine could not. Fuel cells were the logical choice for NASA's Gemini and Apollo Programs because they could use the same fuel and oxidant that was already available for rocket propulsion, and could generate high-quality electricity and drinking water in a relatively lightweight system. Although this application enabled a small fuel cell industry to emerge, the requirements were so specific, and NASA's cost objectives were so lenient, that fuel cells remained, and still remain, a minor power source.

This situation is beginning to change because recent years have seen impressive progress in reducing the size and cost of fuel cell systems to the point where they are now considered one of the most promising "engines" for the future. Simultaneously, increased concerns over climate change and air quality have stimulated many organizations to fund and develop technologies that offer significant environmental benefits. Their high efficiency and low emissions make fuel cells a prime candidate for research funding in many major industries, particularly as major growth is expected in developing nations, where energy is currently produced with low efficiency and with few emissions controls. In fact, if fuel cells are given the right fuel, they can produce zero emissions with twice the efficiency of heat engines.

Although environmental trends are helping to drive fuel cell development, they are not the only drivers. Another is the shift toward decentralized power, where many small power sources replace one large powerplant; this favors fuel cells since their costs tend to be proportional to power output, whereas the cost per kilowatt ($/kW) for gas turbines increases as they are shrunk. Moreover, the premium on high-quality electricity is likely to increase in the future as the cost of power outages rises. Fuel cells that generate electricity on-site (in hotels, hospitals, financial institutions, etc.) from natural gas promise to generate electricity without interruption and, unlike generators, they can produce this electricity very quietly and cleanly.

At the other end of the power spectrum, there is increasing interest in fuel cells for small electronic appliances such as laptop computers, since a high value can be placed on extending the time period between power outages. It is possible that a small ambient-pressure fuel cell mounted permanently in the appliance would allow longer-lasting hydrogen cartridges or even methanol ampoules to replace battery packs.

ELECTRICITY PRODUCTION BY A FUEL CELL

Fuel Cell Stack

As with a battery, chemical energy is converted directly into electrical energy. However, unlike a battery, the chemical energy is not contained in the electrolyte, but is continuously fed from an external source of hydrogen.

In general, a fuel cell converts gaseous hydrogen and oxygen into water, electricity (and, inevitably, some heat) via the following mechanism, shown in Figure 2:

i. The anode (positive pole) is made of a material that readily strips the electron from the hydrogen molecules. (This step explains why hydrogen is so important for fuel cell operation; with other fuels, it is difficult to generate an exchange current because multiple chemical bonds must be broken before discrete atoms can be ionized).
ii. Free electrons pass through an external load toward the cathode—this is dc electric current—while the hydrogen ions (protons) migrate through the electrolyte toward the cathode.
iii. At the cathode, oxygen is ionized by incoming electrons, and then these oxygen anions combine with protons to form water.

Since a typical voltage output from one cell is around 0.4–0.8 V, many cells must be connected together in series to build up a practical voltage (e.g., 200 V). A bipolar plate performs this cell-connecting function and also helps to distribute reactant and product gases to maximize power output.

Fuel cell stack voltage varies with external load. During low current operation, the cathode's activation overpotential slows the reaction, and this reduces the voltage. At high power, there is a limitation on how quickly the various fluids can enter and

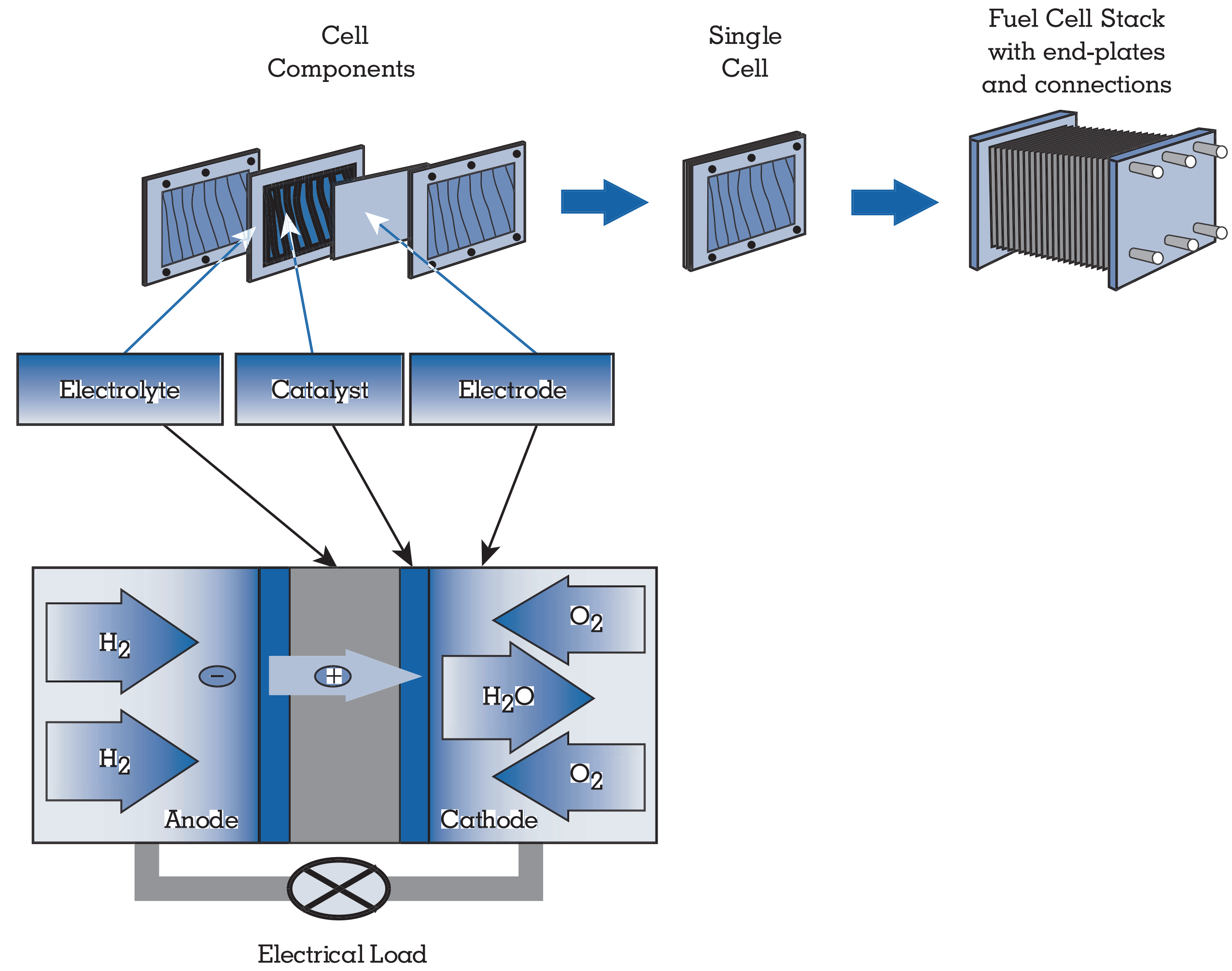


Figure 2.
Fuel cell stack.

exit the cell, and this limits the current that can be produced. In most of the operating range, however, it is ohmic polarization, caused by various electrical resistances (e.g., inside the electrolyte, the electrolyte-electrode interfaces, etc.) that dominates cell behavior. Continued research into superior electrodes and electrolytes promises to reduce all three types of losses.

Because of these losses, fuel cells generate significant heat, and this places a limit on the maximum power available because it is very difficult to provide adequate cooling to avoid formation of potentially dangerous "hot spots." This self-heating, however, helps to warm up the system from cold-start (mainly a concern for transportation applications).

A complete fuel cell system, even when operating on pure hydrogen, is quite complex because, like most engines, a fuel cell stack cannot produce power without functioning air, fuel, thermal, and electrical systems. Figure 3 illustrates the major elements of a complete system. It is important to understand that the sub-systems are not only critical from an operational standpoint, but also have a major effect on system economics since they account for the majority of the fuel cell system cost.

Air Sub-System

The reaction kinetics on the cathode (air) side are inherently slow because oxygen's dissociation and subsequent multi-electron ionization is a more complex sequence of events than at the (hydrogen) anode. In order to overcome this activation barrier (and hence increase power output) it is necessary to raise the oxygen pressure. This is typically accomplished by compressing the incoming air to 2–3 atmospheres. Further compression is self-defeating because, as pressurization increases, the power consumed by compression more than offsets the increase in stack power. In addition, air compression consumes about 10–15 percent of the fuel cell stack output, and this parasitic load causes the fuel cell system efficiency to drop rapidly at low power, even though the stack efficiency, itself, increases under these conditions. One way to recover some of this energy penalty is to use some of the energy of the hot exhaust gases to drive an expansion turbine mounted on the shaft compressor. However, this substantilly raises system cost and weight.

Alternatively, the fuel cell stack can be operated at ambient pressure. Although this simplifies the system considerably and raises overall efficiency, it does reduce stack power and increase thermal management challenges.

Fuel Sub-System

On the fuel side, the issues are even more complex. Hydrogen, although currently it is made in relatively large amounts inside oil refineries for upgrading petroleum products and for making many bulk chemicals (e.g., ammonia), it is not currently distributed like conventional fuels.

Moreover, although there are many ways to store hydrogen, none is particularly cost effective because hydrogen has an inherently low energy density. Very high pressures (e.g., 345 bar, or 5,000 psi) are required to store practical amounts in gaseous form, and such vessels are heavy and expensive. Liquid hydrogen is more energy-dense and may be preferred for transportation, but it requires storage at –253°C (–423°F) and this creates handling and venting issues. Moreover, hydrogen is mostly made by steam reforming natural gas, and liquefying is so energy-intensive that roughly half of the energy contained in the natural gas is lost by the time it is converted into liquid hydrogen. A third approach is to absorb hydrogen into alloys that "trap" it safely and at low pressure. Unfortunately, metal hydrides are heavy and are prohibitively expensive.

Because it is so difficult to economically transport and store hydrogen, there is interest in generating hydrogen on-demand by passing conventional fuels through catalysts; such an approach is called fuel processing. The challenge for fuel cell application is to convert available, relatively impure, fossil fuels into a hydrogen-rich gas (hydrogen-feed) without contaminating the various catalysts used in the fuel processor and fuel cell stack. Depending on the type of fuel cell stack and its operating temperature, there can be as many as four sequential stages involved in fuel processing:

i. fuel pre-treatment (such as desulfurization and vaporization)
ii. reforming (conversion of fuel into hydrogen-rich feed, sometimes called synthesis gas, or syngas)
iii. water-gas shift for converting most (>95%) of the carbon monoxide byproduct into carbon dioxide
iv. preferential oxidation for final removal of the remaining carbon monoxide.

Most of the differentiation between various fuel processor strategies comes in the second stage. One reforming method uses steam produced by the fuel cell stack reaction to reform the fuel into hydrogen. This process, steam reforming, is endothermic (absorbs heat), may require high temperatures (depending on the fuel) and, because the catalysts that enable the reaction are selective, different fuels cannot be used in the same catalytic reactor. However, a major advantage is that the product is

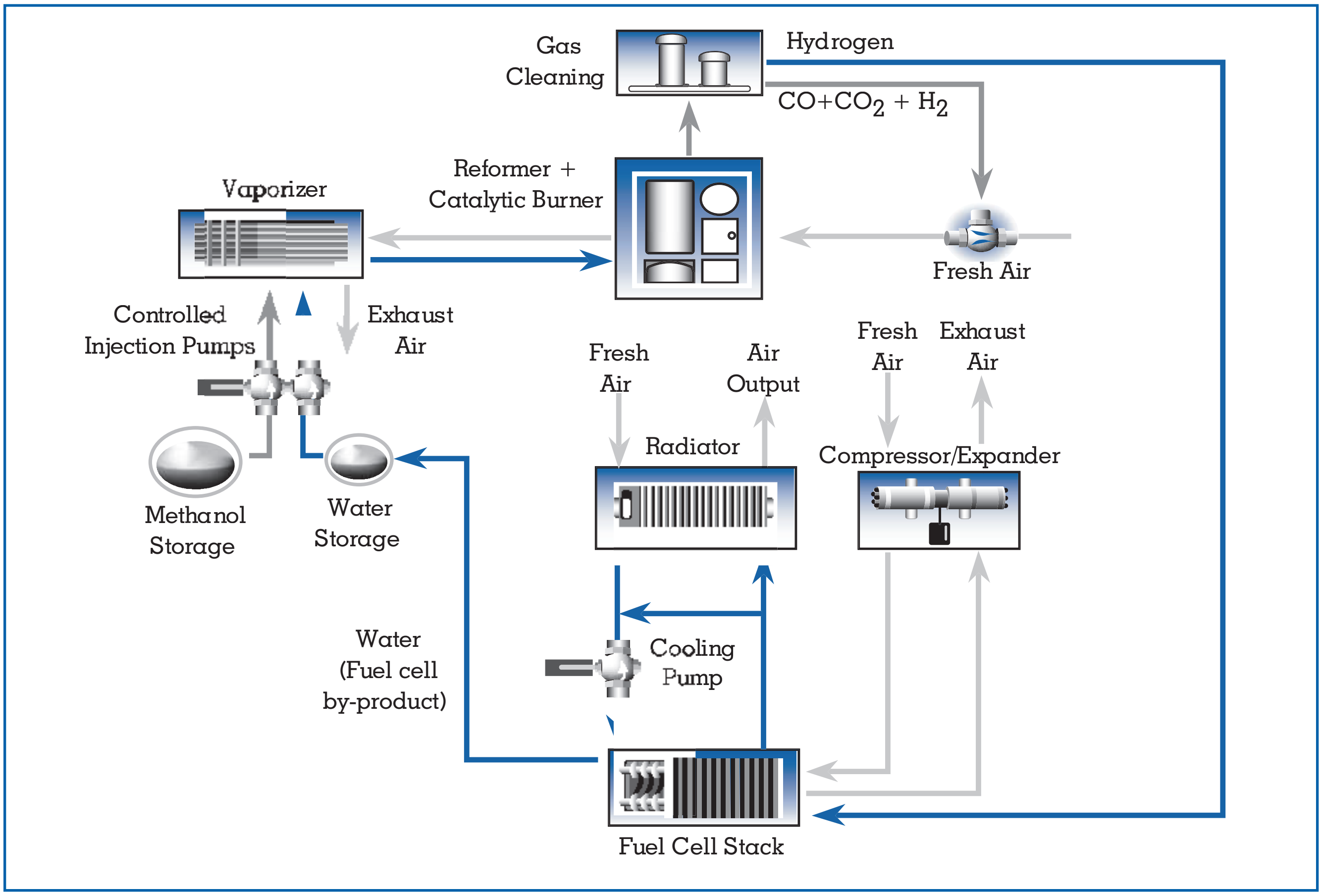


Figure 3.
Fuel Cell System.

hydrogen-rich; for example, when methane is steam reformed, the hydrogen concentration is ~75% ($CH_4 + H_2O \rightarrow 3H_2 + CO$). This process is favored for stationary applications, where a single fuel and steady-state operation are typical.

In contrast to steam reforming, partial oxidation (POX) uses air instead of steam and, as its name implies, burns the fuel in restricted amounts of air so that it generates partially combusted products, including hydrogen. POX generates heat and can, therefore, potentially respond faster than a steam reformer. This is beneficial for load-following applications (e.g., transportation).

Moreover, because all fuels burn, POX does not demand a catalyst, although advanced designs often use one to lower flame temperatures, which helps to relax materials requirements and to improve efficiency and emissions. The hydrogen concentration, however, is considerably lower (~40%) because none comes from steam and there is about 80 percent nitrogen diluent in the air ($CH_4 + \frac{1}{2}O_2 + 2N_2 \rightarrow 2H_2 + CO + 2N_2$).

Autothermal reforming, ATR, combines steam reforming and POX. Since the heat released from POX is consumed by steam reforming, the reactor can be adiabatic (or autothermal). For some applications, ATR may offer the best of both worlds: fuel-flexibility, by partly breaking the fuel down into small HC fragments using air, and relatively high hydrogen yield, by steam reforming these HC fragments.

Thermal Sub-System

Cooling strongly depends on fuel cell operating temperature and also depends on the fuel cell's external environment. For low temperature fuel cells, cooling imposes a significant energy debit because pumps need to force coolant out to a heat

	Operating Temperature (˚C)	*Advantages*	*Disadvantages*	*Potential Application*
Alkaline	25-100	• Mature technology • No precious metals	• Must use pure hydrogen	• Space
Proton Exchange Membrane	0-85	• Can operate at ambient temperature • High power density	• Sensitive to CO-poisoning • Need for humidification	• Transportation • Distributed Power
Phosphoric Acid	170-220	• Mature • Reformate-intolerant	• Bulky • Cannot start from ambient	• Heavy-duty transportation • Distributed Power
Molten Carbonate	~650	• Some fuel flexibility • High-grade waste heat	• Fragile electrolyte matrix • Electrode sintering	• Distribute power • Utilities
Solid Oxide	800-1000	• Maximum fuel flexibility • Highest co-generation efficiency	• Exotic materials • Sealing and cracking issues	• Distribute power • Utilities

Table 1.
Comparison of various fuel cells.

exchanger, from which heat must be rejected to the air. Operating the fuel cell at maximum efficiency reduces heat loads, but also reduces power output, forcing an increase in fuel cell stack size and cost. For high-temperature fuel cells, however, waste heat can be utilized by expanding the off gases through a turbine to generate additional electricity; such co-generation efficiencies can reach 80 percent. In some applications, even the remaining 20 percent can provide value (e.g., by warming the building's interior).

Heat rejection is only one aspect of thermal management. Thermal integration is vital for optimizing fuel cell system efficiency, cost, volume and weight. Other critical tasks, depending on the fuel cell, are water recovery (from fuel cell stack to fuel processor) and freeze-thaw management.

Electrical Sub-System

Electrical management, or power conditioning, of fuel cell output is often essential because the fuel cell voltage is always dc and may not be at a suitable level. For stationary applications, an inverter is needed for conversion to ac, while in cases where dc voltage is acceptable, a dc-dc converter may be needed to adjust to the load voltage. In electric vehicles, for example, a combination of dc-dc conversion followed by inversion may be necessary to interface the fuel cell stack to a 300 V ac motor.

TYPES OF FUEL CELLS

There are five classes of fuel cells. Like batteries, they differ in the electrolyte, which can be either liquid (alkaline or acidic), polymer film, molten salt, or ceramic. As Table 1 shows, each type has specific advantages and disadvantages that make it suitable for different applications. Ultimately, however, the fuel cells that win the commercialization race will be those that are the most economical.

The first fuel cell to become practical was the alkaline fuel cell (AFC). In space applications, liquid hydrogen and liquid oxygen are already available to provide rocket propulsion, and so consumption in the AFC, to create on-board electricity and potable water for the crew, is an elegant synergy. It therefore found application during the 1960s on the Gemini manned spacecraft in place of heavier batteries. Their high cost ($400,000/kW) could be tolerated because weight reduction is extremely valuable; For example, it could allow additional experimental equipment to be carried on-board.

The AFC has some attractive features, such as relatively high efficiency (due to low internal resistance and high electrochemical activity), rapid start-up, low corrosion characteristics, and few precious metal requirements.

However, the AFC's corrosive environment demands that it uses some rather exotic materials, and the alkaline (potassium hydroxide solution) concentration must be tightly controlled because it has poor tolerance to deviations. Critically, the alkali is readily neutralized by acidic gases, so both the incoming fuel and air need carbon dioxide clean-up. This limits AFC applications to those in which pure hydrogen is used as the fuel, since a fuel processor generates large amounts of carbon dioxide. The small amount of carbon dioxide in air (~0.03%) can be handled using an alkaline trap upstream of the fuel cell and, consequently, is not as much of a problem.

Because of this extreme sensitivity, attention shifted to an acidic system, the phosphoric acid fuel cell (PAFC), for other applications. Although it is tolerant to CO_2, the need for liquid water to be present to facilitate proton migration adds complexity to the system. It is now a relatively mature technology, having been developed extensively for stationary power usage, and 200 kW units (designed for co-generation) are currently for sale and have demonstrated 40,000 hours of operation. An 11 MW model has also been tested.

In contrast with the AFC, the PAFC can demonstrate reliable operation with 40 percent to 50 percent system efficiency even when operating on low quality fuels, such as waste residues. This fuel flexibility is enabled by higher temperature operation (200°C vs. 100°C for the AFC) since this raises electro-catalyst tolerance toward impurities. However, the PAFC is still too heavy and lacks the rapid start-up that is necessary for vehicle applications because it needs preheating to 100°C before it can draw a current. This is unfortunate because the PAFC's operating temperature would allow it to thermally integrate better with a methanol reformer.

The PAFC is, however, suitable for stationary power generation, but faces several direct fuel cell competitors. One is the molten carbonate fuel cell (MCFC), which operates at ~650°C and uses an electrolyte made from molten potassium and lithium carbonate salts. High-temperature operation is ideal for stationary applications because the waste heat can enable co-generation; it also allows fossil fuels to be reformed directly within the cells, and this reduces system size and complexity. Systems providing up to 2 MW have been demonstrated.

On the negative side, the MCFC suffers from sealing and cathode corrosion problems induced by its high-temperature molten electrolyte. Thermal cycling is also limited because once the electrolyte solidifies it is prone to develop cracks during reheating. Other issues include anode sintering and elution of the oxidized nickel cathode into the electrolyte.

These problems have led to recent interest in another alternative to PAFC, the solid oxide fuel cell (SOFC). As its name suggests, the electrolyte is a solid oxide ceramic. In order to mobilize solid oxide ions, this cell must operate at temperatures as high as 1,000°C. This ensures rapid diffusion of gases into the porous electrodes and subsequent electrode reaction, and also eliminates the need for external reforming. Therefore, in addition to hydrogen and carbon monoxide fuels, the solid oxide fuel cell can even reform methane directly. Consequently, this fuel cell has attractive specific power, and cogeneration efficiencies greater than 80 percent may be achievable. Moreover, the SOFC can be air-cooled, simplifying the cooling system, although the need to preheat air demands additional heat exchangers. During the 1980s and 1990s, 20–25 kW "seal-less" tubular SOFC modules were developed and tested for producing electricity in Japan. Systems producing as much as 100 kW have recently been demonstrated.

Because this design has relatively low power density, recent work has focused on a "monolithic" SOFC, since this could have faster cell chemistry kinetics. The very high temperatures do, however, present sealing and cracking problems between the electrochemically active area and the gas manifolds.

Conceptually elegant, the SOFC nonetheless contains inherently expensive materials, such as an electrolyte made from zirconium dioxide stabilized with yttrium oxide, a strontium-doped lanthanum manganite cathode, and a nickel-doped stabilized zirconia anode. Moreover, no low-cost fabrication methods have yet been devised.

The most promising fuel cell for transportation purposes was initially developed in the 1960s and is called the proton-exchange membrane fuel cell (PEMFC). Compared with the PAFC, it has much greater power density; state-of-the-art PEMFC stacks can produce in excess of 1 kW/l. It is also potentially less expensive and, because it uses a thin solid polymer electrolyte sheet, it has relatively few sealing and corrosion issues and no problems associated with electrolyte dilution by the product water.

Since it can operate at ambient temperatures, the PEMFC can startup quickly, but it does have two significant disadvantages: lower efficiency and more stringent purity requirements. The lower efficiency is due to the difficulty in recovering waste heat, whereas the catalyzed electrode's tolerance toward impurities drops significantly as the temperature falls. For example, whereas a PAFC operating at 200°C (390°F) can tolerate 1 percent CO, the PEMFC, operating at 80°C (175°F) can tolerate only ~0.01% (100 ppm) CO. The membrane (electrolyte) requires constant humidification to maintain a vapor pressure of at least 400 mmHg (~0.5 bar), since failure to do so produces a catastrophic increase in resistance. Operation at temperatures above 100°C would greatly simplify the system, but existing membranes are not sufficiently durable at higher temperatures and will require further development.

Fuel cells can run on fuels other than hydrogen. In the direct methanol fuel cell (DMFC), a dilute methanol solution (~3%) is fed directly into the anode, and a multistep process causes the liberation of protons and electrons together with conversion to water and carbon dioxide. Because no fuel processor is required, the system is conceptually very attractive. However, the multistep process is understandably less rapid than the simpler hydrogen reaction, and this causes the direct methanol fuel cell stack to produce less power and to need more catalyst.

FUTURE CHALLENGES

The biggest commercial challenge facing fuel cells is cost, and mass production alone is insufficient to drive costs down to competitive levels. In the stationary power market, fuel cell systems currently cost ~$3,000/kW and this can be split into three roughly equal parts: fuel cell stack, fuel processor, and power conditioning. With mass production, this cost might fall to below $1,500/kW but this barely makes it competitive with advanced gas turbines. Moreover, for automotive applications, costs of under $100/kW are necessary to compete with the internal combustion engine. Bringing the cost down to these levels will require the development of novel system designs, materials, and manufacturing processes, in addition to mass production. However, even if PEM fuel cells fail to reach the stringent automotive target, they are still likely to be far less expensive than fuel cells designed specifically for other applications.

Even in a "simple" hydrogen fuel cell system, capital cost reduction requires improvements in many diverse areas, such as catalyst loadings, air pressurization, cell thermal management, and sealing. Compounding the challenge is the need for durability and reliability. For example, electrodes and seals must be resistant to corrosion, stress, temperature fluctuations, and fuel impurities. Unfortunately, stable materials tend to be more expensive, and so a trade-off between life and cost can be expected as the technology nears the market stage.

This trade-off may not even occur in some cases. Membranes used in the PEMFC have been developed for the chlor-alkali industry and have 40,000-hour durability (shutdowns are prohibitively expensive in stationary applications), require only 5,000-hour durability (corresponding to 100,000 miles) for automotive applications. Hence, it may be possible to develop less expensive membranes that still meet automotive requirements.

Operating costs, in contrast, are more straightforward to determine because they depend on system efficiency, which, in turn, is related to voltage and current density (the current generated per unit area of electrolyte). Fuel savings are expected since the fuel cell operates more efficiently than a heat engine, and there may be lower maintenance and repair costs because fuel cells have fewer moving parts to wear out.

In addition to cost, a major technical challenge is in the fuel processor sub-system. Improvements are still needed in reducing size, weight, and, for transportation applications, cold-starting. Vaporizers also rely on heat generated from unused fuel leaving the fuel cell, and this combustion must be emission-free if the fuel cell system is to be environmentally attractive. This leads to the use of catalytic burners and, although such burners reduce NO_x emissions, they have yet to demonstrate sufficient durability. For the PEMFC, the final CO clean-up stage, PROX, uses precious metal catalysts and requires very fine temperature control in order to maintain the catalyst's selectivity toward oxidizing a small amount (1%) of CO in the presence of large amounts (>40%) of hydrogen. Such precision is difficult to achieve under conditions where the load varies continuously.

Full system integration is a major challenge since system design often must accommodate contradictory objectives. For example, it is relatively straightforward to design a fuel cell for high efficiency by maximizing thermal integration, but this is likely to increase complexity and degrade dynamic response. It may also increase cost and, given the dollar value of

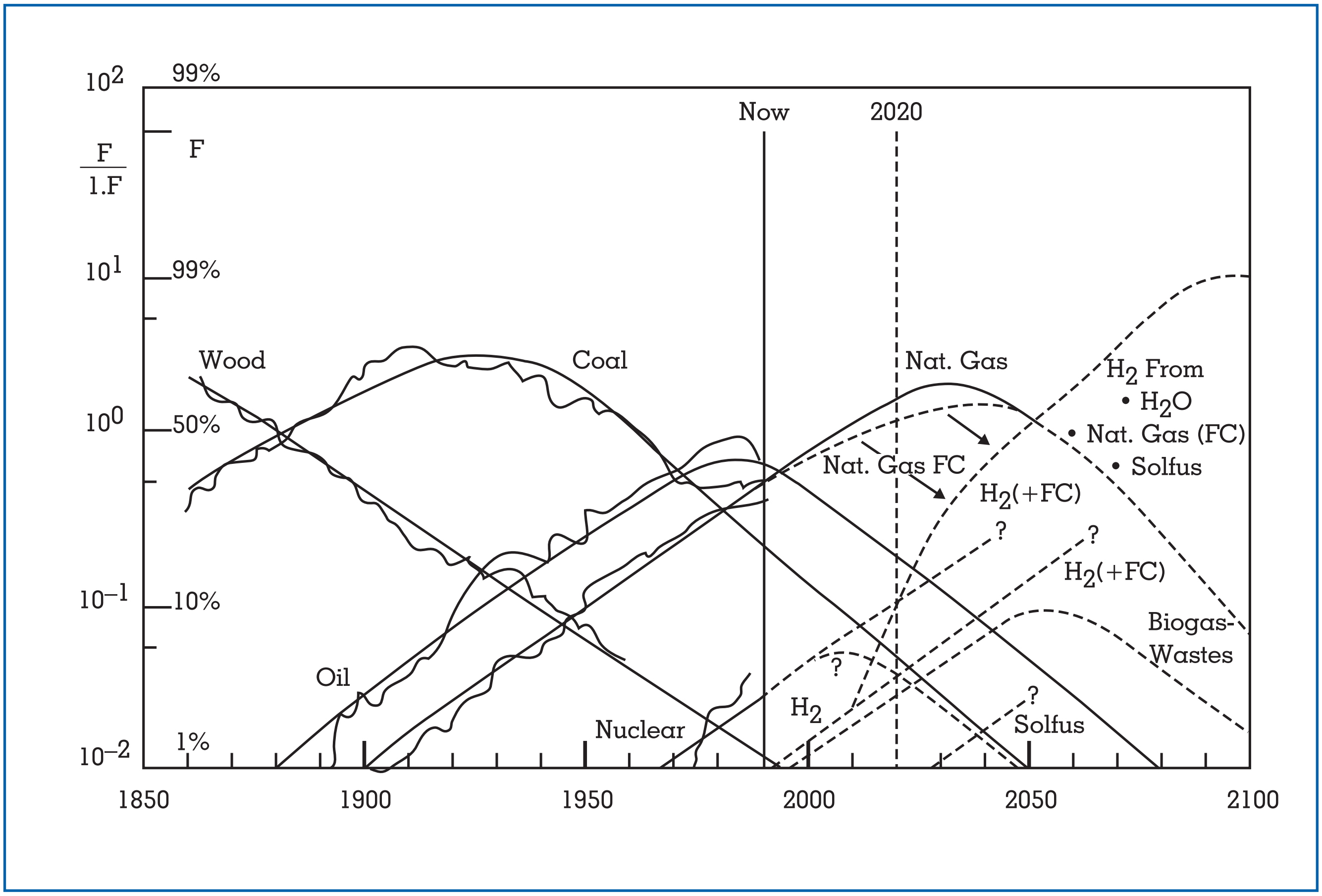


Figure 5.
Shift towards a hydrogen economy?

each efficiency percentage gain, this may not be economically justifiable.

System integration involves numerous miscellaneous development activities, such as control software to address system start-up, shut-down and transient operation, and thermal sub-systems to accomplish heat recovery, heat rejection and water recovery within the constraints of weight, size, capital and operating costs, reliability, and so on. Depending on the application, there will be additional key issues; automotive applications, for example, demand robustness to vibrations, impact, and cold temperatures, since if the water freezes it will halt fuel cell operation.

CONCLUSION

The same environmental drivers that are stimulating fuel cell development are also causing increased interest in alternatives. For example, in stationary power applications, microturbines are being developed that might compete with fuel cells for distributed power generation. In transportation, there is renewed interest in diesel engines, hybrid propulsion systems, and alternative-fueled vehicles. Advances in solar cells may also eliminate potential markets for fuel cells. Despite the strong, entrenched competition, there are reasons to believe that fuel cell commercialization is inevitable. Perhaps the strongest energy trend is the gradual shift toward renewable hydrogen. For example, wind-generated electricity can be used for electrolyzing water and, despite the extra step, the *potential* advantage of hydrogen over electricity is its easier transmission and storage. As Figure 5 indicates, the process of using fuels with ever-increasing H:C ratio has been developing for over 100 years. Recently, such a shift has been reinforced by environmental arguments. Should hydrogen evolve to be the fuel of the future, there is a compelling case to convert this into end-use electricity using fuel cells.

It must be recognized, however, that hydrogen usage for transportation will always create a trade-off between energy efficiency (fuel economy) and energy density (range), whereas this trade-off is non-existent for other applications.

In summary, fuel cell development is being accelerated both by the wide variety of applications and by the search for cleaner and more efficient utilization of primary energy and, ultimately, renewable energy. Because these forces for change are unlikely to disappear, it is quite likely that fuel cells will emerge as one of the most important and pervasive power sources for the future.

Christopher E. Borroni-Bird

BIBLIOGRAPHY

Appleby, A. J., and Foulkes, F. R. (1989). *Fuel Cell Handbook*. New York: Van Nostrand Reinhold Co.

FUEL CELL VEHICLES

ENVIRONMENTAL CONCERNS CONVERGE ON TECHNOLOGY PROGRESS

International concerns about the release of greenhouse gas emissions, deteriorating air quality, and reliance on oil sources in the Middle East oil continue to place pressure on the automobile industry to develop cleaner and more efficient vehicles. In practice there are three main ways this can be achieved: vehicle weight reduction, deployment of more efficient propulsion systems, and cleaner fuels. Fuel cell technology, by virtue of its unique means of operation, is well placed to address the power train limitations of both conventional internal combustion (IC), engines (i.e., nonzero emissions and relatively low efficiency) and batteries (i.e., inadequate range and long “recharging” time).

Many of the world’s major automakers, prompted by both this consumer demand and progress in reducing the inherent cost and size of fuel cells, are now committed to developing and commercializing fuel cell vehicles.

FUEL CELL BASICS

A fuel cell creates electricity directly from a fuel, usually hydrogen. The hydrogen is fed into one side of the fuel cell (the anode) where electrons are stripped off it to produce hydrogen ions (protons). In a vehicle, these electrons energize a motor to turn the wheels, and then they return to the cathode to combine with incoming oxygen from the air to produce oxygen anions. Meanwhile, the protons pass through the electrolyte and link up with the oxygen ions to produce water. The net effect is identical to the combustion of hydrogen in air except that the transfer of electrons has been channeled directly into making electricity, as opposed to heat.

As with batteries, differences in electrolytes create several types of fuel cells. The automobile’s demanding requirements for compactness and fast start-up have led to the Proton Exchange Membrane (PEM) fuel cell being the preferred type. This fuel cell has an electrolyte made of a solid polymer.

CHALLENGES TO FUEL CELL VEHICLE COMMERCIALIZATION

The biggest challenge facing fuel cells is cost. Current fuel cell systems probably cost about ten times as much as the $25/kW target set by the IC engine. Novel materials and manufacturing processes must be developed because mass production alone will not drive fuel cell stack (series of fuel cells combines to generate a useable voltage) costs down to the required level. Inexpensive bipolar plate materials, more effective catalyst utilization, and refined flowfields for efficiently channelling the input gases and exhaust water will contribute to the cost reduction.

It is expected that the fuel cell should be able to compete with an IC engine in terms of size and weight. As an added advantage, many fuel cell components can be configured into a relatively wide array of shapes to take advantage of space onboard the vehicle.

A fuel cell system also needs ancillaries to support the stack, just as an IC engine has many of the same type of ancillary subsystems. Major subsystems are needed for providing adequate humidification and cooling, and for supplying fuel and oxidant (air) with the correct purity and appropriate quantity.

These subsystems profoundly affect the fuel cell system performance. As an example, the inherently slow air (oxygen) electrode reaction must be accelerated using an air compressor, but this creates noise and exacts high power consumption, typically 10 to 15 percent of the fuel cell stack output. Fuel cell *system* efficiency, therefore, drops drastically when the fuel cell is operating at light load (below 10% rated power) even though the *stack* efficiency increases under these conditions. Low-cost, highly efficient air compressors that exploit the hot compressed exhaust gases may have more impact on increasing system efficiency than stack improvements would have.

Although the fuel cell is more efficient than an IC engine, it does produce waste heat at low temperature, and this poses problems from a heat rejection standpoint. Because there are vehicle styling constraints on how large the radiator can be, there is significant research into raising fuel cell stack operating temperatures by up to 10°C. If this can be achieved without degrading the electrolyte's durability, then it may almost halve the temperature differential between waste heat and ambient air in extremely hot conditions, with corresponding reductions in radiator size.

An automobile must also operate reliably in freezing conditions, which means that the humidifying, de-ionized water may need draining from the fuel cell system at key-off and reinjection during start-up. The coolant must also be freeze tolerant. These requirements for immediate power will probably force the vehicle to contain an energy storage component, such as a battery, flywheel, or ultracapacitor. This hybridization should also boost fuel economy because it enables regenerative braking and reduces the time that the fuel cell system spends below 10% rated load (where it is least efficient and where the driven car spends much of its time).

The vehicle also puts constraints on the choice of fuel. A combination of factors, such as the anticipated fuel and vehicle cost, both vehicular and "well-to-wheels" emissions and efficiency, perceived or real safety, and infrastructure will determine the choice of optimum fuel.

FUEL SELECTION AFFECTS PERFORMANCE AND COMMERCIALIZATION

Environmentally benign hydrogen is clearly the ideal fuel for use with fuel cells. The question is whether to store the hydrogen on-board or to generate it on board from liquid fuels that are easier to store and distribute.

The problem is that hydrogen, even at 10,000 psi (or 690 bar), requires five to ten times the volume of today's gasoline tank, depending on the fuel cell vehicle's real world efficiency. Packaging volume is compromised even further because pressurized tanks require thick carbon fiber walls and are, therefore, nonconformable. Moreover, they may cost several thousand dollars more than a conventional gasoline tank.

Although liquifying the hydrogen to 20K (–253°C) does increase energy density, it also creates new problems, such as cryogenic handling and venting (as ambient heat passes through the walls of the insulated vessel and evaporates the liquid). In addition, liquefaction is highly energy intense.

The third method of storing hydrogen is to absorb it into a metal. The only metal hydrides that can currently liberate their absorbed hydrogen at 80°C (allowing exploitation of the waste heat from the PEM fuel cell) are low-temperature ones and these can only store around 2% hydrogen by weight. Since 5 kg (11 lb) of hydrogen may be required to provide 380 miles (610 km) range in an 80 mpg (3 l/100 km) vehicle, the hydride will weigh more than 250 kg (550 lb). This weight not only reduces fuel economy but adds significant cost.

Hydrogen can also be adsorbed onto activated carbons; storage occurs in both gaseous and adsorbed phases. Existing carbon adsorbents outperform compressed gas storage only at relatively low pressures, when most of the gas is stored in the adsorbed phase. At pressures above 3,000 psi (207 bar), which are needed to store significant amounts, the adsorbent tends to be counterproductive since it blocks free space. Carbon adsorbents and metal hydrides introduce several control issues, such as risk of poisoning, and heat management during refueling and discharging.

Improved hydrogen storage is the key to a hydrogen economy. Without adequate storage, hydrogen may remain a niche transportation fuel which in turn, could limit the development of a hydrogen infrastructure. Of course, hydrogen could still play a leading role in *stationary* power generation without the need for dramatic improvements in storage.

Another concern with hydrogen is safety. Reputable organizations with extensive experience in using hydrogen have concluded that hydrogen is often safer than gasoline because it has a tendency to rapidly disperse upwards. However, its invisible flame and wide flammability range cause legitimate concerns. Reversing hydrogen's image requires a combination of public education and real world vehicle demonstrations. Automakers must design safe ways to store hydrogen on board. Despite these storage and safety concerns, hydrogen is considered the most attractive fuel for fuel cell buses because central refueling is possible and there is adequate space for hydrogen storage on the roof, which is also probably the safest location. However, for cars and light-duty trucks used by individuals, there is great interest in generating hydrogen on board the vehicle from fuels that are easier to store and transport.

Natural gas (CH_4) is also bulky to store, but reacting it with steam generates methanol (CH_3OH), an ambient-temperature liquid fuel. Having already been manufactured from natural gas outside the vehicle and containing no C-C bonds, methanol can generate hydrogen relatively easily and efficiently on board the vehicle using the steam produced by the fuel cell ($CH_3OH + H_2O \rightarrow 3H_2 + CO_2$). This hydrogen is then fed into the fuel cell to generate electricity. Unlike alkaline fuel cells used in the Apollo space missions, the carbon dioxide diluent can be fed into the PEM fuel cell without harm. Another attraction of methanol is the potential to ultimately eliminate the complexity of a reformer by using a direct methanol fuel cell where methanol is converted into hydrogen inside the cell. Because of these advantages, methanol is currently the leading candidate to propel mainstream fuel cell vehicles, absent a hydrogen storage breakthrough.

However, there are several issues with widespread methanol usage. Methanol production from natural gas is relatively inefficient (~67%), and this largely offsets the vehicular improvement in efficiency and carbon dioxide reduction (since gasoline can be made with ~85% efficiency from oil). Additionally, the PEM fuel cell demands very pure methanol, which is difficult to deliver using existing oil pipelines and may require a new fuel distribution infrastructure.

Compared with methanol, gasoline is more difficult to reform into hydrogen because it contains C-C (carbon-carbon) bonds. Breaking these down to generate hydrogen requires small amounts of air (called partial oxidation or POX) and this lowers the efficiency and makes heat integration more difficult. Also, gasoline is not a homogenous compound like methanol. Its sulfur poisons the fuel processor and fuel cell stack catalysts, and its aromatics have a tendency to form soot. Finally, gasoline's hydrogen-carbon (H:C) ratio is under two whereas methanol's is four. Gasoline, therefore, it yields less hydrogen and this reduces fuel cell stack output and efficiency. The net effect is that gasoline fuel cell vehicles are likely to be less efficient and more complex than methanol fuel cell vehicles. However, use of gasoline keeps the refueling infrastructure intact, and even gasoline fuel cell vehicles offer potential to be cleaner and more efficient than gasoline engines.

A fuel closely related to gasoline is naphtha, which is also a potential fuel cell fuel. Naphtha is already produced in large quantities at refineries and is a cheaper fuel than gasoline, which must have octane-boosting additives blended into it. Unlike methanol, naphtha can be distributed in the same pipelines as gasoline. From the fuel cell's perspective, it has a higher H:C ratio and lower sulfur and aromatics content than gasoline.

Given the technical challenges facing gasoline reforming, it is possible that gasoline fuel cells may be introduced to vehicles as an Auxiliary Power Unit (APU) feature. A 5 kW APU that provides electricity to power air conditioning and other electrical loads can operate with less demanding transient performance than is necessary for propulsion. Moreover, lower efficiency can be tolerated since it competes with a gasoline engine-alternator combination at idle and not with a gasoline engine directly.

The transition from the IC engine to fuel cells will not alleviate energy supply insecurity. All of these fuel options—hydrogen, methanol, gasoline, and naphtha will be generated most economically from fossil fuels (natural gas or oil) in the near to mid-term. However, fuel cells are still expected to reduce carbon dioxide emissions since propulsion is not dependent on combustion. With respect to local air quality, fuel cell vehicles should significantly reduce pollutants, particularly if they use hydrogen. The only emissions from a hydrogen fuel cell vehicle would be pure water, thus allowing it to qualify as a Zero Emission Vehicle. In addition, fuel cell vehicles should stimulate development of cleaner fuels more quickly than IC engines because they are less tolerant of fuel impurities and benefit more from hydrogen-rich fuels.

Christopher Borroni-Bird

See also: Fuel Cells.

BIBLIOGRAPHY

Borroni-Bird, C. E. (1996). “Fuel Cell Commercialization Issues for Light-Duty Vehicle Applications.” *Journal of Power Sources* 61:33–48.

FUEL ECONOMY

See: Automobile Performance

FUEL OIL

See: Residual Fuels

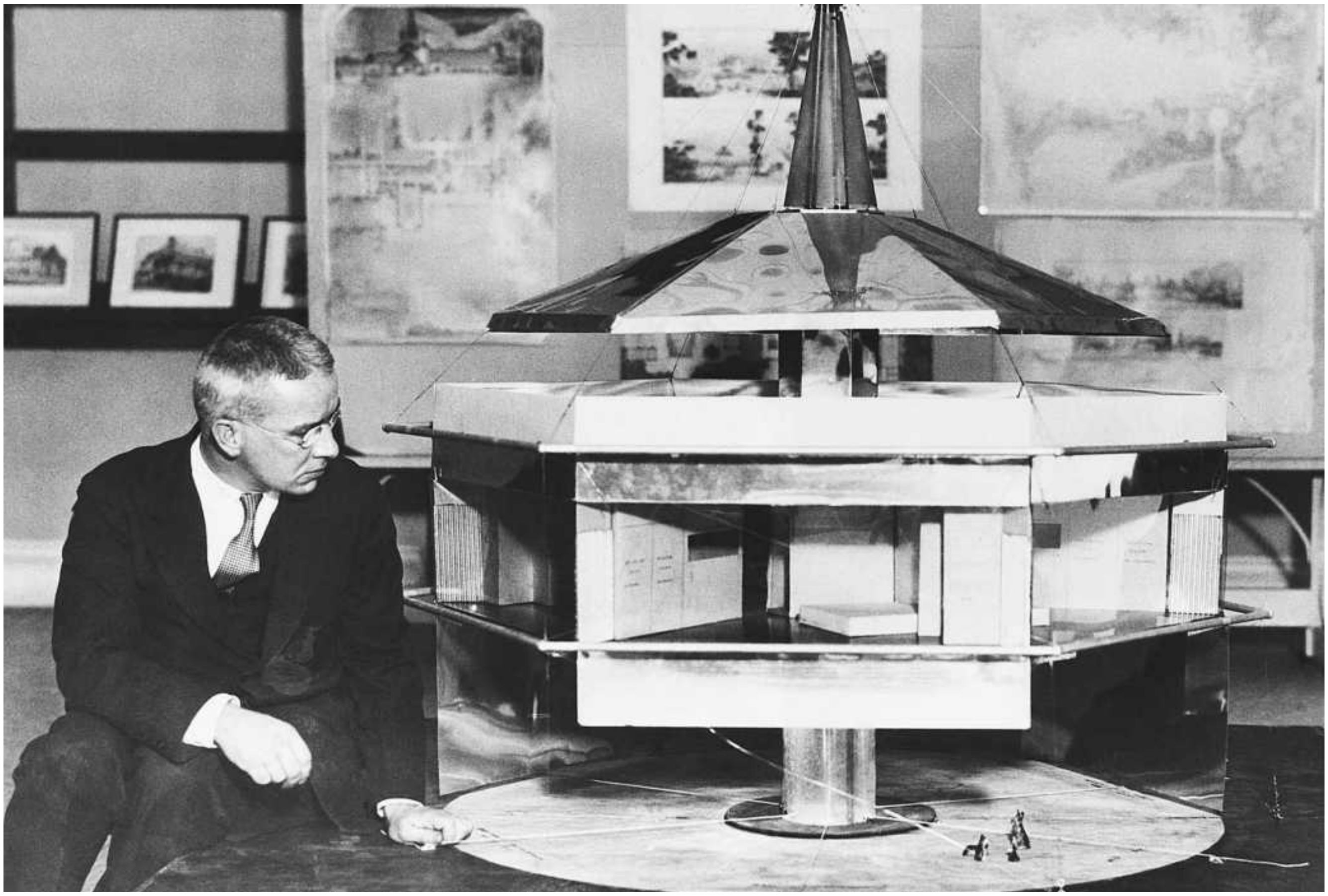
FULLER, R. BUCKMINSTER, JR. (1895–1983)

Richard Buckminster Fuller, Jr., best known as the architect of Houston’s Astrodome and other geodesic structures, enjoyed a long and varied career as a structural engineer and unconventional humanistic thinker. A colorful and gregarious individual, Fuller was first embraced by government officials for his innovative designs and later cherished by the 1960s counterculture. He patented more than twenty new inventions, authored twenty-five books and dozens of articles, lectured globally on energy issues and the wise use of world resources, and dabbled in both art and science. Never one to be modest, Fuller called himself “an engineer, inventor, mathematician, architect, cartographer, philosopher, poet, cosmologist, comprehensive designer and choreographer.” He especially liked the self-description “anticipatory comprehensive design scientist,” because he saw himself as a scientist who anticipated human needs and answered them with technology in the most energy-efficient way.

Fuller’s “more with less” philosophy first gained the attention of Americans in the 1920s with his “Dymaxion” inventions. Fuller employed this term—a combination of “dynamic,” “maximum,” and “ion”—to describe inventions that do the most with the least expenditure of energy. His Dymaxion house was self-sufficient in that it generated its own power, recycled water, and converted wastes into useable energy. It also featured air conditioning and built-in laborsaving utilities including an automatic laundry machine, a dishwasher that cleaned, dried, and reshelved the dishes, and compressed air and vacuum units. The Dymaxion bathroom was a one-piece aluminum unit containing a “fog gun”—an economical showerhead using a mixture of 90 percent air and 10 percent water. Fuller’s Dymaxion car ran on three wheels, had front-wheel drive with rear-wheel steering, and registered a reported forty to fifty miles per gallon in 1933. None of these projects were mass-produced, but his futuristic designs, and later his geodesic buildings, World Game workshops, and startling questions and proposals about natural resources and human survival stimulated imaginations and encouraged others to explore ways to create a more energy-efficient environment.

Born in Milton, Massachusetts, in 1895 to Richard and Caroline Fuller, Bucky, as most knew him, spent his summers at the family retreat, Bear Island, in Penobscot Bay, Maine. As Fuller remembered it, his interest in building better “instruments, tools, or other devices” to increase the “technical advantage of man over environmental circumstance” began there. One of his tasks each day as a young boy was to row a boat four miles round trip to another island for the mail. To expedite this trip, he constructed his “first teleologic design invention,” a “mechanical jelly fish.” Noting the structure of the jellyfish and attending to its movement through the water, Fuller copied nature and produced a boat of greater speed and ease. Observing natural phenomena remained his lifelong source of inspiration which was not surprising, given that Fuller was the grandnephew of transcendentalist Margaret Fuller.

Following his Brahmin family’s tradition, Fuller went to Harvard, but instead of graduating as a lawyer or Unitarian minister as his ancestors had, he was expelled twice for cutting classes, failing grades, and raucous living. He never did gain a bachelor’s degree. Instead, Fuller tutored himself in the arts and



Buckminster Fuller, sitting beside a model of his "Dymaxion" house, seen in the 1930 World's Fair in Chicago. (Corbis Corporation)

sciences, and the Navy and apprenticeships at a cotton mill machinery plant and a meat-packing factory provided him with a practical education. Eventually, Fuller garnered multiple honorary doctorates and awards, including a Presidential Medal of Freedom shortly before his death in 1983.

In 1917, Fuller married Anne Hewlett, the daughter of the respected New York architect James M. Hewlett. A year later their daughter Alexandra was born, only to die a few years later from infantile paralysis. From 1922 to 1926, Fuller and his father-in-law founded and ran the Stockade Building System that produced lightweight construction materials. Fuller failed miserably at the business and, in 1927, stung by the death of his daughter, years of carousing, and financial failure, he considered suicide. Impoverished and living in the gangster region of Chicago with his wife and newborn daughter, Allegra, Fuller walked to the shore of Lake Michigan with the intent of throwing himself in. Instead, he had a revelation: "You do not have the right to eliminate yourself, you do not belong to you. You belong to the universe. . . apply yourself to converting all your experience to the highest advantage of others." At this point, Fuller made it his ambition to design "tools for living," believing that human life would improve if the built environment was transformed.

Fuller approached his mission both philosophically and practically. He explored mathematics in search of "nature's coordinate system" and invented his own "energetic-synergetic geometry" as his expression of the underlying order he saw. He also studied philosophy and physics for understanding of time and motion. But his philosophical musings always had a practical bent. For example, Fuller, taken with Albert Einstein's theories about time and motion, sought to employ physics in his design initiatives.

Fuller devoted his early years to the problem of building energy-efficient and affordable housing, but his projects and ideas were largely confined to a few students in architectural schools, executives of small corporations, and readers of *Shelter* and *Fortune* magazines. His most successful enterprise was the manufacture of Dymaxion Deployment Units for use by the Army in World War II. Fuller tried to convert these military units into civilian housing, but failed to raise production funds.

In 1949, Fuller's luck changed with the construction of his first geodesic dome at Black Mountain College. Motivated by a desire to create an energy-efficient building that covered a large space with a minimal amount of material, Fuller fashioned a sphere-shaped structure out of triangular pyramids. He joined these tetrahedrons together by pulling them tight in tension rather than relying on compression to raise the dome. In doing so, Fuller reconceptualized dome engineering and created a practical, efficient, cost-effective, strong, and easy-to-transport-and-assemble building.

Several hundred thousand geodesics function today worldwide as auditoriums, aviaries, banks, churches, exposition halls, greenhouses, homes, industrial plants, military sheds, planetariums, playground equipment, and sports arenas. Some of these geodesics, including the clear, thin-skinned "Skybreak" and "Garden-of-Eden" structures, were experiments in using renewable energy sources. These transparent spheres, primarily functioning as greenhouses or as experimental biodomes by environmentalists before the much-publicized Biosphere projects, relied on solar power to regulate temperature. Fuller also had more fantastic ideas to reduce energy losses in summer cooling and winter heating: in 1950, for example, he proposed building a dome over Manhattan to regulate the environment.

Fuller's vision was more expansive than creating better shelter systems, however. His scope was global and his ideals utopian. At the same time that he was creating "machines for living," he was surveying Earth's resources, designing maps, and plotting strategies for an equitable distribution of goods and services. In 1927 Fuller started his "Inventory of World Resources, Human Trends and Needs." In the 1930s he began his Dymaxion map projects to gain a global perspective, and by the 1960s, in the heart of the Cold War, Fuller was busy devising ways to ensure the survival of the earth.

For Earth to continue functioning for the maximum gain of people everywhere, Fuller believed that resources must be used wisely and shared equally. He scorned reliance on fossil fuels and encouraged the development of renewable energy sources, including solar, wind, and water power. To convey his ideas, Fuller employed the metaphor of Earth as a spaceship. This spacecraft, he explained in his best-selling book, *Operating Manual for Spaceship Earth*, was finite and in need of careful management. With limited materials on board, "earthians" must work out an equitable and just distribution to keep the spaceship operating smoothly and efficiently.

Intent upon channeling human energy and resources into projects for "livingry," as opposed to "weaponry," Fuller participated in cultural exchanges between the Soviet Union and the United States, worked on United Nations projects, and traveled the world to communicate his vision. He conducted marathon "thinking-aloud" sessions before large audiences, taking on a cult status for some and the role of crackpot for others. He criticized political solutions to world problems and promoted technological design to reallocate wealth, labor, and resources.

The culmination of these thoughts can be found in Fuller's strategic global planning organization, the World Game Institute. Counter to military war games, World Game was both a tool for disseminating information and an exercise to engage others in problem-solving. Through a set of simulated exercises, participants used Fuller's inventory of resources, synergetic geometry, and Dymaxion maps to plot strategies to "make the world work for 100% of humanity in the shortest possible time, through spontaneous cooperation, and without ecological offense or the disadvantage of anyone." Begun in small college classrooms in the 1960s, World Game attracted a dedicated following. In the 1990s, the World Game Organization established itself on the Internet to facilitate innovative thinking about world resources.

Fuller's lifelong engagement in global energy issues and his lasting contributions in engineering design make him a noteworthy study in twentieth century debates on energy supply and use. Fuller believed that world problems of war, poverty, and energy allocation could be eradicated by cooperation between nations and through technological innovation. His optimistic celebration of technology pitted him against individuals such as historian Lewis Mumford who questioned the salutary effects of technology, and placed him at odds with Paul Ehrlich and other environmentalists who believed that Earth's resources could not support the world's growing population. Fuller, a modernist at heart, maintained that the trouble was in distribution, and that continual employment of technology would not destroy but save the planet.

Linda Sargent Wood

BIBLIOGRAPHY

Baldwin, J. (1996). *BuckyWorks: Buckminster Fuller's Ideas for Today*. New York: Wiley.

Fuller, R. B. (1963). *Ideas and Integrities: A Spontaneous Autobiographical Disclosure*, edited by R. W. Marks. Englewood Cliffs, NJ: Prentice-Hall.

Fuller, R. B. (1963). *Operating Manual for Spaceship Earth*. New York: E. P. Dutton.

Fuller, R. B. (1969). *Utopia or Oblivion: The Prospects for Humanity*. New York: Bantam.

Fuller, R. B. (1975). *Synergetics: Explorations in the Geometry of Thinking*. New York: Macmillan Publishing Company.

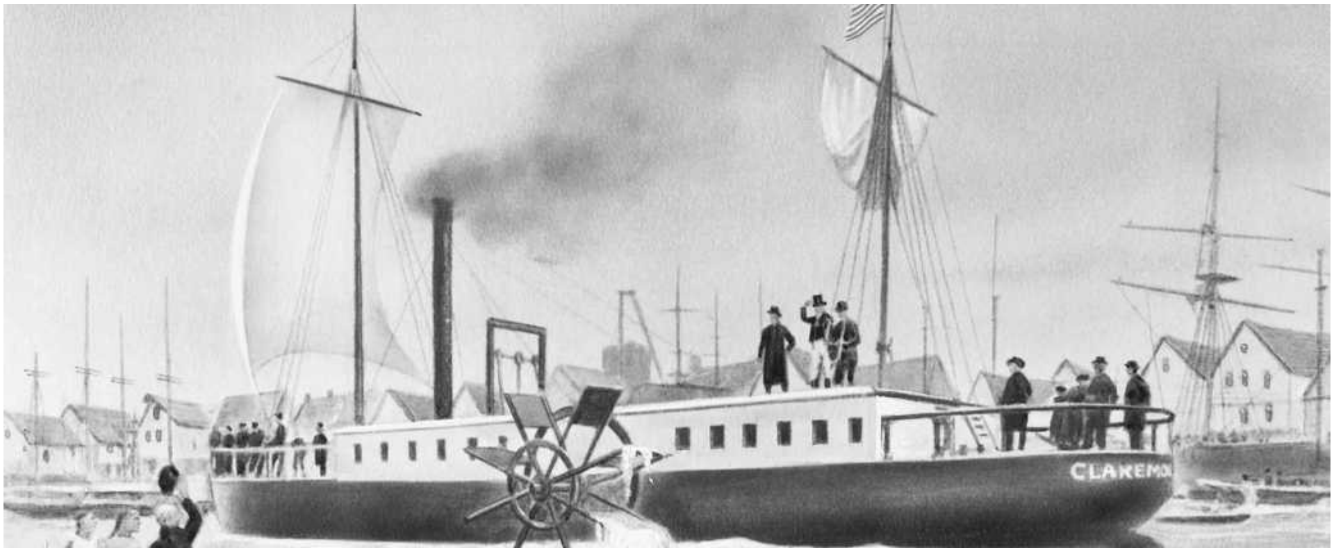
Gabel, M. (1975). *Energy Earth and Everyone: A Global Energy Strategy for Spaceship Earth*. San Francisco: Straight Arrow Books.

McHale, J. (1962). *R. Buckminster Fuller*. New York: George Brazillier, Inc.

Sieden, L. S. (1989). *Buckminster Fuller's Universe: An Appreciation*. New York: Plenum Press.

FULTON, ROBERT (1767–1815)

Robert Fulton was unique among inventors of his time, since he was born in America, apprenticed in Europe and returned to his native country to perfect his greatest invention, the steamboat. Fulton was born in 1765 of a respectable family in Little Britain, Pennsylvania, in Lancaster County. Unlike many inventors, he was not motivated by paternal leadership since he grew up fatherless after the age of three. His mother is credited for his soaring interest in the world of painting. Despite, or perhaps because of, the loss of his father, Robert's free time was spent with local mechanics or alone with his drawing pencil. By age seventeen, he was accomplished enough as a landscape and portrait artist that he earned income from it in the city of Philadelphia for four years. He earned enough to build a home for his mother in Washington County before receiving a generous



Robert Fulton's "Clermont." The illustrator mistakenly used the name "Claremont." (Corbis Corporation) offer to take his artistic talents to England in 1786, at age 21, where they were equally well received.

Over the next seven years, he made numerous acquaintances, including two who were instrumental in his transition from artist to civil engineer—the Duke of Bridgewater, famous for work with canals, and Lord Stanhope, who specialized in works of science and mechanical arts. By 1793, Robert was experimenting with inland navigation, an area that remained of interest throughout his life. A year later, he filed a patent in Britain for a double inclined plane, and spent several years in Birmingham, the birthplace of the Industrial Revolution that came about thanks to the powerful engines built by James Watt as early as 1763.

Remarkable for someone so talented in the arts, Fulton was able to use his accomplished drawing skills to express his designs. As a now noted draftsman, he invented numerous pieces of equipment such as tools to saw marble, spin flax, make rope and excavate earth. Sadly, he once lost many of his original manuscripts during a shipping accident.

By the beginning of the nineteenth century, Fulton turned his attention to his obsession with submarines and steamships. He made no secret of his goal for submarines—he intended to build them in order to destroy all ships of war so that appropriate attention could be devoted by society to the fields of education, industry and free speech. In 1801, he managed to stay under water for four hours and twenty minutes in one of his devices, and in 1805 he demonstrated the ability to utilize a torpedo to blow up a well built ship of two hundred tons. Unfortunately for Robert, neither the French nor British government was particularly impressed with the unpredictable success, nor the importance of his innovations; so he packed his bags and returned to America in December of 1806.

By the time Robert Fulton arrived in America, he had been studying steam navigation for thirteen years. Five years earlier, he had met Chancellor Livingston, who was partly successful in building a steam vessel, but yielded to Fulton's skill by allowing him to take out a patent in the latter's name to begin his improvements. By 1807, there was much skepticism and derision expressed along the banks of the East River in New York when Fulton prepared to launch his steamship, "The Clermont," from the ship yard of Charles Brown. A few hours later, there were nothing but loud applause and admiration at the spectacular achievement that had been termed "Fulton's Folly." Finding his own design flaws, Fulton modified the water wheels and soon launched the ship on a stunning three-hundred-mile round-trip run to Albany. The total sailing time was sixty-two hours, a remarkable endurance record for so new an invention. Soon there were regular runs being made on the Hudson River and by rival ships around the country, with many disputes over the patent rights of different vessels in different states. But the industry was clearly set in motion by Fulton's uncanny designs.

By 1812, he added a double-hulled ferryboat to his credits, designing floating docks to receive them. In 1814, the citizens of New York vindicated his years of war vessel experiments by demanding a protective ship in defense of their New York Harbor. In March of that year, the President of the United States, James Madison, was authorized to enlist Fulton to design a steam-powered frigate, with full explosive battery, as a military defense weapon. He also empowered Fulton to complete a design for the defensive submarine Fulton longed to finish. Unfortunately, just three months before the completion of the steam frigate, Fulton fell victim to severe winter exposure on the river, dying on February 24, 1815, in New York City at the age of 50.

The public mourning was said to be equal to that expressed only for those who held public office of considerable acclaim. But by this time, the word 'folly' had long been dropped from his name. Just as George Stephenson followed the inventors of locomotives to gain credit for creating Britain's railways, Robert Fulton was not the inventor of the steam boat. What he *was*, however, was the brilliant creator of such an improved version that it made possible the many advancements of steamships in succeeding years. He took theory out of the experimentation process and designed machinery that opened the door to unlimited practical use from the technology available at the time.

Dennis R. Diehl

BIBLIOGRAPHY

Howe, H. (1858). *Memoirs of the Most Eminent American Mechanics: Also, Lives of Distinguished European Mechanics.* New York: Harper & Brothers.

FURNACES AND BOILERS

Furnaces and boilers are devices that burn fuel to space heat homes, offices, and industrial facilities. Natural gas, liquefied petroleum gas, and heating oil are the dominant fuels used for furnaces and boilers. In the United States, furnaces and boilers burning gas and oil take care of over 75 percent of all space heating.

EARLY HISTORY

The first oil burning devices for heating appeared in the oil-rich Caucasus region of Russia as early as 1861. But because of the remoteness of this region, those devices remained in obscurity, out of the flow of marketable goods. An event almost 115 years ago marks the birth of the modern oil burner. On August 11, 1885, the U.S. Patent office granted a patent to David H. Burrell of Little Falls, New York, for a "furnace apparatus for combining and utilizing oleaginous matters." As noted in the June 1985 centennial issue of *Fuel Oil News*, Burrell's invention "...was the forerunner of today's modern oil burner, and is generally accepted as the one that started the oil heat industry."

Gas burners are different from oil burners in that they control the mixing of air and gas and are referred to as *aerated burners*. This type of burner was invented around 1855 by Robert Wilhelm von Bunsen and is often called a *Bunsen burner* or *blue flame burner*. Natural gas is mostly methane and typically found underground in pockets. Liquified petroleum gas (LPG) occurs in "wet" natural gas and crude oil and must be extracted and refined before use.

GAS AND OIL HEAT BASICS

Liquid fuels, including heating oil after it is refined from crude oil, are transported by pipelines, oil tankers, barges, railroads, and highway tanker trucks to local bulk storage facilities, and delivered by tank trucks to the homeowner's tank, which is located either underground or in the basement, garage, or utility room. Figure 1 shows a typical residential oil heating system. These tanks typically store 275 to 500 gallons of fuel that can supply heat for a typical house for a month, and usually several oil tank truck deliveries to the home are needed during the year. LPG is similarly transported under pressure as a liquid by tanker trucks to the homeowner's above-ground tank. LPG may also be delivered in portable tanks typically 100-250 lbs, or as small as the 20-lb. tanks, used for portable gas grills. LPG is heavier than air and should not be stored inside a house for safety reasons.

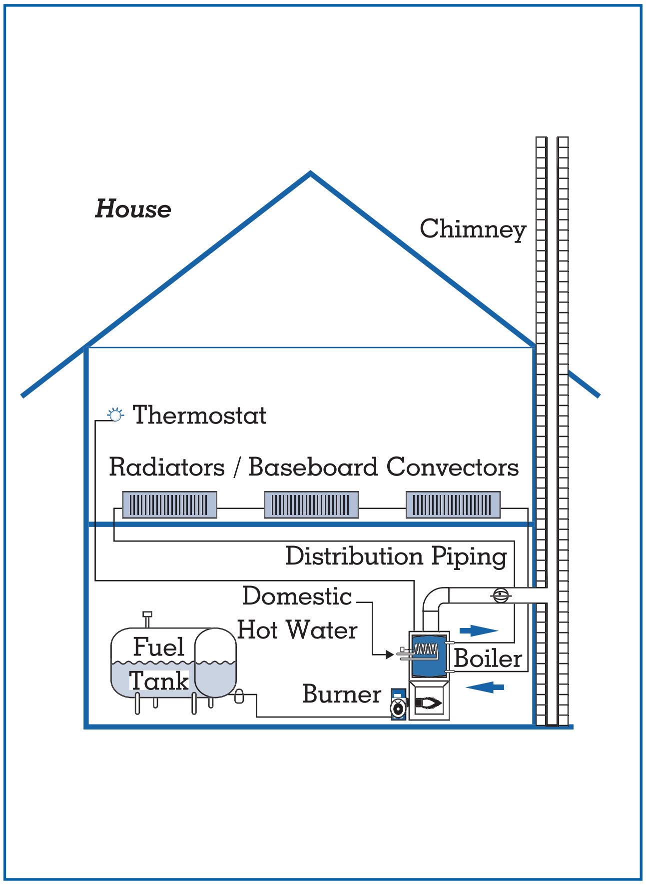


Figure 1.
Oil heating system.

Natural gas is transmitted in pipelines first across the country under high pressure in transcontinental pipe lines to local gas utility companies. The gas utility company reduces the pressure and distributes the gas through underground mains, and again reduces pressure from the street main into the home.

Gas or oil fuel is burned in a furnace or boiler (heat exchanger) to heat air or water, or to make steam that carries the heat absorbed from the combustion process to the rooms in the house, as shown in Figure 2. This is accomplished by circulating the warm air through ducts directly to the rooms, or by circulating hot water or steam through pipes to baseboard hot water convector units or radiators in the room. The boiler loop is completed with warm water returned through pipes to the boiler, or return air through the return register to air ducts back to the furnace to be reheated. In recent years, radiant floor heating systems that also rely on hot water circulation have come back into use. Boilers may also serve the additional function of heating water for showers and baths. A water heating coil immersed within the boiler is often used to heat this water as needed at a fixed rate. Another option is to have the coil immersed in a very well insulated water tank. The tank can provide a large supply of hot water at a very high rate of use. This is accomplished by exchanging heat from boiler water that passes through the inside of the coil to the domestic hot water stored in the tank. This method can provide for more hot water since a full tank of hot water is standing by for use when needed. The boiler operates only when the tank temperature drops below a certain set point. This results in fewer burner start-ups and is a more efficient way to heat water during the non-heating season months, enhancing year-round efficiency.

A recent alternative to the furnace or boiler for space heating is to use the hot water heater to heat the house. A separate air handler is sold along with a pump that pumps hot water from the water heater to the air handler where hot water flows through a finned tube heat exchanger coil. A blower in the air handler pushes air over the hot water coil and then through ducts to the rooms. Water out of the coil returns back to the water heater to be reheated. This way the water heater does double duty: it operates as a combination space heater and water heater. Gas water heaters can be used this way in mild climates and for homes or apartments with smaller heating requirements. Oil-fueled water heaters can also be used this way to handle homes in colder climates because they have higher firing rates and re-heat water faster. The installation cost is much lower for this approach in new construction, but units with high efficiency are not available for the gas water heaters as they are for gas furnaces or boilers.

Furnaces and boilers sold today must by law have annual fuel utilization efficiency of at least 78 to 80 percent. Gas water heaters operating this way as space heaters are equivalent to the efficiency of pre-1992 furnaces and boilers which had space heating efficiencies typically in the mid-60 percent range. However, the combined efficiency for space and water heating together may partly make up for the inefficiency of the water heater alone. There is also a concern regarding the life expectancy of water heaters, typically 5 to 10 years as compared to a boiler that will last anywhere from 20 to 40 years (or longer) depending on its design.

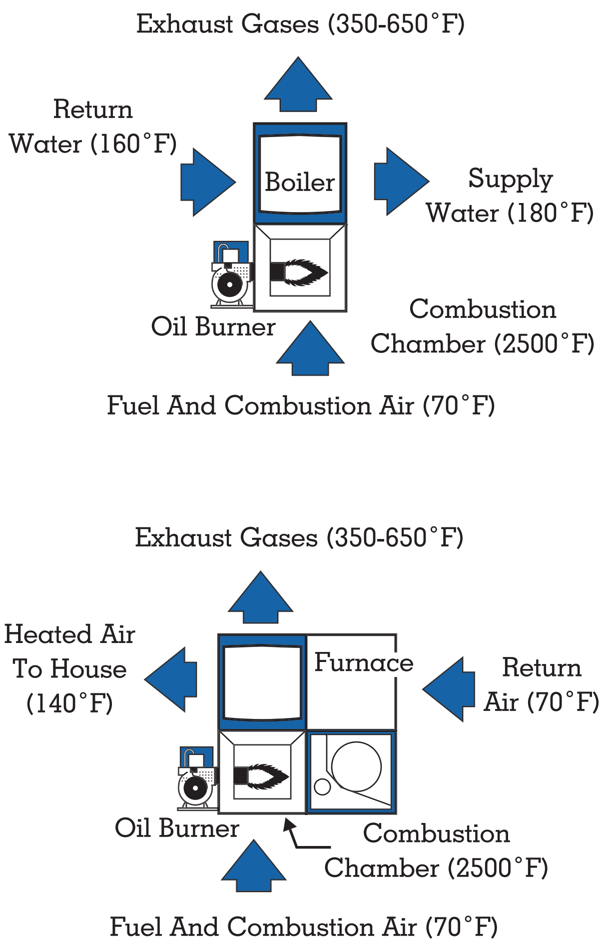


Figure 2.
Typical boiler or furnace operation
NOTE: Temperatures can vary.

Both gas and oil-fired heating systems consist of several subsystems. The oil burner pump draws fuel from the tank; the fuel is atomized and mixed with just the right amount of air for clean combustion; an electric spark provides the ignition of the fuel/air mixture; and the flame produces the heat in a combustion chamber. The heat is then released to the heat exchanger to heat air in the furnace and water in the boiler or to make steam. The process is similar for gas that is supplied under regulated gas pressure into the house piping to the furnace. The sequence of operation for the system starts at the wall thermostat. When the room temperature drops below the set point, a switch closes allowing the gas valve to send gas to the burners. With almost all new gas furnaces today, a spark or glowing hot wire igniter typically lights a gas pilot, or a spark system ignites the oil spray. The presence of the ignition source or pilot is proven, and the gas valve opens, sending gas mixed with air to the burner for ignition. The flame sensing system must be satisfied that the flame ignites or else it will quickly shut the burner off to prevent significant amounts of unburned fuel from accumulating in the heating appliance. The flue pipe connects the unit to a chimney to exhaust the products of combustion from the furnace into the outside ambient environment. A side wall vented flue pipe made of a plastic is used for some high efficiency condensing furnaces. The combustion of natural gas and fuel oils, which are primarily hydrocarbon molecules, results primarily in the formation of water vapor and carbon dioxide. The process also results in very minor amounts of sulfur dioxide (which is proportionate to the sulfur content of fuel oil, typically less than one half of one percent by weight), and trace amounts of other sulfur oxides, nitrogen oxides, and particulate matter (unburned hydrocarbon-based materials) in the range of parts per million to parts per billion or less.

The furnace blower or boiler circulator pump starts up to send heated room air or hot water through the ducts or hot water pipes to the steam radiators, or hot water baseboard units, or individual room air registers located throughout the house. When the thermostat is satisfied that the room temperature has reached the set point, the burner shuts off. In furnaces, the blower continues to run a few seconds until the air temperature drops to about 90° F, then the blower also shuts off. The furnace blower may come on again before the next burner start-up to purge heat out of the furnace, particularly if the fan has a low turn-on set point. The cycle repeats when the room temperature drops below the set point. In boilers, heat may be purged from the boiler by running the circulator for a controlled amount of additional time, delivering either more heat to the living space or hot water to a storage tank. Heat will continue to be supplied to the room from the radiators or hot water baseboard units until they cool down to room temperature.

U.S. HOUSEHOLD ENERGY CONSUMPTION AND EXPENDITURES

There were 10.8 million U.S. households that used fuel oil for space and/or water heating in 1993. The average household using fuel oil typically comsumed 684 gallons a year for space heating. In addition, 3.6 million U.S. homes used kerosene. Together they represented 14.4 million homes or 15 percent of the 96.6 million households in the United States. These households consumed a total of 7.38 billion gallons of heating fuel and 340 million gallons of kerosene.

There were 52.6 million U.S. households that used natural gas for space and/or water heating and/ or use in other appliances (mainly cooking ranges) in 1993. In addition there were 5.6 million U.S. homes that used LPG. Together they represented 58.2 million homes or 60 percent of the 96.6 million households in the United States. These households consumed a total of 4,954 billion cubic feet of natural gas and 3.84 billion gallons of LPG for space and water heating. The average household using natural gas typically uses about 70 million Btu a year for space heating.

Whereas heated homes are mostly located in the colder regions of the United States like the north and northeast, natural gas and LPG are used for heating homes through out the United States in both warmer and colder climates.

EFFICIENCY

Most new gas and oil-fueled furnaces and boilers have similar efficiencies. The range of efficiency has narrowed with the introduction of minimum efficiency standards for new products sold since 1992. New gas and oil heating equipment currently available in the marketplace have Annual Fuel Utilization Efficiency (AFUE) ratings of at least 78 to 80 percent. AFUE is a measure of how efficient a furnace operates on an annual basis and takes into account cycling losses of the furnace or boiler. It does not include the

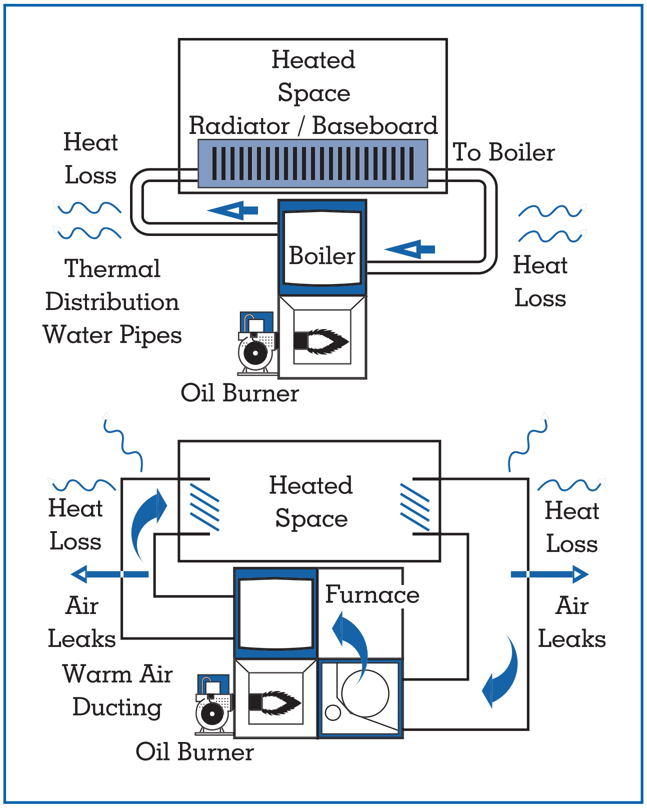


Figure 3.
Heat loss from distribution pipes and air ducts.

distribution losses in the duct system, which can be as much as 30 percent or more, especially when ducts are located in unheated areas such as the attic or unheated crawl space (see Figure 3). The AFUE also does not include the electricity to operate the blowers and any other electrical motors, controls or ignitors. The AFUE for gas furnaces are clustered in two groups: those in the range of 78 to 82 percent AFUE and those in the range of 90 to 96 percent. Few if any gas furnaces are available between these ranges because the upper range is a condensing furnace that requires an additional heat exchanger of a special alloy steel, a condensate pump and drain, and a special flue pipe material.

Because of the price diffferential between low- and high-efficiency condensing furnaces, only 22 percent of gas furnaces sold in the mid-1990's were high-efficiency condensing-type furnaces. Condensate is water that forms as a result of the combustion process. When the hydrogen in the fuel combines with oxygen from the combustion air, it forms water vapor (steam) that is cooled and condensed in a condensing furnace during the heat exchange process. Carbon dioxide in the flue gases makes this water act like soda water and becomes acidic. Fuel oil contains less hydrogen than natural gas and is therefore less likely to condense water vapor in the flue gases when operated in the mid-80 percent range. Therefore oil furnaces with up to 87 percent AFUE are available without condensing. The availability of condensing oil-fueled furnaces is limited since there is little efficiency benefit to be gained from the additional cost of a special alloy heat exchanger.

Some new features of furnaces available today include variable speed blowers, which deliver warm air more slowly and more quietly when less heat is needed, and variable heat output from the burner, which when combined with the variable speed blower allows for more continuous heating than the typical fixed firing rate. Distribution system features can be sophisticated with zoned heating which employs a number of thermostats, a sophisticated central controller, and a series of valves or dampers that direct airflow or water to different parts of the home only when needed in those areas.

The average AFUE of all installed furnaces is 65 to 75 percent, much lower than post-1992 efficiency standards due to the different vintages of furnaces and boilers. Systems that are 40 years old or older are even less efficient (55-65%), but these represent a very small fraction of furnaces operating in the United States today.

The U.S. Department of Energy develops test procedures for efficiency measurements and sets minimum efficiency standards for furnaces, boilers, and water heaters. Information on energy efficiency of buildings and equipment is available from the DOE.

Esher Kweller
Roger J. McDonald

BIBLIOGRAPHY

American Council for an Energy Efficient Economy. (1989). *The Most Energy Efficient Appliances.* Washington, DC: Author.

American Council for an Energy Efficient Economy. (1989). *Oil and Gas Heating Systems.* Washington, DC: Author.

Beckett Corporation. (1979). *The Professional Serviceman's Guide to Oil Heat.* Elyria, OH: Author.

Energy Efficiency and Renewable Energy Network (EREN) Home Page. Washington DC: U.S. Department of

Energy, Energy Efficiency and Renewable Energy Network. August 4, 2000 <http://www.eren.doe.gov/>.

Lanthier, George. (1995). *Combustion and Oil Burning Equipment: An Advanced Primer*. Arlington, MA: Firedragon Enterprises.

Petroleum Marketer Association of America. (1995). *Advanced Oil Heat: A Guide to Improved Efficiency*. Arlington, VA: Kendall/Hunt Publishing.

Trethwey, Richard. (1994). *This Old House: Heating, Ventilation, and Air Conditioning*. Boston: Little, Brown.

Weinberger, Herbert. (1992). *Residential Oil Burners*. Albany, NY: Delmar Publishers.

FUSES

See: Electricity

FUSION, BOMB

See: Nuclear Fusion

FUSION, POWER

See: Nuclear Fusion

FUTURES

OVERVIEW

A futures contract is an agreement that calls for a seller to deliver to a buyer a specified quantity and quality of an identified commodity, at a fixed time in the future, at a price agreed to when the contract is made. An option on a commodity futures contract gives the buyer of the option the right to convert the option into a futures contract. Energy futures and options contracts are used by energy producers, petroleum refiners, traders, industrial and commercial consumers, and institutional investors across the world to manage their inherent price risk, to speculate on price changes in energy, or to balance their portfolio risk exposure.

With very limited exceptions, futures and options must be executed on the floor of a commodity exchange through persons and firms registered with regulatory authorities. The contracts are traded either by open outcry, where traders physically transact deals face to face in specified trading areas called pits or rings, or electronically via computerized networks. The futures market provides a standardized trading environment such that all users know exactly what they are trading and where their obligations and risks lie. By entering into a standard futures and/or options contract, a certain amount of price assurance can be introduced into a world of uncertainty and price volatility.

Most futures contracts assume that actual delivery of the commodity can take place to fulfill the contract. However, some futures contracts require cash settlement instead of delivery. Futures contracts can be terminated by an offsetting transaction (i.e., an equal and opposite transaction to the one that opened the position) executed at any time prior to the contract's expiration. The vast majority of futures contracts are terminated by offset or a final cash payment rather than by physical delivery.

HISTORICAL PERSPECTIVE

Futures contracts for agricultural commodities have been traded in the United States since the nineteenth century and have been under federal regulation since the 1920s. Starting in the late 1970s, futures trading has expanded rapidly into many new markets, beyond the domain of traditional physical and agricultural commodities such as metals and grains. Futures and options are now offered on many energy commodities such as crude oil, gasoline, heating oil, natural gas, and electricity, as well as on a vast array of other commodities and financial instruments, including foreign currencies, government securities, and stock indices.

TERMS AND CONDITIONS

A typical futures contract might call for the delivery of 1,000 barrels (42,000 U.S. gallons) of unleaded gasoline meeting defined specifications at petroleum product terminals in New York Harbor during the next twelve months at an agreed price in dollars and cents per gallon. All terms and conditions other than the price are standardized. Gasoline is sold through hundreds of wholesale distributors and thousands of retail outlets, and is the largest single volume refined product sold in the United States. It accounts for almost half of national oil consumption. A market that diverse is often subject to intense competition, which in turn breeds price volatility and the need for reliable risk management instruments for energy producers and users.

OPTIONS

There are two types of options — call options and put options. A call option on a futures contract gives the buyer the right, but not the obligation, to purchase the underlying contract at a specified price (the strike or exercise price) during the life of the option. A put option gives the buyer the right to sell the underlying contract at the strike or exercise price before the option expires. The cost of obtaining this right to buy or sell is known as the option's "premium." This is the price that is bid and offered in the exchange pit or via the exchange's computerized trading system. As with futures, exchange-traded option positions can be closed out by offset.

HOW FUTURES AND OPTIONS DIFFER

The major difference between futures and options arises from the different obligations of buyer and seller. A futures contract obligates both buyer and seller to perform the contract, either by an offsetting transaction or by delivery. Both parties to a futures contract derive a profit or loss equal to the difference between the price when the contract was initiated and when it was terminated. In contrast, an option buyer is not obliged to fulfill the option contract. Buying an options contract is similar to buying insurance. The buyer is typically paying a premium to remove risk, while the seller earns the premium and takes on risk. The option buyer's loss is limited to the premium paid, but in order for the buyer to make a profit, the price must increase above (call option) or decrease below (put option) the option's strike price by more than the amount of the premium paid. In turn, the option seller (writer or grantor), in exchange for the premium received, must fulfill the option contract if the buyer so chooses. Thus, the option's exercise takes place if the option has value (is "in the money") before it expires.

FEATURES NEEDED FOR A WELL-FUNCTIONING MARKET

Whatever the item, such as crude oil, underlying the futures or options contract, every market needs certain ingredients to flourish. These include

- Risk-shifting potential—the contract must provide the ability for those with price risk in the underlying item to shift that risk to a market participant willing to accept it. In the energy world, commercial producers, traders, refiners, distributors and consumers need to be able to plan ahead, and frequently enter into commitments to buy or sell energy commodities many months in advance.
- Price volatility—the price of the underlying item must change enough to warrant the need for shifting price risk. Energy prices are subject to significant variance due to factors affecting supply and demand such as level of economic activity, weather, environmental regulations, political turmoil, and war. Market psychology also plays its part.
- Cash market competition—the underlying cash (or physicals) market for the item must be broad enough to allow for healthy competition, which creates a need to manage price risk and decreases the likelihood of market corners, squeezes, or manipulation. The physical market in energy commodities is the largest such market in the world.
- Trading liquidity—active trading is needed so that sizable orders can be executed rapidly and inexpensively. Popular markets such as crude oil, heating oil, and unleaded gasoline have thousands of contracts traded daily.
- Standardized underlying entity—the commodity or other item underlying the futures contract must be standardized and/or capable of being graded so that it is clear what is being bought and sold. Energy commodities are fungible interchangeable goods sold in accordance with strict specifications and grades.

Energy is considered to be a well-functioning market because it satisfies these criteria. The existence of such a market has a significant modifying effect on short-term price volatility, and will temper the impact of any future disruptions such as those that occurred in the pre-futures market 1970s. However, even a well-functioning futures market cannot be expected to eliminate the economic risks of a massive physical supply interruption.

MARKET PARTICIPANTS

Most of the participants in the energy futures and option markets are commercial or institutional energy producers, such as petroleum producers, refiners, and electric utilities; traders; or users, such as industrial and transportation companies. The energy producers and traders, most of whom are called "hedgers," want the value of their products to increase and also want to limit, if possible, any loss in value. Energy users, who are also hedgers, want to protect themselves from cost increases arising from increases in energy prices. Hedgers may use the commodity markets to take a position that will reduce their risk of financial loss due to a change in price. Other participants are "speculators" who hope to profit from changes in the price of the futures or option contract. It is important to note that hedgers typically do not try to make a killing in the market. They use futures to help stabilize their revenues or their costs. Speculators, on the other hand, try to profit by taking a position in the futures market and hoping the market moves in their favor. Hedgers hold offsetting positions in the cash market for the physical commodity but speculators do not.

THE MECHANICS OF TRADING

The mechanics of futures and options trading are straightforward. Typically, customers who wish to trade futures and options contracts do so through a broker. Both, buyers and sellers, deposit funds—traditionally called margin, but more correctly characterized as a performance bond or good-faith deposit—with a brokerage firm. This amount is typically a small percentage—less than 10 percent—of the total value of the item underlying the contract.

The New York Mercantile Exchange (NYMEX) is the largest physical commodity futures exchange in the world. The exchange pioneered the concept of risk management for the energy industry with the launch of heating oil futures in 1978, followed by options and/or futures for sweet and sour crude oil, unleaded gasoline, heating oil, propane, natural gas, and electricity. The NYMEX is owned by its members and is governed by an elected board of directors. Members must be approved by the board and must meet strict business integrity and financial solvency standards.

The federal government has long recognized the unique economic benefit futures trading provides for price discovery and offsetting price risk. In 1974, Congress created the Commodity Futures Trading Commission (CFTC), replacing the previous Commodity Exchange Authority, which had limited jurisdiction over agricultural and livestock commodities. The CFTC was given extensive authority to regulate commodity futures and related trading in the United States. A primary function of the CFTC is to ensure the economic utility of futures markets as hedging and price discovery vehicles.

The London-based International Petroleum Exchange (IPE) is the second largest energy futures exchange in the world, listing futures contracts that represent the pricing benchmarks for two-thirds of the world's crude oil and the majority of middle distillate traded in Europe. IPE natural gas futures may also develop into an international benchmark as the European market develops larger sales volume.

Besides NYMEX and IPE, there are a number of other exchanges offering trading opportunities in energy futures. These include the Singapore International Monetary Exchange (North Sea Brent crude), The Chicago Board of Trade (electricity), Kansas City Board of Trade (western U.S. natural gas), and the Minneapolis Grain Exchange (Twin Cities' electricity). Domestic energy futures trading opportunities have arisen due to deregulation of the electricity and natural gas industries introducing many new competitors prepared to compete on the basis of price.

There has been discussion of the possibility of a futures market in emission credits arising from domestic regulations or international treaties to reduce energy use-related greenhouse gas emissions. These so-called pollution credits would be generated when Country A (or Corporation A) reduces its emissions below a specific goal, thereby earning credits for the extra reductions. At the same time, Country/Corporation B decides that emission controls are too expensive, so it purchases A's emission reduction credits. A declining cap on allowable emissions would reduce the available number of credits over time. The controversial theory is that market forces would thereby reduce emissions. Although there is some U.S. experience with the private sale and barter of such emission credits (a cash or "physicals" market), it remains to be seen if a true exchange-traded futures market in emission credits will arise.

ENERGY AND FUTURES PRICES

In addition to providing some control of price risk, futures and options markets are also very useful mechanisms for price discovery and for gauging market sentiment. There is a world-wide need for accurate, real-time information about the prices established through futures and options trading, that is, a need for price transparency. Exchange prices are simultaneously transmitted around the world via a network of information vendors' terminal services directly to clients, thereby allowing users to follow the market in real time wherever they may be. Energy futures prices are also widely reported in the financial press. These markets thus enable an open, equitable and competitive environment.

Frank R. Power

BIBLIOGRAPHY

Battley, N. (1991). *Introduction to Commodity Futures and Options.* New York: McGraw-Hill.

Brown, S. L., and Errera, S. (1987). *Trading Energy Futures: A Manual for Energy Industry Professionals.* New York: Quorum Books.

Clubley, S. (1998). *Trading in Energy Futures and Options.* Cambridge, Eng.: Woodhead.

Commodity Futures Trading Commission. (1997). *Futures and Options: What You Should Know Before You Trade.* Washington, DC: Author.

Commodity Futures Trading Commission. (1997). *Glossary: Guide to the Language of the Futures Industry.* Washington, DC: Author.

Futures Industry Institute. (1998). *Futures and Options Fact Book.* Washington, DC: Author.

Hull, J. (1995). *Introduction to Futures and Options Markets.* Englewood Cliffs, NJ: Prentice-Hall.

Kolb, R. W. (1991). *Understanding Futures Markets.* New York: New York Institute of Finance.

Kozoil, J. D. (1987). *A Handbook for Professional Futures and Options Traders.* New York: John Wiley.

Razavi, H., and Fesharaki, F. (1991). *Fundamentals of Petroleum Trading.* New York: Praeger.

G

GASOLINE AND ADDITIVES

Gasoline is the primary product made from petroleum. There are a number of distinct classes or grades of gasoline. Straight-run gasoline is that part of the gasoline pool obtained purely through distillation of crude oil. The major portion of the gasoline used in automotive and aviation is cracked gasoline obtained through the thermal or catalytic cracking of the heavier oil fractions (e.g., gas oil). A wide variety of gasoline types are made by mixing straight-run gasoline, cracked gasoline, reformed and synthesized gasolines, and additives.

Motor fuels account for about one-quarter of all energy use in the United States. The energy content of gasoline varies seasonally. The average energy content of regular gasoline stands at 114,000 Btu/gal in the summer and about 112,500 Btu/gal in the winter. The energy content of conventional gasolines also varies widely from batch to batch and station to station by as much as 3 to 5 percent between the minimum and maximum energy values.

Gasoline can be made from coal as well as petroleum. In the 1930s and 1940s, Germany and other European countries produced significant quantities of gasoline from the high-pressure hydrogenation of coal. But to convert the solid coal into liquid motor fuels is a much more complex and expensive process. It could not compete with the widely available and easily refined petroleum-based motor fuels.

In the United States, all gasoline is produced by private commercial companies. In many cases, they are vertically integrated so that they drill for, find, and transport oil, process the oil into gasoline and other products and then sell the gasoline to a network of retailers who specialize in certain brand name gasolines and gasoline blends. These large refiners may also market these products to other refiners, wholesalers, and selected service stations.

GROWING DEMAND FOR GASOLINE

At the end of the nineteenth century, virtually all of the gasoline produced (around 6 million barrels) was used as a solvent by industry, including chemical and metallurgical plants and dry cleaning establishments, and as kerosene for domestic stoves and space heaters. But by 1919, when the United States produced 87.5 million barrels of gasoline, 85 percent was consumed by the internal combustion engine (in automobiles, trucks, tractors, and motorboats).

Between 1899 and 1919, as demand for gasoline grew, the price increased more than 135 percent, from 10.8 cents/gal to 25.4 cents/gal. From 1929 to 1941, gasoline use by passenger cars increased from 256.7 million barrels to 291.5 million barrels. Consumption of aviation fuel went from only 753,000 barrels in 1929 to over 6.4 million barrels at the start of World War II. By 1941, gasoline accounted for over one-half of petroleum products with 90 percent of gasoline output used as fuel for automotive and aircraft engines.

Between 1948 and 1975, per capita consumption of gasoline in the United States increased from about 150 gal/yr to a little less than 500 gal/yr. A growing trend after the war was the increasing use of jet fuel for aircraft and the decline in use of aviation gasoline. After 1945, oil production increased in other parts of the world, especially the Middle East and Latin America. By the 1970s, the Middle East became a dominant oil producing region. The cartel formed by the major Middle Eastern oil producing countries, known as OPEC, became a major force in setting oil prices internationally through the control of oil production.

Since the mid-1970s, the rate of growth of per capita gasoline consumption has slowed. An important factor in causing this moderation in demand was the trend to improve automobile fuel economy that was initiated by worldwide fuel shortages. Fuel economy hovered around 14.1 mpg between 1955 and 1975; it rose sharply over the next 15 years, reaching around 28.2 mpg in 1990.

An aging population and continued improvements in engine technology and fuel economy may slow U.S. gasoline demand in the early part of the twenty-first century from the 2 percent annual growth rate of the 1990s.

KNOCKING AND OCTANE RATING

The process of knocking has been studied extensively by chemists and mechanical engineers. Knocking is rapid and premature burning (i.e., preignition) of the fuel vapors in the cylinders of the engine while the pistons are still in the compression mode. Research on knocking was carried out prior to World War I, but it was only with the increase in the size and power of automotive engines after 1920 that significant attempts were made to deal with the problem on a commercial basis.

Knocking, which has a distinctive metallic “ping,” results in loss of power and efficiencies and over time causes damage to the engine. Knocking is a great energy waster because it forces the automobile to consume greater quantities of gasoline per mile than do engines that are functioning properly. The problem of engine knocking was an important factor in the U.S. push for a gasoline rating system. Around the time of World War I, there was no single standard specification or measure of gasoline performance. Many states developed their own specifications, often conflicting with those promulgated by the automotive and petroleum industries and the federal government. Even the various branches of government had their own specifications. The specifications might be based on the boiling point of the gasoline fraction, miles allowed per gallon of fuel, or the chemical composition of the gasoline.

The octane numbering system was developed in the late 1920s and was closely linked to the federal government’s program of measurement standards, designed jointly by the Department of the Army and the National Bureau of Standards.

The octane number of a fuel is a measure of the tendency of the fuel to knock. The octane scale has a minimum and maximum based on the performance of reference fuels. In the laboratory, these are burned under specific and preset conditions. One reference fuel is normal heptane. This is a very poor fuel and is given an octane rating of zero. On the opposite end of the scale is iso-octane (2,2,4 trimethyl pentane). Iso-octane is a superior fuel and is given a rating of 100.

The octane rating of fuels is derived by simple laboratory procedures. The fuel being tested is burned to determine and measure its degree of “knocking.” Then the two reference fuels are blended together until a reference gasoline is formed that knocks to the same degree as the tested fuel. The proportion of iso-octane present in the reference fuel is then the octane number of the tested gasoline. Some compounds, like methanol and toluene, perform better than iso-octane and, by extrapolation, their octane numbers are over 100. A higher octane number is important from a very practical consideration: it gives better engine performance in the form of more miles per gallon of gasoline.

From the 1920s to the 1940s, catalytic cracking processes were developed that not only increased processing efficiencies, but progressively raised the octane number of gasoline. In 1913, prior to the devising of the octane scale, the commercialization of the Burton Process, a noncontinuous thermal technology, produced gasoline with an estimated octane number of between 50 and 60. Continuous thermal cracking, first operated in the early 1920s, produced gasoline with an octane number of close to 75. With the first catalytic process in the form of the Houdry technology (1938), cracked gasoline reached the unprecedented octane level in the high 80s. Fluid catalytic cracking, the culmination of the cracking art that came on line in 1943, pushed the quality of gasoline to an octane level of 95.

While octane rating provided an objective and verifiable measure of performance across all grades of gasoline, it did not immediately lead to unified standards. It was not until the 1930s, when both the octane rating and new types of octane-boosting additives entered the industry, that automotive fuel began to center around two major types of gasoline—regular and premium—each operating within its own octane range. Over the next 60 years, octane rating of gasoline increased due to improved refining practices and the use of additives. In the 1970s and 1980s, the use of additives became increasingly tied to environmental concerns (i.e., clean air), as well as higher octane ratings. Gasoline has come a long way since the Model T, and it is important to note that, in terms of constant dollars, it is cheaper today than it was in 1920.

Table 1 shows the major gasoline additives that were introduced from the 1920s through the 1980s. The increase in octane number of gasoline with use of these additives is shown.

Additive	*Octane Number*
Tetraethyl lead (TEL)	100
Methanol	107
Ethanol	108
Methyl-t-butyl ether (MTBE)	116
Ethyl-t-butyl ether (ETBE)	118

Table 1.
Octane Numbers of Common Gas Additives

TETRAETHYLEAD: THE FIRST GASOLINE ADDITIVE

One source of knocking was related to the vehicle engine. All else being equal, an automobile engine with a higher compression ratio, advanced spark schedule, or inefficient combustion is more likely to experience knocking. Within the United States, research into knocking has focused on the chemical aspects of gasoline, which is a complex hydrocarbon mixture of paraffins, naphthenes, and aromatics.

Chemical additives first entered the industry in the first decade of the century. These additives served a number of uses. For example, they lessened the capacity of gasoline to vaporize out of the gas tank or to polymerize (i.e., produce gummy residues) in the engine. In the early 1920s, the most important application for these substances was to eliminate knocking. Tetraethylead (TEL) was the first major gasoline additive to be commercialized for this purpose.

Charles F. Kettering, the inventor of the self starter, the Delco battery, and other major components of modern automotive engineering, started to work on the problem in 1916 at his Dayton Engineering Laboratories Company (DELCO). Kettering was induced into this research not by the problems faced by the automobile but by gasoline-powered electric lighting systems for farms. These systems employed generators utilizing internal combustion engines. These engines, which burned kerosene and not gasoline, knocked badly. Kettering and his team addressed this nonautomotive concern as a profitable research project and one of potentially great benefit to the agricultural sector.

Kettering hired Thomas Midgley Jr., a mechanical engineer from Cornell, to work with him on the project. Studying the combustion process in more detail, Kettering and Midgley determined that low volatility in the fuel caused knocking to occur. This conclusion led them to search for metallic and chemical agents to blend with the gasoline to increase volatility and reduce knocking.

A promising line of research led Midgley to the halogen group of chemicals and specifically iodine and its compounds. General Motors purchased DELCO Labs in 1919 and the search for an antiknock agent came under GM management, with Kettering and Midgley remaining on board to continue the work, but with the focus now on automotive application.

Using iodine as their starting point, they experimented with a series of compounds including the anilines and a series of metals near the bottom of the periodic table. Lead turned out to be the most effective of the additives tested. But lead alone caused a number of problems, including the accumulation of its oxide in engine components, and particularly the cylinders, valves, and spark plugs.

Experiments continued to find an appropriate form of lead that could at the same time prevent the formation of oxide deposits. Ethylene was found to combine with lead to form tetraethyllead (TEL), a stable compound that satisfied this requirement.

Kettering and Midgley were the first to identify it as a prime antiknock agent, though the compound had been known since 1852. They estimated that only a very small amount of TEL—a few parts per thousand—would result in a 25 percent increase in horsepower as well as fuel efficiency.

The next stage of development was to design a production process to link the ethyl group to lead. GM attempted to make TEL from ethyl iodide. They built an experimental plant, but the process proved too expensive to commercialize.

An alternative source of the ethyl component was ethyl bromide, a less expensive material. It was at this point that GM called upon DuPont to take over process development. DuPont was the largest U.S. chemical company at the time. It had extensive experience in the scale-up of complex chemical operations, including explosives and high-pressure synthesis. The manufacturing process was undertaken by DuPont's premier department, the Organic Chemical section. GM contracted with DuPont to build a 1,300 pound per day plant. The first commercial quantities of TEL were sold in February 1923 in the form of ethyl premium gasoline.

In 1923, GM set up a special chemical division, the GM Chemical Co., to market the new additive. However, GM became dissatisfied with DuPont's progress at the plant. In order to augment its TEL supply, and to push DuPont into accelerating its pace of production, GM called upon the Standard Oil Company of New Jersey (later Esso/Exxon) to set up its own process independently of DuPont. In fact, Jersey Standard had obtained the rights to an ethyl chloride route to TEL. This turned out to be a far cheaper process than the bromide technology. By the mid-1920s, both DuPont and Jersey were producing TEL.

GM brought Jersey in as a partner in the TEL process through the formation of the Ethyl Corporation, each party receiving a 50 percent share in the new company. All operations related to the production, licensing, and selling of TEL from both DuPont and Jersey were centralized in this company.

Soon after production began, TEL was held responsible for a high incidence of illness and deaths among production workers at both the DuPont and Jersey Standard plants. The substance penetrated the skin to cause lead poisoning. Starting in late 1924, there were forty-five cases of lead poisoning and four fatalities at Jersey Standard's Bayway production plant. Additional deaths occurred at the DuPont Plant and at the Dayton Laboratory. This forced the suspension of the sale of TEL in 1925 and the first half of 1926.

These incidents compelled the U.S. Surgeon General to investigate the health effects of TEL. The industry itself moved rapidly to deal with the crisis by instituting a series of safety measures. Now, ethyl fluid was blended at distribution centers and not at service stations (it had been done on the spot and increased the chances of lead poisoning to service station attendants). Also, ethyl gasoline was dyed red to distinguish it from regular grade gasoline. DuPont and Jersey placed tighter controls over the production process. The federal government placed its own set of restrictions on TEL. It set the maximum limit of 3 cc of TEL per gallon of gasoline. By 1926, TEL was once again being sold commercially.

Ironically, this episode proved beneficial to DuPont. DuPont became the dominant source of TEL after the mid-1920s because they perfected the chloride process and were far more experienced than Jersey Standard in producing and handling toxic substances.

The Ethyl Corp. and DuPont held the TEL patent, and controlled the TEL monopoly. The company held the sole right to the only known material that could eliminate automotive knocking. And it used its influence in the gasoline market to manipulate prices. Over the next few years, the company wielded its monopoly power to maintain a 3–5 cent differential between its "ethyl" gasoline and the regular, unleaded gasoline sold by the rest of the industry.

Throughout the 1930s TEL proved itself a profitable product for DuPont, which remained virtually the only TEL producer into the post–World War II period. With no advantage to be gained in further collaboration, DuPont severed its ties with Ethyl Corp. in 1948 and continued to manufacture TEL independently.

COMPETING AGAINST TEL: ALTERNATIVE ANTIKNOCK TECHNOLOGY

As the automotive industry continued to introduce higher compression engines during the 1920s and 1930s, refiners increasingly relied on TEL to meet gasoline quality. By 1929, fifty refiners in the United States had contracted with the Ethyl Corp. to incorporate TEL in their high test gasoline.

TEL was not the only way to increase octane number. Those few companies who did not wish to do business with Jersey Standard, sought other means to produce a viable premium gasoline. TEL represented the most serious threat to the traditional gasoline product. It was cheap, very effective, and only 0.1 percent of TEL was required to increase the octane number 10 to 15 points. In contrast, between 50 to 100 times this concentration was required of alternative octane enhancers to achieve the same effect.

Benzol and other alcohol-based additives improved octane number, up to a point. Experiments using alcohol (ethanol, methanol) as a replacement for gasoline began as early as 1906. In 1915, Henry Ford announced a plan to extract alcohol from grain to power his new Fordson tractor, an idea that never achieved commercial success.

The shortage of petroleum after World War I induced an intense search for a gasoline substitute in the form of alcohol. The trade press felt alcohol would definitely replace gasoline as a fuel at some point. The advantages of alcohol cited in the technical press included greater power and elimination of knocking.

The push to use alcohol as a fuel surfaced at various times coinciding with real or perceived gasoline shortages and often directed by the farm lobby during periods of low grain prices. The great discoveries of oil in the Mid-Continental fields in the1920s reduced the incentive for the use of alcohol as a fuel. But in the 1930s the severe agricultural crisis brought back interest in alcohol. Alcohol distillers, farmers, and Midwest legislators unsuccessfully attempted to regulate the blending into gasoline of between 5–25 percent ethanol. It took the oil supply disruptions of the 1970s for farm state legislators to pass legislation to highly subsidize ethanol. The subsidies, which remain in effect today, are the reason ethanol continues to play a notable role as a fuel additive.

As experiments at Sun Oil Co. in the early 1930s indicated, there were serious disadvantages associated with alcohol. While alcohol did in fact appear to increase the octane number, it left large amounts of deposits in the engine. Alcohol also vaporized out of the gas tank and engine at rapid rates. And the combustion temperature of the alcohol group is lower than for hydrocarbons because it is already partially oxidized.

The most effective competitive approach for the more independent refiners was in developing new types of cracking technologies. Companies like Sun Oil, one of the few companies who remained independent of Jersey and Ethyl, continued to expand the limits of thermal cracking, notably by employing higher pressures and temperatures. Sun's gasoline reached octane levels close to those achieved by gasolines spiked with TEL (i.e., between 73 and 75). Sun Oil continued to compete with additives purely through advanced cracking technology, a path that would lead by the late 1930s to the first catalytic cracking process (i.e., the Houdry Process). But by this time, more advanced refining processes were coming on line and competing with the Houdry Process. By the early 1940s, Jersey Standard developed fluidized bed catalytic cracking technology. Fluidized cracking proved superior to Houdry's fixed bed process with respect to both production economies and the quality of the product (i.e., octane rating of the gasoline). Fluidized cracking quickly displaced Houdry's catalytic cracking technology as the process of choice.

Competition did not center on quality alone. Price and packaging were called into play as weapons against the onslaught of TEL. For example, as a marketing tool, Sun Oil dyed its gasoline blue to more easily identify it as a high premium fuel (customers actually saw the gasoline being pumped in a large clear glass reservoir on top of the gas pump). Sun then competed aggressively on price. Whereas TEL-using refiners sold two grades of gasoline, regular and premium, Sun marketed only its premium "blue Sunoco" at regular grade prices. Sun could do this because it was not burdened, as TEL was, with such additional costs as blending and distribution expenses that cut into profit margins.

By the late 1920s and into the 1940s, with the use of either TEL, other additives, or advanced cracking technology, a number of premium grade gasolines appeared on the market. In addition to Sun's premium, there was Gulf's "NoNox," Sinclair's premium "H.C.'s" gasoline, and Roxana Petroleum's "Super-Shell." The use of TEL has plummeted since the government's mandate in 1975 to install catalytic converters for reducing the carbon monoxide and unburned hydrocarbons in automotive exhaust gases. This is because lead poisons the noble metal (chiefly platinum) catalysts used. In addition the lead bearing particulates in the emissions from engines burning leaded fuel are toxic in their own rights.

POSTWAR DEVELOPMENTS: GASOLINE AND THE ENVIRONMENT

Additives and the blending process became an increasingly important part of gasoline manufacture after World War II. Refiners had to balance such factors as customer specifications, regulatory requirements, and probable storage (i.e., nonuse) time. The industry became more precise in how, when, and how many of components should be added to gasolines. The large, modern refinery increasingly incorporated complex computer programs to help plan and effect blending requirements. Critical factors that had to be factored into these calculations included seasonal adjustments, current and anticipated demand, regulatory levels, and supply schedules of the various components.

Since the 1950s, an increasing portion of a refiner's R&D has gone into new and improved additives. Beyond their role as antiknock agents, additives and blending agents have taken on an ever broadening range of functions to improve the performance of fuels in automotive and aircraft engines.

Sulfur and Gasoline

In recent years, there has been a greater understanding of the role of automotive emissions as environmental pollutants. Sulfur dioxide, nitrogen oxides, and carbon monoxide degrade the earth's atmosphere and are health hazards. Carbon dioxide adds to the atmospheric buildup of greenhouse gases and in turn accelerates the process of global warming.

Sulfur is a particular problem as an environmental hazard. It occurs naturally in various concentrations in petroleum, and it is difficult and costly to remove all of it. Distillation and cracking removes some, but small amounts survive the distillation and cracking processes and enter into the gasoline. The average level of sulfur in gasoline has not changed much since 1970, remaining at 300 parts per million (ppm) with a range between 30 and 1,000 ppm.

High levels of sulfur not only form dangerous oxides, but they also tend to poison the catalyst in the catalytic converter. As it flows over the catalyst in the exhaust system, the sulfur decreases conversion efficiency and limits the catalyst's oxygen storage capacity. With the converter working at less than maximum efficiency, the exhaust entering the atmosphere contains increased concentrations, not only of the sulfur oxides but also, of hydrocarbons, nitrogen oxides, carbon monoxides, toxic metals, and particulate matter.

In the 1990s, the EPA began controlling sulfur through its reformulated gasoline program. It developed regulations in 1999 that would sharply reduce the sulfur content in gasoline from 300 ppm to a maximum of 80 ppm.

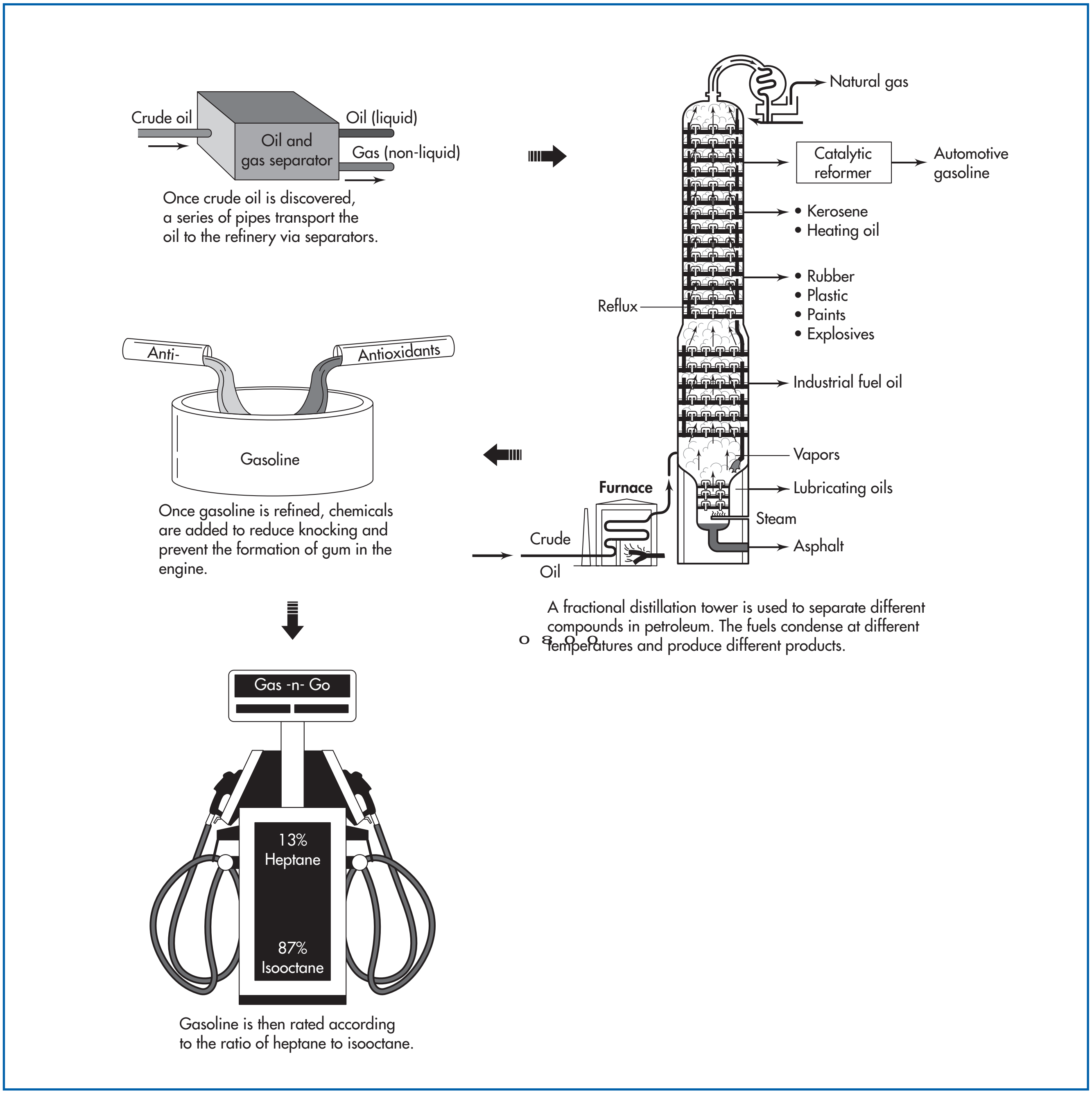
The new regulations, scheduled to go into effect in 2004, are compelling certain refiners to purchase low-sulfur content ("sweet") crude oil. This is the strategy being pursued by Japanese refiners. However, the Japanese are not major oil producers but import oil from other producing countries. U.S. refiners, in contrast, consume oil from a wide range of ("sour") petroleum sources that have a high-sulfur content, including Venezuela, California, and parts of the Gulf Region. U.S. companies own and operate oil producing infrastructures (i.e., derricks, pipelines), within the United States and overseas. They are committed to working these oil fields, even if producing high-sulfur oil. U.S. refineries thus need to continue dealing with high-sulfur crude oil. Imported crude from the Middle East, while historically low in sulfur, is also becoming increasingly less sweet.

Petroleum refiners will have to reduce sulfur content at the refineries. This will require the costly retooling of some of their plant operations in order to achieve a suitable fuel mix. Removing additional amounts of sulfur at the refinery will entail installation of separate catalyst-based process such as hydrosulfurization. Another possible approach is the removal of sulfur in liquid oil or gasoline by the use of both organic and inorganic scavenger agents added to the oil or gasoline to seek out, combine with, and precipitate out sulfur and its compounds.

Reformulated Gasoline and MTBE

Prior to the Clean Air Act of 1990, environmental regulations were aimed at reducing emissions as they left the exhaust system. The catalytic converter has been the primary means of attacking air pollution in this way. After 1990, regulations for the first time undertook to alter the composition of the fuel itself. Reformulated gasoline applies to gasoline that is sold in the nine metropolitan areas designated by the EPA with the highest level of ozone pollution. About 48 million people reside in areas where ozone concentrations exceed federal standards.

Reformulation refers to the transformation of gasoline to make it cleaner with respect to emissions. Beginning in 1995, specifications for reformulated gasoline included a 2 percent minimum oxygen content and a maximum content of various organic and inorganic pollutants. In addition, heavy metal additives in gasoline are prohibited. A disadvantage of



Once crude oil is discovered, a series of pipes transport the oil to the refinery via separators.

A fractional distillation tower is used to separate different compounds in petroleum. The fuels condense at different temperatures and produce different products.

Once gasoline is refined, chemicals are added to reduce knocking and prevent the formation of gum in the engine.

Gasoline is then rated according to the ratio of heptane to isooctane.

The gasoline-making process. (Gale Group)

reformulated gasoline is that it contains 1 to 3 percent less energy per gallon than traditional gasoline.

Many reformulated gasolines use oxygenated compounds as additives. Clean Air regulations specify the need for oxygenated fuel in 39 metropolitan areas with high carbon monoxide concentrations. The regulations for oxygenated fuel are seasonal: during the winter season, gasoline must contain a minimum of 27 percent oxygen. The oxygen helps engines to burn the fuel more completely which, in turn, reduces monoxide emissions. The major additive to supply the additional oxygen to reformulate gasoline to satisfy these requirements is the methanol derivative, methyl tertiary butyl ether (MTBE).

Currently, this additive is used in over 30 percent of U.S. gasoline.

MTBE was first used as a fuel additive in the 1940s and was a popular additive in Europe in the 1970s and 1980s. In the late 1970s, MTBE began replacing lead in this country to enhance octane number. In the late 1980s, California led the way in the United States for its use as an oxygenate for cleaner burning fuel. The consumption of MTBE in the United States increased rapidly between 1990 and 1995 with the passage of the Clean Air Act and, a few years later the implementation of the federal reformulated gasoline program. Currently, MTBE is produced at 50 U.S. plants located in 14 states. About 3.3 billion gallons of MTBE, requiring 1.3 billion gallons of methanol feedstock, are blended annually into reformulated gasoline.

In the late 1990s, MTBE came under serious attack on grounds of both efficacy and safety. A report by the National Research Council (1999) stated that the addition of oxygen additives in gasoline, including MTBE and ethanol, are far less important in controlling pollution than emission control equipment and technical improvement to vehicle engines and exhaust systems.

Moreover, MTBE has been found in groundwater, lakes and reservoirs used for drinking water, and it has been linked to possible serious disease. The probable occurrence of cancerous tumors in laboratory rats injected with MTBE alerted federal agencies as to its possible health hazards. In 1999, the EPA reversed itself, recommending the phasing out of MTBE as an additive to gasoline.

During the first half of 2000, MBTE production in the Unites States averaged 215,000 barrels per day. In the same six-month period, the average production offuel ethanol was 106,000 barrels per day. In light of the EPA's 1999 recommendation, ethanol will most likely replace MTBE as an effective oxygenate additive. In addition to its use as an oxygenate, ethanol enhances octane ratings and dilutes contaminants found in regular gasoline.

New and Emerging Gasoline Additives

The development and blending of additives is undertaken for the most part by the petroleum refining industry. Additives are essential to the economic well-being of the industry because they tend to boost sales for gasoline and diesel fuel. In most cases, additives do not differ in price by more than three to four cents a gallon. The recently developed additives do not necessarily sacrifice fuel efficiency for higher octane numbers. They are multifunctional. In addition to boosting octane ratings they may also clean the engine, which, in turn, leads to greater fuel efficiency.

Beyond their role in enhancing octane numbers and reducing emissions, the group of more recent fuel additives performs a growing range of functions: antioxidants extend the storage life of gasoline by increasing its chemical stability; corrosion inhibitors prevent damage to tanks, pipes, and vessels by hindering the growth of deposits in the engine and dissolving existing deposits; demulsifiers or surface active compounds prevent the formation of emulsions and the dirt and rust entrained in them that can foul the engine and its components.

Beginning in the 1970s, gasoline additives increasingly took on the role of antipollutant agent in the face of government attempts to reduce automotive emissions into the atmosphere. Despite the advances made in cracking and reforming technologies and in the development and blending of additives (not to mention enhancements in the engine itself), the use of automotive gasoline has increased the level of air pollution. This is so because modern distillates, blends of straight run and cracked or chemically transformed product, tend to have a higher aromatic content. The result is longer ignition delays and an incomplete combustion process that fouls the engine and its components and increases particulate and oxide emissions.

Continued implementation of clean air legislation, especially within the United States, is expected to accelerate the consumption of fuel additives. In 1999, the EPA proposed wide-ranging standards that would effectively reformulate all gasoline sold in the United States and significantly reduce tailpipe emissions from trucks and sports utility vehicles. These regulations require potentially expensive sulfur-reducing initiatives from both the oil industry and the automakers. For refiners, it will require significant redesign and retooling of plant equipment and processes will be required in order to achieve suitable changes in the fuel mix because the U.S. oil industry is committed to continue development of its sour petroleum reserves. The DOE expects that the more complex processing methods will add six cents to the cost of a gallon of gasoline between 1999 and 2020.

In the United States alone, the demand for fuel additives is expected to reach over 51 billion pounds by 2002. Oxygenates are anticipated to dominate the market, both within the United States and internationally. Nonpremium gasoline and diesel fuel represent the fastest growing markets for fuel additives.

A recently marketed fuel additive is MMT (methylcyclopentadienylmanganese tricarbonyl). MMT was first developed by the Ethyl Corporation in 1957 as an octane enhancing agent and has experienced a growth in demand in the 1990s. MMT was Ethyl Corporation's first major new antiknock compound since TEL.

However, in 1997, the EPA blocked the manufacture of MMT. The Agency took this action for two reasons. It determined that MMT had the potential for being hazardous to humans, and in particular to children. The EPA is especially concerned about the toxic effects of the manganese contained in MMT. Also, the EPA discovered that MMT was likely interfering with the performance of the catalytic converters in automobiles and in turn causing an increase in exhaust emissions in the air. In 1998, the EPA decision was overturned by a federal appeals court in Washington. The court's decision allows the Ethyl Corp. to test MMT while it is selling the additive. The decision set no deadline for the completion of tests. In addition to its use in the United States, MMT is consumed as an additive in unleaded gasoline in Canada.

In addition to MTBE and MMT, other kinds of additives are being developed. Some of these are derivatives of alcohol. Variations of MTBE are also being used, especially the ether-derived ETBE (ethyl-t-butyl ether).

A new generation of additives specifically designed for aircraft gasoline are also being developed. These additives address such problems as carbon buildup, burned and warped valves, excessive cylinder head temperatures, stuck valves and piston rings, clogged injectors, rough idle, and detonation. Aviation fuel additives often act as detergents (to remove deposits), octane enhancers, and moisture eliminators.

Competition from Alternative Fuels

Most alternative fuel vehicles on the road today were originally designed for gasoline, but converted for use with an alternative fuel. Because the petroleum industry has successfully responded to the competitive threats of alternative fuels by developing reformulated gasolines that burn much cleaner, the conversions are typically performed more for economic reasons (when the alternative fuel is less expensive, which has occurred with propane) rather than environmental reasons. It is likely that technical advances will continue to permit petroleum refiners to meet the increasingly more stringent environmental regulations imposed on gasoline with only minor increases in the retail price. And since petroleum reserves will be abundant at least through 2020, gasoline promises to dominate automotive transportation for the foreseeable future.

However, fuel cell vehicles, which are designed to generate their power from hydrogen, pose a major long-term threat to the preeminence of gasoline. Automakers believe the best solution is to extract hydrogen from a liquid source because hydrogen has a low energy density and is expensive to transport and store. All the major automakers are developing fuel cell vehicles powered by hydrogen extracted from methanol because reforming gasoline into hydrogen requires additional reaction steps, and a higher operating temperature for the reformer. Both requirements are likely to make the gasoline reformer larger and more expensive than the methanol reformer. Moreover, the sulfur content of gasoline is another major reason that automakers are leery of developing gasoline reformers for fuel cell vehicles. Quantities as low as a few parts per million can be a poison to the fuel cell stack. There are no gasoline reformer fuel cell vehicles in operation, so an acceptable level of sulfur has not been determined. If it is determined that an ultralow-sulfur gasoline blend can be developed specifically for fuel cell vehicles, it would be a far less expensive solution than developing the fuel production, delivery and storage infrastructure that would be needed for methanol-powered fuel cell vehicles.

Sanford L. Moskowitz

See also: Efficiency of Energy Use, Economic Concerns and; Engines; Fuel Cells; Fuel Cell Vehicles; Hydrogen; Methanol; Synthetic Fuel.

BIBLIOGRAPHY

Alexander, D. E., and Fairbridge, R. W., eds. (1999). *The Chapman & Hall Encyclopedia of Environmental Science*. Boston, MA: Kluwer Academic.

Buell, P., and Girard, J. (1994). *Chemistry: An Environmental Perspective*. Englewood Cliffs, NJ: Prentice-Hall.

Encyclopedia of Chemical Technology, 4th ed. (1994). New York: John Wiley & Sons, Inc.

Enos, J. (1962). *Petroleum, Progress and Profits: A History of Process Innovation*. Cambridge, MA: MIT Press.

Environmental Protection Agency. (1998). "Press Release: EPA Announces Blue-Ribbon Panel to Review the Use of MTBE and Other Oxygenates in Gasoline." Washington, DC: Author.

Environmental Protection Agency. (1999). "Press Release: Statement by the U.S. EPA Administrator on the Findings of EPA's Blue Ribbon MTBE Panel." Washington DC: Author.

"Ford Expects Output of Flexible Fuel Vehicles to Represent 10% of its U.S. Automotive Production." *Purchasing*. 125(9):9.

Giebelhaus, A. W. (1980). *Business and Government in the Oil Industry: A Case Study of Sun Oil*, 1876–1945. Greenwich, CT: JAI Press.

Hounshell, D. A., and Smith, J. K. (1988). *Science and Corporate Strategy: DuPont R&D, 1902–1980*. New York: Cambridge University Press.

Larson, H. M., et al. (1971). *History of Standard Oil Company (New Jersey): New Horizons, 1927–1950*. New York: Harper and Row.

"MTBE Phaseout Will Spur Ethanol Demand." *Purchasing* 127(3):32C1.

National Research Council. (1990). *Fuels to Drive Our Future*. Washington, DC: National Academy Press.

"New Fuel May Meet Emissions Rules." *Purchasing*. 118(10):67.

Speight, J. G., ed. (1990). *Fuel Science and Technology Handbook*. New York: Marcel Dekker.

Spitz, P. (1988). *Petrochemicals: The Rise of an Industry*. New York: John Wiley and Sons.

U.S. International Trade Commission (ITC), Office of Industries. (1998). "Industry and Trade Summary: Refined Petroleum Products," USITC Publication 3147. Washington, DC: Author.

Williamson, H. F., et al. (1963). *The American Petroleum Industry, Vol. II: The Age of Energy, 1899–1959*. Evanston, IL: Northwestern University Press.

GASOLINE ENGINES

The gasoline engine is a device to convert the chemical energy stored in gasoline into mechanical energy to do work—to mow a lawn; chainsaw a tree; propel a car, boat, or airplane; or to perform myriad other tasks. The energy in the gasoline is transformed into heat within the engine through combustion, so the gasoline engine is an internal combustion engine. Because combustion is normally initiated by an electric spark, the gasoline engine is also frequently known as a spark-ignition engine.

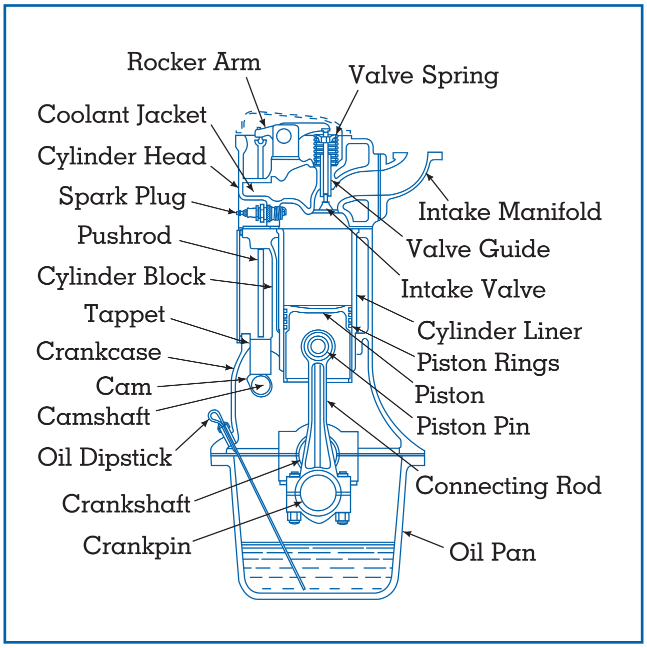


Figure 1.
Cross section through gasoline engine using push-rod valve actuation.

A number of different kinematic mechanisms have been used to extract mechanical work from the heated products of combustion. The preferred option is the slider-crank mechanism, which is incorporated into the gasoline-engine cross section of figure 1. In the slider-crank mechanism, the piston reciprocates up and down within a cylinder, alternately doing work on and extracting work from the gas enclosed by the piston, cylinder walls, and cylinder head. A poppet-type intake valve in the cylinder head opens during part of the engine cycle to admit a fresh charge of air and fuel. A spark plug ignites the mixture at the appropriate time in the cycle. A poppet-type exhaust valve (hidden behind the intake valve in Figure 1) opens later in the cycle to allow the burned products of combustion to escape the cylinder.

The reciprocating motion of the piston is transformed into rotary motion on the crankshaft by two of the links in the slider-crank mechanism—the connecting rod and the crank (hidden from view in this cross section). The connecting rod joins the piston pin to the crank pin. The crank connects the crank

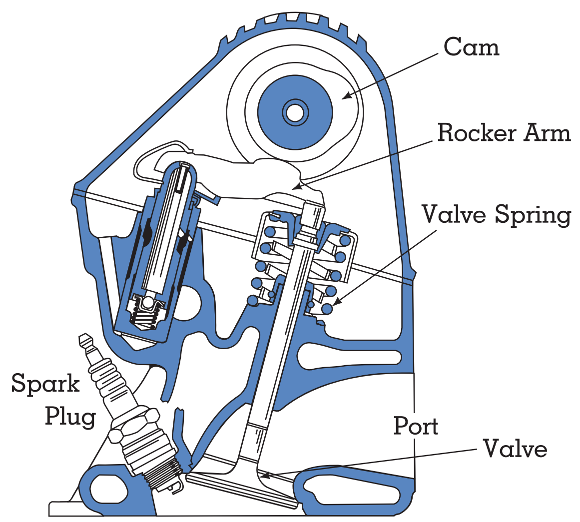


Figure 2.
Overhead-cam valve actuation.

pin to the crankshaft. The crank and the crank pin are usually integral parts of the crankshaft.

In Figure 1, the intake and exhaust valves are actuated by a camshaft chain, or a gear-driven from the crankshaft at half engine speed. The camshaft, operating through a tappet and push rod, tilts one end of a pivoted rocker arm at the appropriate time during the cycle. The opposite end of the rocker arm, working against the force of a valve spring, opens and closes a poppet valve. Each intake and exhaust valve has its own such mechanism.

This arrangement of valves defines a push-rod engine. In the alternative approach of Figure 2, the camshaft is moved to a position above the cylinder head, eliminating the push rod. This configuration defines an overhead-cam engine. The overhead camshaft is driven from the crankshaft by either a belt or a chain.

Other essential elements of a gasoline engine are also evident in Figure 1. The oil pan, fastened to the bottom of the crankcase, contains a reservoir of oil that splashes up during engine operation to lubricate the interface between the cylinder wall and the piston and piston rings. These rings seal the gas within the space above the piston. The dipstick used to verify an adequate oil supply is evident. Liquid in the coolant jacket seen in the cylinder head and surrounding the cylinder wall maintains the engine parts exposed to combustion gases at an acceptable temperature.

HISTORICAL BACKGROUND

The first practical internal-combustion engine is attributed to the Frenchman Jean Lenoir in 1860. His single-cylinder engine completed its cycle in but two strokes of the piston. On the first stroke, the piston drew a fresh charge of air and a gaseous fuel into the cylinder. Near midstroke, the intake valve closed, a spark ignited the trapped charge, the ensuing combustion quickly raised cylinder pressure, and the remainder of the piston stroke involved expansion of the combustion products against the piston to produce useful output work for transmission to the crankshaft. On the return stroke of the piston, the combustion products were expelled from the cylinder through an exhaust valve.

Four-Stroke Piston Engine

In France in 1862, Beau de Rochas outlined the principles of the four-stroke engine so common today. However, he never transformed those principles into hardware. Among the improvements proposed by de Rochas was compression of the charge prior to combustion. In contrast, the charge in the Lenoir engine was essentially at atmospheric pressure when combustion was initiated.

The operating principles of the four-stroke cycle are illustrated in Figure 3. In Figure 3a the descending piston draws a fresh charge into the cylinder through the open intake valve during the intake stroke. In Figure 3b, the intake valve has been closed, and the ascending piston compresses the trapped charge. As the piston approaches top dead center (TDC) on this compression stroke, the spark plug ignites the mixture. This initiates a flame front that sweeps across the chamber above the piston. Combustion is normally completed during the expansion stroke (see Figure 3c) well before the piston reaches bottom dead center (BDC). Before BDC is reached, the exhaust valve begins to open, releasing pressurized combustion products from the cylinder.

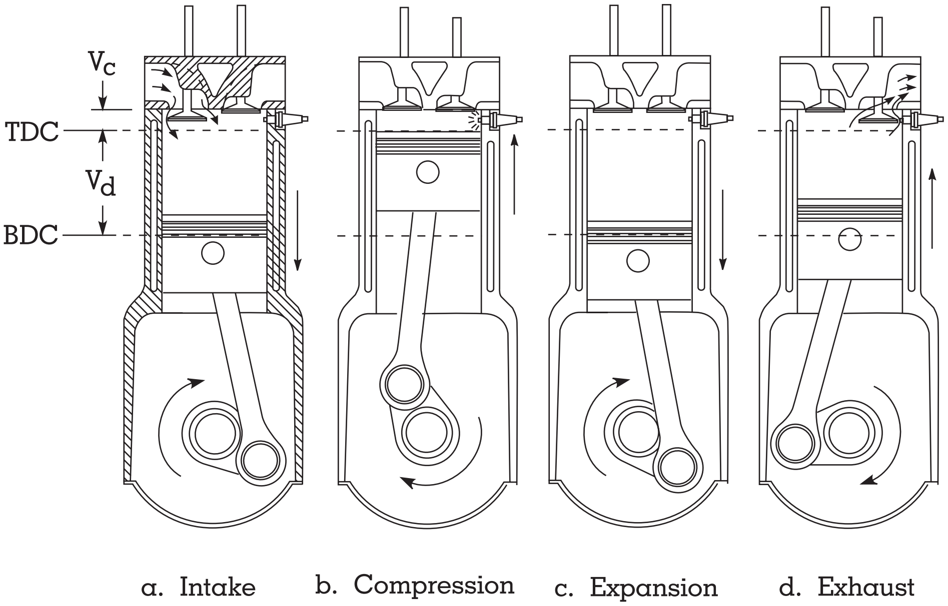


Figure 3.
Four-stroke cycle. TDC and BDC = top dead center and bottom dead center positions of the piston, respectively. V_d = displacement. V_c = clearance volume. Compression ratio = $(V_d + V_c)/V_c$.

During the exhaust stroke (see Figure 3d) the ascending piston expels most of the remaining products through the open exhaust valve, in preparation for a repetition of the cycle.

In 1876 in Germany, Nikolaus Otto built the first four-stroke engine, even though he was apparently unaware of the proposals of de Rochas. An idealized version of the cycle on which his 1876 engine operated is represented on coordinates of cylinder pressure versus volume in Figure 4. From 1 to 2, the piston expends work in compressing the fresh charge as it moves from BDC to TDC. At 2, combustion releases chemical energy stored in the fuel, raising cylinder pressure to 3. From 3 to 4, the products expand as the piston returns to BDC, producing useful work in the process. From 4 to 5, the burned gas is expelled from the cylinder as pressure drops to its initial value. These four events comprise what has come to be known in engineering thermodynamics as the Otto cycle.

The horizontal line at the bottom of the pressure-volume diagram of Figure 4 traces the other two strokes of the four-stroke cycle. On the exhaust stroke, from 5 to 6, the rising piston expels most of the remaining combustion products from the cylinder. On the intake stroke, from 6 to 7 (= 1), the descending piston inducts a fresh charge for repetition of the cycle. The net thermodynamic work developed in this cycle is proportional to the area enclosed by the pressure-volume diagram. In the ideal case, both the exhaust and intake strokes occur at atmospheric pressure, so they have no effect on the net output work. That justifies their exclusion from the thermodynamic representation of the ideal Otto

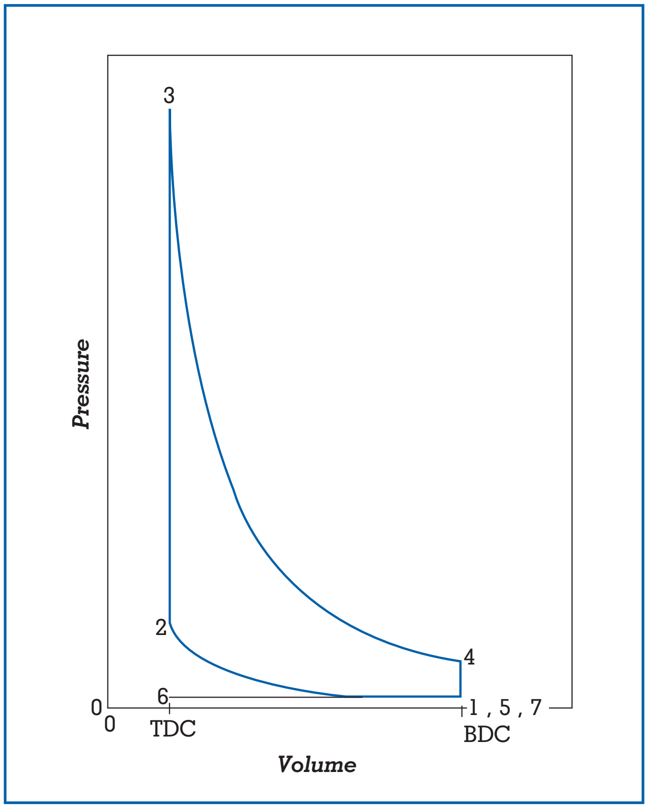


Figure 4.
Pressure-volume diagram for ideal Otto cycle (1-2-3-4-5), with exhaust and intake of four-stroke cycle (5-6-7) added .

cycle. The pressure-volume diagram for an actual operating engine deviates somewhat from this idealization because of such factors as noninstantaneous combustion, heat loss from the cylinder, and pressure loss across the valves during intake and exhaust.

Two-Stroke Piston Engine

Toward the end of the nineteenth century, successful two-stroke engines operating on the Otto cycle were developed by Dugald Clerk, James Robson, Karl Benz, and James Day. In this engine, the intake, combustion, expansion, and exhaust events all occur with but two piston strokes, or one crankshaft revolution. In principle this should double the output of a four-stroke engine of equal piston displacement. However, instead of the intake and exhaust events taking place during sequential strokes of the piston, they occur concurrently while the piston is near BDC. This impairs the ability of the engine to induct and retain as much of the fresh charge as in an equivalent four-stroke engine, with its separate intake and exhaust strokes. Thus the power delivered per unit of piston displacement is somewhat less than twice that of a comparable four-stroke engine, but still translates into a size and weight advantage for the two-stroke engine.

To deliver a fresh charge to the cylinder in the absence of an intake stroke, the two-stroke engine requires that the incoming charge be pressurized slightly. This is often accomplished by using the underside of the piston as a compressor, as illustrated in Figure 5. In Figure 5a, the piston is rising to compress the charge trapped in the cylinder. This creates a subatmospheric pressure in the crankcase, opening a spring-loaded inlet valve to admit a fresh charge. In Figure 5b, the mixture has been ignited and burned as the piston descends on the expansion stroke to extract work from the products. Later during the piston downstroke, Figure 5c, the top of the piston uncovers exhaust ports in the cylinder wall, allowing combustion products to escape. During descent of the piston, Figures 5b and c, the air inlet valve has been closed by its spring, and the underside of the piston is compressing the charge in the crankcase. In Figure 5d, the piston is just past BDC. Intake ports in the cylinder wall have been uncovered, and the piston, which has just completed its descent, has transferred the fresh charge from the crankcase into the cylinder through a transfer passage.

Note in Figure 5 that with the piston near BDC, both intake and exhaust ports are open concurrently. This provides a pathway whereby some of the incoming charge can “short-circuit” the cycle and exit with the exhaust gas. If the engine uses an upstream carburetor to mix fuel into the air before the charge enters the crankcase, then a fraction of the fuel leaves with the exhaust gas. That penalizes fuel economy and increases exhaust emissions. This escape path for unburned fuel can be eliminated by injecting fuel directly into the cylinder after both ports are closed, but at the cost of increased complexity.

If the crankcase compression illustrated in Figure 5 is used, the reservoir of lubricating oil normally contained in the crankcase of a four-stroke engine (see Figure 1) must be eliminated. Cylinder lubrication is then usually accomplished by mixing a small quantity of oil into the fuel. This increases oil consumption. An alternative allowing use of the

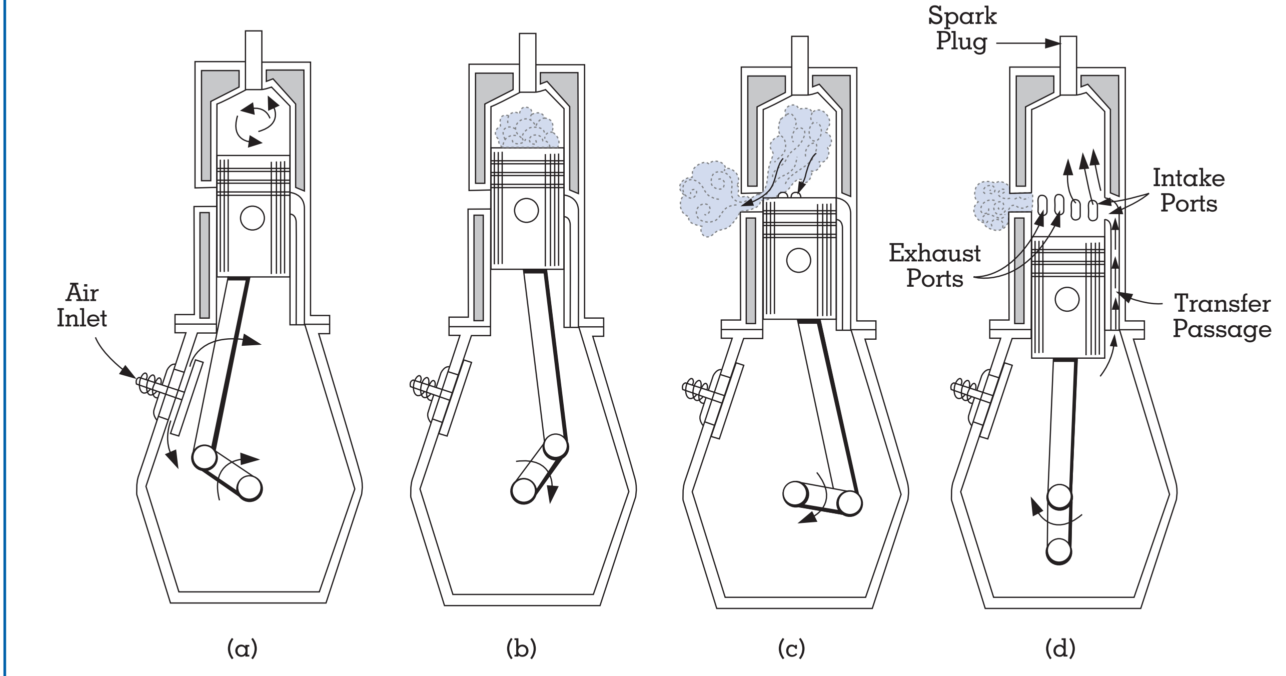


Figure 5.
Two-stroke cycle with crankcase compression.

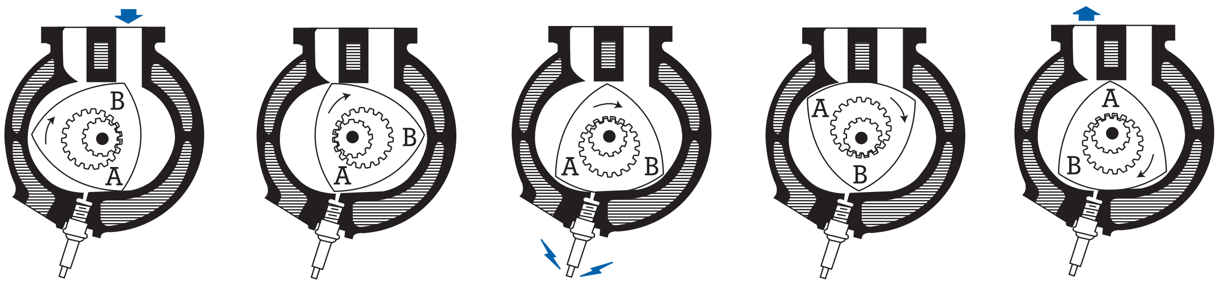
crankcase as an oil reservoir is to eliminate the transfer passage and add an engine-driven blower to provide the pressurized fresh charge directly through the intake ports. Again this increases engine complexity.

Rotary Engine

The four-stroke and two-stroke engines described above both use the slider-crank mechanism to transform piston work into crankshaft torque, but other intermittent-combustion engines have been conceived that use different kinematic arrangements to achieve this end. The only one that has realized significant commercial success is the rotary engine first demonstrated successfully in Germany by Felix Wankel in 1957.

Illustrated in Figure 6, this engine incorporates a flat three-sided rotor captured between parallel end walls. The rotor orbits and rotates around the central shaft axis, and within a stationary housing that is specially shaped so that the three apexes of the rotor always remain in close proximity to the inner wall of the housing. Linear apex seals separate three chambers, each enclosed by a rotor flank, the stationary housing, and the two end walls. The chambers are further sealed by rotor side seals that rub against the end walls. As the rotor orbits and rotates, the volume of each chamber periodically increases and decreases, just as the cylinder volume above a piston of the slider-crank mechanism changes throughout the engine cycle. A crankshaft-mounted eccentric transmits the output work from the rotor to the engine output shaft.

The engine cycle can be understood by following the flank AB in Figure 6. In Figure 6a, the fresh charge is entering the chamber through a peripheral inlet port. As the rotor rotates clockwise, in Figure 6b the volume of the chamber is decreasing to compress the charge. In Figure 6c, the chamber volume is near its minimum and the spark plug has ignited the mix-



a. Intake b. Compression c. Ignition d. Expansion e. Exhaust

Figure 6.
Wankel rotary combustion engine.

ture. In Figure 6d, the burned products are expanding. In Figure 6e, apex A has uncovered the peripheral exhaust port to release the expanded products. Thus each chamber mimics the four-stroke cycle of the slider-crank mechanism.

Given the three flanks of the rotor, there are three power pulses per rotor revolution. The output shaft rotates at three times rotor speed. Thus there is one power pulse for each revolution of the output shaft, just as in a two-stroke piston engine. This, combined with the absence of connecting rods and the ability of the engine to run smoothly at high speeds, contributes to the compactness of the Wankel rotary engine. However, the segments of the housing exposed to the heat of the combustion and exhaust processes are never cooled by the incoming charge, as in the reciprocating piston engine. The high surface-to-volume ratio of the long, thin combustion chamber promotes high heat loss. The chambers have proved difficult to seal. Such factors penalize the fuel economy of the Wankel engine relative to its slider-crank counterpart, which is one of the factors that has impaired its broader acceptance.

MULTICYLINDER ENGINES

Otto's single-cylinder engine of 1876 had a nominal cylinder bore of half a foot and a piston stroke of a foot. Operating at 180 rpm, it developed about 3 horsepower from its 6-liter displacement. Today a 6-liter automotive engine would likely have eight cylinders and deliver more than a hundred times as much power.

The most common arrangements for multicylinder engines are illustrated in Figure 7. Configurations employing four cylinders in line (I-4), and vee arrangements of either six (V-6) or eight (V-8) cylinders, currently dominate the automotive field. However, I-3, I-5, I-6, V-10, V-12, and six-cylinder horizontally opposed (H-6) arrangements are used as well.

Also represented in Figure 7 is the radial arrangement of cylinders common in large aircraft engines before the advent of jet propulsion. Five, seven, or nine cylinders were arranged in a bank around the crankshaft. Larger engines used two banks, one behind the other. One of the last, most powerful radial aircraft engines employed twenty-eight cylinders in four banks of seven cylinders each. These radial aircraft engines were air-cooled.

Individual-cylinder piston displacement in contemporary gasoline engines generally ranges from 0.15 to 0.85 liter, with the bore/stroke ratio varying between 0.8 and 1.3. Mean piston speed, which is twice the product of stroke and crankshaft rotational speed, generally falls between 8 and 15 meters/per second. Thus engines with larger cylinders are designed for slower rotational speeds. Choosing the number of cylinders

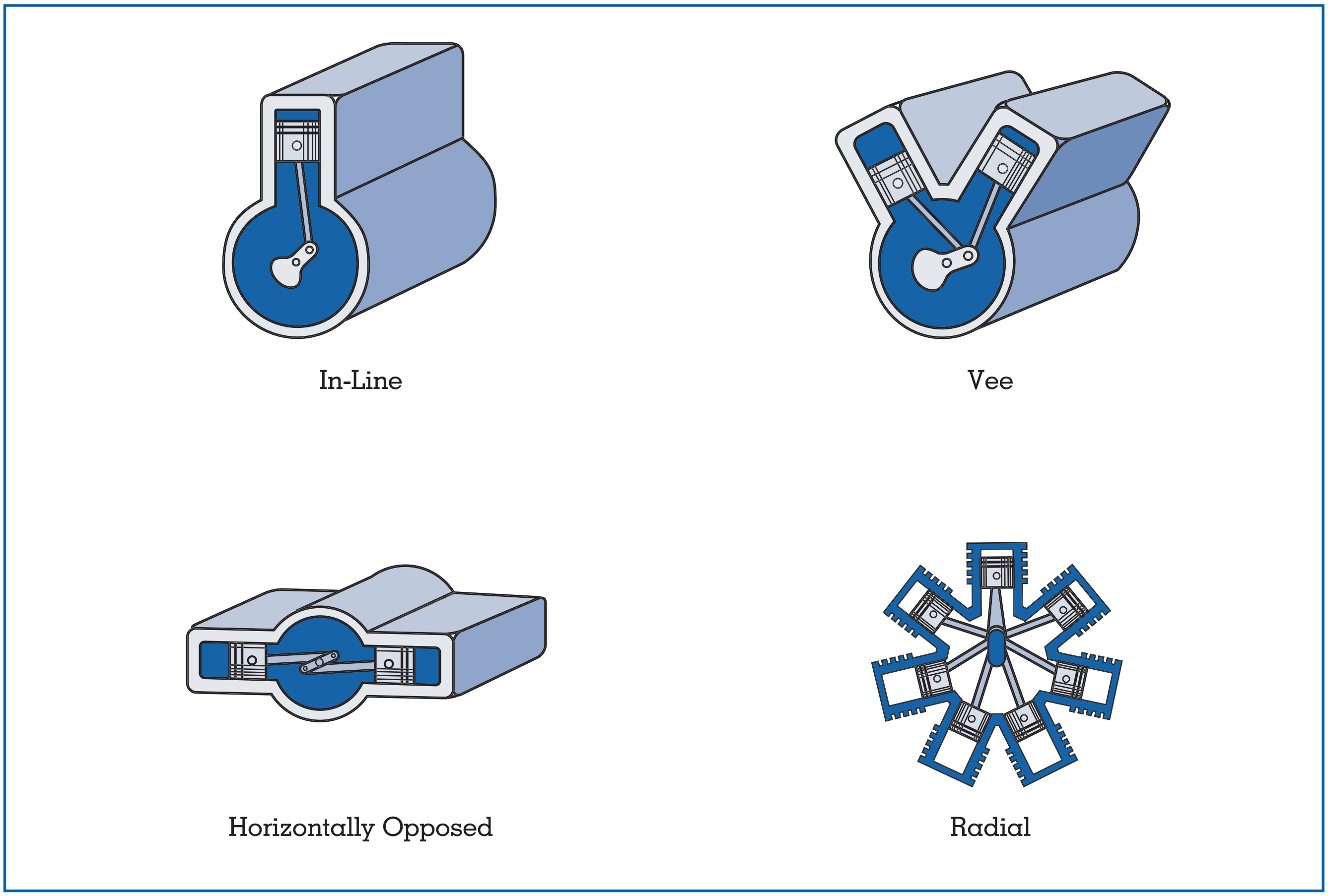

In-Line
Vee
Horizontally Opposed
Radial

Figure 7.
Some multicylinder engine arrangements.

and their physical arrangement in engine design are often influenced significantly by such issues as the dimensions of the engine compartment or the bore-centerline spacing built into existing tooling.

ENGINE EFFICIENCY

It is common practice to measure the effectiveness of the engine as an energy conversion device in terms of the fraction of the fuel energy consumed that is actually delivered as useful output work. This fraction is customarily termed "thermal efficiency." In vehicular use, fuel economy (e.g., miles/per gallon) is directly proportional to thermal efficiency, although vehicle fuel economy also depends on such parameters as vehicle weight, aerodynamic drag, rolling resistance, and the energy content of a gallon of fuel.

The efficiency of the 1860 Lenoir engine was no more than 5 percent. An efficiency of 14 percent was claimed for Otto's original four-stroke engine of 1876. The best efficiency of modern automobile engines burning gasoline is in the 30 to 35 percent range. Small engines are generally somewhat less efficient than large ones. As discussed above, the simplicity of the crankcase-scavenged two-stroke engine, commonly employed in such low-power engines as those used in gardening equipment, carries an additional efficiency penalty. At the other extreme, large, slow-speed marine engines have demonstrated efficiencies in the 50 percent range, but these are compression-ignition engines operated on diesel fuel rather than gasoline.

The best efficiency attainable from a gasoline engine of specified power rating depends heavily on four parameters: compression ratio, air/fuel ratio, spark timing, and the fraction of the mechanical energy developed in the cylinder or cylinders devoted to overcoming parasitic losses. The compression ratio is the ratio of cylinder volume at BDC to volume at TDC. The air/fuel ratio is simply the mass of air inducted by the cylinders divided by the mass of fuel added to it. The spark timing is the number of degrees of crankshaft rotation before TDC at which the ignition spark is discharged. Parasitic losses include (1) engine friction, (2) the energy consumed by such essential components as the oil pump, fuel pump, coolant pump, generator used for charging the battery, and radiator cooling fan, and (3) the pumping loss associated with drawing the fresh charge into the cylinder during intake and expelling the combustion products from the cylinder during the exhaust.

Compression Ratio

Raising the compression ratio generally increases the thermal efficiency of the gasoline engine, as can be demonstrated by thermodynamic analysis of the engine cycle. The simplest such analysis is of the air-standard Otto cycle (see Figure 3, path 1-2-3-4-5). This technique assumes that the cylinder is filled with air behaving as a perfect gas (i.e., with constant specific heat), that from 2 to 3 heat is added to the air from an external source (no internal combustion), that from 4 to 5 heat is rejected from the air to the surroundings (no exhaust process), and that during compression (1 to 2) and expansion (3 to 4) the air exchanges no heat with its surroundings. This constitutes a thermodynamic cycle because the composition, pressure, and temperature of the cylinder contents are the same at the end of the cycle, point 5, as at the beginning, point 1. The thermodynamic efficiency (η) of this air-standard Otto cycle is given by $\eta=1\text{-}R^a$ where R = compression ratio and $\alpha = -0.4$.

A thermodynamically closer approximation to the actual engine is provided by fuel-air cycle analysis. In it the cylinder contents during compression are assumed to consist of air, fuel, and residual burned gas from the previous cycle. The specific heat of the cylinder gas is varied with temperature and composition. Constant-volume combustion is assumed, with the products in chemical equilibrium. At the end of the expansion stroke and during the exhaust stroke, the combustion products are allowed to escape from the cylinder, being replaced by a fresh charge during the intake stroke. No exchange of heat between the cylinder gas and its surroundings is considered.

In contrast to the air-standard cycle, the fuel-air cycle is a cycle in the mechanical sense only. It is not a true thermodynamic cycle because neither the composition nor the temperature of the cylinder contents returns to its initial state at point 1. This, of course, is also true of a real internal-combustion engine. With the more realistic assumptions of the fuel-air cycle, exponent $\alpha \equiv 0.28$ in $\eta=1\text{-}R^a$ for the normal range of air-fuel ratios used with a gasoline engine. This relationship defines an upper limit for efficiency of an otto-cycle engine burning premixed air and gasoline.

According to $\eta=1\text{-}R^a$, the efficiency of the ideal Otto cycle increases indefinitely with increasing compression ratio. Actual engine experiments, which inherently include the real effects of incomplete combustion, heat loss, and finite combustion time neglected in fuel-air cycle analysis, indicate an efficiency that is less than that given by $\eta=1\text{-}R^a$ when $\alpha = 0.28$. Furthermore, measured experimental efficiency reached a maximum at a compression ratio of about 17 in large-displacement automotive cylinders but at a somewhat lower compression ratio in smaller cylinders.

Otto's engine of 1876 had a compression ratio of 2.5. By the beginning of the twentieth century, the efficiency advantage of higher compression ratios was recognized, but liquid-fueled engines built with higher compression ratios often experienced an unexplained, intolerable, explosive combustion noise. Over a couple of decades, researchers came to realize that the noise was caused by autoignition in the unburned mixture ahead of the normal flame front as it advanced from the spark plug. This autoignition is caused by the increase in pressure and temperature within the cylinder once normal combustion has been initiated. The phenomenon can be likened to the compression ignition of fuel in the diesel engine. Experience taught that a fuel such as kerosene was more prone to this "combustion knock" than more volatile petroleum products, such as gasoline.

Thomas Midgley, working with a group of researchers under Charles F. Kettering, discovered a fuel-additive, tetraethyl lead that enhanced the knock resistance of existing fuels. By the time this additive entered commercial use in 1923, the average compression ratio of new U.S. cars had advanced to 4.3.

An octane rating scale was devised for fuels to quantify their knock resistance. Further research led to cataloguing the antiknock qualities of the myriad individual hydrocarbon species found in gasoline.

Fuel refiners learned how to rearrange the atoms in fuel molecules to enhance the octane number of the fuel available at gasoline stations. Studying the combustion process led to combustion-chamber designs with lower octane requirements. All of these activities fostered a nearly continuous increase in the average compression ratio of the new U.S. car, as follows: 4.3 in 1923; 5.1 in 1930; 6.3 in 1940; 7.0 in 1950; 9.0 in 1960; 9.8 in 1970; and 8.6 in 1980. The drop from 1970 to 1980 resulted from the imposition of increasingly severe exhaust emission standards in the United States, which included the need to design engines to run on unleaded gasoline. This was necessitated by the introduction of the catalytic converter to the exhaust system. The catalyst is poisoned by lead in the fuel.

Since 1980, further improvements in engine design have allowed the average compression ratio to creep back up to the range of 9 to 10, despite the absence of tetraethyl lead. Barring an unanticipated increase in the octane number of commercial gasoline, a further major increase in the average compression ratio of the contemporary gasoline engine is unlikely. A variable compression ratio engine that would retain the currently acceptable ratio at high engine loads but increase it at light load, where combustion knock is less likely, is a concept that has long existed but has yet to be implemented commercially on a significant scale.

Air-Fuel Ratio

The second important parameter affecting efficiency is air/fuel ratio. For every hydrocarbon fuel, there is an air/fuel ratio that, in principle, causes all the hydrogen in the fuel to burn to water vapor and all the carbon in the fuel to burn to carbon dioxide. This chemically correct proportion is called the stoichiometric ratio.

If the engine is fed a mixture containing more fuel than the stoichiometric amount, the mixture is said to be rich, and carbon monoxide (CO) and hydrogen are added to the combustion products. Because these two gases are fuels themselves, their presence in the exhaust signifies incomplete combustion and wasted energy.

If the engine is fed a mixture containing more air than the stoichiometric amount, the mixture is said to be lean, and unconsumed oxygen appears in the combustion products. This signifies that the full power-producing potential of the air inducted by the engine has not been utilized.

Among the lesser constituents of engine exhaust are unburned hydrocarbons (HC) and oxides of nitrogen (NOx). Exhaust HC concentration increases rapidly as the mixture is enriched from stoichiometric. Exhaust NOx concentration peaks at a mixture ratio slightly leaner than stoichiometric and falls off as the mixture is made either richer or leaner. These two exhaust-gas constituents react in the atmosphere in the presence of sunlight to form ozone, a major ingredient of photochemical smog. Consequently their mass emissions, along with that of CO, are regulated for environmental reasons.

For conventional gasoline, the stoichiometric ratio is approximately 14.7. Its precise value varies slightly with the composition of the gasoline. Maximum power is achieved with a slightly rich air/fuel ratio—say, 12.5. Maximum efficiency is achieved with a slightly lean mixture—say, 16—although this best-economy mixture ratio is somewhat dependent on combustion quality.

For most of the nineteenth century, the fuel and air were mixed upstream of the engine in a carburetor. In the automobile, the carburetor contains an inlet throttle valve linked to the accelerator pedal. The throttle valve is fully opened when the accelerator pedal is depressed to the floorboard, but barely open when the pedal is released so that the engine runs at its idle condition. The intent of the carburetor is to supply the engine with its highest air-fuel ratio during midspeed cruise conditions. As the throttle approaches its wide-open position, the mixture is enriched to maximize engine power. The mixture prepared by the carburetor is also temporarily enriched when the throttle is opened suddenly, as during a sharp vehicle acceleration, because the resultant fuel-flow increase lags behind the increase in airflow. As engine idle is approached, the mixture is enriched to compensate for poor combustion quality at the low-speed, low-pressure idle condition. For cold starting, the carburetor includes a choke valve that enriches the mixture so that enough of the fuel is vaporized near the spark plug to ensure ignition.

In early days, the mixture supplied to the engine was often quite rich in order to ensure smooth engine operation. This, of course, wasted fuel energy through incomplete combustion. Fleet surveys in early years showed marked improvement in the percentage of fuel that was wasted: 15.5% in 1927, 7.5%

in 1933, and only 1.5% by 1940. (The vehicles were operated at 40 miles per hour, with gasoline having a stoichiometric air-fuel ratio of 15).

Following World War II, engine improvements included further leaning of the mixture, especially toward the leaner-than-stoichiometric best-economy ratio during cruise. Then in the early 1980s, U.S. emission standards became so stringent that the catalytic converter, which had been added in about 1975 to oxidize HC and CO in the exhaust, also had to reduce NOx. This has led to the widely used three-way catalyst that controls all three emissions, but only if the air/fuel ratio is kept in a narrow range about the stoichiometric ratio.

Such tight mixture control is beyond the capability of the traditional carburetor. Consequently, after sorting through a number of alternatives, industry has settled on closed-loop-controlled port-fuel injection. Typically, an electronically controlled fuel injector is mounted in the intake port to each cylinder. A sensor in the air intake system tells an on-board computer what the airflow rate is, and the computer tells the fuel injectors how much fuel to inject for a stoichiometric ratio. An oxygen sensor checks the oxygen content in the exhaust stream and tells the computer to make a correction if the air/fuel ratio has drifted outside the desired range. This closed-loop control avoids unnecessary use of an inefficient rich mixture during vehicle cruise.

An old concept for increasing thermal efficiency at part load that is the subject of renewed interest is the use of a much leaner average mixture ratio at part load. To burn such an overall-lean mixture in the time made available by the contemporary gasoline engine, this concept calls for stratifying the fresh charge. Thus a combustible richer-than-average cloud of air and fuel is segregated near the spark plug and surrounded by fuel-free air. Although gains in part-load efficiency have been demonstrated with this concept, emissions control issues must be resolved before this concept is suitable for automotive use.

Ignition Timing

The third important parameter affecting engine efficiency is ignition timing. For each engine operating speed and throttle position, there is a spark timing giving best efficiency. In the early days of the automobile, timing was adjusted manually by the driver, so fuel economy was somewhat dependent on driver skill.

Eventually, automatic control of spark advance evolved. The conventional controller included two essential elements. Mechanically, spring-restrained rotating flyweights responding to centrifugal force advanced the spark as engine speed increased. Pneumatically, a diaphragm that sensed the increasing vacuum in the intake manifold as the intake throttle was closed added additional spark advance as the resulting reduction in cylinder pressure slowed flame speed.

With the advent of electronic engine controls, the on-board computer now manages ignition timing. It offers greater versatility in timing control, being able to integrate signals from sensors monitoring engine speed, airflow rate, throttle position, intake-manifold pressure, engine temperature, and ambient pressure.

Combustion knock can still be experienced in the engine near full throttle if the octane rating of the fuel is too low for the compression ratio, or vice versa. When ignition timing was controlled manually, the driver could retard the spark when knock was heard. In the era of mechanical/pneumatic control, the built-in timing schedule typically included slight retard from best-economy spark advance near full throttle to allow use of a slightly higher compression ratio for greater engine efficiency at part load. With the advent of electronic control, many engines are now equipped with knock sensors that detect incipient knock and maximize efficiency by retarding the spark just enough to avoid it. This technology has contributed to the modest increase in average compression ratio in automobiles since the 1980s.

Parasitic Losses

When parasitic losses are measured as a fraction of the power developed in the cylinder or cylinders, that fraction can be improved in two ways. Either the parasitic losses themselves can be reduced, or the power developed in the cylinder or cyinders can be increased. Over the years, both paths have been pursued.

Energy lost to mechanical friction has been decreased significantly through a large number of small, incremental improvements. In the typical modern engine, about half of the friction occurs between the cylinder wall and the piston/piston ring assembly. The sidewall contact surfaces of the piston have been specially shaped to minimize engine friction. Piston ring tension has been reduced, commensurate with adequate sealing of the combustion chamber. Oil viscosity has been chosen, and oil composition altered, to reduce friction. Analysis of the effects of operating pressure and temperature on engine dimensional changes has led to improved conformity between piston and cylinder during engine operation. Bearing dimensions and clearance have been selected to decrease friction. Reciprocating elements of the slider-crank mechanism have been lightened to decrease inertia loads on the bearings. In the valvetrain, sliding contact between adjacent members has often been replaced with rolling contact. Valves have been lightened, and valve-spring tension reduced.

Improved aerodynamic design of intake and exhaust ports has contributed to lower pumping losses. Paying greater attention to pressure drop in the catalytic converter also has helped. When recirculating exhaust gas into the intake system at part load to decrease NOx production, a common practice since 1973, the throttle has to be opened further to produce the same power at a given speed. This reduces intake pumping loss at a given part-load power.

Power requirements for oil, coolant, and fuel pumps are generally minor. Coolant-system power requirement increased when the thermo-siphon cooling system of early automobiles was superseded by an engine-driven coolant pump as engine power rating increased. Avoiding the use of unnecessarily high oil pressure has been beneficial. On the other hand, the higher pressure used in fuel injectors, compared to the fuel-pressure requirement of the carburetor, has not. In passenger cars, perhaps the largest reduction in auxiliary-power consumption has come by changing from an engine-driven cooling fan, running continuously, to an electrically driven fan that is inoperative as long as the radiator provides adequate cooling from ram air flowing through it as a result of vehicle forward motion.

The power that can be produced in each cylinder depends on the amount of air it can induct. Recent interest in replacing the traditional single-intake valve per cylinder with two smaller ones, and in a few cases even three, has increased total intake-valve area, and hence engine airflow at a given speed. At the same time, engine speed capability has been increased. These changes have raised the air capacity of the engine, resulting in increased power delivery for a given displacement. If the engine displacement is then decreased to hold maximum power constant, the parasitic losses are smaller for the same output power capability. This can be translated into improved vehicle fuel economy.

Charles A. Amann

See also: Automobile Performance; Combustion; Diesel Cycle Engines; Drivetrains; Engines; Otto, Nikolaus August; Steam Engines; Stirling Engines; Tribology; Turbines, Gas; Turbines, Steam.

BIBLIOGRAPHY

Heywood, J. B. (1998). *Internal Combustion Engine Fundamentals*. New York: McGraw-Hill Book Company.

Obert, E. F. (1973). *Internal Combustion Engines and Air Pollution*. New York: Indext Educational Publishers.

Stone, R. (1999). *Introduction to Internal Combustion Engines*, 3rd ed. Warrendale, PA: Society of Automotive Engineers.

GEAR

See: Drivetrains; Mechanical Transmission of Energy

GEAR BOX

See: Drivetrains

GENERATOR, ELECTRICAL

See: Electric Motor Systems

GEOGRAPHY AND ENERGY USE

Geography looks for patterns in the distribution of diverse phenomena, and it tries to explain their variation by examining a wide range of underlying environmental and socioeconomic factors. Many fundamental realities of modern civilization cannot be fully appreciated without the basic understanding of the geography of energy use. There is a very high degree of inequality in the use of energy resources throughout the word: this is true for aggregate use of energy, for every individual fuel or renewable energy flow, as well as for the consumption of electricity. These consumption disparities are inevitably accompanied by large differences in energy self-sufficiency and trade.

A country's dependence on energy imports, its size and climate, its stage of economic development, and the average quality of life of its inhabitants are the key variables determining the patterns of fuel and electricity consumption among the world's 180 countries. Large countries also have distinct regional patterns of energy consumption, but the progressing homogenization of energy use has been reducing some of these differences. For example, in the United States sports utility vehicles are now favored by both urban and rural drivers, and, unlike a few decades ago, utilities in both southern and northern states have summer peak loads due to air conditioning.

ENERGY SELF-SUFFICIENCY AND IMPORTS

Energy self-sufficiency is not just a matter of possessing substantial domestic resources; it is also determined by the overall rate of consumption. The United States, following the sharp decline of Russian's output, is now the world's second largest producer of crude oil (after Saudi Arabia); it is also the world's largest importer of crude oil, buying more than 50 percent of it in order to satisfy its high demand for transportation energy. The only affluent economies self-sufficient in fossil fuels are Canada (due to its relatively small population and vast mineral resources), and the United Kingdom and Norway (thanks to the North Sea oil and gas fields).

Russia is the most notable case of a middle-income country with large surpluses of energy. Its huge oil and gas exports are now surpassed only by Saudi Arabia, the largest OPEC oil producer. OPEC produces about 40 percent of the world's crude oil output and it supplies about 45 percent of all traded petroleum. In total, almost 60 percent of the world's crude oil extraction is exported from about forty-five hydrocarbon-producing countries—but the six largest exporters (Saudi Arabia, Iran, Russia, Norway, Kuwait, and the United Arab Emirates) sell just over 50 percent of the traded total. In contrast, more than 130 countries import crude oil and refined oil products; besides the United States, the largest buyers are Japan, Germany, France, and Italy.

About 20 percent of the world's natural gas production was exported during the late 1990s, three-quarters of it through pipelines, and the rest by LNG tankers. The former Soviet Union, Canada, the Netherlands, and Norway are the largest pipeline exporters, while Indonesia, Algeria, and Malaysia dominate the LNG trade. The largest importers of piped gas are the United States, Germany, Italy, and France; Japan and South Korea buy most of the LNG.

In comparison to hydrocarbons, coal trade is rather limited, with only about a tenth of annual extraction of hard coals and lignites exported, mainly from Australia, United States, South Africa, and Canada. Because of their lack of domestic resources and large iron and steel industries, Japan and South Korea are the two biggest buyers of steam and metallurgical coal. Although the development of high-voltage networks has led to rising exports of electricity, less than 4 percent of the global generation is traded internationally; France (mainly due to its large nuclear capacity), Canada, Russia, and Switzerland are the largest exporters. This trade will rise as the markets for electricity grow, and electricity transmission and distribution improves. Only 33 of the 180 countries are net energy exporters, and about 70 countries do not export any commercial energy. Self-sufficiency in energy supply was a major goal of many nations following the first "Energy Crisis" in 1973, but the collapse of OPEC's high crude oil prices in 1985 and the subsequent stabilization of the world's oil supply have greatly lessened these concerns during the 1990s.

AGGREGATE ENERGY USE AND ITS COMPOSITION

During the late 1990s annual consumption rates of commercial energy ranged from less than 25 kgoe (or

	Energy	*Consumption*			
Source	***1950***	***1973***	***1985***	***1990***	***1998***
Solid fuels	1040	1700	2170	2330	2220
Liquid fuels	460	2450	2520	2790	3010
Natural gas	170	1100	1460	1720	2020
Primary Electricity	30	130	300	360	440
Total	1700	5380	6450	7200	7650

Table 1.
Global Energy Consumption
Note: All fuel conversions according to the UN rates; all primary electricity expressed in terms of its thermal equivalent.

less than 1 GJ) per capita in the poorest countries of sub-Saharan Africa to nearly 8 toe (or more than 300 GJ) per capita in the United States and Canada. The global mean was close to 1.5 toe (or 60 GJ) per capita—but the sharply bimodal distribution of the world's energy use reflecting the rich-poor divide meant that only a few countries (including Argentina and Portugal) were close to this level. Affluent countries outside North America averaged close to 3.5 toe per capita, while the mean for low-income economies was just 0.6 toe, close to the Chinese average. This huge gap in aggregate energy consumption has been narrowing slowly. In 1950 industrialized countries consumed about 93 percent of the world's commercial primary energy. Subsequent economic development in Asia and Latin America reduced this share, but by 1998 industrialized countries containing just one fifth of global population still consumed about 70 percent of all primary energy.

The United States alone, with less than 5 percent of the world population, claims about 25 percent of the world in total commercial energy use. Among the world's affluent countries only Canada has a similarly high per capita use of fossil fuels and primary electricity (about 8 t of crude oil equivalent per year). In spite of its huge fuel and electricity production, the United States imports more than a fifth of its total energy use, including more than half of its crude oil consumption. Almost two fifths of all commercial energy is used by industry, a quarter in transportation, a fifth by households, and a bit over one sixth goes into the commercial sector.

In contrast, during the late 1990s the poorest quarter of humanity—made up of about fifteen sub-Saharan African countries, Nepal, Bangladesh, Indochina, and most of rural India—consumed a mere 2.5 percent of all commercial energy. The poorest people in the poorest countries, including mostly subsistence peasants but also millions of destitute people in large cities, do not directly consume any commercial fuels or electricity at all!

All of the world's major economies, as well as scores of smaller, low-income nations, rely mainly on hydrocarbons. Crude oil now supplies two-fifths of the world's primary energy (Table 1). There are distinct consumption patterns in the shares of light and heavy oil products: the United States burns more than 40 percent of all its liquid fuels as gasoline, Japan just a fifth; and the residual fuel oil accounts for nearly a third of Japanese use, but for less than 3 percent of the U.S. total. Small countries of the Persian Gulf have the highest per capita oil consumption (more than 5 t a year in the United Arab Emirates and in Qatar); the U.S. rate is more than 2.5 t a year; European means are around 1 t; China's mean is about 120 kg, and sub-Saharan Africa is well below 100 kg per capita.

Natural gas supplies nearly a quarter of the world's primary commercial energy, with regional shares ranging from one half in the former Soviet Union to less than 10 percent in the Asia Pacific. The United States and Russia are by far the largest consumers, followed by the United Kingdom, Germany, Canada, Ukraine, and Japan; leaving the small Persian Gulf emirates aside, Canada, the Netherlands, the United States, Russia, and Saudi Arabia have the highest per capita consumption. Coal still provides 30 percent of the world's primary commercial energy, but outside China and India—where it is still used widely for heating, cooking, and in transportation, and where it supplies, respectively, about three quarters and two

Global Electricity Production (in TWh)

Source	1950	1973	1985	1990	1997
Fossil Fueled	620	4560	6260	7550	8290
Hydro	340	1320	2000	2210	2560
Nuclear	—	190	1450	1980	2270
Geothermal	—	—	30	40	50
Wind, Solar	—	—	—	50	80
Total	960	6070	9740	11,830	13,250

Table 2.
Global Electricity Production

thirds of commercial energy consumption—it has only two major markets: electricity generation and production of metallurgical coke.

When converted at its heat value (1 kWh = 3.6 MJ) primary electricity supplied about 6 percent of global commercial energy consumption during the late 1990s. Hydro and nuclear generation account for about 97 percent of all primary electricity, wind, geothermal, and solar—in that order—for the rest (Table 2). Canada, the United States, Russia, and China are the largest producers of hydroelectricity; the United States, Japan, and France lead in nuclear generation; and the United States, Mexico, and the Philippines are the world's largest producers of geothermal electricity.

ENERGY CONSUMPTION AND ECONOMIC DEVELOPMENT

On the global level the national per capita rates of energy consumption (Table 3) correlate highly ($r \geq 0.9$) with per capita gross domestic product (GDP): the line of the best fit runs diagonally from Nepal (in the lower left corner of a scattergram) to the United States (in the upper right corner). This commonly used presentation has two serious shortcomings: the exclusion of biomass fuels and the use of exchange rates in calculating national GDPs in dollars. Omission of biomass energies substantially underrates actual fuel use in low-income countries where wood and crop residues still supply large shares of total energy demand (more than 90% in the poorest regions of sub-Saharan Africa; about a fifth in China), and official exchange rates almost invariably undervalue the GDP of low-income countries.

Inclusion of biomass energies and comparison of GDPs in terms of purchasing power parities (PPP) weaken the overall energy-economy correlation, and disaggregated analyses for more homogeneous regions show that energy-GDP correlations are masking very large differences at all levels of the economic spectrum. Absence of any strong energy-GDP correlation is perhaps most obvious in Europe: while France and Germany have very similar PPP-adjusted GDPs, Germany's per capita energy use is much higher; similar, or even larger, differences can be seen when comparing Switzerland and Denmark or Austria and Finland.

Another revealing look at the energy-economy link is to compare national energy/GDP intensities (i.e., how many joules are used, on the average, to produce a unit of GDP) expressed in constant monies. These rates follow a nearly universal pattern of historical changes, rising during the early stages of economic development and eventually commencing gradual declines as economies mature and become more efficient in their use of energy. This shared trend still leaves the economies at very different energy-intensity levels. The U.S. energy intensity fell by more than a third since the mid-1970s—but this

Countries	*Energy consumption*			
	1950	*1973*	*1985*	*1995*
World	700	1450	1330	1300
Developed countries				
U.S.	5120	8330	6680	7500
France	1340	3150	2810	2650
Japan	390	2760	2650	2330
Largest oil exporters				
Saudi Arabia	110	690	4390	4360
Russia	1120	3540	4350	4630
Developing countries				
Brazil	140	440	500	640
China	60	300	490	680
India	70	130	180	270
Poorest economies				
Bangladesh	10	20	40	70
Ethiopia	—	20	10	20

All fuel conversions according to the UN rates; all primary electricity expressed in terms of its thermal equivalent.

Table 3.
Per Capita Consumption of Commercial Energy

impressive decline has still left the country far behind Japan and the most affluent European countries. China cut its energy intensity by half since the early 1980s—but it still lags behind Japan.

Weak energy-GDP correlations for comparatively homogeneous groups of countries and substantial differences in energy intensities of similarly affluent economies have a number of causes. A country's size plays an obvious role: there are higher energy burdens in integrating larger, and often sparsely inhabited, territories by road and rail, and in affluent countries the need to span long distances promotes air travel and freight, the most energy-intensive form of transportation. Not surprisingly, Canadians fly two to three times more frequently than most Europeans do. Even within countries, there are wide consumption disparities. In the United States, gasoline and diesel fuel consumption by private cars is highest in Wyoming where an average car travels nearly 60 percent more miles annually than the national mean; Montana and Idaho are the other two thinly populated states with considerably longer average car travel.

Climate is another obvious determinant of a country's energy use. For example, Canada averages annually about 4,600 heating degree days compared to Japan's 1,800, and this large difference is reflected in a much higher level of household and institutional fuel consumption. But climate's effects are either partially negated or highly potentiated by different lifestyles and by prevailing affluence. The Japanese and British not only tolerate much lower indoor temperatures (below 15°C) than Americans and Canadians (typically above 20°C), but they also commonly heat only some rooms in the house or, in the case of Japan, merely parts of some rooms (using *kotatsu* foot warmers).

Larger, overheated and often poorly insulated American houses mean that the U.S., with the national annual mean of 2,600 heating degree days, uses relatively more fuel and electricity for heating than does Germany with its mean of 3,200 heating degree days. Space heating takes half of U.S. residential consumption, water heating (with about 20%) comes second, ahead of air conditioning. Widespread adoption of air conditioning erased most of the previously large differences in residential energy consumption between the U.S. snowbelt and the sunbelt: now Minnesota and Texas, or Nevada and Montana have nearly identical per capita averages of household energy use. The most notable outliers are Hawaii (40% below the national mean: no heating, limited air conditioning) and Maine (almost 30% above the mean due to heating).

Composition of the primary energy consumption makes a great deal of difference. Because coal combustion is inherently less efficient due to the presence of combustible ash than the burning of hydrocarbons, the economies that are more dependent on coal (China, United States) are handicapped in comparison with nations relying much more on liquid fuels, natural gas, and primary electricity (Japan, France). So are the energy exporting countries: energy self-sufficiency (be it in Russia or Saudi Arabia) is not conducive to efficient conversions, but high dependence on highly taxed imports (as in Japan or Italy) promotes frugality.

Differences in industrial structure are also important: Canada is the world's leading producer of energy-intensive aluminum—but Japan does not smelt the metal at all. And as the only remaining superpower, the United States still invests heavily in energy-intensive production of weapons and in the maintenance of military capacity. Annual energy consumption of all branches of the U.S. military averaged about 25 million t of oil equivalent during the 1990s (more than half of it as jet fuel): that is more than the total primary commercial consumption of nearly two-thirds of the world's countries! In contrast, the size of Japan's military forces is restricted by the country's constitution.

But no single factor is more responsible for variations in energy intensity among high-income countries than the level of private, and increasingly discretionary (as opposed to essential), energy use. During the late 1990s, the Japanese used only about 0.4 toe in their generally cramped and poorly heated apartments and houses, but U.S. residential consumption was about 1 toe. Ubiquitous air-conditioning, overheating of oversized houses, and heating of large volumes of water explain the difference. Refrigerator and washing machine ownership is nearly universal throughout the rich world, but the appliances are smaller outside of North America where clothes dryers, dishwashers, and freezers are also much rarer.

There are even greater disparities in energy used for transportation: North American commuters commonly consume three times as much gasoline per year as do their European counterparts who rely much more on energy-efficient trains; and Americans and

Canadians also take many more short-haul flights, the most energy-intensive form of transportation, in order to visit families and friends or to go for vacation. They also consume much more fuel during frequent pastime driving and in a still growing range of energy-intensive recreation machines (SUVs, RVs, ATVs, boats, snowmobiles, seadoos).

In total, cars and light trucks consume nearly three-fifths of all fuel used by U.S. transportation. In spite of a more than 50 percent increase in average fuel efficiency since 1973, the U.S. cars still consume between 25 to 55 percent more fuel per unit distance than the average in European countries (11.6 l per 100 kilometers compared to 9.1 l in Germany, and 7.4 l in Denmark in 1995). All forms of residential consumption and private transportation thus claim more than 2 toe per capita in the United States, compared to less than 1 toe in Europe and about 0.75 toe in Japan.

ENERGY USE AND THE QUALITY OF LIFE

But does the higher use of energy correlate closely with the higher quality of life? The answer is both yes and no. There are obvious links between per capita energy use and the physical quality of life characterized above all by adequate health care, nutrition, and housing. Life expectancy at birth and infant mortality are perhaps the two most revealing indicators of the physical quality of life. The first variable subsumes decades of nutritional, health-care and environmental effects and the second one finesses these factors for the most vulnerable age group. During the 1990s average national life expectancies above 70 years required generally annual per capita use of 40 to 50 GJ of primary energy—as did the infant mortality rate below 40 (per thousand newborn).

Increased commercial energy use beyond this range brought first rapidly diminishing improvements of the two variables and soon a levelling-off with hardly any additional gains. Best national achievements—combined male and female life expectancies of 75 years and infant mortalities below ten—can be sustained with energy use of 70 GJ per year, or roughly half of the current European mean, and less than a quarter of the North American average. Annual commercial energy consumption around 70 GJ per capita is also needed in order to provide ample opportunities for postsecondary schooling—and it appears to be the desirable minimum for any society striving to combine a decent physical quality of life with adequate opportunities for intellectual advancement.

On the other hand, many social and mental components of the quality of life—including such critical but intangible matters as political and religious freedoms, or satisfying pastimes—do not depend on high energy use. Reading, listening to music, hiking, sports, gardening and craft hobbies require only modest amounts of energy embodied in books, recordings, and appropriate equipment or tools—and they are surely no less rewarding than high-energy pastimes requiring combustion of liquid fuels.

It is salutary to recall that the free press and the ideas of fundamental personal freedoms and democratic institutions were introduced and codified by our ancestors at times when their energy use was a mere fraction of ours. As a result, contemporary suppression or cultivation of these freedoms has little to do with overall energy consumption: they thrive in energy-rich United States as they do in energy-poor India, and they are repressed in energy-rich Saudi Arabia as they are in energy-scarce North Korea. Public opinion polls also make it clear that higher energy use does not necessarily enhance feelings of personal and economic security, optimism about the future, and general satisfaction with national or family affairs.

The combination of abundant food energy supplies and of the widespread ownership of exertion-saving appliances has been a major contributor to an epidemic of obesity (being at least a 35% over ideal body weight) in North America. National health and nutrition surveys in the United States how that during the 1990s every third adult was obese, and an astonishing three-quarters of all adults had body weights higher than the values associated with the lowest mortality for their height.

Vaclav Smil

See also: Agriculture; Energy Economics; Energy Intensity Trends; Reserves and Resources.

BIBLIOGRAPHY

Biesiot, W., and Noorman, K. J. (1999). "Energy Requirements in Household Consumption: A Case Study of the Netherlands." *Ecological Economics* 28:367–383.

BP Amoco. (2000). *BP Statistical Review of World Energy*. BP Amoco, London. <http://www.bpamoco.co.il/worldenergy>.

Darmstadter, J. (1977). *How Industrial Societies Use Energy: A Comparative Analysis*. Baltimore, MD: Johns Hopkins University Press.

Energy Information Administration. (2000). *Annual Energy Review*. EIA, Washington, DC. <http://www.eia.doe.gov>.

Energy Information Administration. (2000). *Annual Energy Outlook*. EIA, Washington, DC. <http://www.eia.doe.gov>.

Energy Information Administration. (2000). *State Energy Data Report, Consumption Estimates*. EIA, Washington, DC. <http://www.eia.doe.gov>.

Energy Information Administration. (2000). *International Energy Outlook*. EIA, Washington, DC. <http://www.eia.doe.gov>.

International Energy Agency. (2000). *Energy Statistics of OECD Countries*. EIA, Paris. <http://www.eia.doe.gov>.

International Energy Agency. (2000). *Energy Balances of OECD Countries*. EIA, Paris. <http://www.eia.doe.gov>.

International Energy Agency. (2000). *Energy Statistics and Balances of Non-OECD Countries*. IEA, Paris. <http://www.ei.doe.gov>.

International Energy Agency. (1997). *Indicators of Energy Use and Efficiency: Understanding the Link Between Energy and Human Activity*. Paris: OECD/EIA.

Meyers, S., and Schipper, L. (1992). "World Energy Use in the 1970s and 1980s: Exploring the Changes." *Annual Review of Energy* 17:463–505.

Sathaye, J., and Tyler, S. (1991). "Transitions in Household Energy Use in Urban China, India, the Philippines, Thailand, and Hong Kong." *Annual Review of Energy* 16:295–335.

Schipper, L. et al. (1989). "Linking Life-styles and Energy Use: A Matter of Time?" *Annual Review of Energy* 14:273–320.

Schipper, L., and Meyers, S. (1992). *Energy Efficiency and Human Activity*. New York: Cambridge University Press.

Smil, V. (1991). *General Energetics*. New York: John Wiley.

Smil, V. (1992). "Elusive Links: Energy, Value, Economic Growth and Quality of life." *OPEC Review* Spring 1992:1–21.

United Nations. (2000). *Yearbook of World Energy Statistics*. UNO, New York.

GEOTHERMAL ENERGY

Geothermal energy is heat energy that originates within Earth itself. The temperature at the core of our planet is 4,200°C (7,592°F), and heat flows outward to the cooler surface, where it can produce dramatic displays such as volcanoes, geysers, and hot springs, or be used to heat buildings, generate electricity, or perform other useful functions. This outward flow of heat is continually being maintained from within by the decay of radioactive elements such as uranium, thorium, and radium, which occur naturally in Earth. Because of its origin in radioactivity, geothermal energy can actually be thought of as being a form of natural nuclear energy.

The U.S. Department of Energy has estimated that the total usable geothermal energy resource in Earth's crust to a depth of 10 kilometers is about 100 million exajoules, which is 300,000 times the world's annual energy consumption. Unfortunately, only a tiny fraction of this energy is extractable at a price that is competitive in today's energy market.

In most areas of the world, geothermal energy is very diffuse—the average rate of geothermal heat transfer to Earth's surface is only about 0.06 watt per square meter. This is very small compared to, say, the solar radiation absorbed at the surface, which provides a global average of 110 watts per square meter. Geothermal energy can be readily exploited in regions where the rate of heat transfer to the surface is much higher than average, usually in seismic zones at continental-plate boundaries where plates are colliding or drifting apart. For example, the heat flux at the Wairakei thermal field in New Zealand is approximately 30 watts per square meter.

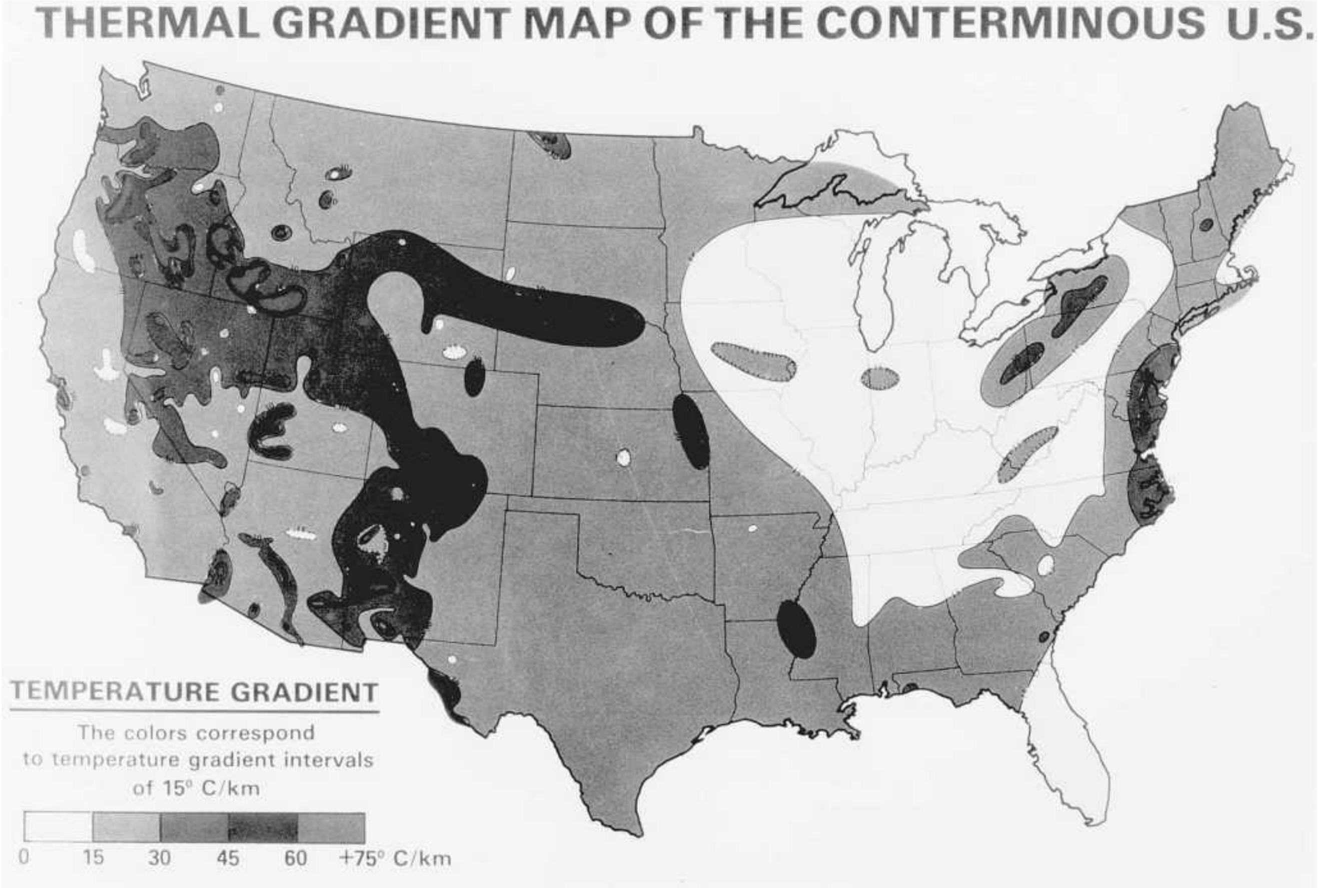
A related aspect of geothermal energy is the thermal gradient, which is the increase of temperature with depth below Earth's surface. The average thermal gradient is about 30°C (54°F) per kilometer, but it can be much higher at specific locations—for instance, in Iceland, where the increase is greater than 100°C (180°F) per kilometer in places.

TYPES OF GEOTHERMAL SOURCES

Geothermal sources are categorized into various types: hydrothermal reservoirs, geopressurized zones, hot dry rock, normal geothermal gradient, and magma.

Hydrothermal Reservoirs

Groundwater can seep down along faults in Earth's crust and become heated through contact with hot rocks below. Sometimes this hot water accumulates in an interconnected system of fractures and becomes a hydrothermal reservoir. The water might remain underground or might rise by convection through fractures to the surface, producing geysers and hot springs.



Map indicating regions of high thermal gradient where HDR geothermal techniques may be applied. (U.S. Department of Energy)

Geothermal power plant located in a lava field in Blue Lagoon, Iceland. (Corbis-Bettmann)

Hydrothermal reservoirs are the only geothermal sources that have been used for commercial energy production. Because of the high pressure deep below Earth's surface, the water in these sources can become heated well above the usual boiling temperature of 100°C (212°F). As the superheated water makes its way to the surface, either by convection or because a geothermal well has been drilled, some or all of the water will vaporize to become steam because of the lower pressures encountered. The most desirable geothermal sources have very high temperatures—above 300°C (572°F)—and all of the water vaporizes to produce dry steam (containing no liquid water), which can be used directly in steam-electric turbines.

Wet steam reservoirs are much more common than the simple dry type. Again, the field is full of very hot water, under such high pressure that it cannot boil. When a lower-pressure escape route is provided by drilling, some of the water suddenly evaporates (flashes) to steam, and it is a steam-water mixture that reaches the surface. The steam can be used to drive a turbine. The hot water also can be used to drive a second turbine in a binary cycle, described in the section "Electricity Generation" in this article.

Many geothermal reservoirs contain hot water at a temperature too low for electricity generation. However, the water can be used to heat buildings such as homes, greenhouses, and fish hatcheries. This heating can be either direct or through the use of heat pumps.

Geopressurized Zones

Geopressurized zones are regions where water from an ancient ocean or lake is trapped in place by impermeable layers of rock. The water is heated to temperatures between 100°C (212°F) and 200°C (392°C) by the normal flow of heat from Earth's core, and because of the overlying rock, the water is held under very high pressure as well. Thus energy is contained in the water because of both the temperature and the pressure, and can be used to generate electricity. Many geopressurized zones also contain additional energy in the form of methane from the decay of organic material that once lived in the water. The U.S. Geological Survey has estimated that about one third of the energy from geopressurized zones in the United States is available as methane.

Hot Dry Rock

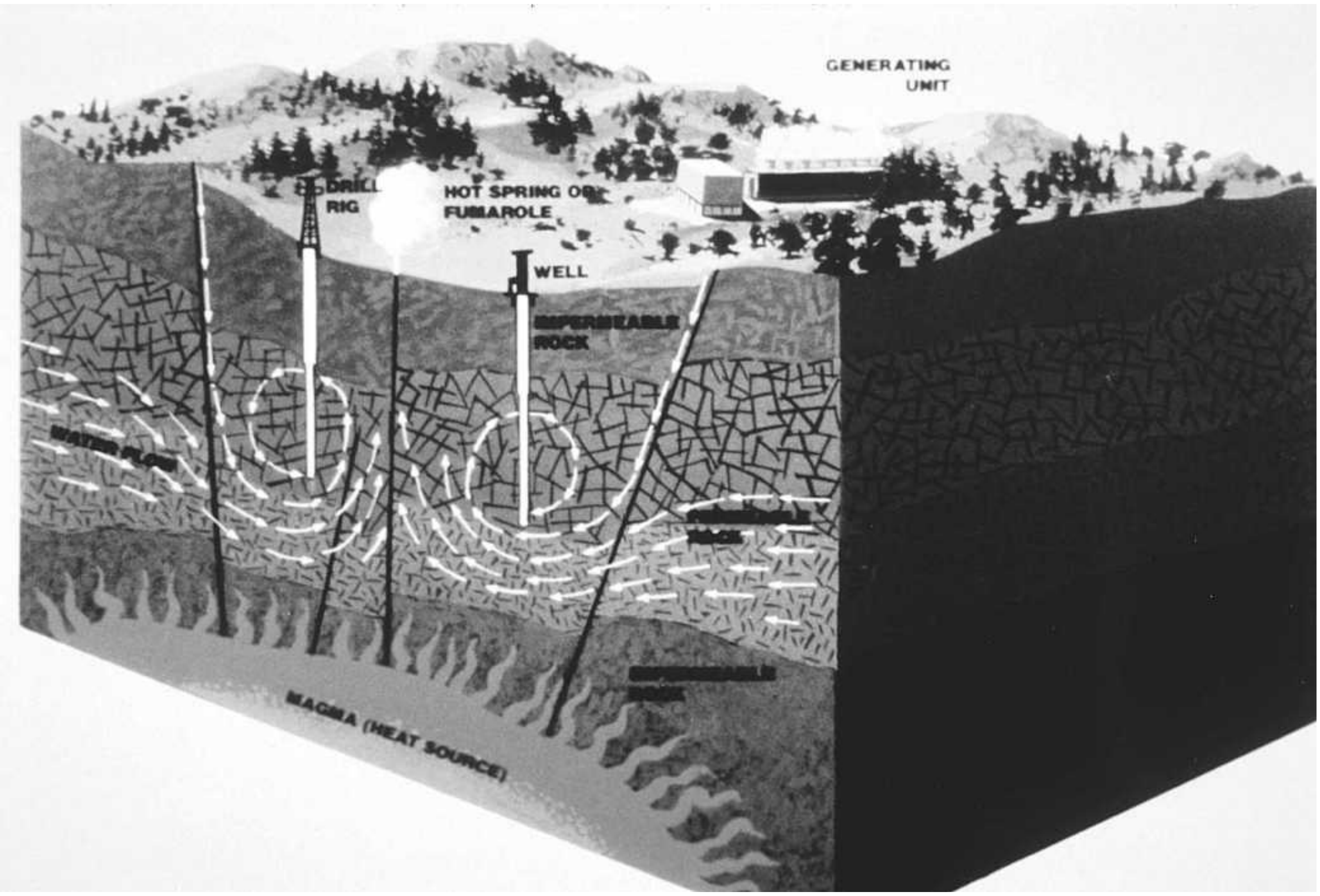
In many regions of the world, hot rocks lie near Earth's surface, but there is little surrounding water. Attempts have been made to fracture such rocks and then pump water into them to extract the thermal energy, but the technical difficulties in fracturing the rocks have proven to be much more troublesome than anticipated, and there has been a problem with water losses. Consequently, progress in extracting energy from hot, dry rocks has been slow. However, experiments are ongoing in the United States, Japan, and Europe because the amount of energy available from hot, dry rocks is much greater than that from hydrothermal resources. The U.S. Department of Energy estimates that the total energy available from high-quality hot, dry rock areas is about 6,000 times the annual U.S. energy use.

Normal Geothermal Gradient

In principle the normal geothermal gradient produces a useful temperature difference anywhere on the globe. If a hole is drilled to a depth of 6 kilometers (which is feasible), a temperature difference of about 180°C (324°F) is available, but no technology has been developed to take advantage of this resource. At this depth, water is unlikely, and the problems of extracting the energy are similar to the difficulties encountered with hot, dry rocks near the surface.

Magma

Magma is subterranean molten rock, and although the potential thermal resource represented by magma pools and volcanoes is extremely large, it also presents an immense technological challenge. The high temperatures produce obvious problems with melting and deformation of equipment. The most promising candidates for heat extraction from magma are young volcanic calderas (less than a few million years old) that have magma relatively close to the surface. There have been some preliminary test wells drilled at the Long Valley caldera, located about 400 kilometers north of Los Angeles. The hope is that heat can be extracted in this area by drilling a well down to the magma level and pumping water into the well to solidify the magma at the bottom of the well. Then more water could be pumped into the well, become heated by contact with the solidified magma, and then returned to the surface to generate steam for electricity production.



Cutaway drawing of the earth showing source of geothermal energy. (U.S. Department of Energy)

ELECTRICITY GENERATION

There are two general ways by which geothermal energy can be utilized: generation of electricity, and space heating. Production of electricity from a geothermal source was pioneered in an experimental program at the Larderello thermal field in Italy in 1904, and a 205-kilowatt generator began operation at this site in 1913. By 1998 the world's geothermal electrical generating capacity was about 8,240 megawatts, which represents only a small portion of the total electricity capacity of 2 million megawatts from all sources. However, the geothermal capacity has been growing steadily, and the 1998 amount represents a 40 percent increase above the 1990 value of 5,870 megawatts. The amount of electrical energy generated geothermally worldwide in 1998 was about 44 terawatt-hours, or 0.16 exajoule, representing approximately 0.4 percent of global electricity generation.

The United States is the world leader in geothermal electricity production, with about 2850 megawatts of capacity. As shown in Table 1, nine other countries each had more than 100 megawatts of electrical capacity in 1998, with this group being led by the Philippines, producing 1,850 megawatts of geothermal electrical power.

The world's most developed geothermal source is at the Geysers plant in California's Mayacamas Mountains, about ninety miles north of San Francisco. Electricity has been generated at this site since 1960, and as of 1999 the total installed generating capacity there was 1,224 megawatts. It has been demonstrated at the Geysers and at other geothermal electric plants that electricity from geothermal resources can be cost-competitive with other sources.

To produce electricity from a geothermal resource, wells are drilled into the reservoir, and as the hot, high-pressure water travels to Earth's surface, some of it vaporizes into steam as the pressure decreases. The hotter the original source, the greater the amount of dry steam produced. For dry-steam sources, the technology is basically the same as for electric plants that use the burning of fossil fuels to produce steam, except that the temperature and pressure of the geothermal steam are much lower. Dry-steam fields are being used in the United States, Italy, and Japan. In the Geysers geothermal area in California, the steam temperature is about 200°C (392°F) and the pressure about 700 kilopascals (7 atmospheres). Because geothermal steam is cooler than fossil-fuel steam, the efficiency of conversion of thermal energy to electricity is less than at a fossil-fuel plant; at the Geysers it is only 15 to 20 percent, compared to 40 percent for fossil fuels. The generating units are also smaller, ranging in size from 55 to 110 megawatts at the Geysers.

For geothermal reservoirs that are at lower temperatures, and hence produce less dry steam, working plants usually employ multiple-vaporization systems. The first vaporization (“flash”) is conducted under some pressure, and the remaining pressurized hot water from the ground, along with the hot residual water from the turbine, can be flashed again to lower pressure, providing steam for a second turbine. Flashed power production is used in many countries, including the United States, the Philippines, Mexico, Italy, Japan, and New Zealand.

If the temperature of the original hot water is too low for effective flashing, the water can still be used to generate electricity in what is referred to as binary cycle (or organic cycle) electricity generation. The hot water is pumped to the surface under pressure to prevent evaporation, which would decrease the temperature, and its heat is transferred to an organic fluid such as isobutane, isopentane, Freon, or hexane, all of which have a boiling temperature lower than that of water. The fluid is vaporized by the heat from the water and acts as the working fluid in a turbine. It has been estimated that geothermal reservoirs with temperatures suitable for binary cycle generation are about fifty times as abundant as sources that provide pure dry steam.

Some geothermal power plants use a combination of flash and binary cycles to increase the efficiency of electricity production. An initial flash creates steam that drives a turbine; then the binary cycle is run, using either the hot water remaining after the initial flash or the hot exhaust from the turbine.

Geopressurized zones, discussed earlier, also are areas where electricity could be generated geothermally, not only from the hot water but also from the associated methane. Three geopressurized well sites in the United States have been developed experimentally for electricity production, but the cost of the electricity generated is considerably higher than that from conventional energy sources.

World Total	**8240**
U.S.A.	2850
Philippines	1850
Italy	770
Mexico	740
Indonesia	590
Japan	530
New Zealand	350
Iceland	140
Costa Rica	120
El Salvador	110
Nicaragua	70
Kenya	40
China	30
Turkey	20
Portugal (Azores)	10
Russia	10

Table 1.
Geothermal Electricity Generation Capacity in 1998 (in megawatts)
Countries with fewer than 10 megawatts of capacity are not listed.

DIRECT GEOTHERMAL ENERGY USE

Geothermal sources at too low a temperature for electricity generation can be utilized for space heating, bathing, and other uses. Hot fluid from the reservoir is piped to the end-site location through insulated pipes to minimize loss of heat energy. Pumps are installed either in the geothermal well itself (if the temperature is low enough) or at the surface to drive the fluid through the piping system. To operate, these pumps require energy, usually supplied by electricity. The hot fluid can be used itself to provide the heating, or it can be pass through a heat exchanger where another working fluid is heated.

Direct geothermal energy is used for space heating of homes, greenhouses, livestock barns, and fish-farm ponds. As well, it is employed as a heat source in some industrial processes, such as paper production in New Zealand and drying diatomite in Iceland. Since the industrial applications usually require high-

	World	Japan	Iceland	China	U.S.A.
Space heating	33	21	77	17	10
Bathing, swimming, therapeutic use of baths	19	73	4	21	11
Greenhouses	14	2	4	7	5
Heat pumps (heating/cooling)	12	0	0	0	59
Fish farming	11	2	3	46	10
Industry	10	0	10	9	4
Snow melting	1	2	2	0	1

Table 2.
Types of Geothermal Direct Use Worldwide and in the Top Four Countries, 1995 (in %)

er temperatures than, say, space heating, it is advantageous to cascade the geothermal fluid, using it first at high temperature in industry and then afterward at lower temperature for another use.

Another way by which heat from the ground can be utilized is through the use of heat pumps. Pipes containing a fluid are buried in the ground, and heat can be extracted from the ground in the winter to heat a building, and dissipated in the ground in the summer to provide air conditioning. Essentially a heat pump acts like a refrigerator, extracting thermal energy from one area and moving it to another area. In the winter, the ground is cooled and the building is heated, and in the summer the heat pump runs in reverse, to cool the building and heat the ground.

An extensive summary of the various direct uses of geothermal energy has been compiled. Worldwide the most common use is for space heating (33%), followed by bathing, swimming, and therapeutic use of baths (19%). Table 2 shows the worldwide percentages for other uses, as well as percentages for the top four direct-use countries. As seen in this table, the specific uses of geothermal energy vary greatly from country to country.

The geothermal direct-use power capacity worldwide in 1997 was close to 10,000 megawatts, with China and the United States leading the way with 1910 megawatts each. Iceland and Japan also had more than 1,000 megawatts each, as shown in Table 3, and more than thirty countries in total were using geothermal heat. Total geothermal direct-use energy in 1997 was about 37 terawatt-hours, or 0.13 exajoule.

One of the countries that makes extensive use of direct geothermal heat is Iceland. This might seem surprising for a country with “ice” as part of its name, but Iceland lies in a region of continental-plate activity. Virtually all the homes and other buildings in Iceland are heated geothermally, and there is also a small geothermal electric plant. In Iceland’s capital city, Reykjavik, geothermally heated water has been used for space heating since 1930, and the cost of this heating is less than half the cost if oil were used.

Other important examples of geothermal heating are in France, both near Paris and in the Southwest. During oil-exploration drilling in the 1950s, hot water was discovered in the Paris region, but exploitation did not begin until the 1970s as a result of rapidly increasing oil prices. In France, the equivalent of 200,000 homes are being provided with space heating and water heating from geothermal sources. One interesting feature of the French geothermal sources is that they do not occur in regions of elevated thermal gradient.

ENVIRONMENTAL EFFECTS OF GEOTHERMAL ENERGY

The most important potential environmental impacts of geothermal energy are water and air pollution. At the largest geothermal plants, thousands of tons of hot water and/or steam are extracted per hour, and these fluids contain a variety of dissolved pollutants. The hot geothermal water dissolves salts from surrounding rocks, and this salt produces severe corrosion and scale deposits in equipment. To prevent contamination of surface water, the geothermal brine must either be returned to its source or discarded carefully in another area. In addition to salts, the geothermal waters sometimes contain high concentrations of toxic elements such as arsenic, boron, lead,

Countries	
World total	**9960**
China	1910
U.S.A.	1910
Iceland	1440
Japan	1160
Hungary	750

Table 3.
Geothermal Direct-Use Capacity in 1997 (megawatts)—Top Eight Countries and World Total

Located in Sonoma and Lake Counties in Calif., the geysers complex produces over 900,000 kW of electricity using steam from geothermal wells 7,000 to 10,000 feet below the surface. (U.S. Department of Energy)

and mercury. The 180-megawatt Wairakei geothermal electrical plant in New Zealand dumps arsenic and mercury into a neighbouring river at a rate four times as high as the rate from natural sources nearby.

Geothermal water often contains dissolved gases as well as salts and toxic elements. One of the gases that is often found in association with geothermal water and steam is hydrogen sulfide, which has an unpleasant odor—like rotten eggs—and is toxic in high concentrations. At the Geysers plant in California, attempts have been made to capture the hydrogen sulfide chemically, but this job has proven to be surprisingly difficult; a working hydrogen sulfide extractor is now in place, but it was expensive to develop and install, and its useful life under demanding operating conditions is questionable. Geothermal brines also are a source of the greenhouse gas carbon dioxide, which contributes to global warming. However, a typical geothermal electrical plant produces only about 5 percent of the carbon dioxide emitted by a fossil-fuel-fired plant generating the same amount of electricity. Many modern geothermal plants can capture the carbon dioxide (as well as the hydrogen sulfide) and reinject it into the geothermal source along with the used geothermal fluids. At these facilities, the carbon dioxide that escapes into the atmosphere is less than 0.1 percent of the emissions from a coal- or oil-fired plant of the same capacity.

Another potential problem, particularly if geothermal water is not returned to its source, is land subsidence. For example, there has been significant subsidence at the Wanaker field in New Zealand. Finally, an annoying difficulty with geothermal heat has been the noise produced by escaping steam and water. The shriek of the high-pressure fluids is intolerable, and is usually dissipated in towers in which the fluids are forced to swirl around and lose their kinetic energy to friction. However, the towers provide only partial relief, and the plants are still noisy.

GEOTHERMAL ENERGY—RENEWABLE OR NOT?

Most people tend to think of geothermal energy as being renewable, but in fact one of the major problems in choosing a geothermal energy site lies in estimating how long the energy can usefully be extracted. If heat is withdrawn from a geothermal source too rapidly for natural replenishment, then the temperature and pressure can drop so low that the source becomes unproductive. The Geysers plants have not been working at full capacity because the useful steam would be depleted too quickly. Since it is expensive to drill geothermal wells and construct power plants, a source should produce energy for at least thirty years to be an economically sound venture, and it is not an easy task to estimate the working lifetime beforehand.

THE FUTURE OF GEOTHERMAL ENERGY

The main advantage of geothermal energy is that it can be exploited easily and inexpensively in regions where it is abundantly available in hydrothermal reservoirs, whether it is used for electricity production or for direct-use heat. Geothermally produced electricity from dry-steam sources is very cheap, second only to hydroelectric power in cost. Electricity from liquid-dominated hydrothermal sources is cost-competitive with other types of electrical generation at only a few sites. It is unlikely that other types of geothermal sources, such as hot, dry rock and the normal geothermal gradient, will soon become economical. Hence, geothermal energy will provide only a small fraction of the world's energy in the foreseeable future.

An important feature of geothermal energy is that it has to be used locally, because steam or hot water cannot be piped great distances without excessive energy loss. Even if electricity is generated, losses also are incurred in its transmission over long distances. As a result, geothermal energy use has geographical limitations.

The future development of geothermal energy resources will depend on a number of factors, such as cost relative to other energy sources, environmental concerns, and government funding for energy replacements for fossil fuels. An important concern that many members of the public have about future developments is centered around whether resources such as hot springs and geysers should be exploited at all. Many of these sources—such as those in Yellowstone National Park in Wyoming—are unique natural phenomena that many people feel are important to protect for future generations.

Ernest L. McFarland

See also: Biofuels; Diesel Fuel; District Heating and Cooling; Economically Efficient Energy Choices; Electric Power, Generation of; Environmental Problems and Energy Use; Fossil Fuels; Heat and Heating; Heat Pumps; Heat Transfer; Hydroelectric Energy; Reserves and Resources; Seismic Energy; Thermal Energy; Thermal Energy, Historical Evolution of the Use of; Thermodynamics; Water Heating.

BIBLIOGRAPHY

Atomic Energy of Canada Limited. (1999). *Nuclear Sector Focus: A Summary of Energy, Electricity, and Nuclear Data*. Mississauga, Ont.: Author.

Fridleifsson, I. B. (1998). "Direct Use of Geothermal Energy Around the World." *Geo-Heat Center Quarterly Bulletin* 19(4):4–9.

Golub, R., and Brus, E. (1994). *The Almanac of Renewable Energy: The Complete Guide to Emerging Energy Technologies*. New York: Henry Holt.

Howes, R. (1991). "Geothermal Energy." In *The Energy Sourcebook*, ed. R. Howes and A. Fainberg. New York: American Institute of Physics.

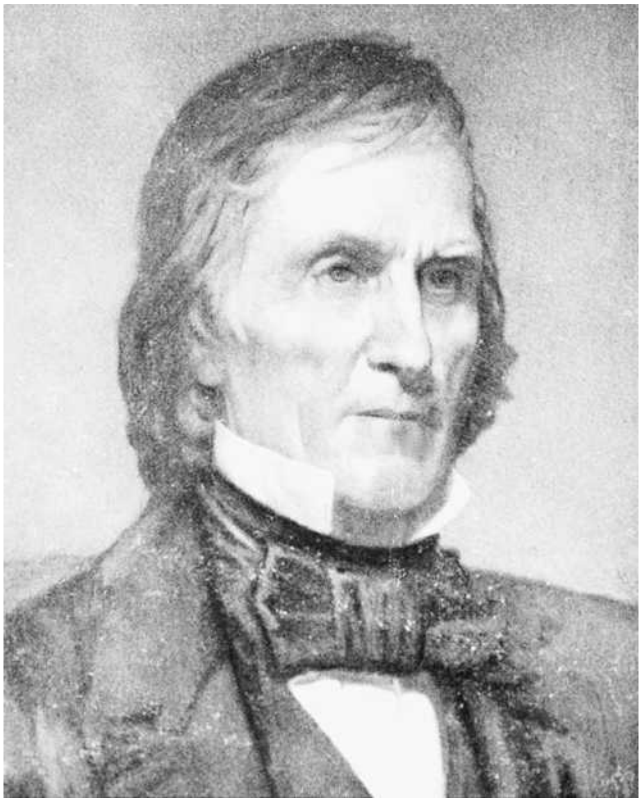
Lund, J. W. (1996). "Lectures on Direct Utilization of Geothermal Energy." United Nations University Geothermal Training Programme Report 1996-1. Reykjavik, Iceland.

McFarland, E. L.; Hunt, J. L.; and Campbell, J. L. (1997). *Energy, Physics and the Environment*, 2nd ed. Guelph, Ont.: University of Guelph.

GIBBS, JOSIAH WILLARD (1839–1903)

Gibbs came from an academic family in New Haven, Connecticut. His father was a noted philologist, a graduate of Yale and professor of sacred literature there from 1826 until his death in 1861. The younger Gibbs grew up in New Haven and graduated from Yale College, having won a number of prizes in both Latin and mathematics. He continued at Yale as a student of engineering in the new graduate school and, in 1863, received one of the first Ph.D. degrees granted in the United States. After serving as a tutor in Yale College for three years, giving elementary instruction in Latin and physics, Gibbs left New Haven for further study in Europe. He spent a year each at the universities of Paris, Berlin, and Heidelberg, attending lectures in mathematics and physics and reading widely in both fields. He was never a student of any of the luminaries whose lectures he attended (the list includes Liouville and Kronecker in mathematics, and Kirchhoff and Helmholtz in physics) but these European studies, rather than his earlier engineering education, provided the foundation for his subsequent scientific work. A qualification was a life-long fondness for geometrical reasoning, evident in Gibbs's scientific writings, but first developed in his dissertation.

Gibbs returned to New Haven in 1869. He never again left America and seldom left New Haven except for annual summer holidays in northern New England and a very occasional journey to lecture or attend a meeting. Gibbs never married and lived all his life in the house in which he had grown up, less than a block from the college buildings. In 1871, two years before he published his first scientific paper, Gibbs was appointed professor of mathematical physics at Yale. He held that position without salary



Josiah Willard Gibbs. (Library of Congress)

for nine years, living on inherited income. It was during this time that he wrote the memoirs on thermodynamics that, in most estimates, constitute his greatest contribution to science. Gibbs declined the offer of a paid appointment at Bowdoin College in 1873, but he was tempted to leave Yale in 1880 when he was invited to join the faculty of the newly-founded Johns Hopkins University in Baltimore. Only then did Yale provide Gibbs a salary, as tangible evidence of the high regard his colleagues had for him and of his importance to the University. Gibbs remained at Yale and continued to teach there until his death, after a brief illness, in 1903.

Gibbs worked on electromagnetism during the 1880s, concentrating on optics and particularly on James Clerk Maxwell's electromagnetic theory of light, and on statistical mechanics from at least the mid-1880s until his death. The latter research resulted in his seminal *Elementary Principles in Statistical Mechanics*, published in 1902. However, it seems more appropriate in this place to briefly describe Gibbs's memoirs on thermodynamics.

In his first work on thermodynamics in 1873, Gibbs immediately combined the differential forms of the first and second laws of thermodynamics for the reversible processes of a system to obtain a single "fundamental equation":

$$\mathrm{d}U = T\mathrm{d}S - p\mathrm{d}V,$$

an expression containing only the state variables of the system in which U, T, S, p and V are the internal energy, temperature, entropy, pressure, and volume, respectively. Noteworthy here is the assumption, which Gibbs made at the outset but which was not common at the time, that entropy is an essential thermodynamic concept. At the same time, the importance of energy was also emphasized. As Gibbs wrote at the beginning of his great memoir, "On the Equilibrium of Heterogeneous Substances," whose first installment appeared in 1876: "The comprehension of the laws which govern any material system is greatly facilitated by considering the energy and entropy of the system in the various states of which it is capable." The reason, as he then went on to explain, is that these properties allow one to understand the interactions of a system with its surroundings and its conditions of equilibrium.

As was usual with him, Gibbs sought to resolve the problem in general terms before proceeding to applications. Again beginning with the differential forms of the first two laws (which, in effect, define the state functions U and S), but this time for any process, whether reversible or irreversible, he combined the two expressions to yield the general condition of equilibrium for any virtual change:

$$\delta U - T\delta S - \delta W > 0$$

where W is the external work. If a system is isolated, so that $\delta W = 0$, this condition becomes:

$$(\delta U)_S \geq 0 \text{ or } (\delta S)_U \leq 0$$

for constant S and U, respectively. The second inequality, which Gibbs showed to be equivalent to the first, immediately indicates that thermodynamic equilibrium is a natural generalization of mechanical equilibrium, both being characterized by minimum energy under appropriate conditions.

The first and probably most significant application of this approach was to the problem of chemical equilibrium. In a heterogeneous system composed of several homogeneous phases, the basic equilibrium condition leads to the requirement that temperature, pressure, and the chemical potential (a new concept introduced by Gibbs) of each independent chemical component have the same values throughout the system. From these general conditions, Gibbs derived the phase rule:

$$\delta = n+2-r,$$

that cornerstone of physical chemistry, which specifies the number of independent variations δ in a system of r different coexistent phases having n independent chemical components.

Among many other valuable results in his memoir on heterogeneous equilibrium is a formulation of the Gibbs free energy, also called the Gibbs function, which is defined by the equation

$$G = H-TS,$$

where H is the enthalpy, that is, the sum of the internal energy of a body orsystem and the product of itspressure and volume. It is useful for specifying the conditions of chemical equilibrium at constant temperature and pressure, when G is a minimum. More generally, Gibbs's memoir greatly extended the domain covered by thermodynamics, including chemical, electric, surface, electromagnetic, and electrochemical phenomena into one integrated system.

Gibbs's thermodynamic writings were not as widely read—much less appreciated—as they deserved to be in the decade following their appearance. One reason is that they were published in the obscure *Transactions of the Connecticut Academy of Sciences.* Gibbs sent offprints of his memoirs to many scientists, but only Maxwell seems to have recognized their importance. That changed after Wilhelm Ostwald published a German translation in 1892. In the meantime, continental scientists such as Helmholtz and Planck independently developed Gibbs's methods and results, unaware of his prior work.

Another reason for lack of interest is the severity of Gibbs's style. Austere and logically demanding, it was a challenge even for mathematicians as distinguished as Poincaré. The same severity extended to Gibbs's lectures, which his few students found clear and well-organized, but not easy to understand, owing to their great generality and meticulous precision.

A third reason is that Gibbs made no effort to promote or popularize his results. He seems to have been a solitary, self-contained, and self-sufficient thinker, confident in his ability, who worked at his own unhurried pace, neither needing nor wanting feedback from others. An attitude of detachment from the work of his students plus his own solitary habit of work is undoubtedly responsible for the fact that Gibbs founded no "school" or group of students to develop his ideas and exploit his discoveries.

Robert J. Deltete

BIBLIOGRAPHY

Deltete, R. J. (1996). "Josiah Willard Gibbs and Wilhelm Ostwald: A Contrast in Scientifc Style." *Journal of Chemical Education* 73(4):289–295.

Klein, M. J. (1972). "Gibbs, Josiah Willard," in *Dictionary of Scientific Biography*, Vol. 5, ed. C. G. Gillispie. New York: Scribner.

Klein, M. J. (1989). "The Physics of J. Willard Gibbs in His Time." In *Proceedings of the Gibbs Symposium. Yale University, May 15–17, 1989*, ed. D. G. Caldi and G. D. Mostow. New York: American Mathematical Society.

GLASS TECHNOLOGY

See: Windows

GLOBAL WARMING

See: Climatic Effects

GOVERNMENT AGENCIES

Energy is the economic lifeblood of all economies. It is an essential gear for the economies of the developed world, and of ever growing importance to developing nations. It is of concern to the largest of nations as well as the smallest. And it is as great a concern of local governments as national ones.

Government action or inaction at all levels in energy policy has varied tremendously. Prior to 1930, involvement in energy issues at all levels of government was minimal in the United States because of relatively moderate demand for energy consuming technology, and an abundant and relatively cheap supply of fossil fuels. The government's approach to energy changed with the Depression and the New Deal. The United States began to subsidize energy by building hydroelectric stations (the Tennessee Valley Authority), supporting rural electrification, and subsidizing of nuclear research. Much of the developing world, as well as the Soviet Union, began subsidizing energy early in the twentieth century as a means to buy votes and strengthen political support. Because of World War II, the strategic role of energy to economic growth and national security became much more apparent. Almost all governments began to take a more active role in energy markets as world energy consumption accelerated, as spending was boosted on all aspects of the energy puzzle—exploration, production, distribution, consumption, and the energy statistics to track all of the above.

Traditionally, the developed world has paid the greatest attention to petroleum (reserves, resources, security, price), and devoted the most resources to nuclear energy research (both fission and fusion), a source that proponents in the 1950s believed would someday turn out to be "too cheap to meter." Planning, funding, and development of renewables and other energy resources has fluctuated much more wildly through the years, usually inversely to the real and perceived availability of oil. During and following an oil crisis, where there is great supply uncertainty and prices skyrocket, planning, funding and development grow, and correspondingly decrease when oil supplies are more secure and stable.

WORLD AGENCIES

The pooling of resources to meet shared objectives have been the factor most responsible for the establishment and growth of international agencies concerned with energy issues. The three most important issues bringing about cooperation have been the need to coordinate production, the pooling of resources for research and development of energy technology, and the coordinating of activities to secure supply.

Organization of Petroleum Exporting Countries (OPEC)

The OPEC cartel was founded by Iraq, Iran, Saudia Arabia, Kuwait, and Venezuela in September 1960 as a way to coordinate petroleum production and pricing among member countries. It was not until the 1970s that the cartel tried to become an effective monopoly.

As of 2000, membership has expanded to thirteen, accounting for over 60 percent of all production. The reserves controlled by Member Countries were much higher in the 1970s, and consequently so was the cartel's monopoly power in controlling prices. But because of exploration discoveries, advances in technology (enhanced recovery of existing wells, improved offshore equipment), and the lack of cooperation from non-OPEC producers, the power of the cartel to control oil production and prices has steadily diminished. During the 1970s, an announcement of an OPEC meeting would be the major news story of the day, but by the 1990s, low oil prices and a secure supply caused the media to only superficially cover OPEC meetings. As most economists predicted, the ability of the cartel to control the price of petroleum did not last long. There will always be an incentive for member states—especially the smaller producers—to "cheat" on their quotas, and the free market will always react swiftly to higher prices by pumping up other sources of oil production and securing other energy resources.

As OPEC's share of the world oil supply market continued to fall in the 1990s, they began taking steps to better coordinate production with non-OPEC producers such as Mexico and other members of the Independent Petroleum Exporting Countries (IPEC). By exchanging information, and undertaking joint studies of issues of common interest, the hope was to stabilize prices and improve the economic outlook for all oil producers. This collaboration between OPEC and major non-OPEC producers helped raise oil prices to over $27 a barrel in 1999 from a low of less than $13 in 1998.

International Energy Agency (IEA)

In response to the Arab embargo, and the attempt of OPEC to dictate the supply and price of petroleum, the Organization for Economic Cooperation and Development (OECD) established the International Energy Agency (IEA) in November 1974. IEA membership included all of Western Europe, Canada, the

United States, Australia, New Zealand, and Japan. Besides compiling useful consumption and production statistics, its mandate was for cooperation and coordination of activities to secure an oil supply in times of supply disruption. IEA's first action to insulate member countries from the effects of supply disruptions was the Emergency Sharing System (ESS). This system was to be put into effect only in cases of serious disruptions, an actual or anticipated loss of 7 percent of expected supply. The System consisted of three parts: a building up of supplies, a reduction of consumption during periods of short supply, and a complex sharing system that would attempt to distribute the loss equitably. During the 1979–1981 oil supply disruption, the system was never really tested because the loss of supplies never reached a level to trigger the ESS. Nevertheless, because of the economic hardship felt by many nations during the 1979–1981 disruption, in 1984 the Coordinated Emergency Response Measures (CERM) was adopted. The CERM were intended as a means to reach rapid agreement on oil stockpile drawdown and demand restraint during oil supply disruptions of less than 7 percent. Because of ESS and CERM, and the agreement of other OPEC Member States to boost production, the supply disruption caused by the Iraqi occupation of Kuwait, and the subsequent United Nations embargo of all oil exports from Iraq and Kuwait, did not have such a large effect.

Twenty-five years after its establishment, the focus of the IEA has changed and expanded significantly. Whereas in 1974, the IEA looked at coal and nuclear energy as the two most promising alternatives to oil, the call for more environmentally acceptable energy sources has pushed to the forefront the development of renewable energy, other nonfossil fuel resources, and the more clean and efficient use of fossil fuels. To avoid duplication of effort and better research and development results, the IEA coordinates cooperation among members to more efficiently use resources, equipment, and research personnel.

United Nations (UN)

The United Nations officially came into existence at the end of World War II on October 24, 1945 to help stabilize international relations and better secure peace. Through the years, the UN has greatly expanded its mission to include other issues such as development and protecting the environment. Because energy is such a key piece of the development puzzle, in 1947 the UN established a statistics division that began tracking energy consumption and production throughout the world in 1992.

The divisions established to take an active role in energy and development issues were the United Nations Development Program (UNDP) and the Commission on Sustainable Development in 1993. Funded by voluntary contributions from member states, with funding directed toward countries with annual per-capita GNP of $750 or less, UNDP focuses on preventing unsustainable production and consumption patterns, and ways of curbing pollution and slowing the rate of resource degradation—economic growth with environmental protection and conservation. The burden of environmental protection is far greater for poorer nations because it diverts resources away from more pressing problems such as poverty, low levels of social development, inadequate energy infrastructure, and a lack of capital for infrastructure. Despite emissions of industrial countries dropping from 1980 to 2000, many emissions, most notably toxic substances, greenhouse gases, and waste volumes are continuing to increase in the developing world.

Global warming is another energy-related problem that the UN has attempted to address through its Intergovernmental Panel of Climate Change (IPCC). The IPCC was very successful in negotiating an agreement to solve the problem of chlorofluorocarbons (January 1987 Montreal Protocol on Substances That Deplete the Ozone Layer), yet the task of getting member countries to agree on carbon emission reductions to combat the perceived threat of global warming has proven to be a much more daunting task. The science is more uncertain, there are no easy solutions (the combustion that provides the majority of the world's energy needs always entails carbon emissions), and to achieve the cuts that climate modelers feel would be necessary to fully address the problem will require a complete restructuring of current civilization. Nevertheless, based on the work of the IPCC, the United Nations Framework Convention on Climate Change adopted the Kyoto Protocol on December 11, 1997, as a first step in the process. The Kyoto Protocol calls for most of the developed nations to reduce carbon emissions by 10 percent of 1990 levels by the year 2010. Most nations ratified the treaty despite no reduction commitments from the developing nations. The poorer nations strongly objected to universal emission reductions. They countered that energy drives development, and the exploitation of cheap and abundant energy, often at the expense of the poorer nations, is how the richer nations achieved their higher standard of living. Burdensome environmental regulations of energy would severely retard development; thus, the poorer nations feel it is only equitable that the richer nations, who created the majority of carbon emissions in the last 150 years, first make the effort to curtail emissions.

The Kyoto negotiations show the tremendous friction energy issues can cause between the richest and poorest countries in the world. Addressing future energy-related conflicts—whether it be wars over oil fields or disputes over carbon emission quotas—will remain a major role of the UN.

The International Atomic Energy Agency (IAEA)

To address the technology, waste, safety and security issues concerning nuclear energy, the UN established the International Atomic Energy Agency (IAEA) in 1957, a few years after U.S. President Dwight D. Eisenhower's famous "Atoms for Peace" speech before the United Nations General Assembly.

As the political, economic, and technological realities of the world evolved, so did IAEA. The scope of products, services, and programs of the IAEA has expanded well beyond its original function as the world's central intergovernmental forum for scientific and technical nuclear cooperation, and the watchdog for civilian nuclear power programs. In the 1990s, the IAEA helped countries carry out comparative cost effective assessments of how to expand electric power generation capacity such as the potential roles of renewable energy, and the different options and costs for reducing atmospheric and greenhouse gas emissions. They also conduct research and provide outreach through the Agency's laboratories. The IAEA tries to go beyond technology transfer and energy capacity building to help provide solutions to problems of sustainable human development.

International Institute for Applied Systems Analysis

Founded in 1972 by the United States, the Soviet Union, and ten other countries, the International Institute for Applied Systems Analysis is a research organization located in Laxenburg, Austria that conducts scientific studies in many energy-related areas of global consequences such as transboundary air pollution, sustainable forest resources, climate change, and environmentally compatable energy strategies.

World Bank

Established in 1946, with a subscribed capital fund of $7.67 million, the primary mission of the World Bank is to combat poverty by securing low cost funding for sustainable development. The largest shareholder is the U.S., followed by the United Kingdom, Japan, Germany, and France.

Although the World Bank provides loans for a variety of purposes, energy-related infrastructure receives over 7 percent, which primarily goes toward electric power infrastructure (hydroelectric, fossil fuel, nuclear) but also oil and natural gas exploration, production and distribution. Other energy-related areas making up a considerable share of the World Bank's loan portfolio are the transportation and agricultural sector. Loans to improve the energy and logistic efficiency of the transportation sector account for 11 percent of funding, followed by 10 percent going toward agriculture to provide assistance in expanding the amount of food energy produced.

By 1999 the World Bank loan portfolio for the energy and mining sectors had grown to $4.1 billion. Its strategy for energy development is to reform and restructure markets to attract private investment to build energy infrastructure, and expand energy access to the rural and low-income populations, and promote an environmentally responsible energy supply and use. One priority is to get more of the 2 billion plus poor people access to modern energy for cooking and lighting. Without modern energy, the world's poor rely on wood, crop residues, and other biofuels that result in environmentally damaging deforestation. Deforestation is of interest to all nations because it is not only a habitat destroyer, but also contributes to the increasing concentrations of carbon dioxide, a gas long suspected of creating global warming.

The World Bank grants financing for fossil fuel electricity generation, yet finances only facilities that have advanced emission control equipment. And although the World Bank has never financed a nuclear power plant, a zero carbon emitter, it is very active in evaluating hydropower projects, helping to establish the World Commission on Large Dams.

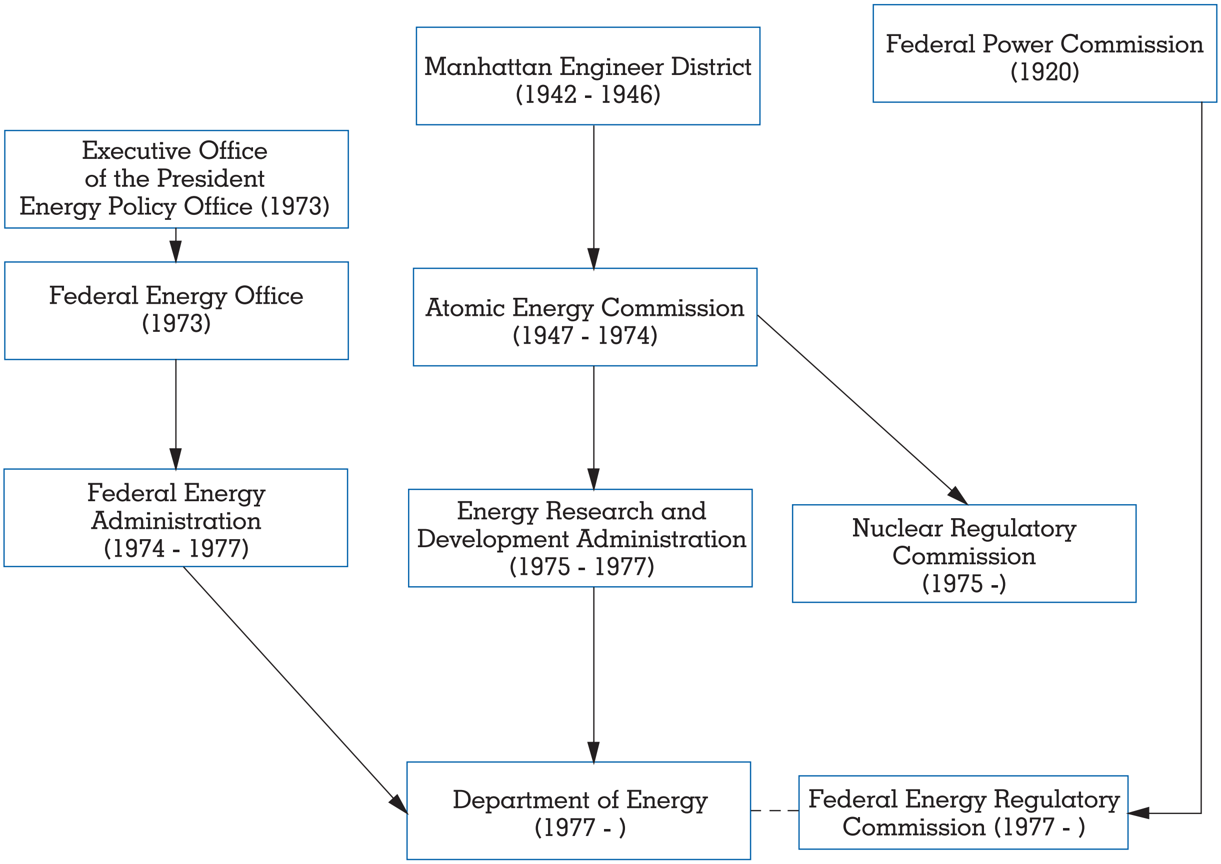


Figure 1.
Precursors and agencies related to the U.S. Department of Energy.

The World Bank also looks for energy efficiency improvement opportunities by pushing for the elimination of fossil fuel subsidies (estimated at over $200 billion a year in the developing world), improving demand-side efficiency, and making the energy supply system more cost conscious.

UNITED STATES AGENCIES

The focus of U.S. energy policy is with the Department of Energy, but because energy issues touch almost every sector of the economy, the Department of Commerce, the Department of Transportation, Department of Agriculture, Department of Defense, and the Environmental Protection Agency also devote a considerable amount of resources toward energy issues.

The Department of Energy (DOE)

The Department of Energy was established on June 3, 1977, unifying offices, laboratories, and staffs from other federal agencies (see Figure 1). Besides replacing and taking on all the responsibilities of the Federal Energy Administration, the Energy Research and Development Administration, and the Atomic Energy Commission, limited functions were transferred from the Departments of Agriculture, Commerce, Housing and Urban Development, and Transportation. The newly formed agency had around 20,000 employees and a budget of $10.4 billion.

Prior to the energy crisis of 1973, the federal government took a limited role in formulating a national energy policy. Markets operated freely, long-range planning was left up to the private sector, and oversight controlled by state, local, and regional authorities. Americans felt comfortable with private industry controlling production, distribution, marketing, and pricing. But in the case of natural monopolies, such as in the interstate transportation of natural gas and electricity, it was generally acknowledged that the Federal Power Commission (established in 1920) was necessary to ensure fair prices.

The regulatory agency that the DOE established for oversight of natural monopolies was the Federal Energy Regulatory Commission (FERC). The five-member commission was set up to control the regulation and licensing of hydroelectric power facilities, the regulation of the transmission and sale of electric power, the transportation and sale of natural gas, and the operation of natural gas and oil pipelines.

After the Oil Embargo, the Economic Regulatory Administration within DOE administered oil pricing, energy import programs, the importing and exporting of natural gas, programs to curtail natural gas consumption, and to supervise the conversion of electric power production from natural gas and oil to coal. These command and control programs continued through the Carter Administration, but ended early in the Reagan Administration, a strong free market advocate.

To provide long-term energy trends to the Department, President, Congress and the public, the DOE set up a statistics division called the Energy Information Administration. The federal government had been gathering and publishing energy statistics since 1949, but under the centralized control of the Energy Information Administration, the scope of data gathering expanded and continued to expand through the 1990s to encompass related areas of interest like carbon emissions.

During the Carter Administration, funding increased to pay for the greater focus on conserving energy, greater oil production including increasing the production of oil, natural gas and synthetic fuel from coal and shale oil reserves, and speeding the development and implementation of solar power. The Carter Administration took a very activist approach to energy policy, and believed that the United States could be getting 20 percent of its energy from the sun or other renewable energy sources by the year 2000. Toward this end, the Carter Administration pushed through Congress very generous subsidies for residential and commercial solar and wind installations.

The Reagan Administration (1981–1989) came in with a completely different vision of the federal role in the energy field. Reagan wanted to abolish the DOE, but with that being politically impossible, he set about restructuring it, letting private industry and the free marketplace set energy priorities. By ending government regulations and price controls, which were detrimental to domestic oil and natural gas production, he felt a free marketplace would prevent or limit the impact of future energy crises. And although Reagan reduced and eliminated subsidies for energy conservation and energy technologies like solar—preferring private capital to demonstrate commercial viability of technology—he continued to support long-term energy research and development, such as fusion, that private industry felt too risky to undertake. Despite these cutbacks, the DOE budget grew during the Reagan era mainly because of the emphasis on defense-related spending and the Strategic Defense Initiative designed to stop incoming nuclear warhead missiles.

The Bush Administration (1989–1993) had a similar free marketplace philosophy as Reagan, but faced the daunting task of having to start directing billions toward cleaning up after forty years of neglect at the contaminated weapons complex, particularly the federal facilities at Savannah River South Carolina, Hanford Washington, and Rocky Flats Colorado. The cleanup plan was fourfold: characterize and prioritize all waste cleanups at departmental sites, confine and correct immediate problems, establish long-term cleanup plans, and mandate compliance with all applicable laws. By the end of the Bush Administration, environmental management of defense-related nuclear waste consumed nearly a third of the budget, which reflected the dauntingly difficult and expensive nature of the cleanup.

Perhaps the most noteworthy energy-related accomplishment of the Bush Administration was the handling of the Persian Gulf Crisis of 1970 and the January 15, 1991 Operation Desert Storm—a military action that many categorized as an energy war. Because Iraqi and Kuwaiti production constituted around 4.3 million barrels per day, or 9 percent of the world total, the role of the DOE during the crisis was to reassure the public and press about oil issues, improve energy coordination with other countries primarily through the IEA, and promote energy conservation and increase energy production. Rapidly rising spot market prices for crude oil as well as gasoline prices did occur, but because of these efforts, along with the short duration of the crisis, the impact on world economies was minimized.

The Clinton Administration (1993–2000) that followed was far more inclined to embrace environmental activism than Reagan or Bush, and far more likely to propose command and control solutions to energy and environmental problems. However, the Clinton Administration also realized the need to allow markets to work, to do otherwise would result in some of the disastrous consequences of intervention policies used in the 1970s.

To pay for boosted spending for conservation grants, energy research and development, and spending on defense waste management, the Administration dramatically cut defense research and development. Many energy policy decisions of the Administration were directly linked to the quality and health of the environment. And unlike the Reagan and Bush Administrations that remained skeptical of global warming, the Clinton Administration made action on the global warming threat a priority. On the first Earth Day of the Administration, Clinton promised to stabilize greenhouse gas emissions at 1990 levels by the year 2000, an extremely ambitious goal. The method proposed was the Btu tax, a 25.7 cents per million Btus tax (4 %) on all forms of energy except solar, geothermal, and wind, and an additional 34.2 cents per million Btus tax for gasoline (4 cents a gallon) and other refined petroleum products. Although projected to reduce emissions by 25 million metric tons, the proposal was rejected by Congress primarily because it was widely viewed as a tax increase to finance increased social program spending. In December 1997, the Clinton Administration signed the Kyoto Protocol calling for dramatic carbon emission reductions, but never sent it to the Senate for ratification because of certain defeat. Almost all Senators felt there should be future carbon emission targets for the developing world, and the inequity of the cuts would give an unfair competitive advantage to the developing economies of the world, such as China, that did not agree to emission reductions.

By the end of the Clinton Administration in 2000, the budget for the agency had nearly doubled, and the number of employees exceeded 170,000, a figure that includes all those employed at DOE's national laboratories, cleanup sites, and other facilities. Despite organizational reshuffling and shifts in funding, much of the DOE mission—energy security, developing new energy sources, and collecting statistics about energy production and consumption—remains the same. The 1997 Budget allocated 38 percent toward nuclear security, followed by nuclear cleanup at 36 percent, basic science 15 percent, and energy subsidies 11 percent. Many critics feel the $6 billion spent annually for nuclear cleanup of DOE facilities is extremely wasteful. Instead of trying to return these sites to pristine conditions, critic feel DOE should renegotiate the cleanup plan with the EPA, and state and local authorities so that the emphasis can be shifted toward containment and neutralization of waste.

As the DOE moves into the twenty-first century, the Department faces significant challenges in defining a mission to warrant its generous funding. As long as fossil fuel supplies remain cheap and abundant, and the public remains skeptical of global warming, there will remain a desire to cut funding.

The Environmental Protection Agency (EPA)

To consolidate the environmental protection activities of the federal government, the Environmental Protection Agency was formed on December 2, 1970 from three federal Departments, three Administrations, three Bureaus, and several other offices. The mission of EPA was to collect data on pollution, conduct research on the adverse effects of pollution, and through grants and technical assistance, develop methods for controlling it.

Much of the EPA's regulatory activity centers around the energy business, and in many ways, its actions have a greater impact on altering production and consumption patterns in the energy sector than the DOE's actions. Oil drilling, production, transportation (pipelines and tankers), and refining are all heavily regulated segments of the energy sector. Refining, in particular, is subjected to periodic EPA audits for hazardous waste by-products and airborne emissions released during the refining process. The stringency of EPA refinery regulations, and the uncertainty about future regulations, partly explains why the industry has not built any new major refineries in the United States since the mid-1970s.

Power plant and auto emissions are two other energy areas touched by EPA regulations. These are regulated through the Clean Air Act of 1970, and the amendments in 1977 and 1990 that set guidelines for acceptable emission levels. The Clean Air Act has been widely praised for improving air quality, significantly lowering emissions of sulfur dioxide, nitrous oxide, and other particulate matter. However, most economists feel that the benefits from the EPA administered safety investments (including nonenergy aspects) are far less cost effective than the safety investments of other regulatory agencies (see Table 1).

Regulatory Agency	*Median Cost/ Life-Year Saved*
Federal Aviation Administration	$23,000
Consumer Product Safety Commission	$68,000
National Highway Traffic Safety Administration	$78,000
Occupational Safety and Health Administration	$88,000
Environmental Protection Agency	$7,629,000

Table 1.
Median Value of Cost/Life-Year Saved For Five Regulatory Agencies
SOURCE: Tengs, 1995

The 1990 amendments, which gave the Agency broad new power to revise Clean Air standards if EPA felt the results of studies warranted changes, has been particularly troublesome for the energy sector. Much of the EPA's air standard actions taken during the 1990s came under intense criticism for not being warranted by science. In particular, the EPA severely toughened new Clean Air particulate matter standards for power plant emissions in 1996, claiming it would save 15,000 lives a year, and reduce hospital admissions and respiratory illness. This was followed, in 1999, by a new lower sulfur content gasoline standard for refiners (effective in 2004) that will add 2 to 6 cents to the price of a gallon of gasoline. The EPA claims an additional 2,400 deaths will be prevented every year from this new standard. It was impossible to dispute the merits of either new standard because the EPA prevented public review of the data upon which these regulations are based (the "Pope" study) as required by the Freedom of Information Act (taxpayer-funded scientific data used to support federal regulations must be made available). Without access to the data, there can be no confirmation that the new standards will save any lives.

On the nuclear energy front, the EPA was instrumental in the late 1980s for bringing DOE nuclear facilities into compliance with the Nuclear Waste Policy Act of 1982 and 1987 (high-level radioactive wastes), the Low-Level Radioactive Waste Policy Act of 1980 and 1985, the Uranium Mill Tailings Radiation Control Act of 1978, and the Superfund statute. The most egregious sites needing cleanup are the nuclear weapons production sites located at Savannah River South Carolina, Rocky Flats Colorado, and Hanford Washington, that are all going to take decades to cleanup and cost billions of dollars. The EPA action at these DOE sites was touted as a sign that the federal government can police itself, and that the environmental laws that apply in the private sector apply equally to the public sector.

To combat the dangers of global warming by reducing carbon emissions, the Clinton Administration formed the Climate Protection Division (CPD), formerly the Atmospheric Pollution Prevention Division. This Division has been directed to find nonregulatory ways to reduce greenhouse gases through energy-efficiency improvements in all sectors of the economy. In collaboration with DOE, the EPA established the Energy Star Labeling Program as a way to develop voluntary energy-efficiency specifications for products such as office equipment, heating and cooling equipment, residential appliances, and computers. This Program allows manufacturers to prominently place an Energy Star label on qualifying products. The hope is for consumers to learn to recognize the label as the symbol for energy-efficiency, and become accustomed to buying only energy efficient products with the Energy Star label.

The Department of Commerce (DOC)

The commercial and industrial sectors account for more than fifty percent of all energy consumption in the United States. The Department of Commerce (DOC), whose mission is to improve the overall competitiveness of the commercial and industrial sector, is very concerned with reducing consumption through conservation and energy efficiency, and in improving the competitiveness of the energy businesses on the production issues.

On the energy use side, the DOC helps promote the DOE and the EPA energy efficiency programs (Energy Star) for the industrial and commercial sectors of the economy. To aid the energy businesses on the production side, the International Trade Administration and the Office of Energy, Infrastructure and Machinery assist all the energy fuel industries in improving their market competiveness and ability to participate in international trade. One of the more aggressive actions of the Department under the Clinton Administration was to sponsor trade missions to promote the export of U.S. electric power production technology.

Another major function of the DOC is the management of energy-related research through the National Institute of Standards and Technologies Laboratories in Gaithersburg, Maryland and Boulder, Colorado and the research laboratories of the National Oceanic and Atmospheric Administration.

The National Institute of Standards and Technology (NIST), formerly the National Bureau of Standards (NBS), was established by Congress in 1901. Its mission is to assist industry in the development of technology needed to improve product quality, to modernize manufacturing processes, to ensureproduct reliability, and to facilitate rapid commercialization of products based on new scientific discoveries. For example, it provides measurements that support the nations' standards for lighting and electric power usage. In 1988, NIST added three major new programs to its measurement andstandards laboratories: The Advanced Technology Program (ATP) which in partnership with industry accelerates innovative, enabling technologies with strong potential to deliver large payoff for the nation; the Manufacturing Extension Partnership (MEP), a network of local centers offering technical and business assistance to smaller manufacturers; and the Malcomb Baldridge National Quality Award that recognizes business performance excellence. Programs the ATP is funding in the energy arena include the development of rapid thermal processing to produce low-cost solar cells with Solarex (a business unit of Amoco/Enron Solar), and ultrathin silicon ribbon for high-efficiency solar cells with Evergreen Solar, Inc.

The Office of Oceanic and Atmospheric Research (OAR) is the division of NOAA that conducts and directs oceanic and atmospheric research. Since carbon dioxide is a greenhouse gas and fossil fuels are the leading generator of carbon dioxide, the work of the twelve Environmental Research Laboratories and eleven Joint Institutes of OAR to describe, monitor, and assess climate trends are of great interest to all parties interested in the affect of energy use on climate change.

The Department of the Interior

The Department of the Interior is the home of the Bureau of Land Management, the Minerals Management Service, the Office of Surface Mining Reclamation and Enforcement, and the U.S Geological Survey (USGS).

The Bureau of Land Management is responsible for the leasing of land for coal, oil, and natural gas exploration and production, and the Minerals Management Service does likewise for offshore leasing. Both divisions rely on resource evaluation information provided by the USGS in negotiating leases. The main responsibilities are the inspection and enforcement of leases, and the collection of royalty payments and other revenues due the Federal Government from the leasing and extraction of energy resources. Besides providing the statistics for the other Department of the Interior divisions, the USGS compiles databases and geologic maps of energy reserves and resources.

The mining industry has taken great steps in limiting environmental problems in developing energy resources, yet much of the mining industry prior to the 1980s abandoned sites leaving huge environmental problems. The Office of Surface Mining Reclamation and Enforcement was established to formulate policy and working plans for the Abandoned Mine Land reclamation programs. This was necessary because the Department cannot sue many of the offending mining companies for cleanup costs because most no longer exist.

The Department of Agriculture (USDA)

Energy issues are a major, yet indirect, focus of the Department of Agriculture (USDA). Energy is involved in every step of agriculture: it takes energy to produce the inputs of nitrogen, phosphate, seeds and pest/weed control, energy to run the machinery to plant, maintain, harvest and process crops, and energy to transport crops by truck, train, barge and container ships. And the even the end result, food itself, is essentially energy. Food is as essential to humans as gasoline is to the automobile.

The USDA has an inherent conflict of interest: promote eating more dairy products and meats to relieve agriculture surpluses, while promote good eating with the ubiquitous USDA Food Pyramid. Out of a budget of over $65 billion in 2000, about two-thirds of the USDA budget goes toward nutritional programs and social programs for the poor, such as food stamps, to ensure that all Americans can afford an ample supply of food energy. For the middle and upper class of society, food prices are artificially low because of the subsidies to producers. An overabundance of food, that is more affordable to more of the population than ever before, has turned the United States into an equal opportunity obesity society. Over 20 percent of the population is obese (over 50 percent overweight), and obesity can be found in large percentages at all income levels. This is unlike most of the world where obesity is far less prevalent and mostly found among the affluent.

Another function of the USDA is the promotion of biofuels like ethanol as an important market for the nation's farm products. Ethanol is an alcohol fuel produced from corn and is blended with gasoline to enhance octane and reduce automobile emissions of pollutants. From 1980 through 2000, ethanol producers have received a subsidy of nearly $10 billion. To prop up the price of corn and help the ethanol industry, in 1986 the USDA Office of Energy started to give away free corn for all ethanol producers, which was extended to even the very largest and most profitable ethanol producers like Archer Daniels Midland Corporation. These huge subsidies are very controversial. Proponents claim ethanol as a fuel provides an additional market for corn farmers, and also claim the fuel is better for the environment and helps reduce imports of foreign oil. Critics counter that the energy content of ethanol is one-third that of gasoline. Moreover, the DOE and Congressional Research Service found it was not better for the environment, and would retail at a price much higher than gasoline if not for the heavy subsidies. And because it takes considerable energy to convert corn to ethanol, there can be a net loss in energy in producing ethanol (a gallon of ethanol contains around 76,000 British Thermal Units (Btus) of energy and estimates of the energy to produce that gallon range from around 60,000 to 90,000 Btus).

Another biofuel of importance is wood. The Forest Service, which is part of USDA, administers national forest lands for the sale of wood for wood fuel. Besides determining the quantity of wood fuel to bring to market by collecting and analyzing statistics on woody biomass supply and use, the Forest Service sponsors forest biomass energy-related research in conjunction with federal and state agencies, as well as universities.

The Department of Transportation (DOT)

Because every means of transportation requires energy for propulsion, how energy is used in transportation is something that is carefully tracked by the Office of Transportation Policy Development within the Department of Transportation (DOT). The transportation sector felt the greatest impact from the oil supply disruptions in the 1970s because it was, and continues to be, the sector most dependent on oil. It is also the sector with the least flexibility to switch fuels. (see also Consumption)

National Aeronautic and Space Administration (NASA)

The National Aeronautic and Space Administration is involved in every aspect of atmospheric and space science. Because of the special energy requirements of spacecraft and satellites, NASA has been the proving grounds for many emerging energy technologies such as fuel cells and photovoltaics. In the area of transportation, the Jet Propulsion Laboratory in Pasadena California is a leading center for bettering jet and rocket engines and developing new technologies such as ion propulsion. NASA is also responsible for building, launching, and collecting the data from satellites trying to detect global warming.

The Department of Defense (DOD)

The DOD consumes more energy than most nations of the world, and four times more energy than all the other federal agencies combined. The DOD took an interest in energy long before there was an energy crisis. Energy in the form of petroleum is what fuels military technology, and almost every military strategy involves, directly or indirectly, energy. The DOD is not only interested in the logistics of petroleum planning (through the Defense Fuel Supply Center), but also in developing an array of nontraditional energy technologies, everything from better nuclear propulsion for ships and submarines to solar photovoltaic applications and battery technologies to power mobile communication technologies.

Nuclear Regulatory Commission

The Nuclear Regulatory Commission (NRC) is an independent federal agency that licenses and decommissions commercial nuclear power plants and other nuclear facilities. NRC inspections and investigations are designed to assure compliance with the Agency's regulations, most notably the construction and operation of facilities, the management of high-level and low-level nuclear wastes, radiation control in mining, and the packaging of radioactive wastes for transportation.

STATE AND LOCAL AGENCIES

Almost every state in the nation has an energy policy and planning agency charged with ensuring a reliable and affordable energy supply. Duties of these agencies include forecasting future energy needs, siting and licensing power plants, promoting energy efficiency, and planning for energy emergencies. Instead of leaving all energy decisions entirely up to the free market, many states actively promote changes in production and consumption. Faced with some of the worst air quality problems, the California Energy Commission is one of the most active state agencies in implementing demand side management and market transformation strategies.

In an effort to coordinate policy, exchange information, and to convey to the federal government the specific energy priorities and concerns of the states, the state agencies formed the National Association of State Energy Officials (NASEO) in 1986. To better ensure that appropriate policy was being implemented, NASEO set up the Association of Energy Research and Technology Transfer Institute to track the successes and failures of different programs.

Because the federal government has taken a more hands-off approach toward energy since the 1970s, this trend of state and local governments becoming more active in making energy and environmental decisions about electricity production and transportation issues is likely to continue, especially for the more populace areas of the country like California and the Northeast.

John Zumerchik

See also: Agriculture; Air Pollution; Biofuels; Climatic Effects; Culture and Energy Usage; Demand-Side Management; Emission Control, Power Plant; Emission Control, Vehicle; Geography and Energy Use; Government and the Energy Marketplace; Market Transformation; Military Energy Use, Historical Aspects of; National Energy Laboratories; Nuclear Waste; Propulsion; Spacecraft Energy Systems.

BIBLIOGRAPHY

Cochrane, R. C. (1966). *Measures for Progress: A History of the National Bureau of Standards*. Gaithersburg, MD: U.S. Department of Commerce.

Energy Information Administration. (1998). *Energy Information Directory 1998*. Washington, DC: Author

Fehner, T., and Holl, J. M. (1994). *Department of Energy 1977–1994 A Summary History*. Oak Ridge, TN: Department of Energy.

Government Printing Office. (1980). *Cutting Energy Costs: The 1980 Yearbook of Agriculture*. Washington, DC: Author.

Greene, D. L. (1996). *Transportation and Energy*. Landsdowne, VA: Eno Transportation Foundation, Inc.

Hewlett, R. G., and Anderson, O. E., Jr. (1991). *History of the United States Atomic Energy Commission*, Vol. 1: 1939–1946. Berkeley, University of California Press.

Hewlett, R. G., and Duncan, F. (1991). *Atomic Shield*, Vol. 2: 1947–1952. Berkeley: University of California Press.

Hewlett, R. G., and Holl, J. (1991). *Atoms for Peace and War: Eisenhower and the Atomic Energy Commission*, Vol. 3: 1953–1961. Berkeley: University of California Press.

Landy, M. K.; Roberts, M. J.; Thomas, S. R.; and Lansy, M. K. (1994). *The Environmental Protection Agency: Asking the Wrong Questions: From Nixon to Clinton*. New York: Oxford University Press.

Scheinman, L. (1998). *International Atomic Energy Agency and World Nuclear Order*. Washington, DC: Resources for the Future.

Scott, R. (1994). *The History of the International Energy Agency 1974–1994*, Vol. 1: *Origins and Structure*, Vol. 2: *Major Policies and Actions*. Paris, France: OECD/IEA.

Tengs, R. O., et al. (1995). "Five-Hundred Life-Saving Interventions and Their Cost-Effectiveness." *Risk Analysis* 15(3):369–389.

Tetreault, M. A. (1981). *The Organizations of Arab Petroleum Exporting Countries.* Westport, CT: Greenwood Publishing Group.
United Nations. (1998). *Basic Facts About the United Nations.* New York: Author.

GOVERNMENT AND THE ENERGY MARKETPLACE

Government intervention in energy markets involves either government ownership of industry resources or, more commonly, planning and regulation of privately held resources. The proper role of government planning and regulation has been a hotly debated issue among economists and politicians for many years. Through the twentieth century, governments of the United States and of most of the rest of the developed world have become increasingly larger and more involved in taking action that affect the outcomes of how energy is produced, transported, and consumed. Some believe government intervention in markets is the best solution; others believe government intervention should be avoided. Given the extensive number of energy markets and the extensive use of government intervention in those markets, it is not surprising that the record is mixed. Markets have proved to be reasonably efficient, yet also can be harsh; government action has provided important protections, yet also has made some good situations bad and bad situations worse.

THE VIRTUES OF THE MARKETPLACE

Markets decide what energy resources shall be produced, how they shall be produced, and who will receive the benefits of the production processes. With millions of different activity options, and millions of individuals making individual and collective decisions, it is an overwhelmingly complex process. At the same time, it is usually very efficient because self-interested market players communicate through the price system. A self-interested rational individual will make decisions based on true preferences, and follow those preferences in a way that will provide the greatest satisfaction. Choices will be made under certain constraints that apply to all consumers: income, energy prices, and the ability to switch fuels or fuel providers.

Economic theory holds that individuals will make decisions based on what is in their best interests. Government intervention distorts these capital investment decisions; it alters constraints by favoring or giving preferences or incentives to one choice or type of behavior over another. Government intervention tends to subtly manipulate supply and demand of energy, which usually increases costs to users, reduces supply, and leads to higher prices. Political revolt to government intervention is rare since the consumer often is unable to separate the interventionist component of price (the price premium attributable to regulation and taxation) from what the price would be in the free market.

REASONS FOR GOVERNMENT INTERVENTION

Government intervention is justified as a way to correct the shortcomings of the marketplace. Proponents of intervention do not necessarily believe that markets in general do not work, but rather that there are dynamics going on in certain markets that require intervention to cause them to work in a more socially desirable manner. First, intervention is justified in the energy arena because of problems arising from the laws regarding the ownership and exchange of property rights and the purchase and sale of energy rights. Property rights are far more troublesome in the oil and gas production business than most other businesses for two reasons: the commodity of value sits below the surface, and the commodity is a liquid or a gas not a solid, which means it can migrate many miles. Since energy resources know no boundaries (oil and gas deposits often straddle the property of several owners) a system is needed to assign energy rights.

Second, intervention is justified because social costs may exceed private costs as well as private benefits. For example, when an individual chooses to take a personal automobile to work instead of mass transit, the individual driver receives the short-term benefits (privacy, comfort, speed, and convenience) while the negative social costs (greater air pollution, highway construction, traffic jams, and resource depletion) are shared by all. Intervention usually is an attempt to lower the social costs. However, the problem is that social costs may be easily identifiable in theory, but much more difficult to accurately quantify in practice.

A third function of intervention is the oversight of energy utility monopolies for electricity and natural gas. Utilities have been considered natural monopolies because it was generally cost effective to have a sole generator, transmitter or distributor of electricity or natural gas for a local area due to the large infrastructure required and economies of scale. Without competitors, government historically felt a duty to protect customers from unfair price gouging. But new technology is challenging the concept of natural monopoly in the utilities, and has resulted in the deregulation of energy generation, transmission and distribution in the 1990s. Government's role is shifting to establishing the rules, guidelines, and procedures that attempt to be equitable to all parties (industry and customers) and minimize the social costs to society by ensuring these markets are also socially desirable.

Finally, government intervention is often called for to establish standards by essentially reducing informational market barriers. This is important because the rational individual can only make a decision in his best interest when the information is at hand to make that decision.

TYPES OF GOVERNMENT INTERVENTION

Governments intervene in all energy markets—exploration, production, distribution, and consumption—and carry out intervention in many different ways. While some impacts of intervention are intended, other impacts occur indirectly as an unintended consequence. Examples of a direct impact are the price-lowering effect of subsidies for the biofuel ethanol, and the increased desirability of conservation resulting from a tax on energy use. Examples of unintended indirect impacts are the encouragement of single passenger driving over mass transit resulting from the provision of free parking, or a dramatic increase in the price of coal-generated electricity resulting from clean air regulations to reduce air pollution.

Taxes

Governments levy taxes to raise revenue and to discourage consumption of what is taxed. The primary purpose of United States federal and state gasoline taxes is to raise revenue for transportation infrastructure, particularly highway construction. In Europe and Japan, where gasoline taxes are many times greater (gasoline retails for over twice as much as in the United States), the purpose is also to discourage consumption. However, when taxes are implemented, rarely are drops in transportation energy consumption immediate. Even very steep increases in gasoline taxes usually take time to result in a reduction in consumption. People who drive fifty miles to work still need to get to work by driving in the short term. Only over the long term can people decide whether to move closer to work, change jobs, rearrange schedules to use mass transit or to carpool, purchase a more fuel efficient car, or accept higher energy costs by maintaining the same lifestyle.

Unlike the gasoline tax that only impacts the transportation sector, carbon taxes affect all sectors of the economy. Implemented by some European countries and proposed in the United States by the Clinton Administration in 1993, the carbon tax makes consumption of fossil fuels more expensive for the energy user. The goals of a carbon tax are to reduce the consumption of energy and to make non-carbon emitting sources like wind and hydroelectric more cost-competitive with fossil fuels.

Subsidies

Subsidies are transfer payments from governments to business interests and individuals. Governments have been involved in the subsidizing of energy since the early part of the twentieth century. For democratic governments, subsidies have been often used as a means to buy votes, and for authoritarian governments, as an important means of placating the masses. Price supports for ethanol, tax credits for renewable energy, depletion allowances for oil and gas production, and grants for energy conservation measures for low income families fit the narrow definition of energy subsidies. But if subsidies are considered in a broader context, trade barriers, regulations, and U.S. military spending to ensure the flow of oil from the Persian Gulf can be considered subsidies as well.

Regulation

Regulations are rules set by governments to protect consumers and to control or direct conduct in the marketplace. Production, transportation, consumption, and prices have all been affected by energy regulations.

Allocation control regulations affecting oil prices were issued in the 1970s in response to what was perceived as a number of market inefficiencies. The primary motivation was fairness to low-income individuals and families, yet price controls ended up exacerbating the inefficiencies—worsening the problem it was supposed to solve. Since prices convey critical information to buyers and sellers of energy, government intervention that tries to alter the message can distort market signals so that everyone ultimately ends up less satisfied. Price controls send the wrong signal by artificially stimulating consumption (more demand at a lower price) and reducing the quantity supplied (producers are willing to produce less at the lower price).

The most-well known energy regulations are probably the ones covering the production, transmission, and distribution of electricity and natural gas. To avoid the inefficiency of having multiple electric cables or gas lines run down each street and into every home, in the past governments awarded the right to serve an area to one firm—a "natural monopoly." Unlike most countries, where the government owns and operates electric and gas monopolies, in the United States most of the power industry remains privately-owned and the government regulates it by setting prices, the amount that can be earned, and the quality of service provided. Although Congress develops legislation that authorizes regulation, identifying and enforcing of the rules for specific cases are carried out by the Federal Energy Regulatory Commission and state public service commissions.

Criticism of the way public utility monopolies have been regulated usually centers around the built-in biases that may be favorable to the utility and the customer but are also socially undesirable. Electric utility regulation was structured so that companies earned more by building more facilities, selling more power, encouraging consumption, and discouraging conservation. In the late 1970s and early 1980s, states instituted demand-side management programs that attempted to shift the incentive away from consumption to conservation.

In the late 1970s, public policies began to focus on opening up power generation markets so that non-utility generators and independent power producers could have access to transmission systems and electricity markets. By the early 1990s, much more aggressive public policies were being considered and adopted. These public policies envisioned giving customers in all customer classes the opportunity to select their electricity suppliers. It was recognized that allowing customer choice would require not only providing open access to transmission systems but also the total "unbundling" of power production from power delivery and customer service functions to prevent discriminatory transactions and self dealing.

While unbundling was seen as a necessary step for deregulation, utility companies pointed out that unbundling could create "stranded assets." The creation of stranded assets could be devastating to those utility companies that had made investments in nuclear power plants and other assets that would be uneconomic or unable to compete in open markets. To address the stranded assets issue, government intervention sought to create transition plans elevating rates above market values and allowing the recovery of stranded costs. In the longer term, it is likely that many electric power marketers will be buying power from many different electricity producers, and then turning around and selling the power to industry, business and the residential customers.

"Cross subsidies" have traditionally been used in establishing rates. Industrial customers pay a higher rate for their electricity than residential customers and, in effect, subsidize the power supply for residential customers. In a different context, "cross subsidies" can be created to select power generation technologies based on their environmental impact. There is political pressure from the environmental interest groups to include ratepayer cross-subsidies so that the less desirable power generation technologies (e.g., coal-fired steam turbines) subsidize the more desirable power generation technologies (e.g., wind turbines). This is being considered as part of the deregulation process. Pollution costs fall largely on third parties. In the case of fossil fuel electric power plants, only a small fraction of air pollution costs are borne by the beneficiaries of the power production. Proponents view cross subsidies as an effective way to raise the cost of fossil fuel energy production, lower the cost of renewable energy, and support energy efficiency and conservation programs. Programs like "Plant a tree whenever you build a new fossil fuel power plant" are, in effect, cross subsidy programs because they balance environmental objectives with economic objectives in power generation planning.

Compared to regulation, cross subsidies are a more market-based approach for discouraging air pollution emissions from energy production. Economists generally agree that economic approaches to pollution control (taxes, subsidies, pollution allowances, and trading programs) have been far more efficient and equitable solutions than "command and control" approaches to regulations that have mandated specific emission levels and specific technology.

Since there is an environmental impact with almost all energy production, environmental regulations and policy are indirectly having a greater impact on the energy industry than direct energy policy. Clean air regulations continue to make it more expensive to burn fossil fuels, especially coal; clean water regulations present expensive challenges for the transportation (oil tankers) and storage (underground fuel tanks) of petroleum products; carbon dioxide reduction policies to prevent global warming favor the growth of renewable energy over fossil fuels.

Safety is another major regulatory area. The Occupational Safety and Health Administration (OSHA) enforces many worker safety regulations covering the operation of refineries and electric power plants, and the Nuclear Regulatory Commission oversees the regulations ensuring the safety of nuclear power plants. There are also numerous safety regulations covering the operations of oil and natural gas pipelines and oil tankers. These protective programs have unquestionably provided major social benefits. However, critics contend that these programs have gone too far, have been too expensive, and have placed too much emphasis on government as the interpreter of safety.

When regulation fails to achieve the intended objective, the question arises of whether it was a failure solely of administration or of theory too. If advocates of regulation can prove failure to be largely administrative, it is easier to make the case for new regulations under improved administration.

Information

There have been considerable efforts made at moral persuasion aimed at the social conscience to convince those involved in high consumption behavior to behave more responsibly. During the 1970s, the United States government embarked on a campaign to persuade the public to conserve energy by, among other things, driving 55 mph, keeping car tires inflated, and turning down thermostats in buildings to 65 degrees. In the 1980s, energy use appliance labels were mandated so that rational consumers could compare the efficiency of models and purchase the more energy efficient models. (This was followed by Energy Star Labeling, a joint action of the EPA and DOE in the 1990s that recognizes products such as computers or windows that meet a given energy efficiency standard. Early in the twenty-first century, an emerging target of moral persuasion is the sports utility vehicles (SUV). Compared to the average automobile, the average SUV consumes about 30 percent more fuel per mile and generates more air pollutants and far greater carbon dioxide emissions. The message is that it is unpatriotic to purchase these energy-guzzling, high-emission vehicles. The combination of moral persuasion and the reclassification of SUVs as automobiles (to meet the tougher emission and Corporate Average Fuel Economy standard for automobiles) could be an effective means of curtailing SUV purchases.

Moral persuasion has been most effective in times of short-term emergencies, like the energy shortages of the 1970s, but is far less effective in the long-term and for problems not universally viewed as problems. The benefits of such efforts are the speed, the inexpensiveness of the approach, and the implicit threat behind the effort: stop the undesirable activity by choice or society will take direct measures to curtail it.

Research and development

Europe, Russia, Japan, and the United States have spent billions for energy research and development since the 1970s. The United States alone spent over $60 billion from 1978 to 1996. Nations liberally spend on energy research to try to expand supply options, to develop indigenous resources for national security reasons, and to speed invention and innovation that will benefit society. When the energy research and development budget was significantly boosted in the mid-1970s, the hope was that nuclear fusion, solar energy and other renewable energy sources could develop into the primary energy sources of the next century. But because there have been no major breakthroughs, and fossil fuels have remained relatively inexpensive, renewables' share of energy production has actually fallen. Fossil fuels research and development have shown better results: cleaner transportation fuels and vehicles, more efficient natural gas turbines, and the clean coal program, which raised efficiency and reduced the harmful emissions from coal burning power plants. Largely because of the disappointing overall results of federal spending on energy research and development—the poor track record of picking winners, the inequitable funding based more on politics than scientific merit, and the draining of research and development capital away from the private sector—the amount of money spent on energy R&D has been falling throughout the 1980s and 1990s as a percent of total U.S. research and development. It is uncertain whether this trend will continue. Many energy experts still advocate government spending because industrial goals are usually short-term and do not necessarily reflect national goals such as energy security and environmental quality.

THE NEXT ENERGY CRISIS

Historically government intervention has been primarily crisis-driven. During the Energy Crisis of the 1970s, it was easy for government leaders to gather popular support for government policies that subsidized alternative energy sources (boosting supply) and promised to lower the price of energy. Yet once the crisis passed, gathering support for new initiatives became much more difficult. That is largely why there have been only minor new energy initiatives since the end of the Carter Administration in 1980.

Freer and more diverse world energy markets make another major energy crisis less likely. If one does occur, it is likely to again trigger strong cries for government action. But because of the valuable economic lessons learned from government actions in the 1970s, any crisis-related intervention should achieve better results; the ineffective and harmful mistakes of the past are unlikely to be repeated. Moreover, because the total value of crude oil to the economy is much smaller, any oil supply disruption is unlikely to be as severe for the American economy as the oil price hikes of 1973 and 1979. The economy can more easily substitute coal and natural gas for petroleum, and the very energy-dependent industries like aluminum, steel and petrochemicals have grown more slowly and make up a smaller share of the economy than high growth industries like pharmaceuticals, computer technology and the service economy.

John Zumerchik

See also: Appliances; Behavior; Capital Investment Decisions; Demand-Side Management; Efficiency of Energy Use, Economic Concerns and; Efficiency of Energy Use, Labeling of; Environmental Economics; Environmental Problems and Energy Use; Government Agencies; Green Energy; Market Imperfections; National Energy Laboratories; Property Rights; Regulation and Rates for Electricity; Subsidies and Energy Costs; Taxation of Energy; True Energy Costs

BIBLIOGRAPHY

Asch, P., and Seneca, R. S. (1988). *Government and the Marketplace*. Fort Worth, TX: HBJ School.

Bradley, R. L. (1996). *Oil, Gas and Government*. Lanham, MD: Rowman & Littlefield Publishers, Inc.

Yergin, D., and Stanislaw, J. (Contributor) (1998). *The Commanding Heights: The Battle Between Government and the Marketplace That Is Remaking the Modern World*. New York: Simon & Schuster.

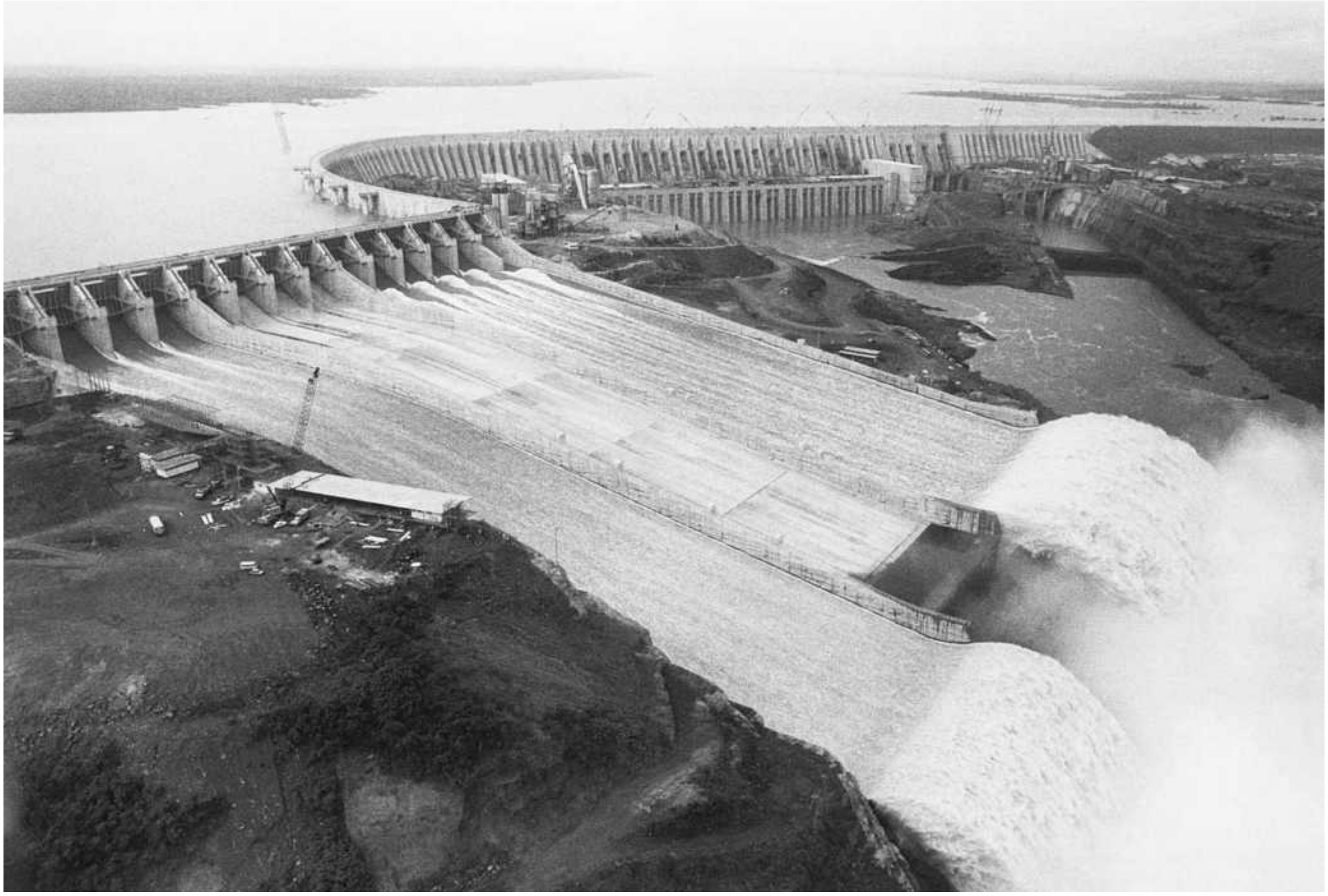
Weidenbaum, M. L. (1995). *Business and Government in the Global Marketplace*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

GRAVITATIONAL ENERGY

Earth's moon is held in orbit by an attractive gravitational force between Earth and the moon. Tides on Earth are due mainly to the gravitational pull of the moon on Earth. Any two masses whether or not Earth and the moon, experience a mutual gravitational force that tends to pull them together. A mass in a position to be pulled to another position by a gravitational force has gravitational potential energy. Anything—water, a book, a parachutist, a molecule in the atmosphere, etc.—has gravitational energy if it is in a position to move closer to the center of Earth. Ordinarily, something has to do work to get the object to the elevated position. A book on the floor of a room has no gravitational energy if it cannot move lower than the floor. But if you lift the book, and place it on top of a table, it has gravitational energy because it is in a position to fall to the floor. The act of lifting the book and doing work produces the gravitational energy of the book. Water atop a dam has gravitational potential energy, but the water did not get into its position at the top of a dam without some agent doing work. The mechanism elevating the water could be a mechanical pump or the natural processes of water evaporating, rising in the atmosphere, condensing to liquid, and falling as rain. Ocean water pulled into a natural basin by the gravitational pulls of the moon and the sun has gravitational potential energy that we refer to as tidal energy.

Electric energy produced by a hydroelectric power plant is derived from gravitational energy. The process starts with gravitational forces pulling the water to the bottom of the dam, where the gravitational energy is converted to kinetic energy (energy due to matter in motion). When the moving water impinges on the blades of a turbine, some of the kinetic energy is converted to rotational energy, causing the turbine to rotate. The turbine is coupled to an electric generator, where the rotational energy of the turbine is converted to electric energy. The electric generator is connected to transmission lines that deliver the electric energy to consumers.

Electricity supplied to consumers is produced primarily by massive generators driven by steam turbines. The system is most efficient and most economical when the generators produce electricity at a constant rate. However, the problem is that demand varies widely by time of day. At night, when consumer demand is low, it would be seem that a utility could keep their most efficient generators running and store electric energy for times when needed. Unfortunately, there is no practical way of storing the quantities of electric energy produced by a large generator. It is possible, however, to convert the electric energy to some other form and store it. Then when electricity is neededm the secondary form of energy is converted back again to electricity. One way of doing this is to keep efficient generators running at night, when the demand is low, and use them to operate electrically driven pumps that store the water as gravitational energy in an elevated reservoir.



The Paraná River meets the Itaipu Dam in Foz Do Iguacu, Brazil. (Corbis Corporation)

When consumer demand picks up during the day, the water is allowed to flow to a lower level to a hydroelectric unit, where the gravitational energy is converted to electric energy. A system like this is called a *pumped-storage facility*. The electric energy recovered in the hydroelectric unit is always less than the electric energy used to store the water as gravitational energy. However, the former is economical because electricity produced at night is relatively cheap. There are a number of these systems throughout the world, the most prominent in the United States being the Robert Moses Plant, near the base of Niagara Falls.

Joseph Priest

See also: Conservation of Energy; Kinetic Energy; Nuclear Energy; Potential Energy.

BIBLIOGRAPHY

Hobson, A. (1995). *Physics: Concepts and Connections.* Englewood Cliffs, NJ: Prentice-Hall.

Priest, J. (2000). *Energy: Principles, Problems, Alternatives*, 5th ed. Dubuque, IA: Kendall/Hunt Publishing Co.

Serway, R. A. (1998). *Principles of Physics*, 2nd ed. Fort Worth, TX: Saunders College Publishing.

GREEN ENERGY

Green pricing for electricity allows consumers to voluntarily pay a premium for more expensive energy sources that are considered better for the environment. The initiative presumes that the buying public believes: (1) the existing corpus of environmental regulation has not fully corrected the negative effects of producing electricity from hydrocarbons (oil, natural gas, orimulsion, and coal); (2) nuclear power is not environmentally friendly despite the absence of air emissions; and (3) certain renewable technologies are relatively benign for the environment.

OPEN-ACCESS RESTRUCTURING: DRIVER OF GREEN PRICING

Electricity typically has been purchased from a single franchised monopolist or municipality. The physical commodity (electrons) and the transmission service were bundled together into one price. The power could be generated (in terms of market share in the United States in 1998) from coal (56%), nuclear (20%), natural gas (11%), hydro (8%), oil (3%), biomass (1.5%), geothermal (0.2%), wind (0.1%), and solar (0.02%). Recently, wholesale and some retail markets have been unbundled, allowing competitors to sell electrons with the monopoly utility or municipality providing the transmission service. Open-access restructuring gives customers choices and creates a commodity market in which the lowest-cost electricity wins market share at the expense of higher-cost alternatives.

The prospect and reality of open-access restructuring is responsible for the current interest in green pricing. Monopoly suppliers became interested in green pricing as a marketing strategy to retain their customer base given the prospect of open-access competition from new commodity suppliers. Independent suppliers operating in open-access states embraced green pricing as an important form of product differentiation. At the same time, environmentalist interest groups campaigned for consumers to use their newfound buying power to absorb a price premium to support those energies perceived to have lower social costs than private costs due to their environmental advantages.

GREEN POWER SUBSIDIES

Electricity generated from biomass, geothermal, solar, and wind facilities (non hydro renewables) have been subsidized by government policies since the 1970s with research and development grants, accelerated depreciation, tax credits, ratepayer cross-subsidies, and must-buy provisions. On an energy production basis, these subsidies have been substantially greater than for conventional generating sources. Without government subsidies today, natural gas technologies would easily dominate the new capacity market in the United States due to their combined economic and environmental characteristics relative to other generation technologies, both conventional and nonconventional.

CHALLENGES FOR GREEN PRICING

The ability of consumers to voluntarily pay a premium to purchase environmentally preferred electric generation would seem to be a happy middle ground between free-market energy proponents and “clean energy” advocates. Yet, on closer inspection there are numerous issues that concern a maturing green-pricing market.

Government Codependence.

Green pricing is not stand-alone philanthropy, but a voluntary addition to mandated subsidies from the entire ratepayer and taxpayer class for the same renewables. Since "green" energy is often intermittent (the sun does not always shine nor the wind always blow), conventional energy with round-the-clock reliability sets the foundation for green pricing. Yet, conventional supply is not compensated for providing the reliability for intermittent "green" energy. Green pricing attempts to get uneconomic renewables "over the top," given upstream subsidies that alone might not be enough to make the renewables economic in a competitive market.

In a free market, the absence of these subsidies might make the green pricing premium prohibitive for consumers. In California, for example, a 1.5-cent per kilowatt hour (kWh) subsidy for green-pricing customers has led to over 100,000 customers for "green" power. There are serious questions about what will happen when the subsidy is reduced or expires. This subsidy is over and above other favors that have led to the construction of high-cost renewable capacity in the last decade—subsidies that may or may not continue in the future.

The codependency of green pricing on government favor is apparent in another way. Green pricing is a short-term consumer commitment, yet needed capacity is capital-intensive and long-lived, requires long lead times, and generates power that may be too expensive to profitably sell in the spot market. To secure financing for such projects, mandates and financial subsidies are required unless qualifying "green" capacity already exists (which predominately has been the case to date). Noneconomic projects built on the strength of short-term green-pricing commitments could become stranded or abandoned assets if the month-to-month consumer premiums cease. This is a reason why ten states as of 1999 have guaranteed a minimum percentage of new capacity to be renewables-based in their electricity restructuring programs. A federal Renewable Portfolio Standard is being sought as well to mandate (nonhydro) renewable quotas nationally.

Subjective Qualifications.

Green pricing assumes that an objective environmental standard exists for consumers to compare competing energies where, in fact, very complicated and subjective tradeoffs exist. A list of issues can be constructed.

Should the programs apply to existing generation, new generation, or both? Applying green pricing to existing projects is a windfall for owners since it is over and above the sustainable support the market originally provided. Many environmentalists have been critical of green pricing programs that utilize existing capacity since no incremental environmental benefits are secured in return for the premium that customers pay. "Greener" for some customers automatically makes the remainder "browner."

Should hydropower be part of the green portfolio since it is renewable and does not emit air pollution, albeit conventional? A green energy primer from the Renewable Energy Policy Project segments hydro into "cleaner" and "polluting" based on river system impacts. Some green pricing programs include projects below 30 megawatts, while other programs do not allow hydro at all.

Should wind projects, like hydro, be bifurcated into "low" and "high" impact categories depending on avian impacts and other nonair emission considerations? Prominent environmental groups have raised concerns over certain bird-sensitive wind sites without generating opposition from other environmental groups pushing windpower as a strategy to reduce air emissions.

Should biomass be included as a "green" energy, given that it has direct air emissions and may not be renewable in its most common application—waste burning?

Should geothermal be included as a "green" energy, given that projects deplete over time and release toxics in many, if not most, applications?

Is nuclear "green" since it produces no air emissions—and in this sense is the cleanest of all energy sources in the aggregate—or should waste disposal and the hazards of plant failure be controlling?

Should "green" energies be blended with "nongreen" sources for genres of green pricing, and if so, in what amounts?

Should power generation from natural gas be included as a "green" energy since it is the cleanest of the hydrocarbons and may compare favorably with some major renewable options in as many as five environmental categories: front-end (embedded) air emissions, wildlife disturbance, land usage, noise, and visual blight? The Renewable Energy Policy

Project delineates between "cleaner" new gas technologies and "polluting" old natural gas technologies—and reserves a third "cleanest" category for energy efficiency/conservation, solar, wind, and geothermal (but not hydro or biomass).

Should "clean" coal plants meeting more stringent new source review standards be differentiated from older plants "grandfathered" to lesser requirements in an energy-environmental matrix? Restated, should operational distinctions used for other renewable sources and natural gas also be made for coal plants for the purpose of environmentally based price differentiation?

Should carbon dioxide (CO_2) emissions from hydrocarbons be rewarded or penalized in green-pricing programs? CO_2 fertilization plays a positive role in plant growth and agricultural productivity. Agricultural economists and scientists have calculated net benefits to a moderately warmer, wetter, and CO_2-enriched world. Other analysis has concluded that unregulated CO_2 emissions, potentially elevating global temperatures enough to destabilize the climate, should be penalized.

Can solar technologies be part of green-pricing programs since they are two or three times more expensive than other renewable options?

Who decides? The above issues point toward a potentially intricate color-coded system differentiating energy not only according to its general type, but also according to operational, site-specific attributes. The categorizations would also have to be revised for new scientific data and pollution-control developments. Can green-pricing programs accommodate this complexity and subjectivity short of resorting to government intervention to define what is "green?"

The leading green-pricing arbiter in the early U.S. experience is the nonprofit Center for Resource Solutions. Their Green-e Certification Seal sets an "objective" standard for green energy under the following criteria:

- Fifty percent or more of the volumetric offering must be from newly constructed, but not mandated, renewable capacity from wind, solar, biomass, geothermal, ocean, or small hydro.
- The remaining 50 percent or less of supply must have an average emission rate below the system average for fossil-fuel-emitted sulfur dioxide, nitrogen oxide, and carbon dioxide.
- No new nuclear supply can be added to the generation mix of a blended product.

State regulators across the country have approved the Green-e labeling program, although marketers do not have to receive the seal to operate. Of greater concern for free green choice is an interest by the Federal Trade Commission and the National Association of Attorneys General to set green power guidelines.

Tracking and Labeling Issues

How should "green" electrons be documented for premium-paying consumers? The industry has proposed that green electrons be certified upon generation, creating a paper record that can be traded and ultimately balanced downstream between sales and purchases. Certain environmentalist interest groups have proposed instead that each "green" electron be physically tracked in the transaction chain so that the final consumer gets actual "green" electrons. Industry groups have resisted this proposal, believing any tracking attempt would reduce flexibility and increase transaction costs when a paper system would result in the same impetus for "green" supply. Physics dictates that electrons flow freely; supplies by necessity are commingled and balanced as they move downstream. In other words, the person buying "green" electrons will not necessarily receive those electrons.

As part of green pricing, environmentalists have proposed a labeling system in which all batches of electricity enter the system with specific information about their generation mix and air-emission characteristics. Industry has proposed that labeling just be made for a positive assertion of green power rather than a blanket reporting of mix and emissions—again for cost and flexibility reasons.

OTHER CONCERNS

The environmental community has raised concerns that green pricing could be construed by lawmakers and the public as a substitute for taxpayer and ratepayer subsidies. A major consumer group has complained that green pricing is only for the affluent, and nonparticipants get to "free ride" on the more environmentally conscious users. Independent power marketers have worried about utility green-pricing programs increasing barriers to entry when the market opens to competition.

CONCLUSION

Green pricing, although nominally a free-market support program for consumer environmental preferences, is closely intertwined with government intervention in energy markets and prone to special interest politics. To pass a free market test, regulators and providers would have to establish open, standalone green-pricing programs that do not contain upstream subsidies or require downstream quotas on market participants. In such a world, voluntary labeling standards and common law protections against fraudulent misrepresentation can ensure the integrity of the program.

Green pricing in a true market would have to respect the heterogeneity and fluidity of consumer preferences. Consumer opinion about what is "green" can be expected to evolve and mature, particularly if environmental programs reduce or remove the perceived negative externalities of less expensive energy, relative economics substantially change between energies, or new information becomes available about environmental characteristics of competing energy choices.

Robert L. Bradley, Jr.

BIBLIOGRAPHY

Bradley, R. L. (1999). "The Increasing Sustainability of Conventional Energy." In *Advances in the Economics of Energy and Resources, Vol. 11: Fuels for the Future*, ed. J. Moroney. Stamford, CT: JAI Press.

Holt, E. A. (1997). "Disclosure and Certification: Truth and Labeling for Electric Power." Issue Brief no. 5. College Park, MD: Renewable Energy Policy Project.

Intergovernmental Panel on Climate Change. (1996). *Climate Change 1995: The Science of Climate Change*. Cambridge, UK: Cambridge University Press.

Mendelsohn, R., and Newman, J., eds. (1999). *The Impact of Climate Change on the United States Economy*. Cambridge, UK: Cambridge University Press.

Morris, J. (1997). *Green Goods? Consumers, Product Labels and the Environment*. London: Institute for Economic Affairs.

Rader, N. (1998). *Green Buyers Beware: A Critical Review of "Green Electricity" Products*. Washington, DC: Public Citizen.

Rader, N., and Short, W. (1998). "Competitive Retail Markets: Tenuous Ground for Renewable Energy." *Electricity Journal* 11(3):72–80.

Renewable Energy Policy Project. (1999). "Your Green Guide to Electricity Choices" (pamphlet).

Wiser, R.; Fans, J.; Porter, K.; and Houston, A. (1999). "Green Power Marketing in Retail Competition: An Early Assessment." NREL/TP.620.25939. Golden, CO: National Renewable Energy Laboratory.

Wittwer, S. (1995). *Food, Climate, and Carbon Dioxide: The Global Environment and World Food Production*. New York: Lewis Publishers.

GREENHOUSE EFFECT

See: Air Pollution; Environmental Problems and Energy Use

GUNPOWDER

See: Explosives and Propellants

H

HEAT AND HEATING

MAINTAINING CONSTANT TEMPERATURE

The purpose of a heating system is to maintain a constant temperature. To do this, all heating systems—whether it is for heating a small room or a domed stadium—must compensate for heat losses with heat gains. Even though the human body continually loses heat through conduction with the surrounding air and evaporation of water from the surface of the skin, the temperature of the human body is remarkably constant. For every joule of heat flowing from the body, there is one joule produced by mechanisms that convert energy from food. If heat loss quickens (e.g., while one is swimming), the heat regulators in the body convert more energy from food to compensate.

It is the same with a room in a building. Heat is lost through many tiny openings to the outside and through conduction through walls and windows. A heating system makes up this loss keeping the temperature constant. The greater the temperature difference between the inside and outside of a building, the greater is the heat loss. This is why the heating system must provide more heat when you want to maintain a higher temperature as well as when the outside temperature is extremely cold.

FORCED-AIR HEATING SYSTEM

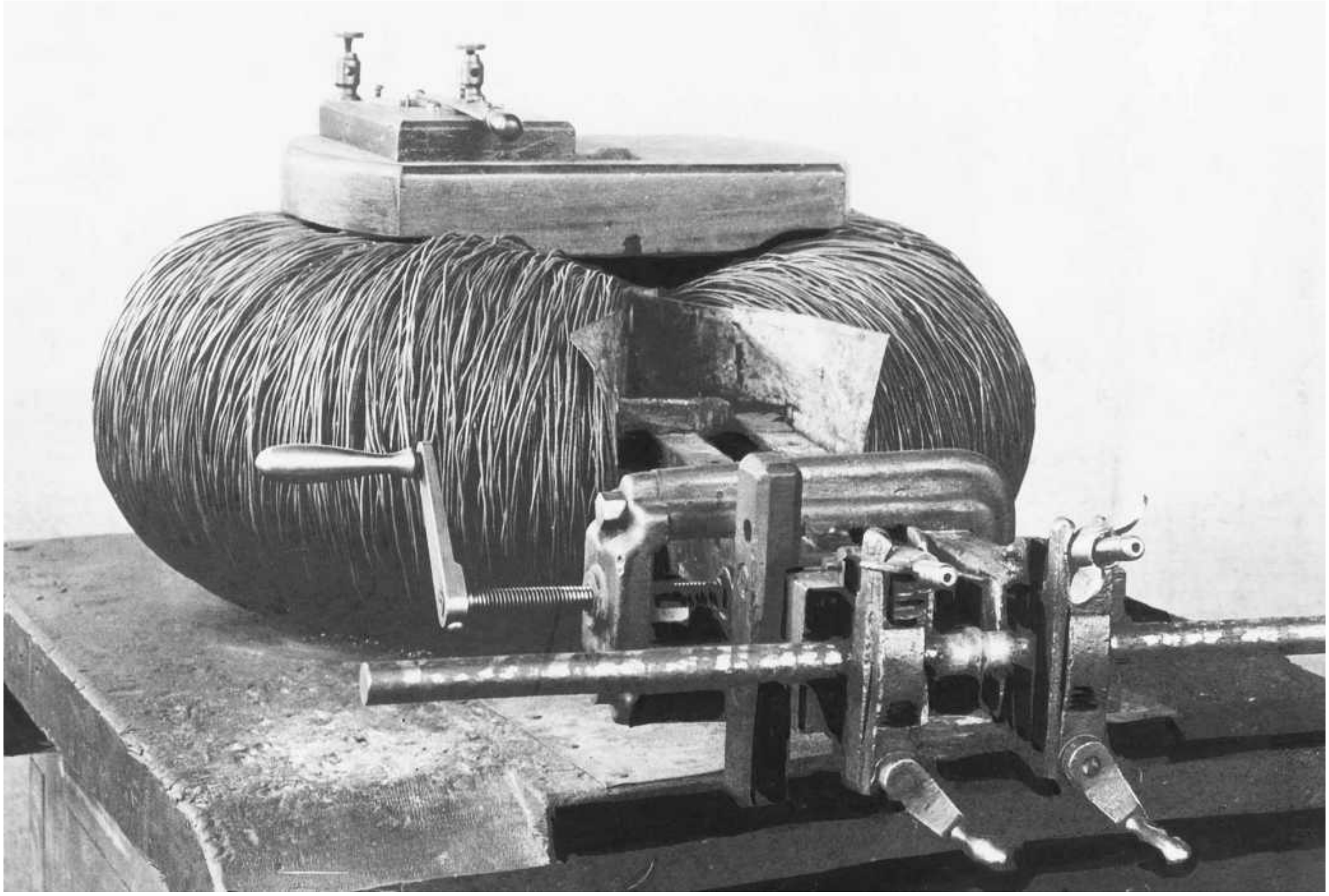
For any substance, the total thermal energy is the sum of the energies of all the molecules. In other words, thermal energy is equal to the energy per molecule multiplied by the number of molecules. A volume of air may have a high temperature, but does not necessarily have a lot of thermal energy because the number of molecules might be small. Conversely, a volume of air may have low temperature and significant thermal energy if many molecules are involved. When a gas is mixed with one at a higher temperature, the mixture will come to a temperature somewhere between the temperatures of the gases being mixed. A forced-air heating system works on this principle by mixing warmer air with the air existing in the room. Sometimes the mixing is uneven in a room producing temperature variations that are annoying to the occupants.

HOT-WATER HEATING SYSTEM

A hot-water heating system forces water into pipes, or arrangements of pipes called registers that warm from contact with warm water. Air in the room warms from contact with the pipes. Usually, the pipes are on the floor of a room so that warmer, less dense air around the pipes rises somewhat like a helium-filled balloon rises in air. The warmer air cools as it mixes with cooler air near the ceiling and falls as its density increases. This process is called convection and the moving air is referred to as convection current. The process of convection described here is pipe-to-air and usually does a better job of heating evenly than in an air-to-air convection system—the circulation of air by fans as in a forced-air heating system.

ELECTRICAL RESISTANCE HEATING

Rather than have pipes warmed by water, some heating systems employ coils of wire or other configurations of metal that are warmed by electric currents. An electric current is a flow of electric charge that may be either positive, negative or a combination of positive and negative. Both positive and negative electric currents are involved in the battery of an



The first electric welding transformer, a device which uses electricity to weld joints of metal using resistance heating. (Corbis Corporation)

automobile. But in a metal the electric current is due to the flow of electrons, which have negative charge. The electrons are components of atoms making up the structure of the metal. In a very similar way, electrons are forced to move through a wire by electrical pressure in somewhat the same way that water molecules are forced to move through a pipe by water pressure. Water molecules encounter resistance as they flow through the pipe and electrons encounter resistance as they flow through a metal wire. The resistance of both the pipe and the wire depends on the dimensions of and obstructions in the pipe and wire. The resistance increases as the length increases but decreases as the diameter increases.

The material selected makes a difference too. For the same length and diameter, a wire made of copper has less resistance than a wire made of aluminum. A current-carrying wire warms because of electrical resistance.

The resistance (R) of a conductor is measured in ohms and the rate of producing heat (power P) in watts. When the conductor is connected to a constant voltage source, as in a household electric outlet, the thermal power produced is given by $P = V^2/R$. Note that if the resistance halves, the power doubles. Some electric toasters produce heat from current-carrying wires that you usually can see. Because the voltage is fixed, typically 120 volts in a house, a desired thermal power is attained by tailoring the resistance of the heating element. The wire chosen must also be able to withstand the high temperature without melting. Typically, the wires are nichrome (an alloy of nickel and chromium) and the electrical resistance is about 15 ohms, making the thermal power about 1,000 watts. One thousand watts means 1,000 joules of heat are produced each second the toaster is operating.

A current-carrying wire converts electric energy to thermal energy with an efficiency of nearly 100 percent. In this sense, electric heating is very efficient. However, the electricity for producing the current in the heater was likely produced by burning coal in a large electric power plant. The efficiency for delivering electric energy to a consumer is about 32 percent due to energy losses in the plant and in transmission to the house. Therefore, the efficiency for converting the thermal energy from burning coal to thermal energy delivered by an electric heater is very low compared to the 80–90 percent efficiency for modern forced-air heating systems burning natural gas. Electric resistance heat is clean and convenient and a good backup for other systems, but as the main source of heat, the economics are very unfavorable.

UNWANTED HEAT PRODUCED IN ELECTRONIC COMPONENTS

Resistors are components of any electronic device such as a computer. If there is need for a resistance of 1,000 ohms, a designer can purchase a resistor having 1,000 ohms resistance. Heat will be produced in the resistor when a current flows through it. If this heat is not transferred away, the temperature will rise and damage the resistor. Usually the designer relies on heat being conducted to surrounding air to keep the resistor from overheating, and the most efficient transfer of heat is by increasing the surface area of the resistor. The physical dimensions of the resistor will be chosen so that the thermal power can be dissipated without damaging the resistor. For example, a 1,000-ohm resistor designed to dissipate 1 watt of heat will be about four times larger than a 1,000-ohm resistor designed to dissipate ¼ watt. In some cases the transfer of heat by direct conduction to the air is not sufficient and the resistor will be mounted onto a metal housing having substantial mass and surface area.

In addition to resistors, the electronic world of computers, calculators, television sets, radios, and high-fidelity systems involves a multitude of electrical components with names like diode, transistor, integrated circuit, and capacitor. Unwanted heat is generated in these components because of ever-present electrical resistance. If the heat is not transferred away, the temperature of the device will increase and, ultimately, be damaged. The heat generated is especially large in transistors in the final stage of a high-fidelity system where the speakers are connected. These so-called power transistors are mounted on a metal structure having substantial mass and surface area to disperse heat transferred to it by the transistor.

Scientists are forever searching for electronic components like transistors that will perform their function with minimum production of heat. In early versions of personal computers the heat produced by the components could not be dissipated by natural means and fans were used to force the heat to the surroundings. Most contemporary personal computers, including laptops, rely on removing heat by natural means and do not require fans.

TRANSMISSION OF ELECTRIC POWER

An electric generator in a modern coal-burning or nuclear-fueled electric power plant converts mechanical energy to electric energy at an impressive efficiency, about 99 percent. The 1 percent of mechanical energy not going into electric energy is lost as heat. When an electrical device is connected to the generator, the power delivered depends on the voltage of the generator and the current in the device. Specifically, electric power (watts) is equal to voltage (volts) times current (amperes). Typically, the voltage is 10,000 volts and the generator is capable of delivering around 1 billion watts of electric power. If the generator were connected directly to wires that transmitted electric power to consumers, the current would be 1,000,000,000 watts per 10,000 volts, which equals 100,000 amperes. The wires through which this current must flow can be tens or hundreds of miles long and have non-negligible electrical resistance. Therefore, heat will be produced and lost. Even if the resistance of the wires was as low as 1 ohm, the energy produced by the generator would be completely lost as heat. The power company uses a device called a transformer to reduce the current in the wires and thereby reduce the heat loss. The transformer is a device that accepts electric power at some voltage and current and then delivers essentially the same power at a different voltage and current. At the power plant the voltage is stepped up from 10,000 volts to around 1,000,000 volts. A voltage of 10,000 volts delivering 100,000 amperes would be changed to a voltage of 1,000,000 volts and current of 1,000 amperes. Because the power lost as heat depends on the square of the electric current, the lower current reduces the heat loss significantly. When the power arrives at a distribution center for consumers, the voltage is stepped down from its transmission value of 1,000,000 volts. For household use this voltage is 120 volts for lights and small appliances and 240 volts for electric stoves, electric clothes dryers, and electric water heaters.

A superconductor is a material having zero electrical resistance. An electric current in a superconductor would produce no heat. To achieve zero resistance the superconductor must be cooled. This temperature has gotten progressively higher over the years. At this writing it is about 100 K, which is about –173°C. As research progresses on superconducting materials there is hope that they can be used in the transmission of electric energy. Even though there would be an energy expenditure involved in cooling the transmission lines, it would be more than offset by the savings accruing from not having power lost to heat generated in the transmission lines.

COOLING TOWERS AND LOW-GRADE HEAT

You can do work by rubbing your hands together and all the work will be converted to heat. A food mixer can churn through water and do work and all the work will be converted to heat. But a heat engine can only convert a portion of the heat it takes in to work. Nowhere is this more evident than in the steam turbines driving electric generators in electric power plants. The clouds seen emanating from the giant towers so noticeable at electric power plants are graphic evidence of heat being transferred to the environment.

High temperature steam pounding on the blades of a turbine causes the blades to rotate. After passing by the blades the cooler steam condenses to liquid water. Condensation liberates energy and if this energy is not removed, the condenser part of the system warms up. The heat liberated at the condenser is removed by circulating water in pipes in contact with the condenser. Each second the water must remove more than a billion joules of heat, so there is a substantial amount of thermal energy absorbed by the cooling water. The purpose of a cooling tower is to transfer the thermal energy in the cooling water to the air environment by splashing the warm water on the floor of the tower. A portion of the water evaporates and is carried into the air environment by a natural draft created by the chimney-like structure of the tower. The vapor cloud seen at the top of a cooling tower is from the water that has been evaporated.

Although there is significant thermal energy in the water circulating from the condenser of a turbine, it is only 10–15°C higher than the input water. It is a case of having a lot of energy by having a lot of molecules. This relatively low-temperature water is not suitable for heating buildings. Furthermore, the electric power plant is usually several miles removed from buildings where the warm water could be used. By the time that the heated water would reach the building, most of the heat energy would be lost to the environment. Locating industries and communities near plants would make it feasible. One suggestion for using the warm water is to stimulate plant growth in nearby greenhouses.

Remember that heat is lost to the environment in virtually all energy conversion processes. The more heat that is lost the less desirable the form of energy and the less efficient the process. A 100-watt fluorescent lightbulb is more efficient than a 100-watt incandescent bulb because more of the electric energy goes into radiant energy and less into heat. Modern electronics using transistor technology is enormously more energy efficient than the original vacuum tube technology. The total elimination of heat loss can never be achieved, but scientists and engineers work diligently to find better ways of minimizing heat loss and successes have been achieved on many fronts.

Joseph Priest

See also: Conservation of Energy; District Heating and Cooling; Furnaces and Boilers; Heat Transfer.

BIBLIOGRAPHY

Incropera, F. P. (1996). *Fundamentals of Heat and Mass Transfer*, 4th ed. New York: John Wiley and Sons.

Killinger, J. (1999). *Heating and Cooling Essentials*. Tinley Park, IL: Goodheart-Willcox Co.

Priest, J. (2000). *Energy: Principles, Problems, Alternatives*, 5th ed. Dubuque, IA: Kendall/Hunt Publishing Company.

Ristinen, R. A., and Kraushaar, J. J. (1999). *Energy and the Environment*. New York: John Wiley and Sons, Inc.

Stevenson, W. (1982). *Elements of Power System Analysis*, 4th ed. New York: McGraw-Hill Publishers.

Woodson, R. D. (1971). *Scientific American* 224(12):70–78.

HEATING OIL

See: Residual Fuels

HEAT PUMPS

A heat pump is a thermodynamic heating/refrigerating system used to transfer heat. Cooling and heating heat pumps are designed to utilize the heat extracted at a low temperature and the heat rejected at a higher temperature for cooling and heating functions, respectively.

The household refrigerator can provide a simple analogy. A refrigerator is actually a one-way heat pump that transfers heat from the food storage compartment to the room outside. In so doing, the inside of the refrigerator becomes progressively cooler as heat is taken out. In a closed room, the heat coming out of the refrigerator would make the room warmer. The larger the refrigerator is, the greater the potential amount of heat there is to transfer out. If the refrigerator door were open to the outdoors, there would be an almost unlimited amount of heat that could be transferred to the inside of a dwelling. Thus it is possible to design a refrigeration system to transfer heat from the cold outdoors (or any other cold reservoir, such as water or the ground) to the insides of a building (or any medium that it is desired to heat).

Heat pumps can move heat energy between any form of matter, but they are typically designed to utilize the common heating and cooling media, i.e., air or water. Heat pumps are identified and termed by transfer media. Thus terms air-to-air, water-to-air, or air-to-water are commonly used. Most heat pumps are designed to transfer heat between outside air and inside air, or, in the case of so-called geothermal or ground-source heat pumps, between the ground or well water and air. The principle application of heat pumps is ambient heating and cooling in buildings.

There are three types of heat pump applications: heating, cooling or heating and cooling. “Heating only” heat pumps are designed to transfer heat in one direction, from a cold source, such as the outdoors, to the inside of a building, or to a domestic hot water plumbing system, or to some industrial process. There is no term “cooling only heat pump,” as the term would simply be describing what is commonly referred to as an air conditioner or cooling system.

“Heating and cooling” heat pumps have refrigeration systems that are reversible, permitting them to operate as heating or cooling systems. An analogy would be a window air conditioner that was turned around in winter, so that it was cooling the outside and blowing hot air inside. Most heat pumps in use are of the heating and cooling type, and are used to heat residences, or commercial and industrial buildings. Residences consume the bulk of those sold: in 1999 about a quarter of new single-family homes were equipped with a heating and cooling heat pump.

Most heat pumps utilize a vapor-compression refrigeration system to transfer the heat. Such systems employ a cycle in which a volatile liquid, the refrigerant, is vaporized, compressed, liquefied, and expanded continuously in an enclosed system. A compressor serves as a pump, pressurizing the refrigerant and circulating it through the system. Pressurized refrigerant is liquefied in a condenser, liberating heat. Liquid refrigerant passes through an expansion device into an evaporator where it boils and expands into a vapor, absorbing heat in the process. Two heat exchangers are used, one as a condenser, and one as an evaporator. In the case of an air-to-air heat pump, one heat exchanger is placed outside, and one inside. In a ground-source heat pump, the outdoor heat exchanger is placed in contact with well water, a pond, or the ground itself. In both cases, the refrigerant flow is made reversible so that each heat exchanger can be used as an evaporator or as a condenser, depending on whether heating or cooling is needed. The entire system is electrically controlled. (See Figures 1 and 2.)

Some heat pumps, called thermoelectric heat pumps, employ the Peltier effect, using thermocouples. The Peltier effect refers to the evolution or absorption of heat produced by an electric current passing across junctions of two suitable, dissimilar metals, alloys, or semiconductors. Presently, thermoelectric heat pumps are used only in some specialized applications. They have not been developed to a point to make them practical for general heating and cooling of buildings.

The energy efficiency of heat pumps is measured by calculating their coefficient of performance (COP), the ratio of the heat energy obtained to the energy input. The capacity of modern heat pumps in the United States is rated in British thermal units per hour (Btu/h). The COP at a given operating point can be calculated by dividing the Btu/h output of the system by the energy input in Btu/h. Since system input for most heat pumps is electricity measured in watt-hours, the watt figure is multiplied by a conver-

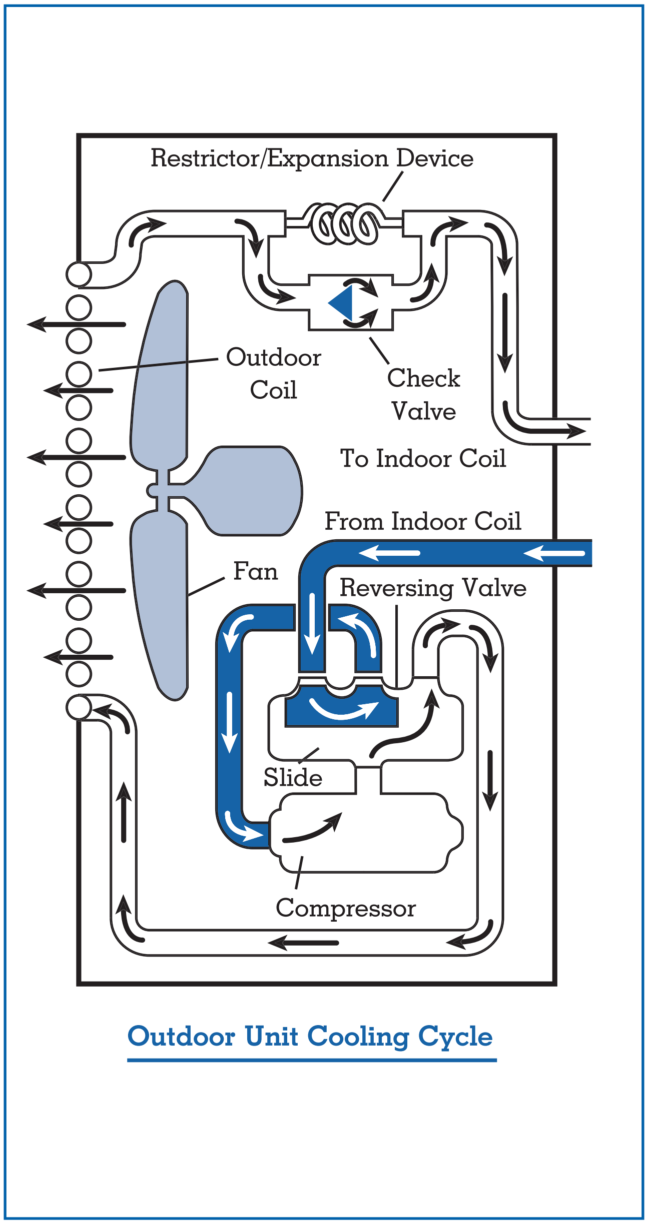


Figure 1.
Schematic of cooling cycle for a heat pump system.
SOURCE: Air Conditioning and Refrigeration Institute

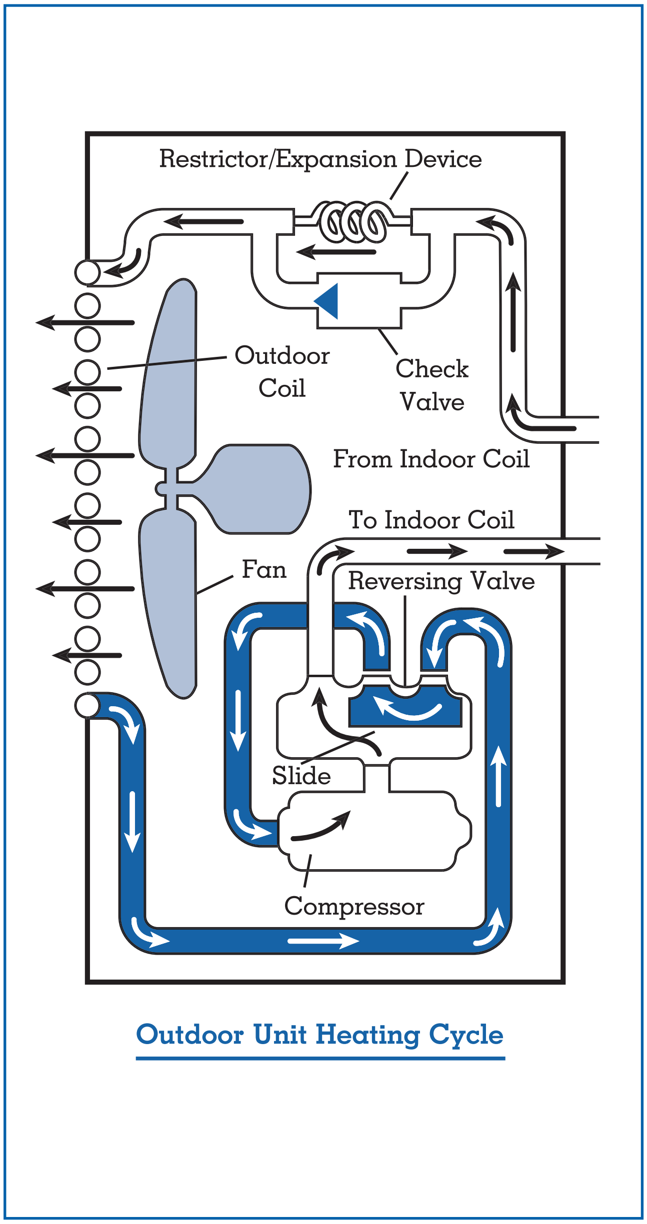


Figure 2.
Schematic of an outdoor heat pump's heating cycle.
SOURCE: Air Conditioning and Refrigeration Institute

sion factor of 3.412 to obtain the energy input in Btu/h.

There are two major factors that impact the COP: temperature difference and system component efficiency. A heat pump requires energy to move heat from a lower temperature to a higher one. As the difference in the two temperatures increases, more energy is required. The COP of a heat pump is higher when the temperature difference is less, and less energy is consumed to transfer a given amount of heat. Thus the COP of a heat pump varies during its operation, and is relative to the system operating point, the temperature of the medium heat is transferred from, and the temperature of the medium it is transferred to. Comparison of the COP of two different systems is meaningless unless the operating points are the same. For this reason, although heat pump efficiency continues to improve, specific historical comparisons of heat pump efficiencies are difficult because system operation data was recorded at different operating points.

Air-to-air heat pumps are particularly prone to varying COPs, due to fluctuating outdoor weather conditions. In the winter, the COP decreases as the outdoor temperature decreases, due to the need for the system to transfer heat energy at greater temperature differences. The heat output decreases as the outdoor temperature drops at the same time that the heating needs of the building are increasing. In fact, the COP may drop so low that the heat pump cannot meet the thermal needs of a building. For this reason an auxiliary heat source such as electric resistance heaters or a fossil-fueled furnace is needed in geographic areas that have cold winters. The COP of air-to-air heat pumps is further reduced by the use of defrost cycles needed to clear the outdoor heat exchanger of frost buildup when the air temperature drops below about 40°F. During defrost, the refrigeration system is reversed into comfort cooling mode, heating the outdoor heat exchanger to melt the frost. Since the defrost mode is cooling the indoor air, supplemental heat is usually necessary to maintain comfort levels, reducing the overall COP of the system. The reduction varies with the duration of the defrost.

In the summer, the COP of an air-to-air heat pump decreases as the outdoor temperature rises, reducing the cooling capacity. Normally the thermal needs of the building are met since it is common practice to size a heat pump so that it will deliver adequate cooling capacity in all but the most extreme summer conditions. The winter heating capacity of the system is then determined by this tradeoff, and if the heating capacity is inadequate, supplemental electric or fossil fuel heat is required.

Because deep ground temperature has little change, geothermal heat pumps using well water function at a fairly stable COP. Geothermal types using the ground for thermal mass will see some variance in COP depending on the dryness of the soil. Wet ground conducts more heat than dry and theoretical soil conductivity may vary up to 1,000 percent. For this reason the ground loop heat exchanger should be buried deep enough to minimize soil moisture fluctuations.

System components can affect the COP because their design and performance can vary. In vapor-compression systems, the transfer efficiency of the heat exchangers and the energy efficiency of the compressor, and how these components are matched, help determine the operating COP. Compressors are performance-optimized for narrow operating ranges. If the optimization is done for heating, cooling operation may suffer. The opposite is also true.

The operating condition of the components also affects the COP. For example, if the heat exchangers become clogged or corroded, or if a homeowner fails to change a furnace filter, the operating COP of a heat pump will decrease. Any heating and cooling system will suffer some performance degradation once it has been put into use. The amount of degradation depends on the conditions under which the system has to operate, as well as how carefully system components are maintained.

Most heat pumps for residences are unitary systems; that is, they consist of one or more factory-built modules. Larger buildings may require built-up heat pumps, made of various system components that require engineering design tailored for the specific building. Sometimes, multiple unitary units are used in large buildings for ease of zone control; the building is divided up into smaller areas that have their own thermostats.

HISTORY

Oliver Evans proposed the closed vapor-refrigeration cycle in 1805 in *The Young Steam Engineer's Guide*. Evans noted: "Thus it appears possible to extract the latent heat from cold water and to apply it to boil other water." By 1852 William Thompson (Lord Kelvin) had proposed that a refrigeration system be used to either cool or heat the air in buildings, and outlined the design of such a machine. In Austria after 1855 Peter Ritter Von Rittenger constructed working heat pumps that were said to be 80 percent efficient. These devices were used to evaporate salt brine, and similar devices were constructed in Switzerland after 1870.

A theoretical discussion of the heat pump appeared in the *Journal of the Franklin Institute* in 1886. T. G. N. Haldane of Scotland comprehensively pursued the application of heat pumps to the heating of buildings after the mid-1920s. Haldane tested air-to-water heat pump systems in his home and concluded that the vapor-compression refrigeration cycle could, under certain conditions, provide a more economical means to heat buildings and swimming pools than fossil fuels. Haldane further proposed using a reversed cycle so the heat pump could be used to cool buildings or make ice. Haldane's heat pump had a coefficient of performance (COP) between 2 and 3, depending on operating conditions.

During the 1930s and 1940s a number of residential and commercial heat pump installations were made in the United States. They were of all types, and the heating COPs of these systems, where results were known, seem to have ranged from 2 to 5. Hundreds of articles and papers were published discussing the theory, application, and installed examples of heat pumps. Despite this activity, most heating systems installed were conventional fossil-fuel furnaces and boilers due to their lower first cost, broad acceptance, and well-established manufacturing, sales, and service infrastructure.

Comfort cooling was one benefit of a heat pump that was typically cited in the literature of the 1930s and 1940s. However there was not a great consumer demand for general comfort cooling at that time, particularly for residences. True, there were isolated pockets of interest, as in movie theaters and some commercial buildings, but genuine mass consumer demand for comfort cooling did not develop until the mid-twentieth century. Thus, there was no financial incentive for mass production of unitary comfort cooling air conditioners in the 1930s and 1940s, and therefore no incentive to produce unitary heat pumps. There were a few isolated attempts, such as a package heat pump marketed by DeLaVergne in 1933, but the efforts were short-lived.

Electric utilities did have a vested financial interest in heat pumps, since most designs used electricity for heating and cooling. But utilities were not manufacturers or consumers. Despite their attempts to promote heat pumps with articles and showcase system installations, they failed to create the demand for the product.

By the 1960s, central residential air conditioning was becoming increasingly popular. Equipment had developed to the point that a number of manufacturers were producing and marketing unitary air conditioning equipment. Some of these manufacturers did attempt to resurrect the idea of applying the heat pump to residential, store, and small office heating and cooling. Systems were designed and marketed, but suffered dismally in the market. System components used did not stand up to the demands of summer and winter operation over wide weather conditions. Compressors and reversing valves in particular saw enough failures that most manufacturers withdrew their products from the marketplace. The air-to-air heat pumps of the 1960s used timed defrost cycles, initiating defrosts even when they were not necessary. Thus the system efficiency was unnecessarily reduced.

Spurred on by the energy crisis of the 1970s interest in heat pumps renewed. Rising price of fossil fuels caused manufacturers to take another look at heat pumps. Owners of "all electric homes," stung by the high cost of electric heat, were looking for a way to reduce heating costs with minimal existing heating system redesign. A heat pump could save 30 to 60 percent of the cost of electric resistance heating.

Remembering the problems of the 1960s, manufacturers redesigned system components. The system efficiency was generally higher than in the 1960s because system components were more efficient. Most air-to-air heat pumps of the period used a demand-controlled defrost cycle, further increasing overall energy efficiency. By the mid-1970s most of the major manufacturers of unitary air conditioning equipment were offering heat pumps. They were particularly popular in areas of moderate winters where air-to-air heat pumps could operate at higher COPs.

Conventional heat pumps continued to be available through the 1970s to the 1990s. The trend has been toward progressively higher energy efficiencies. The average efficiency of new heat pumps increased 60 percent (based on cooling performance) between 1976 and 1998. Minimum efficiencies were mandated for residential heat pumps in 1987 and for commercial equipment in 1992, a significant factor contributing to the rise in efficiency over the past twenty years. For heat pumps used in commercial buildings, the efficiency standards were derived from the model standards published by the American Society of Heating, Refrigerating and Air Conditioning Engineers (ASHRAE). The U.S. Department of Energy may review these standards in the next few years and setting more stringent standards (for both residential and commercial heat pumps) if technically and economically feasible.

CURRENT PRACTICE AND THE FUTURE

Heat pump systems are now rated with a Heating Seasonal Performance Factor (HSPF) and a seasonal energy efficiency ratio (SEER). The HSPF is calculated by the annual heating system output in BTU by the heating electricity usage in watt-hours. The SEER is an estimate of annual cooling output in Btu divided by the cooling electricity usage in watt-hours. The HSPF and SEER are calculated at specific rating points, standardized by the Air Conditioning and Refrigeration Institute (ARI), an industry trade organization that performs testing advocacy, and education. ARI conducts a certification program, participated in by almost all manufacturers of heat pumps. The program gives the consumer access to unbiased and uniform comparisons among various systems and manufacturers. Rating various systems at the same operating point allows accurate comparisons between systems. Heat pumps introduced in 1999 have an NSPF of 6.8 or greater and a SEER of 10.0 or greater, with high efficiency units having an HSPF of as much as 9 and a SEER of 13 or greater.

There is more interest than ever in the geothermal heat pumps. The cost of such systems has decreased with use of plastic piping for water or ground loops. Geothermal systems are proving particularly advantageous in colder winter areas, since ground temperature is much higher than outdoor temperatures. In addition, there is no efficiency-robbing defrost cycle that is necessary in air-to-air heat pumps.

New compressor technology employing rotary scroll type compressors is replacing previously-used reciprocating technology. Scroll compressors operate at higher efficiencies over wider operating conditions. Heat pumps employing scroll compressors use less supplemental heat at low outdoor temperatures. Electrical and electronic technology are making variable-capacity compressors cost-effective for the next century. Compressor performance can be optimized for a wider operation range, and matching compressor capacity to the actual demand for heating or cooling increases the system efficiency, saving energy. Application of microprocessors to control systems further permits fine-tuning of system operation. Energy efficiency is also being increased by development of higher efficiency electric motors for compressors and fans.

Increased environmental awareness will no doubt spur increasing interest in solar-assisted heat pumps in the future. Such systems can operate at higher heating COPs than more conventional heat pumps. A solar-assisted heat pump system uses a solar heating system in parallel with a conventional heat pump. The solar heating system reduces the operating time of the heat pump, and also reduces the need for supplemental electric or fossil-fuel heating during colder weather.

Bernard A. Nagengast

See also: Air Conditioning; Heat Transfer.

BIBLIOGRAPHY

Air Conditioning and Refrigeration Institute. (1999) "How a Heat Pump Works." Consumer Information section. <http://www.ari.org>.

Air Conditioning and Refrigeration Institute. (1999) "Heat, Cool, Save Energy with a Heat Pump." Arlington, VA: ARI.

Air Conditioning and Refrigeration Institute. (1999) *ARI Unitary Directory*. <http://www.ari.org>.

American Society of Heating, Refrigerating and Air Conditioning Engineers. (1978). *ASHRAE Composite Index of Technical Articles 1959–1976*. Atlanta: ASHRAE.

American Society of Heating, Refrigerating and Air Conditioning Engineers. (1996). *Absorption/Sorption Heat Pumps and Refrigeration Systems*. Atlanta: ASHRAE.

American Society of Heating, Refrigerating and Air Conditioning Engineers. (1996). *ASHRAE Handbook: Heating, Ventilating, and Air-Conditioning Systems and Equipment*. Atlanta: ASHRAE.

American Society of Heating, Refrigerating and Air Conditioning Engineers. (1999). *ASHRAE Handbook: Heating, Ventilating, and Air-Conditioning Applications*. Atlanta: ASHRAE.

Bose, J.; Parker, J.; and McQuiston, F. (1985). *Design/Data Manual for Closed-Loop Ground-Coupled Heat Pump Systems*. Atlanta: American Society of Heating, Refrigerating and Air Conditioning Engineers.

Evans, O. (1805). *The Young Steam Engineer's Guide*, p. 137. Philadelphia: H. C. Carey and I. Lea.

Haldane, T. G. N. (1930). "The Heat Pump: An Economical Method of Producing Low-Grade Heat from Electricity." *Journal of the Institution of Electrical Engineers* 68: 666–675.

Howell, R.; Sauer, H.; and Coad, W. (1997) *Principles of Heating, Ventilating and Air-Conditioning*. Atlanta: American Society of Heating, Refrigerating and Air Conditioning Engineers.

Kavanaugh, S., and Rafferty, K. (1997). *Ground Source Heat Pumps*. Atlanta: American Society of Heating, Refrigerating and Air Conditioning Engineers.

Lorsch, H. G. (1993). *Air Conditioning Design*. Atlanta: American Society of Heating, Refrigerating and Air Conditioning Engineers.

Southeastern Electric Exchange. (1947). *Heat Pump Bibliography*. Birmingham, AL: Southern Research Institute.
Sporn, P.; Ambrose, E. R.; and Baumeister, T. (1947). *Heat Pumps*. New York: John Wiley & Sons.
Thompson, W. (1852). "On the Economy of the Heating or Cooling of Buildings by Means of Currents of Air." *Proceedings of the Philosophical Society of Glasgow* 3: 269–272.
University of Pennsylvania. (1975). *Proceedings of Workshop on Solar Energy Heat Pump Systems for Heating and Cooling Buildings*. University Park, PA: University Press.

HEAT TRANSFER

Heat transfer is the energy flow that occurs between bodies as a result of a temperature difference. There are three commonly accepted modes of heat transfer: conduction, convection, and radiation. Although it is common to have two or even all three modes of heat transfer present in a given process, we will initiate the discussion as though each mode of heat transfer is distinct.

CONDUCTION

When a temperature difference exists in or across a body, an energy transfer occurs from the high-temperature region to the low-temperature region. This heat transfer, q, which can occur in gases, liquids, and solids, depends on a change in temperature, ΔT, over a distance, Δx (i.e., ΔT/Δx) and a positive constant, k, which is called the thermal conductivity of the material. In equation form, the rate of conductive heat transfer per unit area is written as

$$q/A = -k\Delta T/\Delta x$$

where q is heat transfer, A is normal (or perpendicular to flow of heat), k is thermal conductivity, ΔT is the change in temperature, and Δx is the change in the distance in the direction of the flow. The minus sign is needed to ensure that the heat transfer is positive when heat is transferred from the high-temperature to the low-temperature regions of the body.

The thermal conductivity, k, varies considerably for different kinds of matter. On an order-of-magnitude basis, gases will typically have a conductivity range from 0.01 to 0.1 W/(m–K) (0.006 to 0.06 Btu/h-ft-°F), liquids from 0.1 to 10 W/(m–K) (0.06 to 6 Btu/h-ft-°F), nonmetallic solids from 0.1 to 50 W/(m–K) (0.06 to 30 Btu/h-ft-°F), and metallic solids from 10 to 500 W/(m–K) (6 to 300 Btu/h-ft-°F). Obviously, gases are among the lowest conductors of thermal energy, but nonmetallic materials such as foamed plastics and glass wool also have low values of thermal conductivity and are used as insulating materials. Metals are the best conductors of thermal energy. There is also a direct correlation between thermal conductivity and electrical conductivity—that is, the materials that have a high thermal conductivity also have a high electrical conductivity. Conductive heat transfer is an important factor to consider in the design of buildings and in the calculation of building energy loads.

CONVECTION

Convective heat transfer occurs when a fluid (gas or liquid) is in contact with a body at a different temperature. As a simple example, consider that you are swimming in water at 21°C (70°F), you observe that your body feels cooler than it would if you were in still air at 21°C (70°F). Also, you have observed that you feel cooler in your automobile when the air-conditioner vent is blowing directly at you than when the air stream is directed away from you. Both of these observations are directly related to convective heat transfer, and we might hypothesize that the rate of energy loss from our body due to this mode of heat transfer is dependent on not only the temperature difference but also the type of surrounding fluid and the velocity of the fluid. We can thus define the unit heat transfer for convection, q/A, as follows:

$$q/A = h(T_i - T_\infty)$$

where q is the heat transfer per unit surface area, A is the surface area, h is the convective heat transfer coefficient W/(m^2 – K), T_i is the temperature of the body, and T_∞ is the temperature of the fluid.

Thus convective heat transfer is a function of the temperature difference between the surface and the fluid; the value of the coefficient, h, depends on the type of fluid surrounding the object, and the velocity of the fluid flowing over the surface. Convective heat transfer is often broken down into two distinct modes: free convection and forced convection. Free convection is normally defined as the heat transfer that occurs in the absence of any external source of velocity—for example, in still air or still water. Forced convection has some external source, such as a pump or a fan, which increases the velocity of the fluid flowing over the surface. Convective heat transfer coefficients range widely in magnitude, from 6W/(m^2 – K) (1 Btu/h-ft^2-°F) for free convection in air, to more than 200,000 W/(m^2 – K) (35,000 Btu/h-ft^2-°F) for pumped liquid sodium. Convective heat transfer plays a very important role in such energy applications as power boilers, where water is boiled to produce high-pressure steam for power generation.

RADIATION

Radiative heat transfer is perhaps the most difficult of the heat transfer mechanisms to understand because so many factors influence this heat transfer mode. Radiative heat transfer does not require a medium through which the heat is transferred, unlike both conduction and convection. The most apparent example of radiative heat transfer is the solar energy we receive from the Sun. The sunlight comes to Earth across 150,000,000 km (93,000,000 miles) through the vacuum of space. Heat transfer by radiation is also not a linear function of temperature, as are both conduction and convection. Radiative energy emission is proportional to the fourth power of the absolute temperature of a body, and radiative heat transfer occurs in proportion to the difference between the fourth power of the absolute temperatures of the two surfaces. In equation form, q/A is defined as:

$$q/A = \sigma\,(T_1^4 - T_2^4)$$

where q is the heat transfer per unit area, A is the surface area, σ is the Stefan-Boltzmann constant 5.67×10^{-8} W/(m^2 – K^4), T_1 is the temperature of surface one, and T_2 is the temperature of surface two.

In this equation we are assuming that all of the energy leaving surface one is received by surface two, and that both surfaces are ideal emitters of radiant energy. The radiative exchange equation has to be modified to account for real situations—that is, where the surfaces are not ideal and for geometrical arrangements in which surfaces do not exchange their energies only with each other. Two factors are usually added to the above radiative exchange equation to account for deviations from ideal conditions. First, if the surfaces are not perfect emitters, a factor called the emissivity is added. The emissivity is a number less than 1, which accounts for the deviation from nonideal emission conditions. Second, a geometrical factor called a shape factor or view factor is needed to account for the fraction of radiation leaving a body that is intercepted by the other body. If we include both of these factors into the radiative exchange equation, it is modified as follows:

$$q/A = F\varepsilon F_G\varepsilon\,(T_1^4 - T_2^4)$$

where $F\varepsilon$ is a factor based on the emissivity of the surface and F_G is a factor based on geometry.

The geometric factor can be illustrated by considering the amount of sunlight (or radiative heat) received by Earth from the Sun. If you draw a huge sphere with a radius of 150 million km (93 million miles) around the sun that passes through Earth, the geometric factor for the Sun to Earth would be the ratio of the area on that sphere's surface blocked by Earth to the surface area of the sphere. Obviously, Earth receives only a tiny fraction of the total energy emitted from the Sun.

Other factors that complicate the radiative heat transfer process involve the characteristics of the surface that is receiving the radiant energy. The surface may reflect, absorb, or transmit the impinging radiant energy. These characteristics are referred to as the reflectivity, absorbtivity, and transmissivity, respectively, and usually are denoted as ρ, α, and τ. Opaque surfaces will not transmit any incoming radiation (they absorb or reflect all of it), but translucent and clear surfaces will transmit some of the incoming radiation. A further complicating factor is that thermal radiation is wavelength-dependent, and one has to know the wavelength (spectral) characteristics of the material to determine how it will behave when thermal radiation is incident on the surface. Glass is typical of a material with wavelength-dependent properties. You have observed that an automobile sitting in hot sunlight reaches a temperature much higher than ambient temperature conditions. The reason is that radiant energy from the Sun strikes the car windows, and the very-short-wavelength radiation readily passes through the glass. The glass, however, has spectrally dependent properties and absorbs almost all the radiant energy emitted by the heated surfaces within the car (at longer wavelengths), effectively trapping it inside the car. Thus the car interior can achieve temperatures exceeding 65°C (150°F). This is the same principle on which a greenhouse works or by which a passively heated house is warmed by solar energy. The glass selectively transmits the radiation from the Sun and traps the longer-wavelength radiation emitted from surfaces inside the home or the greenhouse. Specially made solar glasses and plastics are used to take advantage of the spectral nature of thermal radiation.

R VALUES

Many everyday heat flows, such as those through windows and walls, involve all three heat transfer mechanisms—conduction, convection, and radiation. In these situations, engineers often approximate the calculation of these heat flows using the concept of R values, or resistance to heat flow. The R value combines the effects of all three mechanisms into a single coefficient.

R Value Example: Wall

Consider the simple wall consisting of a single layer of Sheetrock, insulation, and a layer of siding as shown in Figure 1. We assume it is a cool autumn evening when the room air is 21°C (70°F) and the outside air is 0°C (32°F), so heat will flow from inside the wall to the outside. Convective and radiative heat transfer occurs on both the inside and outside surfaces, and conduction occurs through the Sheetrock, insulation, and siding. On the outside, the convective heat transfer is largely due to wind, and on the inside of the wall, the convective heat transfer is a combination of natural and forced convection due to internal fans and blowers from a heating system. The room surfaces are so close to the outside wall temperature that it is natural to expect the radiative heat transfer to be very small, but it actually accounts for more than half of the heat flow at the inside surface of the wall in this situation. When treated using the R value concept, the total R value of the wall is the sum of individual R values due to the inside surface, the Sheetrock, the insulation, the siding, and the outer surface. Hence the heat flow through this wall may be written as

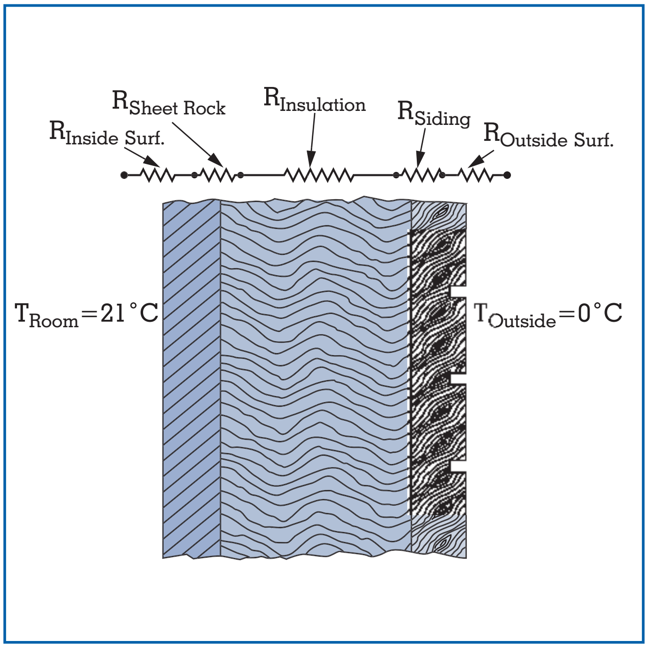


Figure 1.
Schematic diagram of a simple wall.

$$Q = A(T_{\text{room}} - T_{\text{outside}})/(R_{\text{inside surf.}} + R_{\text{Sheetrock}} + R_{\text{insulation}} + R_{\text{siding}} + R_{\text{outside surf.}})$$

The resistances of the Sheetrock, siding, and insulation may be viewed as conductive resistances, while the resistance at the two surfaces combine the effects of convection and radiation between the surface and its surroundings. Typical values for these resistances in units of $(\text{W/m}^2 - \text{W/m}^2\text{-°C})^{-1}$ $((\text{Btu/h-ft}^2\text{-°F})^{-1})$ are as follows:

$R_{\text{inside surf}}$	= 0.12	(0.68)
$R_{\text{Sheetrock}}$	= 0.08	(0.45)
$R_{\text{insulation}}$	= 1.94	(11.0)
R_{siding}	= 0.14	(0.80)
$R_{\text{outside surf}}$	= 0.04	(0.25)
R_{total}	= 2.32	(13.18)

For this case, the heat flow through the wall will be (21–0)°C/2.32 $(\text{W/m}^2\text{-°C})^{-1}$ = 9.05 W/m^2 (2.87

Btu/h-ft^2). While the surfaces, Sheetrock, and siding each impede heat flow, 80 percent of the resistance to heat flow in this wall comes from the insulation. If the insulation is removed, and the cavity is filled with air, the resistance of the gap will be 0.16 (W/m^2-°C)$^{-1}$ (0.9 (Btu/h-ft^2-°F)$^{-1}$) and the total resistance of the wall will drop to 0.54 (W/m^2-°C)$^{-1}$ (3.08 (Btu/h-ft^2-°F)$^{-1}$) resulting in a heat flow of 38.89 W/m^2 (12.99 Btu/h-ft^2). The actual heat flow would probably be somewhat different, because the R-value approach assumes that the specified conditions have persisted long enough that the heat flow is "steady-state," so it is not changing as time goes on. In this example the surface resistance at the outer wall is less than half that at the inner wall, since the resistance value at the outer wall corresponds to a wall exposed to a wind velocity of about 3.6 m/s (8 mph), which substantially lowers the resistance of this surface to heat flow.

If the wall in the example had sunlight shining on it, the heat absorbed on the outer surface of the wall would reduce the flow of heat from inside to outside (or could reverse it, in bright sunshine), even if the temperatures were the same.

R Value Example: Window

A window consisting of a single piece of clear glass can also be treated with R-value analysis. As with the wall, there is convective and radiative heat transfer at the two surfaces and conductive heat transfer through the glass. The resistance of the window is due to the two surface resistances and to the conductive resistance of the glass, R_{glass}. For typical window glass, R_{glass} = 0.003 (W/m^2-°C)$^{-1}$ (0.02 (Btu/h-ft^2-°F)$^{-1}$) so the total resistance of the window is R_{window} = (0.12 + 0.003 + 0.04) (W/m^2-°C)$^{-1}$ = 0.163 (W/m^2-°C)$^{-1}$ (0.95 (Btu/h-ft^2-°F)$^{-1}$). Thus the heat flow will be q = (21 – 0)°C/0.163(W/m^2-°C)$^{-1}$ = 128.8 W/m^2 (40.8 Btu/h-ft^2), or fourteen times as much as that through the insulated wall. It is interesting to note that the heat flow through an ancient window made from a piece of oilskin, or even a "window" made from a piece of computer paper, would not increase by more than 2 percent from that of the glass window, because the resistance of the glass is so small.

When sunlight is shining through a window, the heat transfer becomes more complicated. Consider Figure 2.

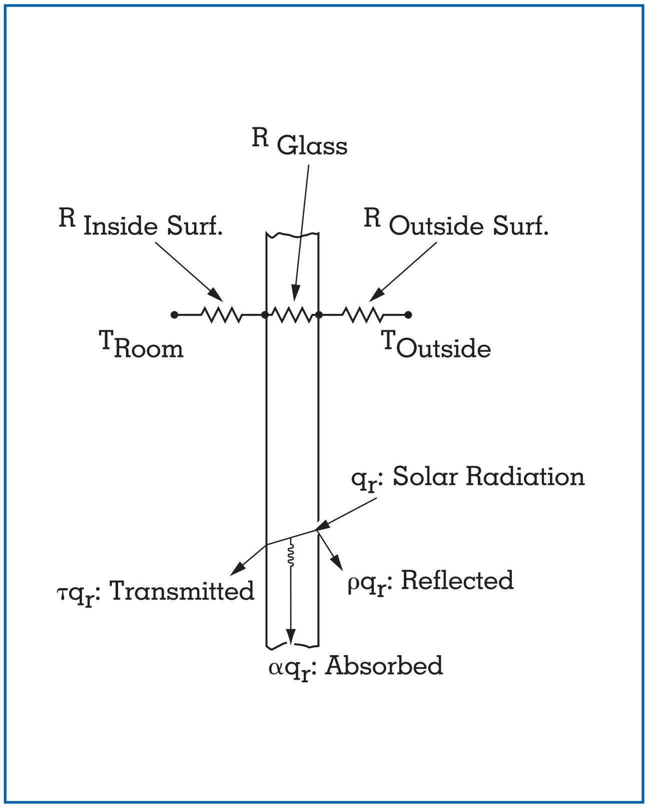


Figure 2.
Schematic diagram of a window.

Suppose the outdoor and indoor temperatures are still 0°C and 21°C, but the window now has 600 W/m^2 (190 Btu/ft^2-h) of sunlight striking it (q_r), typical of a fairly sunny window on an autumn day. About 8 percent (ρq_r) will be reflected back into the atmosphere, about 5 percent (αq_r), or 30 W/m^2, will be absorbed in the glass, and the remaining 87 percent (τq_r), or 522W/m^2, will be transmitted into the room, providing light before it strikes the walls, floor, ceiling, and furnishings in the room. Again, some will be reflected, providing indirect lighting, and the remainder will be absorbed and converted to heat. Eventually, all—except any that may be reflected back out the window—will be absorbed and converted to heat in the room. Thus the amount of sunlight coming through the window is about four times as great as the amount of heat flowing outward. The 30 W/m^2 that is absorbed in the glass heats the glass slightly, reducing the conductive heat flow through the glass to 121 W/m^2, so the net effect of the sunlight is to result in the window providing 401 W/m^2 of heating (522–121) to the room instead of losing 128 W/m^2.

In this example, we have assumed that the outside temperature is lower than inside; therefore the heat flow due to the temperature difference is from inside to outside. In the summer it will be hotter outside, and heat will flow from outside to inside, adding to the heat gained from the sunlight. Because windows represent a large source of heat gain or the heat loss in a building, a number of schemes are used to reduce the heat gains in summer and heat losses in winter. These include the use of double (two glass panes) or triple glazing for windows. The glass is separated by an air space, which serves as an added insulating layer to reduce heat transfer. Reflective films are used to reduce heat gain from sunlight. Some of these films also serve to reduce the heat transfer by longwave (nonsolar) radiation as well.

There are numerous other examples where heat transfer plays an important role in energy-using systems. One is the production of steam in large boilers for power production, where steam is boiled from water. Hot combustion gases transfer heat to the water by radiation, conduction through the pipe walls, and convection. The boiling of the water to produce steam is a special case of convective heat transfer where very high heat transfer coefficients are obtained during the boiling process. Another very practical example of convective cooling is in automobiles. The “radiator” is, in fact, a “convector” where pumped water is forced through cooling tubes and is cooled by air forced over the tubes and fins of the radiator. Forced convective heat transfer occurs at both the air side and the water side of the “radiator.” Cooling towers and air-conditioning coils are other practical examples of combined heat transfer, largely combined conduction and convection.

While heat transfer processes are very useful in the energy field, there are many other industries that rely heavily on heat transfer. The production of chemicals, the cooling of electronic equipment, and food preparation (both freezing and cooking) rely heavily on a thorough knowledge of heat transfer.

W. Dan Turner
David E. Claridge

See also: Air Conditioning; Furnaces and Boilers; Heat and Heating; Heat Pumps; Insulation; Refrigerators and Freezers; Solar Energy; Solar Energy, Historical Evolution of the Use of; Thermodynamics; Windows.

BIBLIOGRAPHY

ASHRAE. (1997). *1997 ASHRAE Handbook: Fundamentals.* Atlanta: Author.

Incropera, F. P., and DeWitt, D. P. (1996). *Fundamentals of Heat and Mass Transfer*, 4th ed. New York: John Wiley & Sons.

Turner, W., ed. (1982). *Energy Management Handbook.* New York: John Wiley & Sons.

HEAVISIDE, OLIVER
(1850–1925)

Oliver Heaviside was born May 13, 1850, in London, England. He was a physicist and mathematician whose theoretical work played a large part in the understanding of radio transmission and long-distance telephony. Educated at Camden House School in London, Heaviside came fifth in the College of Preceptors examination out of five hundred pupils. He left school in 1866 and continued to study. He learned Morse code and studied electricity and languages. In 1868 he went to Denmark and became a telegraph operator, and in 1870 he was appointed Chief Telegraph Operator.

In 1871 Heaviside returned to England to take up a post with the Great Northern Telegraph Company dealing with overseas traffic. In 1875 he had to leave the job due to increasing deafness. Heaviside was encouraged to continue his electrical research by his uncle, Charles Wheatstone, who with W. F. Cooke patented the electric telegraph in 1837, and who later devised the Wheatstone bridge, an electrical network for measuring resistance.

In 1872 Heaviside's first paper, “Comparing of Electromotive Forces,” was published. Heaviside's second paper was published in 1873 and attracted the attention of Scottish physicist James Clerk Maxwell. In 1873 Heaviside was inspired by Maxwell's treatise on electricity and magnetism. It took Heaviside several years to fully understand Maxwell's book, which he then set aside to follow his own course of thinking. Finding Maxwell's conventional mathematics difficult to apply to practical matters, Heaviside introduced a simpler treatment that, in his opinion, did not impair accuracy. It proved to be controversial. His proofs did not conform to the standards of the time and his methods on partial differential equations had varying success.

In 1874 Heaviside established the ordinary symbolic method of analyzing alternating current circuits in common use today. It was a technique developed about fifteen years before AC came into commercial use. He emphasized the role of metallic circuits as guides, rather than conductors of AC currents. He discussed in detail causes of line distortion and suggested ways of alleviating it.

Between 1880 and 1887 Heaviside developed his operational form of calculus, a branch of mathematics that permits the manipulation of varying quantities applicable to practical problems, including such matters as electrical circuits with varying voltages and currents. Heaviside received a great honor in 1891 when he was elected a Fellow of The Royal Society in recognition of his work on electromagnetic waves. He was later the first recipient of the Society's Faraday Medal. The first volume of his great work "Electromagnetic Theory" was published in 1893 and a second volume in 1899. A third volume was published in 1912, but he died before completing the fourth.

In 1902 Heaviside's famous prediction about an ionized layer in the atmosphere that would deflect radio waves was published in an article titled "Telegraphy," in the tenth edition of "Encyclopaedia Britannica." The idea came when he was considering the analogy between the movement of electric waves along a pair of conducting wires and over a conducting earth. Discussing the possibility of radio waves being guided around a curved path he suggested: "There may possibly be a sufficiently conducting layer in the upper air. If so, the waves will, so to speak, catch on to it more or less. Then guidance will be by the sea on one side and the upper layer on the other." The layer was first named the Heaviside Layer and later the Kennelly-Heaviside Layer, as a similar prediction had been made around the same time by Arthur Kennelly at Harvard University. The hypothesis was proved correct in 1924 when radio waves received from the upper atmosphere showed that deflection of upward waves took place at a height of approximately 100 kilometers.

Heaviside made a significant contribution to electrical communications when he advocated the introduction of additional inductance in long-distance telephony cables although there was then no practical means to add it. His idea was eventually patented in 1904 by Michael Campbell of AT&T after Heaviside and George Pupin of Columbia University had shown it was possible to apply inductance in the form of uniformly spaced loading coils. By 1920 engineers had installed such loading on thousands of miles of cable, particularly in the United States.

During his lifetime, Heaviside made extensive contributions to pure mathematics and its practical applications to alternating current, vector analysis and telegraphy. He introduced new concepts that later became commonplace thinking and expressions for every electrical engineer, and invented much of the language now basic to communication engineering, including such words as "capacitance," "inductance," "impedance," and "attenuation."

Heaviside's last years were spent as an embittered recluse at Torquay, Devon where he allowed only a few people to visit him. For much of his life he suffered from recurring jaundice that was to prove fatal. He died on February 3, 1925. Heaviside's work has been an inspiration to countless electrical engineers and mathematicians. Time has enhanced the esteem in which he is held; succeeding generations have spent many hours studying his writings. As a lasting honor, craters on Mars and the Earth's moon were named after him.

Alan S. Heather

See also: Wheatstone, Charles.

BIBLIOGRAPHY

Institution of Electrical Engineers. (1950). *The Heaviside Centenary Volume*. London: Author.

Josephs, H. J. (1963). *Oliver Heaviside: A Biography*. London: Author.

Nahin, P. (1988). *Oliver Heaviside: Sage in Solitude: The Life, Work, and Times of an Electrical Genius of the Victorian Age*. New York: I.E.E.E.

Searle, G. F. C. (1987). *Oliver Heaviside, the Man*. St Albans: CAM Publishing.

Yavetz, I. (1995). *From Obscurity to Enigma: The Work of Oliver Heaviside, 1872–1889*. Basel: Birkhauser Verlag.

HELMHOLTZ, HERMANN VON (1821–1894)

One of the most versatile scientists who ever lived, Hermann von Helmholtz was born in Potsdam, Germany, in 1821, the son of a "Gymnasium" (high school) teacher. From an early age he wanted to be a physicist, but his family could not afford the money required for his education. Instead, his father persuaded him to take up medicine, since his son's education as a physician would be subsidized by the state on the condition that he serve as a doctor in the Prussian army after he received his degree. Helmholtz attended the Institute for Medicine and Surgery in Berlin from 1838 to 1842 and fulfilled his obligation as an army surgeon from 1843 to 1848. His real interest, however, was always research; even in the army barracks he set up a small laboratory for research in physiology and physics.

On July 23, 1847, Helmholtz presented a paper on the conservation of energy at a meeting of the Berlin Physical Society. It was a talk at a reasonably high level of mathematical sophistication intended to convince physicists that energy was always conserved in any physical process. Although it was rejected by J. C. Poggendorf, the editor of the *Annalen der Physik*, as being too long and too mathematical for his readers, it eventually appeared in pamphlet form and was soon recognized as one of the most important papers in nineteenth-century science. This bold and path-breaking paper, written when he was only twenty-six, was Helmholtz's first and most fundamental statement of the conservation-of-energy principle. It came at a critical moment in the history of science when scientists and philosophers were waging a battle over whether conservation of energy was a truly universal principle. In his 1847 paper Helmholtz had shown convincingly that it was.

Helmholtz stated that his only purpose in his 1847 paper was to provide a careful investigation and arrangement of accepted facts about energy conservation for the benefit of both the physicists and the physiologists who attended his lecture. He never claimed priority for himself in this discovery, but later conceded that honor to J. R. Mayer and James Joule. When Helmholtz delivered this paper, however, he knew little about Joule's research, which was going on at about the same time, and nothing at all about Robert Mayer's 1842 paper, which had appeared in a journal not normally read by many scientists.

Some years later, in 1861, when Helmholtz was much involved in physiology, he realized that he had discussed physiology in only two paragraphs of his 1847 paper, and so supplemented it by a fuller consideration of energy transformations in organic systems. This he included in a talk before the Royal Society of London on April 12, 1861. Here he generously referred to the important contributions of Sadi Carnot, Mayer and Joule in establishing the principle of conservation of energy on a firm foundation.

On the basis of this work and other research, in 1849 Helmholtz was appointed professor of physiology at the University of Koönigsberg. There he devoted himself to the physiology of the eye, first explaining the mechanism of lens accommodation. In 1851 his invention of the ophthalmoscope, still the basic instrument used by eye doctors to peer at the retina of the eye, immediately made Helmholtz famous. In 1852 he also became the first experimentalist to measure the speed of nerve impulses in the human body.

Helmholtz did important research on another sense organ, the ear, and explained how it was able to detect differences in pitch. He showed how the quality of a sound depended on the number, nature and relative intensities of the harmonics present in the sound.

On the basis of his studies on the eye and ear, in 1858 Helmholtz was appointed Professor of Anatomy in Heidelberg. His thirteen years in Heidelberg gave him the opportunity to work closely with other gifted scientists including physicist Gustav Kirchhoff and chemist Robert Bunsen, in what was then called "an era of brilliance such as has seldom existed for any university and will not readily be seen again." In 1871 Helmholtz abandoned anatomy and physiology in favor of physics. He accepted the most prestigious chair of physics in Germany at the University of Berlin and spent the rest of his life there.

As he grew older, Helmholtz became more and more interested in the mathematical side of physics and made noteworthy theoretical contributions to classical mechanics, fluid mechanics, thermodynamics and electrodynamics. He devoted the last decade of his life to an attempt to unify all of physics under one fundamental principle, the principle of least action. This attempt, while evidence of Helmholtz's philosphical bent, was no more successful than was Albert Einstein's later quest for a unified field theory. Helmholtz died in 1894 as the result of a fall suffered on board ship while on his way back to Germany from the United States, after representing Germany at the Electrical Congress in Chicago in August, 1893.

It is difficult to exaggerate the influence Helmholtz had on nineteenth-century science, not only in Germany but throughout the world. It was during his lifetime that Germany gained its preeminence in science, which it was not to lose until World War II. His own research contributions, together with the impetus he gave to talented students by his teaching, research guidance, and popular lectures, had much to do with the scientific renaissance that Germany experienced during his lifetime.

Helmholtz was a sensitive and sickly man all his life, plagued by severe migraine headaches and fainting spells. He sought relief from his pain in music and the other arts, and in mountain climbing in the Alps. It is intriguing to imagine what he might have accomplished had he been in good health for all his seventy-three years. On the occasion of Helmholtz's death, Lord Kelvin stated: "In the historical record of science the name of Helmholtz stands unique in grandeur, as a master and leader in mathematics, and in biology, and in physics." Many scientists felt that Kelvin had shortchanged Helmholtz by this statement; Medicine, physiology, chemistry, and philosophy certainly deserved to be added to the list of fields Helmholtz had mastered during his long and productive scientific career.

Joseph F. Mulligan

BIBLIOGRAPHY

Cahan, D., ed. (1993). *Hermann von Helmholtz and the Foundations of Nineteenth-Century Science.* Berkeley, California: University of California Press.

Cahan, D., ed. (1995). *Science and Culture. Popular and Philosophical Essays of Hermann von Helmholtz.* Chicago: University of Chicago Press.

Helmholtz, H. v. (1847). "The Conservation of Force: A Physical Memoir." In *Selected Writings of Hermann von Helmholtz* (1971), ed. R. Kahl, pp. 3–55. Middletown, CT: Wesleyan University Press.

Helmholtz, H. v. (1861). "The Application of the Law of Conservation of Force to Organic Nature." In *Selected Writings of Hermann von Helmholtz* (1971), ed. R. Kahl, pp. 109–121. Middletown, CT: Wesleyan University Press.

Kahl, R., ed. (1971). *Selected Writings of Hermann von Helmholtz.* Middletown, CT: Wesleyan University Press.

Koenigsberger, L. (1965). *Hermann von Helmholtz* (One-volume condensation of three-volume 1902 German edition), tr. Frances A. Welby. New York: Dover Publications.

Mulligan, J. F. (1989). "Hermann von Helmholtz and His Students." *American Journal of Physics* 57:68–74.

Turner, R. S. (1972). "Helmholtz, Hermann von." In *Dictionary of Scientific Biography,* ed. Charles Coulston Gillispie, Vol. 6, pp. 241–253. New York: Scribner.

HERTZ, HEINRICH RUDOLF (1857–1894)

Heinrich Hertz was born into a well-to-do German family in Hamburg on February 22, 1857. He was an exceptionally talented student, doing equally well in the humanities and sciences. He read extensively, tried his hand at sculpting, and even built scientific apparatus on a lathe at home. After a year of military service, Hertz studied structural engineering in Munich, but gave it up for physics when he realized he had the ability to contribute something substantial to that field. At the age of 20 he wrote to his parents: "… I would rather be a great scientific investigator than a great engineer, but I would rather be a second-rate engineer than a second-rate investigator."

In 1878 Hertz enrolled at the University of Berlin to study under Hermann von Helmholtz, the leading German physicist of the time. He obtained his degree *magna cum laude* in 1880 with a theoretical dissertation on the electromagnetic induction of currents in

Heinrich Rudolf Hertz. (Library of Congress)

conducting spheres rotating in a magnetic field. He remained at the Berlin Physics Institute as assistant to Helmholtz until 1883; during these years he published fifteen research papers on a great variety of topics in physics.

In 1883 Hertz was appointed *Privatdozent* for mathematical physics at Kiel, and after two years became a full professor at the *Technische Hochschule* in Karlsruhe. In 1889 Hertz left Karlsruhe to assume his last academic post as Professor of Physics at the Friedrich-Wilhelm University in Bonn. Five years later, following a long period of declining health and many painful operations, Heinrich Hertz died in Bonn of blood poisoning on January 1, 1894, a few months before his thirty-seventh birthday.

After Hertz's death Helmholtz paid tribute to his former student as a "consummate physicist," who uniquely combined mathematical ability, theoretical insight, and experimental skill. These qualities enabled Hertz to make many important contributions to physics, of which only those relating more directly to energy are outlined below.

Hertz's most direct involvement with all aspects of the energy question was the research he did to prepare his inaugural lecture to the faculty at the *Technische Hochschule* in Karlsruhe, delivered on April 20, 1885 and entitled: "On the Energy Balance of the Earth." The manuscript for this lecture has only recently been found and published in both German and English. Although written more than a century ago, this impressive document records both Hertz's insightful view of the Earth's energy situation at that time and his remarkably good order-of-magnitude estimates of the energy sources then known to be available to the Earth.

The most important contribution Hertz made in this inaugural lecture was his prediction, based on his estimates of the energy sources available, that ultimately the Earth was completely dependent on the Sun for the light and heat it needed to support life. Of course, this picture would change after Henri Becquerel discovered radioactivity in 1896, and thus introduced the nuclear age of physics.

The research that brought Hertz undying fame as a physicist was that on electromagnetic waves, performed in 1886–1889 in Karlsruhe. By his elegant experiments he confirmed the theoretical prediction of James Clerk Maxwell that electromagnetic waves in what are now called the microwave and radiowave regions of the spectrum travel through a vacuum at the speed of light. He also demonstrated that microwaves of 66-cm wavelength exhibit the same properties of reflection, refraction, interference and polarization as do light waves. Hertz's research also provided conclusive evidence that electromagnetic energy cannot be transmitted from place to place instantaneously, but only at a finite velocity, that of light. Hertz never considered the possibility of using electromagnetic waves for wireless communication over long distances. His sole interest was in understanding the world about him — "the intellectual mastery of nature," in the words of Helmholtz.

In the course of his research on electromagnetic waves Hertz discovered the photoelectric effect. He showed that for the metals he used as targets, incident radiation in the ultraviolet was required to release negative charges from the metal. Research by Philipp Lenard, Wilhelm Hallwachs, J. J. Thomson, and other physicists finally led Albert Einstein to his famous 1905 equation for the photoelectric effect, which includes the idea that electromagnetic energy is "quantized" in units of hν, where h is Planck's constant and ν is the frequency of the "bundle of energy" or "photon." Einstein's equation is simply a conservation-of-energy equation, stating that the energy of the incident photon (hν) is used partially to provide the energy needed to extract the negative particle (electron) from the metal, with the rest of the photon's energy going into the kinetic energy of the extracted electron. Hertz's discovery of the photoelectric effect in 1887 therefore led eventually to Einstein's energy-conservation equation for submicroscopic systems.

Hertz's last piece of experimental work was done when his health was deteriorating and he was devoting most of his research time to intensive theoretical work on the logical foundations of mechanics. In 1892 in his laboratory in Bonn, he discovered that cathode rays could pass through thin metallic foils. He published a short paper on the subject, but did not pursue the matter further. Instead he handed his apparatus and his ideas over to Philipp Lenard (1862–1947), his assistant in Bonn. Lenard pushed Hertz's suggested research so far that Lenard received the Nobel Prize in Physics in 1905 "for his work on cathode rays."

During his brief life Hertz moved back and forth with extraordinary ease and great dedication between intense theoretical study at his desk and equally demanding experimental work in his laboratory. As Robert S. Cohen wrote of Hertz in 1956, "His like is rare enough within science ... but his fusion of theory and experiment with a creative interest in philosophical and logical foundations [as revealed particularly in his *Principles of Mechanics*] is nearly unique."

Joseph F. Mulligan

BIBLIOGRAPHY

Fölsing, A. (1997). *Heinrich Hertz. Eine Biographie.* Hamburg: Hoffmann und Campe.

Hertz, H. G. and Mulligan, J. F., eds. (1998). "Der Energiehaushalt der Erde," von Heinrich Hertz, *Fridericiana* (Zeitschrift der Universität Karlsruhe), Heft 54:3–15.

McCormmach, R. (1972). "Hertz, Heinrich Rudolf." In *Dictionary of Scientific Biography,* ed. Charles Coulston Gillispie, Vol. 6, 340–350. New York: Scribner.

Mulligan, J. F., ed. (1994). *Heinrich Rudolf Hertz: A Collection of Articles and Addresses*; with an Introductory Biography by the editor. New York: Garland.

Mulligan, J. F., and Hertz, H. Gerhard. (1997). "An Unpublished Lecture by Heinrich Hertz: 'On the Energy Balance of the Earth.'" *American Journal of Physics* 65: 36–45.

Planck, M. (1894). *Heinrich Rudolf Hertz: A Memorial Address.* English translation in *Heinrich Rudolf Hertz: A Collection of Articles and Addresses* (1994), ed. J. F. Mulligan, pp. 383–403. New York: Garland.

Susskind, C. (1995). *Heinrich Hertz. A Short Life.* San Francisco: San Francisco Press, Inc.

HIGH OCCUPANCY VEHICLE LANES

See: Traffic Flow Management

HISTORICAL PERSPECTIVES AND SOCIAL CONSEQUENCES

As with so many other phenomena that define our civilization, we owe the idea of energy—from *en* (in) and *ergon* (work)—to ancient Greeks. In his *Metaphysics*, Aristotle gave the term a primarily kinetic meaning: "The term "actuality" (*energeia*) … has been extended to other things from motions, where it was mostly used; for *actuality* is thought to be motion most of all" (*Metaphysics*, Theta 3, p. 149).

For Greeks the word and its cognate terms filled a much larger conceptual niche than they do in modern scientific usage. In some of Aristotle's writings *energeia* stands in opposition to mere disposition, *hexis*; in others it carries the vigor of the style. The verb *energein* meant to be in action, implying constant motion, work, production, and change. The classical concept of *energeia* was thus a philosophical generalization, an intuitive expression embracing the totality of transitory processes, the shift from the potential to the actual. Although the perception was clearly holistic, it did not embrace the modern notion of the underlying commonality of diverse energies, the fact that their conversions can perform useful work.

This understanding was clearly formulated only by the middle of the nineteenth century; conceptualization of energy thus made hardly any advances during more than two thousand years following Aristotle's writings. Interestingly, even many founders of modern science held some dubious notions concerning energy. To Galileo Galilei, heat was an illusion of senses, an outcome of mental alchemies; Francis Bacon thought that heat could not generate motion or motion, heat.

But neither the absence of any unified understanding of the phenomenon nor the prevalence of erroneous interpretations of various energy conversions prevented a great deal of empirical progress in harnessing diverse energies and in gradually improving efficiencies of some of their conversions. Seen from a biophysical point of view, all human activities are nothing but a series of energy conversions, and so it is inevitable that different energy sources and changing conversion techniques have imposed obvious limits on the scope of our action—or opened up new possibilities for human development.

PREHISTORIC AND ANCIENT CULTURES

From the perspective of general energetics, the long span of human prehistoric development can be seen as the quest for a more efficient use of somatic energy, the muscular exertions used primarily to secure a basic food supply and then to gradually improve shelters, acquire more material possessions, and evolve a variety of cultural expressions. This quest was always limited by fundamental bioenergetic considerations: Fifty to ninety watts is the limit of useful work that healthy adults can sustain for prolonged periods of time (of course, short bursts of effort could reach hundreds of watts).

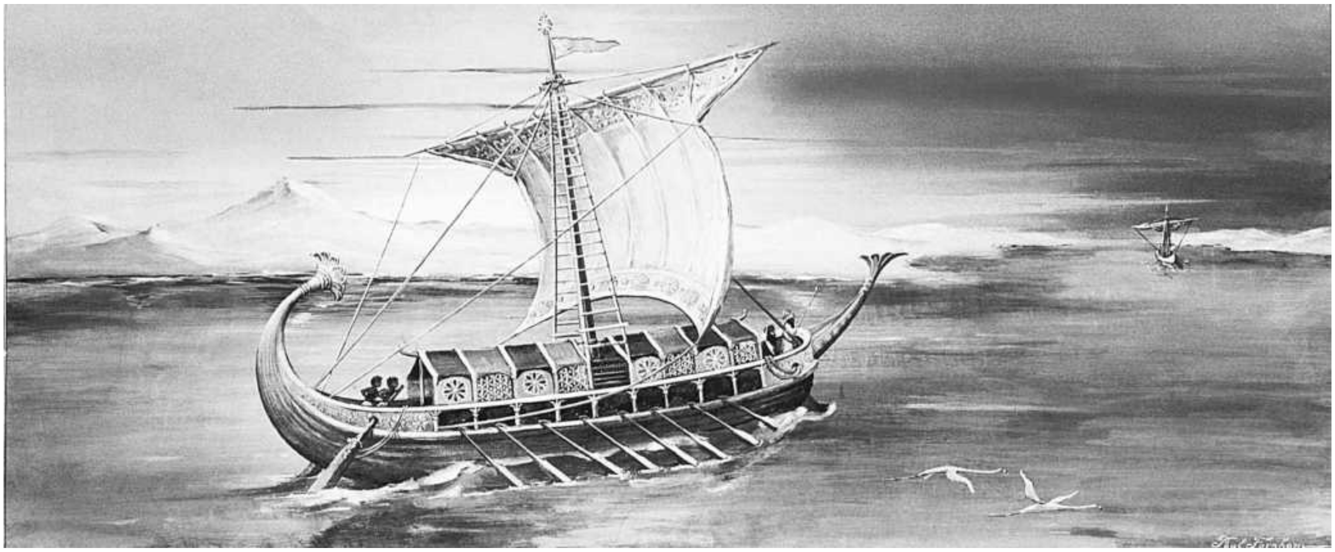
Human labor dominated all subsistence foraging activities, as the food acquired by gathering and hunting sufficed merely to maintain the essential metabolic functions and to support very slow population growth. Societies not very different from this ancestral archetype survived in some parts of the world (South Africa, Australia) well into the twentieth century: Because they commanded very little energy beyond their subsistence food needs, they had very few material possessions and no permanent abodes.

Simple wooden, stone, and leather tools—including digging sticks, bows, arrows, spears, knives and scrapers—were used to increase and extend the inherently limited musclepower in collecting, hunting, and processing tasks. The only extrasomatic energy conversion mastered by foraging societies was the use of fire for warmth and cooking. The earliest dates for controlled use of fire for warmth and cooking remain arguable: it may have been nearly 500,000 years ago, but a more certain time is about 250,000 years ago.

Energy returns in more diversified foraging (energy in food/energy spent in collecting and hunting) varied widely: they were barely positive for some types of hunting (particularly for small arboreal animals), high for gathering tubers (up to fortyfold), and very high for coastal collecting and hunting of marine species ranging from shellfish to whales. Some foraging societies able to secure high energy returns built permanent dwellings and also channeled the surplus energies into more elaborate tools and remarkable artistic expressions.

Energy returns actually declined when some foragers began practicing shifting cultivation, and they declined further with sedentary agriculture—but these modes of subsistence were gradually adopted because they made it possible to support much larger population densities. Carrying capacities of foraging societies were as low as 0.01 person per square kilometer in arid regions and almost as high as 1 person per square kilometer in species-rich coastal sites. In contrast, shifting agricultures would commonly support 20 to 30 people per square kilometer, and even the earliest, extensive forms of settled farming could feed 100 to 200 people per square kilometer (or one to two people from a hectare of cultivated land). Shifting agriculturalists extended the use of fire to the removal of forest vegetation and acquired a larger assortment of tools.

Human labor—with varying contributions by slave, corvee, and free work—continued to be the dominant source of mechanical energy throughout antiquity, and combustion of biomass fuels remained the most important conversion of extrasomatic energy in all ancient civilizations. Indeed, these two energy sources remained critically important until the creation of modern industrial civilization. But, in contrast to the prehistoric era, energetic perspectives on human history reveal a fascinating quest for harnessing extrasomatic energies combined with ingenious efforts to increase the efficiency of available energy conversions. Both the collective achievements



Painting of a Phoenician merchant galley, c. 7th century B.C.E., crossing the Red Sea. (Corbis Corporation)

of human societies and individual standards of living have always been highly correlated with success in diversifying energy sources and increasing efficiencies of their use.

With more ingenious tools (based on such fundamental mechanical principles as the lever and the inclined plane) and with thoughtful organization of often very complex tasks, societies commanding just musclepower were able to accomplish such feats as the building of admirably planned megalithic structures (such as Stonehenge and the colossi of Easter Island), pyramids (Egypt, Mesoamerica), and stone temples, walls, and fortresses on four continents. In contrast to this concerted deployment of human labor in construction—amounting to one to five kilowatts of sustained power in work gangs of ten to fifty people—no Old High culture took steps to the really large-scale manufacture of goods.

The use of fire was extended during antiquity to produce bricks and to smelt various metals, starting with copper (before 4000 B.C.E.) and progressing relatively rapidly to iron (the metal was used extensively in parts of the Old World after 1400 B.C.E.). Conversion of wood to charcoal introduced a fuel of higher energy density (about 50% higher than air-dried wood) and a superior quality. Being virtually pure carbon, charcoal was nearly smokeless, and hence much better suited for interior heating, and its combustion could produce high temperatures needed for smelting metal ores. But the inefficient method of charcoal production meant that a unit of charcoal's energy required commonly five to six units of wood.

ENERGY CONVERSIONS IN TRADITIONAL SOCIETIES

All ancient societies were eventually able to harness a variety of extrasomatic energies other than the combustion of biomass fuels and the use of pack animals for transport. Capturing wind by simple sails made seagoing transport much cheaper than moving goods or people on land, and it enabled many societies (most notably the Phoenicians and the Greeks, and later, of course, the Romans) to extend their cultural reach far beyond their core regions. However, simple square or rectangular sails were rather inefficient airfoils unsuitable for sailing in the open ocean. Transoceanic voyages had to wait for many centuries not only for the adoption of better sails but also for better keels, rudders, and compasses.

Cattle were the first animals used as sources of draft power in agriculture, and the coincidence of the first clearly documented cases of cattle domestication and plow farming are hardly surprising. Cattle were used in many regions also for lifting irrigation water and for processing harvested crops, and everywhere for transportation. Individual small animals could deliver no more than 100 to 300 watts of useful power, and a pair of good oxen (harnessed by head or neck yokes) could work at a sustained rate of 600 to 800 watts. A peasant working with a hoe would need 100 to 200 hours to prepare a hectare of land for planting cereals. Even with a simple wooden plow pulled by a single medium-size ox, that task could be done in just over 30 hours. Hoe-dependent farming could have never attained the scale of cultivation made possible by domesticated draft animals.

Because of their size and anatomy, horses are inherently more powerful draft animals than cattle: A good horse can work at a rate of 500 to 800 watts, an equivalent of eight to ten men. But this superiority was not translated into actual performance until the invention and general adoption of an efficient harness. The oldest preserved images of working horses do not show them laboring in fields, but rather pulling light ceremonial or attack carriages while using a throat-and-girth harness that is not suited for heavy fieldwork. The breastband harness was better, but only the invention of the collar harness (in China, just before the beginning of the common era) turned horses into superior draft animals.

By the ninth century an improved version of the collar harness reached Europe, where its use remained largely unchanged until horses were replaced by machines more than 700 years later. Millions of working horses worldwide use it still. Two other improvements that made horses into superior draft animals were the diffusion of horseshoes, and better feeding provided by cultivation of grain feeds. A pair of good horses could sustain one and a half kilowatts for many hours, and peak exertions of individual large animals surpassed two kilowatts. Horse teams of the late nineteenth century (with four to forty animals) were powerful enough to open up heavy grassland soils for pioneer farming and to pull the first mechanical harvesters and, later, the first modern combines.

Harnessing of wind by simple sails aside, the diversification of inanimate energy sources began about 2,000 years ago, and their slowly expanding contributions were limited by low efficiencies of prevailing conversion techniques as well as by their cost. Roman civilization, Islam, dynastic China, and premodern Europe came up with many ingenious solutions to convert kinetic energies of wind and water into useful work, easing many everyday tasks as well as allowing for previously impossible mechanical feats.

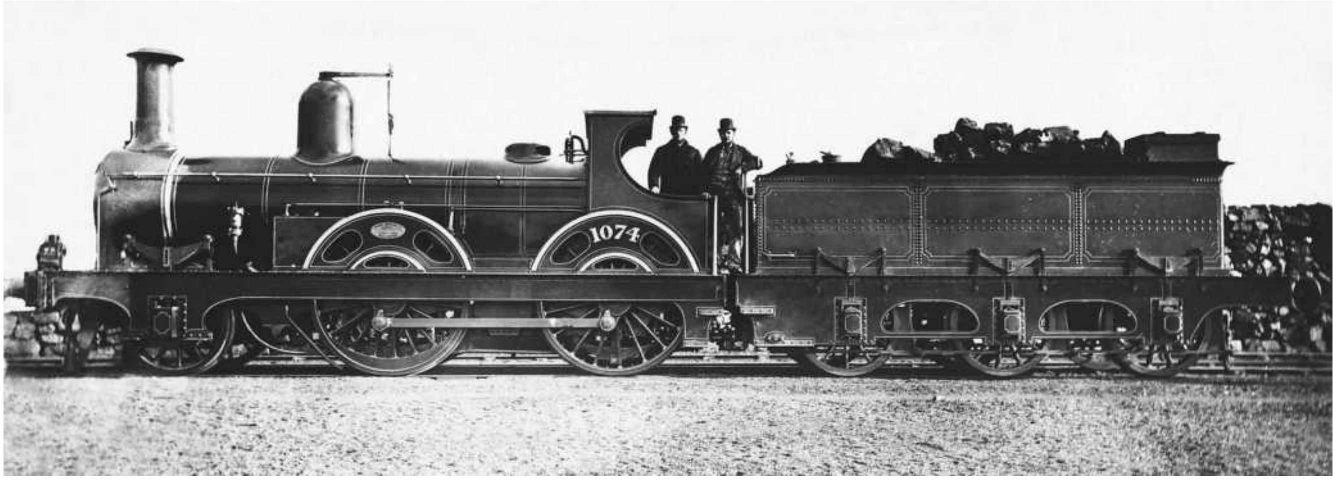
The first inanimate source of energy harnessed by a machine was flowing water. The origin of the earliest waterwheels remains uncertain. Vertical wheels, favored by the Romans, were much more efficient than the horizontal ones, and their first use was turning the millstones by right-angle gears. Use of waterwheels gradually expanded far beyond grain milling, and their power was used for sawing, wood turning, oil pressing, paper making, cloth fulling, tanning, ore crushing, iron making, wire pulling, stamping, cutting, metal grinding, blacksmithing, wood and metal burning, majolica glazing, and polishing.

Most of these tasks were eventually also done by windmills, whose diffusion began about a millennium after the wider use of waterwheels (the first clear European records come from the last decades of the twelfth century). While simple Asian machines had horizontally mounted sails, European mills were vertically mounted rotaries whose driving shafts could be turned into the wind. Useful power of both waterwheels and windmills continued to increase only slowly: Even during the early decades of the eighteenth century, European waterwheels averaged less than four kilowatts, and a typical large eighteenth-century Dutch mill with a thirty-meter span could develop seven and a half kilowatts (an equivalent of ten good horses).

Per capita consumption of biomass fuels in these premodern societies was commonly well below twenty gigajoules a year (an equivalent of less than one metric ton of wood). Radical changes came only with the widespread use of coal (and coke) and with the adoption of steam engines for many stationary and mobile uses. These changes were gradual and decidedly evolutionary: The notion of the eighteenth-century coal-based Industrial Revolution is historically inaccurate. Both in Europe and in North America it was waterpower, rather than coal combustion, that was the prime mover of rapidly expanding textile and other manufacturing industries, and in many countries charcoal-based smelting dominated iron production until the last decades of the nineteenth century.

EMERGENCE OF FOSSIL-FUELED ECONOMIES

In parts of Europe and northern China coal was mined, and used directly as household fuel and in small-scale manufacturing, for centuries preceding the first commercial use of Newcomen's inefficient



Two men stand on the Victoria Express Engine Series 1070, which was designed by Matthew Kirtly, the locomotive superintendent of Midland Railway in England during the early 1870s. (Corbis Corporation)

steam engine in the early decades of the eighteenth century. After 1770 James Watt transformed the existing steam engine into a prime mover of unprecedented power: Although his improved design was still rather inefficient, average capacity of steam engines built by Watt's company was about twenty kilowatts, more than five times higher than the mean for typical contemporary watermills, nearly three times larger than that for windmills, and twenty-five times that of a good horse.

During the nineteenth century the size, and the efficiency, of steam engines rose rapidly. The largest stationary machines designed during the 1890s were about thirty times more powerful (about one megawatt) than those in 1800, and the efficiency of the best units was ten times higher than during Watt's time. Railways and steamships greatly expanded and speeded up long-distance transport and international trade.

As the twentieth century began, two new prime movers were greatly extending the power of fossil-fueled civilization. Internal-combustion engines (Otto and Diesel varieties), developed and perfected by a number of French and German engineers between 1860 and 1900, opened the possibilities of unprecedented personal mobility, first when installed in cars, trucks, and buses, and later when used to propel the first airplanes. The steam turbine, invented by Charles Parsons, patented in 1884 and then rapidly commercialized by his company, made large-scale generation of electricity affordable. The closing decades of the nineteenth century also saw the emergence of a new fossil fuel industry: exploration, drilling, transportation (by pipelines and tankers), and refining of crude oil.

But at that time it was only North America and most of Europe that were in the midst of a radical shift from the combination of animate energies and low-power inanimate prime movers, to the dominance of fossil fuels converted by new, much more powerful, and much more efficient machines. Average annual per capita energy consumption had more than doubled, compared to the pre–fossil fuel era, with most of the fuel spent on industrial production and transportation of people and goods. Acting via industrialization and urbanization, growing rates of energy consumption were reflected in a higher standard of living.

In the early stages of economic growth, these benefits were limited because the fossil fuels were overwhelmingly channeled into building up an industrial base. Slowly increasing consumption of household goods—better cookware, clothes and furniture—were the first signs of improvement. Afterward, with better food supply and improving health care, a lower infant mortality and longer life expectancy.

Eventually, basic material and health advantages started spilling into the countryside. The educational levels of urban populations began to rise, and there were increasing signs of incipient affluence. The emergence of electricity as the most versatile and most convenient form of energy greatly aided the whole process of modernization.

These socioeconomic transformations were accompanied by a deepening understanding of basic principles underlying energy conversions. Practical advances in new machine design as well as new observations in the chemistry of plants and animals were behind some of the most fundamental theoretical breakthroughs of the late eighteenth and the nineteenth centuries that eventually created a unified understanding of energy.

THEORETICAL UNDERSTANDING OF ENERGY

James Watt's invention of a simple indicator, a recording steam gauge, opened the way for detailed studies of engine cycles that contributed immeasurably to the emergence of thermodynamics during the following century. During the 1820s Sadi Carnot used a purely theoretical approach to investigate the ways of producing kinetic energy from heat to set down the principles applicable to all imaginable heat engines, regardless of their working substance and their method of operation.

Not long afterward Justus von Liebig ascribed the generation of carbon dioxide and water to food oxidation, offering a basically correct view of human and animal metabolism. During the 1840s Julius Robert Mayer, a German physician, was the first researcher to note the equivalence of food intake and oxygen consumption. Mayer saw muscles as heat engines energized by oxidation of the blood and offered calculations proving the sufficiency of food's chemical energy to supply the mechanical energy necessary for work as well as to maintain constant body temperature. All of this led him to estimate that mammals are about 20 percent efficient as machines and to establish the mechanical equivalent of heat, and thus to formulate, in 1851, the law of conservation of energy, commonly known as the first law of thermodynamics.

Independent, and more accurate, quantification of the equivalence of work and heat came from an English physicist, James Prescott Joule, whose first publication in 1850 put the conversion rate at 838 foot-pounds and a later revision at 772 foot-pounds, a difference of less than 1 percent from the actual value. Soon afterward William Thomson (Lord Kelvin) identified the sun as the principal source of kinetic energy available to man and wrote about nature's universal tendency toward the dissipation of mechanical energy. Rudolf Clausius sharpened this understanding by showing in 1867 that heat energy at low temperature is the outcome of these dissipations; he named the transformational content entropy, a term derived from the Greek *trope*, for transformation. As the energy content of the universe is fixed, but its distribution is uneven, its conversions seek uniform distribution, and the entropy of the universe tends to maximum.

This second law of thermodynamics—the universal tendency toward heat death and disorder—became perhaps the most influential, and frequently misunderstood, cosmic generalization. Only at the absolute zero (-273°C) is the entropy nil. This third law completes the set of thermodynamic fundamentals. Josiah Willard Gibbs applied the thermodynamic concepts to chemistry and introduced the important notion of free energy. This energy actually available for doing work is determined by subtracting the product of temperature and entropy change from the total energy entering a chemical reaction.

The second law exercised a powerful influence on scientists thinking about energetic foundations of civilization during the closing decades of the nineteenth century. Edward Sacher viewed economies as systems for winning the greatest possible amount of energy from nature and tried to correlate stages of cultural progress with per capita access to fuels. The contrast of rising fuel demands and inexorable thermodynamic losses led to anxious calls for energy conservation. Wilhelm Ostwald, the 1909 Nobel laureate in chemistry, formulated his energetic imperative, admonishing to waste no energy but to value it as mankind makes the inevitable transition to a permanent economy based on solar radiation. Another Nobel laureate, Frederick Soddy (in 1921, in chemistry), investigated the magnitude of the earth's natural resources of energy and was the first scientist to make the often-quoted distinction between utilizing natural energy flows (spending the interest on a legacy) and fossil fuels (spending the legacy itself).

The twentieth century brought a fundamental extension of the first law, with Albert Einstein's follow-up of his famous special-relativity paper published in 1905. Soon after his publication Einstein, writing to a friend, realized that the principle of relativity requires that the mass of a body is a direct measure of its energy content, which means that light transfers mass. During the next two years Einstein formalized this “amusing and attractive thought” in a series of papers firmly establishing the equivalence of mass and energy. In the last of these papers, in 1907, he described a system behaving like a material point with mass

$$M_o = \mu + E_o c^2$$

and noted that this result is of extraordinary importance because the inertial mass of a physical system is equivalent with an energy content μc^2.

HIGH-ENERGY CIVILIZATION

The central role played by abundant, affordable, and varied energy conversions in sustaining and improving modern civilization is thus a matter of indisputable scientific understanding. That the public often ignores or overlooks this role is, in a way, a great compliment to the success of modern science and engineering, which have managed to make the conversions of extrasomatic energies so ubiquitous and so inexpensive that most people hardly give much thought to the centrality of fossil fuels and electricity in our civilization.

A variety of coals, crude oils, and natural gases supplies the bulk of the world’s primary energy, and these flows of fossil energies are complemented by primary electricity derived from flowing water, nuclear fission, wind, and solar radiation. Increasing shares of fossil fuels have been used indirectly, after being converted to electricity, the most versatile form of energy. While most of our prime movers—internal-combustion engines, steam and gas turbines, and electric motors—reached their growth plateaus in terms of their unit power, their overall capacity, and their efficiency, keep increasing.

Most of the citizens of affluent economies have been enjoying numerous benefits resulting from these large, and more efficient, energy flows for several generations, but the successive waves of the energy-driven revolution are only now diffusing throughout low-income countries. Although the process has been uneven in its spatial and temporal progress, there is no doubt that we are witnessing the emergence of a global civilization marked by mass consumption, with its many physical comforts (and frequent ostentatious displays), high personal mobility, longer periods of schooling, and growing expenditures on leisure and health.

Correlations of this sequence with average per capita energy consumption have been unmistakable. National peculiarities (from climatic to economic singularities) preclude any simple classification, but three basic categories are evident. No country with average annual primary commercial energy consumption of fewer than one hundred kilograms of oil equivalent can guarantee even basic necessities to all of its inhabitants. Bangladesh and Ethiopia of the 1990s were the most populous nations in this category, and China belonged there before 1950.

As the rate of energy consumption approaches 1 metric ton of oil equivalent (or 42 gigajoules), industrialization advances, incomes rise, and quality of life improves noticeably. China of the 1980s, Japan of the 1930s and again of the 1950s, and Western Europe and the United States between 1870 and 1890 are outstanding examples of this group. Widespread affluence requires, even with efficient energy use, at least 2 metric tons of oil equivalent (more than 80 gigajoules) per capita a year. France made it during the 1960s, Japan during the 1970s. During the 1990s mean per capita consumption rates were below 150 gigajoules (fewer than 4 metric tons of crude oil equivalent) in Europe and Japan, but more than 300 gigajoules in North America, reflecting the profligate use of liquid fuels for transport and the wasteful use of household energy.

A typical premodern family—that is, a peasant household of five to eight people—controlled no more than three to four kilowatts, roughly split between the animate energy of the working members of the family and a couple of draft animals, and the energy in wood used for heating and cooking. By flipping switches of their electric appliances and setting thermostats of their heating and cooling units, a modern American household of four people controls power of about fifteen to twenty-five kilowatts—and internal-combustion engines in their cars add ten times as much installed power. The average person in modern North American society thus controls about ten times as much power as his or her preindustrial ancestor—and given the much higher conversion efficiencies of modern processes (e.g., a simple wood

A Qantas Boeing 747–400 commercial airplane flies above the opera house in Sydney, Australia. (Corbis Corporation)

stove converts 10 to 15% of fuel to useful heat compared to 80 to 90% for high-performance natural-gas furnaces), the differences in terms of useful power are easily twenty to fiftyfold.

Maximum differences are even more stunning: Even when holding reins of a dozen large horses, a well-to-do traditional farmer controlled no more than about ten kilowatts of useful power. In contrast, a pilot of a Boeing 747 commands some sixty megawatts in the plane's four engines, a duty engineer of a large electricity-generating plant may be in fingertip control of more than one gigawatt, and a launch order by a U.S. president could unleash exawatts (10^{18} watts) in thermonuclear warheads. This immense concentration of power under individual control has been one of the hallmarks of modern high-energy society.

The grand transformation wrought by this surfeit of easily controllable energies defines modern civilization. Greatly increased personal mobility, a change that began with the application of steam power to land and water transport, has been vastly expanded thanks to the mass production of vehicles and airplanes powered by internal-combustion engines and gas turbines (the only notable twentieth-century addition to common prime movers). Wide-bodied jet airplanes in general, and the Boeing 747 in particular, have been among the leading agents of globalization. The latest stage of this energy-driven globalization trend is vastly enlarged access to information made possible by mass diffusion of personal computers, by electronic storage of data and images, and by high-volume wireless and optical fiber transmissions.

The easing of women's household work in the Western world is another particularly noteworthy social transformation wrought by efficient energy conversions. For generations rising fuel consumption made little difference for everyday household work; indeed, it could make it worse. As the standards of hygiene and social expectations rose with better education, women's work in Western countries often got harder. Electricity was the eventual liberator. Regardless of the availability of other energy forms, it was only its introduction that did away with exhausting and often dangerous labor: Electric irons, vacuum cleaners, electric stoves, refrigeration, washing machines and electric dryers transformed housework, beginning in the early decades of the twentieth century. Hundreds of millions of women throughout the poor world are still waiting for this energy-driven liberation.

Air conditioning has been another revolutionary application of electricity. The technique was first patented by William Carrier in 1902 but widespread adoption came only after 1960, opening up first the American Sunbelt to mass migration from northern states, then increasing the appeal of subtropical and tropical tourist destinations, and since the 1980s becoming also a common possession for richer urbanites throughout the industrializing world.

And yet the energy-civilization link should not be overrated. Basic indicators of physical quality of life show little or no increase with average per capita energy consumption rising above one hundred gigajoules per year—and it would be very unconvincing to argue that North American consumption levels, twice as high as in Western Europe, make Americans and Canadians twice as content, happy, or secure. Indeed, international and historic comparisons show clearly that higher energy use will not assure reliable food supply, does not confer political stability, does not necessarily enhance personal security, does not inevitably lead to a more enlightened governance, and may not bring any widely shared improvements in the standard of living.

CHALLENGES AHEAD

During the past generation, a number of poor countries moved to the intermediate energy consumption category. Still, in terms of total population, the distribution of global energy use remains extremely skewed. In 1950 only about 250 million people (one-tenth of the global population) consumed more than two metric tons of oil equivalent a year per capita, yet they claimed 60 percent of the world's primary energy (excluding biomass). By 1999 such populations were about a fifth of all mankind, and they claimed nearly three-quarters of all fossil fuels and electricity. In contrast, the poorest quarter of humanity used less than 5 percent of all commercial energies.

Stunning as they are, these averages do not capture the real difference in living standards. Poor countries devote a much smaller share of their total energy consumption to private household and transportation uses. The actual difference in typical direct per capita energy use among the richest and the poorest quarters of the mankind is thus closer to being fortyfold rather than "just" twentyfold. This enormous disparity reflects the chronic gap in economic achievement and in the prevailing quality of life and contributes to persistent global political instability.

Narrowing this gap will require higher substantially increased output of fossil fuels and electricity, but an appreciable share of new energy supply should come from more efficient conversions. Energy intensities (i.e., the amount of primary energy per unit of the GDP) of all affluent economies have been declining during most of the twentieth century, and some modernizing countries have been moving in the same direction even faster; perhaps most notably, China has more than halved its energy intensity since 1980!

Low energy prices provide little incentive for sustaining the efficiency revolution that flourished after OPEC's crude-oil price rises of the 1970s, but higher conversion efficiencies should be pursued regardless; in the absence of (highly unlikely) voluntary frugality, they are the only means of reducing overall energy throughput and hence minimizing the impact of energy use on the biosphere. Energy industries and conversions have many environmental impacts, but it now appears that the threat of relatively rapid global warming, rather than the shrinking resource base, will be the most likely reason for reducing our civilization's high dependence on fossil fuels.

So far we have been successful in preventing the use of the most destructive energy conversion, the explosion of thermonuclear weapons, in combat. Although the risk of armed superpower conflict has decreased with the demise of the Soviet empire, thousands of nuclear weapons remain deployed around the world. Our challenge during the twenty-first century will be fourfold: avoiding nuclear conflict; extending the benefits of high-energy society to billions of people in low-income countries; decoupling the development of rich societies from continuous growth of energy consumption; and preserving the integrity of the biosphere.

Vaclav Smil

See also: Carnot, Nicolas Leonard Sadi; Clausius, Rudolf Julius Emmanuel; Culture and Energy Usage; Ethical and Moral Aspects of Energy Use; Gibbs, Jonah Willard; Industry and Business, History of Energy Use and; Joule, James Prescott; Kinetic Energy, Historical Evolution of the Use of; Mayer, Julius Robert von; Refining, History of; Thomson, William; Watt, James.

BIBLIOGRAPHY

Alexander, R. McN. (1992). *The Human Machine.* New York: Columbia University Press.

Aristotle. (1966). *Metaphysics*, tr. by H. H. Apostle. Bloomington: Indiana University Press.

Cardwell, D. S. L. (1971). *From Watt to Clausius.* Ithaca, NY: Cornell University Press.

Clausius, R. (1867). *Abhandlungen uüber die mechanische Waärmetheorie*. Braunschweig, Ger.: F. Vieweg.

Cook, E. (1976). *Man, Energy, Society*. San Francisco: W. H. Freeman.

Cottrell, F. (1955). *Energy and Society*. New York: McGraw-Hill.

Cowan, C. W., and Watson, P. J., eds. (1992). *The Origins of Agriculture*. Washington DC: Smithsonian Institution Press.

Dickinson, H. W. (1939). *A Short History of the Steam Engine*. Cambridge, Eng.: Cambridge University Press.

Düring, I. (1966). *Aristoteles Darstellung und Interpretation seines Denken*. Heidelberg, W. Ger.: Carl Winter.

Durnin, J. V. G. A., and Passmore, R. (1967). *Energy, Work, and Leisure*. London: Heinemann Educational Books.

Einstein, A. (1905). "Zur Elektrodynamik bewegter Koörper." *Annalen der Physik* 17:891–921.

Einstein, A. (1907). "Relativitaätsprinzip und die aus demselben gezogenen Folgerungen." *Jahrbuch der Radioaktivitaät* 4:411–462.

Finniston, M., et al., eds. (1992). *Oxford Illustrated Encyclopedia of Invention and Technology*. Oxford, Eng.: Oxford University Press.

Forbes, R. J. (1964–1972). *Studies in Ancient Technology*. Leiden, Neth.: E. J. Brill.

Georgescu-Roegen, N. (1971). *The Entropy and the Economic Process*. Cambridge, MA: Harvard University Press.

Goudsblom, J. (1992). *Fire and Civilization*. London: Allen Lane.

Joule, J. P. (1850). *On Mechanical Equivalent of Heat*. London: R. and J. E. Taylor.

Langdon, J. (1986). *Horses, Oxen and Technological Innovation*. Cambridge, Eng.: Cambridge University Press.

Lindsay, R. B., ed. (1975). *Energy: Historical Development of the Concept*. Stroudsburg, PA: Dowden, Hutchinson, & Ross.

Mayer, J. R. (1851). *Bemerkungen über das mechanische Aequivalent der Waärme*. Heilbronn, Ger.: J. V.Landherr.

McLaren, D. J., and Skinner, B. J., eds. (1987). *Resources and World Development*. Chichester, Eng.: John Wiley & Sons.

National Research Council. (1986). *Electricity in Economic Growth*. Washington, DC: National Academy Press.

Nef, J. U. (1932). *The Rise of the British Coal Industry*. London: Routledge.

Odum, H. T. (1971). *Environment, Power, and Society*. New York: Wiley-Interscience.

Ostwald, W. (1909). *Energetische Grundlagen der Kulturwissenschaften*. Leipzig, Ger.: Alfred Kroöner.

Pacey, A. (1990). *Technology in World Civilization*. Cambridge, MA: MIT Press.

Reynolds, J. (1970). *Windmills and Watermills*. London: Hugh Evelyn.

Rose, D. J. (1986). *Learning about Energy*. New York: Plenum Press.

Schipper, L., and Meyers, S. (1992). *Energy Efficiency and Human Activity*. New York: Cambridge University Press.

Singer, Charles, et al., eds. (1954–1984). *A History of Technology*. Oxford, Eng.: Clarendon Press.

Smil, V. (1991). *General Energetics*. New York: John Wiley & Sons.

Smil, V. (1994). *Energy in World History*. Boulder, CO: Westview Press.

Soddy, F. (1926). *Wealth, Virtual Wealth, and Debt: The Solution of the Economic Paradox*. New York: E. P. Dutton.

World Energy Council Commission. (1993). *Energy for Tomorrow's World*. London: Kogan Page.

HORSEPOWER

See: Units of Energy

HOUDRY, EUGENE JULES (1892–1962)

EDUCATION AND WAR EXPERIENCE

Eugene Jules Houdry, one of the fathers of petroleum refining, was born in Dumont, France, outside of Paris, on April 18, 1892, and grew up during an era of rapid technological change and innovation. Houdry was blessed with considerable emotional and financial support from his family and encouraged to reach for lofty and risky goals.

With the urging of his father, a wealthy steel manufacturer, Houdry studied mechanical engineering at the École des Arts et Métiers in Chalons-sur-Marne near Paris. He earned the government's gold model for highest scholastic achievement in his class and captained its national champion soccer team. Graduating in 1911, Houdry briefly worked in his father's metalworking business, Houdry et Fils, but joined the army prior to the outbreak of World War I. As a lieutenant in the tank corps, Houdry was seriously wounded in 1917 in the first great tank battle, the Battle of Javincourt. He was awarded the Croix de Guerre, the French military decoration created in 1915 to reward feats of bravery, and membership in the Legion d'Honneur for extraordinary bravery.

FROM TANKS TO AUTOMOBILE RACING TO CATALYTIC CRACKING AND CATALYST DEVELOPMENT

After the war he rejoined Houdry et Fils, but his interest drifted from steel to car-racing and then to improving the fuel performance of car racing engines. Houdry recognized the need for more efficient motor fuels during a visit to Ford and while attending the Memorial Day Indianapolis 500 Race. In 1922 Houdry learned of a superior gasoline being produced from lignite by E. A. Prudhomme, a Nice chemist. Undeterred by unfamiliarity with catalysis or chemical engineering, Houdry enlisted Prudhomme in developing a workable lignite-to-gasoline process in a laboratory privately financed by a group of Houdry's friends. In their scheme, solid lignite would be initially broken down by heat (thermally cracked) to produce a viscous hydrocarbon oil and tars. The oil was then somehow further converted (actually, catalytically cracked) to gasoline by contacting the oil with various clays. Houdry's wartime experiences impressed upon him France's need for indigenous petroleum resources, which provided further impetus to his pursuing the project. With government support, the alliance built a plant that could convert sixty tons of lignite per day to oil and gasoline. The process was not economic, and the unit shut down in 1929. Across the ocean, a similar fate befell Almer McAfee at Gulf Oil in the same year. McAfee is credited with developing the first commercially viable catalytic cracking process using aluminum chloride as the catalyst.

The poor economics for McAfee's catalytic process lay in the substantial improvements being made in the technology of thermally cracking crude oil. Notable among these competing efforts were those of William Burton of Standard Oil (Indiana), whose high-pressure and high-temperature thermal technology more than doubled the potential yield of gasoline from crude oil and thereby substantially reduced the cost of thermal cracking.

Houdry was dejected by the failure of his process to compete against the more mature thermal cracking technology, but his seven-year effort was not in vain. Houdry had been devoting his attention initially to using a naturally occurring aluminosilicate clay, fuller's earth, as the catalyst and later to using more catalytically active hydrosilicates of aluminum prepared by washing clays, such as bentonite with acid. This approach ran contrary to the mainstream of catalytic cracking investigators, who were working unsuccessfully with nickel-containing catalysts, or to McAfee, who was using anhydrous aluminum chloride as the catalyst and making it practical by effecting tenfold reductions in its cost of manufacture.

The catalytic cracking activity of Houdry's catalysts and McAfee's aluminum chloride were puzzling to researchers at the time: It worked, but they could not understand why. As later determined by Hans Tropsch of UOP Research Laboratories, both silica-alumna and aluminum chloride behave as acid catalysts capable of donating protons or accepting electron pairs to form "carbonium ion" reaction intermediates. The concept of such active intermediates explained the propensity of catalytic cracking to favor formation of higher-octane branched chain paraffins and aromatics as compared with thermal cracking's disposition toward producing lower-octane straight chain paraffins. It also explained the greater yields of gasoline resulting from other hydrocarbon reactions occurring in the process, namely isomerization (e.g., conversion of zero octane n-pentane to isopentane), alkylation (e.g., conversion of isobutane and isobutene to "isooctane"), polymerization (e.g., propylene to hexene) and dehydrogenation (e.g., hexene to benzene).

Remarkably, seventy years after Houdry's utilization of the catalytic properties of activated clay and the subsequent development of crystalline aluminosilicate catalysts that are a magnitude more catalytically active, the same fundamental principles remain the basis for the modern manufacture of gasoline, heating oils, and petrochemicals.

COMMERCIALIZATION AND IMPACT

Houdry concentrated his personal efforts on developing a viable processing scheme, solving the engineering problems, scaling the process to commercial size, and developing requisite equipment. In 1930,

H. F. Sheets of Vacuum Oil Company, who learned of Houdry's work and shared his vision for converting vaporized petroleum to gasoline catalytically, invited him to the United States. After a successful trial run, Houdry moved his laboratory and associates from France to Paulsboro, New Jersey, to form a joint venture, Houdry Process Corporation, with Vacuum Oil Company. In that year Vacuum Oil Company merged with Standard Oil of New York to become Socony-Vacuum Company (much later Mobil Oil Corporation).

In 1933, a 200-barrel-per-day Houdry unit was put into operation. The financial strains of the Great Depression coupled with formidable technical problems necessitated inclusion of Sun Oil Company in a fruitful three-way codevelopment partnership. In the next few years the Houdry process underwent considerable changes, especially in developing an innovative method for periodically burning off a buildup of coke within the catalyst to restore catalyst activity lost after only ten minutes of usage and employing a molten salt heat exchanger to replace heat absorbed by the endothermic cracking reactions.

In 1936, Socony-Vacuum built a 2,000-barrel-per-day semicommercial Houdry unit at Paulsboro, closely followed the next year by Sun Oil's 15,000-barrel-per-day commercial plant in its Marcus Hook, Pennsylvania, refinery. With a capability to produce a higher-octane and thus better-performing gasoline, the new development was touted as "the new miracle of gasoline chemistry." The timing was fortuitous in giving the Allies a major advantage in World War II. During the first two years of the war Houdry units produced 90 percent of all catalytically cracked gasoline. Minimal additional refining, notably washing with sulfuric acid to remove low-octane straight-chain olefins and deleterious sulfur components, allowed the manufacture of exceedingly scarce high-test aviation gasoline. Allied planes were 10 to 30 percent superior in engine power, payload, speed and altitude during these critical early years of the war.

The Houdry fixed-bed cyclic units were soon displaced in the 1940s by the superior Fluid Catalytic Cracking process pioneered by Warren K. Lewis of MIT and Eger Murphree and his team of engineers at Standard Oil of New Jersey (now Exxon). Murphree and his team demonstrated that hundreds of tons of fine catalyst could be continuously moved like a fluid between the cracking reactor and a separate vessel for catalyst regeneration. Nevertheless, the fixed bed remains an important reactor type in petrochemicals, and the Houdry reactor remains a classic example of their high state of development.

BEYOND CATALYTIC CRACKING

Houdry's entire career was characterized by foresight, imagination, leadership, boldness, and persistence. He was a demanding individual, single-minded of purpose and a workaholic, yet with considerable charm. Houdry was one of twenty whose opposition to the Vichy government resulted in their French citizenship being revoked. Later during the war, when natural rubber supplies were cut off, Houdry invented a catalytic process for producing butadiene, an essential reactant in manufacturing synthetic rubber. Following World War II, motivated by a concern for the environment, Houdry showed foresight in turning his attention to the health hazards of automobiles and industrial pollution, well before they were widely acknowledged as major problems. He founded Oxy-Catalyst Company and holds one of the early U.S. patents for catalytically reducing the amounts of carbon monoxide and unburned hydrocarbons in automobile exhausts, anticipating the development of the catalyst processes now used with modern automobiles. He died on July 18, 1962, at age seventy.

Barry L. Tarmy

See also: Catalysts; Combustion; Gasoline and Additives; Heat and Heating; Petroleum Consumption; Refineries; Refining, History of.

BIBLIOGRAPHY

American Chemical Society. (1996). *A National Historic Chemical Landmark: The Houdry Process for the Catalytic Conversion of Crude Petroleum to High-Octane Gasoline.* Washington, DC: American Chemical Society.

Buonora, P. T. (1998). "Almer McAfee at Gulf Oil." *Chemical Heritage* 16(2):5–7, 44–46.

Enos, J. L. (1962). *Petroleum Progress and Profits: A History of Process Innovation.* Cambridge, MA: MIT Press.

Mosely, C. G. (1984). "Eugene Houdry, Catalytic Cracking and World War II Aviation Gasoline." *Journal of Chemical Education* 61:655–656.

Spitz, P. H. (1988). *Petrochemicals: The Rise of an Industry.* New York: John Wiley & Sons.

Thomas, C. L.; Anderson, N. K.; Becker, H. R.; and McAfee J. (1943). "Cracking with Catalysts." *Petroleum Refiner* 22(1):95–100, 365–370.

HUB AND SPOKE SYSTEMS

See: Air Travel

HYBRID VEHICLES

INTRODUCTION

By incorporating into the driveline of the vehicle the capability to generate electricity onboard the vehicle from a chemical fuel, a hybrid-electric vehicle has the characteristics of both an electric vehicle and a conventional internal combustion engine (ICE) vehicle and can be operated either on wall-plug electricity stored in a battery or from a liquid fuel (e.g., gasoline) obtained at a service station. This essay discusses technologies for hybrid-electric vehicles that can attain significantly higher fuel economy and lower emissions than conventional ICE vehicles of the same size, performance, and comfort.

HYBRID-ELECTRIC VEHICLE DESIGN OPTIONS

There are a large number of ways an electric motor, engine, generator, transmission, battery, and other energy storage devices can be arranged to make up a hybrid-electric driveline. Most of them fall into one of two configurations—series and parallel. In the series configuration (Figure 1, Top), the battery and engine/generator act in series to provide the electrical energy to power the electric motor, which provides all the torque tot he wheel of the vehicle. In a series hybrid, all the mechanical output of the engine is used by the generator to produce electricity to either power the vehicle or recharge the battery. This is the driveline system used in diesel-electric locomotives. In the parallel configuration (Figure 1, Bottom), the engine and the electric motor act in parallel to provide torque to the wheel of the vehicle. In the parallel hybrid, the mechanical output of the engine can be used to both power the vehicle directly and to recharge the battery or other storage devices using the motor as a generator. In recent years, a third type of hybrid configuration, the dual mode, is being developed that combines the series and hybrid configurations. As shown in Figure 1b, the engine output can be split to drive the wheel (parallel mode) and to power a generator to produce electricity (series mode). This configuration is the most flexible and efficient, but it is also likely to be the most complex and costly.

A range-extended electric vehicle would most likely use the series configuration if the design is intended to minimize annual urban emissions. It would be designed for full-performance on the electric drive alone. The series hybrid vehicle can be operated on battery power alone up to its all-electric range with no sacrifice in performance (acceleration or top speed) and all the energy to power the vehicle would come from the wall plug. This type of hybrid vehicle is often referred to as a “California hybrid” because it most closely meets the zero emission vehicle (ZEV) requirement. The engine would be used only on those days when the vehicle is driven long distances.

Hybrid vehicles designed to maximize fuel economy in an all-purpose vehicle could use the series, parallel, or dual configurations depending on the characteristics of the engine to be used and acceptable complexity of the driveline and its control. Parallel hybrid configurations will require frequent on-off operation of the engine, mechanical components to combine the engine and motor torque, and complex control algorithms to smoothly blend the engine and motor outputs to power the vehicle efficiently. The parallel hybrid would likely be designed so that its acceleration performance would be less than optimum on either the electric motor or engine alone and require the use of both drive components together to go zero to sixty mph in 10–12 sec. Such a hybrid vehicle would not function as a ZEV in urban/freeway driving unless the driver was willing to accept reduced acceleration performance. The parallel configuration can be designed to get better mileage than the series configuration. A fuel cell hybrid vehicle would necessarily be a series hybrid because the fuel cell produces only electricity and no mechanical output. The dual mode hybrid is intended to maximize fuel economy and thus be designed like a parallel hybrid, but with a relatively small generator that could be powered by the engine. The engine in the dual mode hybrid would operate in

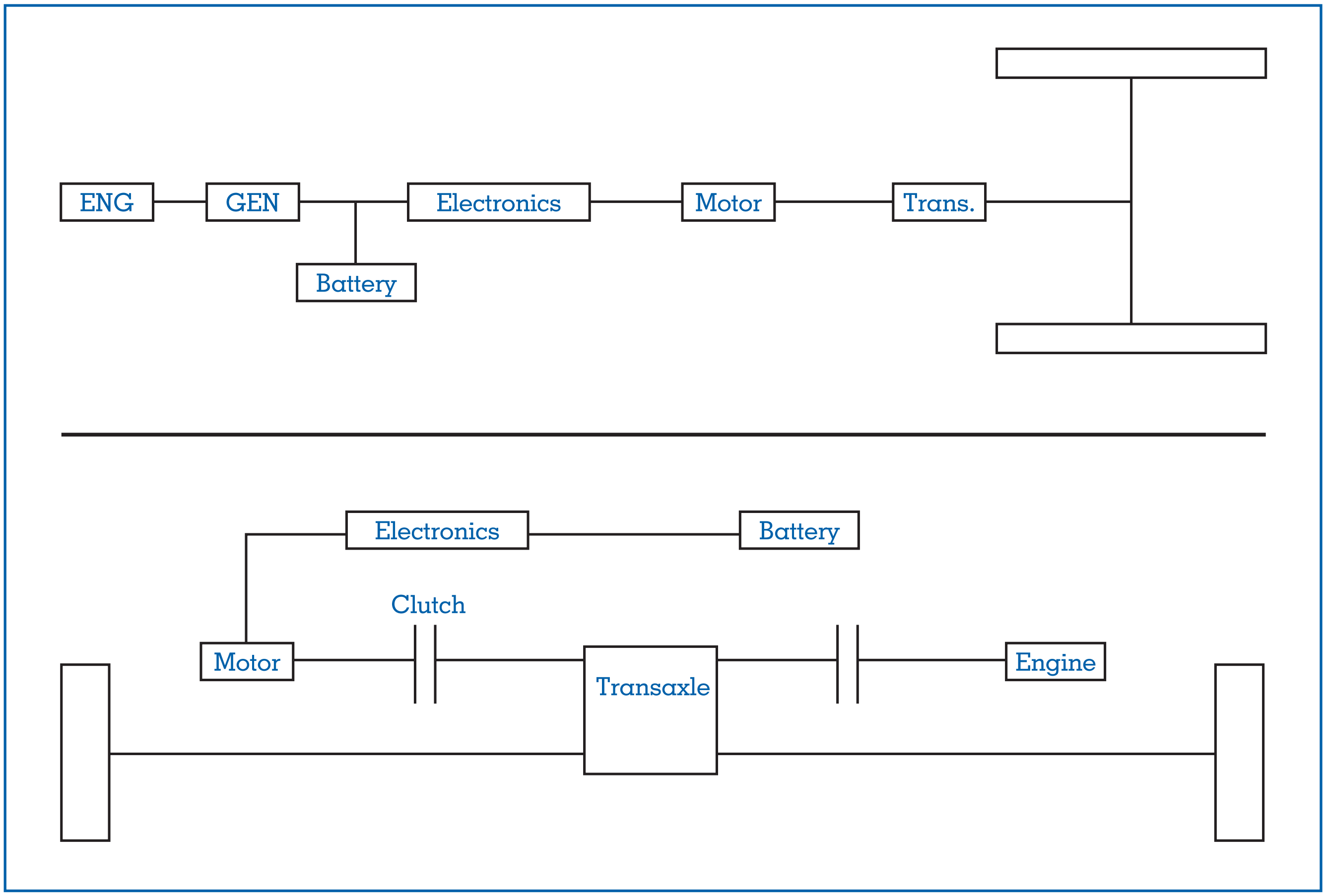


Figure 1.
Hybrid Vehicle Driveline Schematics.
NOTE: Top: Series; Bottom: Parallel

the on/off mode, but be cycled on and off less frequently than in the parallel configuration.

COMPONENT TECHNOLOGIES: STATUS AND PROSPECTS

Motor and Electronics

Recent advances in both motor and inverter electronics technologies have important implications for the design of high performance hybrid vehicles. In 2000 the size and weight of the motor and the electronics combined are significantly smaller (by a factor of two to three—see Table 1) than that of an engine and transmission of the same peak power in 1998. The size advantage of the electrical components will be even greater as the operating voltage of the electrical drive system is increased above the 300–400V that is common today. This means that in packaging the hybrid driveline, the electrical drive components take up only a fraction of the space available and finding room for the mechanical components, such as the engine, transmission or torque coupler, can present a difficult challenge.

Electrical Energy Storage

The key component in the hybrid driveline that permits it to operate more efficiently than the engine/transmission in a conventional car is the electrical energy storage unit. It must store sufficient energy (kWh) to provide the all-electric range of the vehicle or to permit the engine or fuel cell to operate near the average power required by the vehicle (in a load-leveled mode). It also must have sufficient power capability to meet on demand the peak power of the motor/electronics for vehicle acceleration or

Engine Type	kg/kW	l/kW	Maximum Efficiency (%)	Emissions (gm/kWh) (Nominal)			Manufacturer/Developer
				HC	CO	NOX	
Spark Ignition							
Valve Injection	2.0	4.0	32	3.0	20	8.0	Auto Companies
Direct Injection	2.5	4.1	38	5.0	4.0	3.5	Auto Companies
Rotary	.8	1.0	30	3.0	4.0	5.0	Moller International
Two-Stroke	1.0	1.75	30	3.0	3.0	6.0	Orbital
Diesel							
Prechamber/Turbo	2.6	4.2	35	.1	1.0	3.0	Auto Companies
Direct Injection/Turbo	3.0	4.4	42	1.0	1.0	15.0	Auto Companies
Gas Turbine							
Metal							
With Recup.	2.5	5.3	30	.08	.30	.03	Capstone
Incl. generator				(catalytic combustion)			
Ceramic							
with Recup.	2.0	4.0	40	.08	.30	.03	Allison
				(catalytic combustion)			
Stirling							
H_2 Working Fluid/ Wobble Plate Drive	2.7-3	2-2.5	30-35	.01	.15	.22	Stirling Thermal Motors

Table 1.
Hybrid Vehicle Engine Characteristics

regenerative braking. In most cases, the energy storage unit in a hybrid vehicle is sized by the peak power requirement. Because the size (weight and volume) of the energy storage unit (often a battery) in the hybrid vehicle is smaller than the battery in a battery-powered electric vehicle (EV), the power density (W/kg and W/liter) requirements of the energy storage unit in the hybrid vehicle are greater than for the battery in an electric vehicle. For example, power densities of 200–300W/kg are satisfactory for use in an EV, but power densities of 500–1,000W/kg are needed for hybrid vehicles. Considerable progress has been made in the development of high power batteries (often referred to as pulse batteries) of various types. In hybrid vehicles, the energy density of the energy storage unit is of secondary importance; compromises in energy density have been made to reach the high power density of pulse batteries. For example, nickel metal hydride batteries designed for EVs have an energy density of 70Wh/kg and a power density of 250 W/kg, while those designed for hybrid vehicles have an energy density of 40–45Wh/kg and a peak power density of 600–700W/kg. Another important consideration for energy storage units for use in hybrid vehicles is the need to minimize the losses associated with transferring energy into and out of the unit, because in the hybrid, a reasonable fraction of the electrical energy produced onboard the vehicle is stored before it is used to power the vehicle. In order to minimize the energy storage loss, the round-trip efficiency (energy out/energy in) should be at least 90 percent. This means that the useable peak power capability of a battery is much less that the power into a match impedance load at which efficiency of the transfer is less than 50 percent. Useable power is only about 20 percent of the peak power capacity. This is another reason that the design of an energy storage units for hybrid vehicles is much more difficult than for battery-powered electric vehicles and energy storage technology is often described as the enabling technology for hybrid vehicles.

A new energy storage technology that is well suited for hybrid vehicles is the electrochemical ultracapacitor, often referred to as the double-layer capacitor. Ultracapacitors for vehicle applications have been under development since about 1990. The construction of an ultracapacitor is much like a battery in that it consists of two electrodes, a separator, and is filled with an electrolyte. The electrodes have a very high surface area (1,000–1,500 m^2/gm) with much of the surface area in micropores 20 angstroms or less in diameter. The energy is stored in the double-layer (charge separation) formed in the micropores of the electrode material. Most of the ultracapacitors presently available use activated carbon as the electrode material. The cell voltage is primarily dependent on the electrolyte used. If the electrolyte is aqueous (sulfuric acid or KOH) the maximum cell voltage is 1V; if an organic electrolyte (propylene carbonate) is used, the maximum cell voltage is 3V. As in the case of batteries, high voltage units (300–400V) can be assembled by placing many ultracapacitor cells in series.

Batteries have much higher energy density and capacitors have much higher power capacity. The technical challenge for developing ultracapacitors for vehicle applications is to increase the energy density (Wh/kg and Wh/liter) to a sufficiently high value that the weight and volume of a pack to store the required energy (500 Wh for a passenger car) is small enough to be packaged in the vehicle. Power density and cycle life are usually not a problem with ultracapacitors. The cost ($/Wh) of ultracapacitors is presently too high—being about $100/Wh. It must be reduced by at least an order of magnitude (a factor of 10) before this new technology will be used in passenger cars. Nevertheless, ultracapacitors are a promising new technology for electrical energy storage in hybrid vehicles.

Engines and Auxiliary Power Units

The characteristics of engines for hybrid vehicles are shown in Table 1. Most of the hybrid vehicles designed and built have used four stroke gasoline or diesel engines. Nearly all gasoline engines are now fuel-injected and both gasoline and diesel engines are computer controlled. Continuing improvements are being made in these engines in terms of size, weight, efficiency, and emissions. These improvements and the common use of computer control make it fairly easy to adapt the conventional engines to hybrid vehicle applications. The major difficulty in this regard is to find an engine of appropriate power rating for hybrid vehicle application. Most automotive engines have power of 60kW (75hp) and greater, which is too large for use in most hybrid vehicle designs. Experience has shown that good sources of engines for hybrid vehicles are the minicars designed for the small car markets on Japan and Europe.

Several advanced engines have been developed especially for hybrid vehicles. These include a high-expansion ration gasoline engine (Atkinson cycle) developed by Toyota for their Prius hybrid vehicle which they started to market in Japan in 1997. Stirling Thermal Motors (STM) under contract to General Motors (GM) designed and fabricated a Stirling engine (30kW) for use on GM's series hybrid vehicle built as part of the Department of Energy (DOE) hybrid vehicle program. Capstone Technology developed a 25kW recuperative gas turbine engine for use in a flywheel hybrid vehicle built by Rosen Motors. The characteristics of these engines are indicated in Table 1. The most successful of these engine development projects was the Prius engine of Toyota. The other two engines were too large and were not efficient enough to warrant further development for hybrid vehicle applications. The new Toyota engine in the Prius is a four cylinder (1.5 liter), four stroke gasoline engine that utilizes variable inlet valve timing to vary the effective compression ratio from 9.3 to 4.8 in an engine having a mechanical compression ratio of 13.5. Varying the effective volume of air during the intake stroke permits operation of the engine at part load with reduced pumping and throttling losses. This results in an increase in engine efficiency. The expansion ratio at all times is set by the high mechanical compression ratio of the engine. This new engine was optimized for operation in the hybrid mode and had a brake specific fuel consumption of about 235 gm/kWh for output powers between 10 and 40kW. This corresponds to an efficiency of 37 percent, which is very high for a four stroke gasoline engine of the same peak power rating.

The engine output in a hybrid vehicle can be utilized to generate electricity on board the vehicle or to provide torque to the driveshaft of the vehicle. In the first case, the engine output torque drives a generator and the combination of the engine and the generator is termed an auxiliary power unit (APU). The generator can be either an ac induction or a brushless dc permanent magnet machine. The size of the generator for a given power rating depends to a large extent on the voltage of the system and the rpm at which the generator rotates. For a 400V system and a maximum of 8,000–10,000, the size and weight of the APU are 0.7kg/kW and 0.8 liter/kW, respectively, including the electronic controls. The efficiency of the generator system will vary between 90 to 95 percent depending on the power output. The losses associated with the production and storage of the electrical energy onboard the vehicle in a series hybrid are significant (10 to 20%) and cannot be neglected in predicting the fuel economy of the vehicle.

Mechanical Components

The transmission, clutch, and other mechanical components needed in a hybrid vehicle to combine the output of the engine with the electric motor and generator and the main driveshaft of the vehicle are critical to the efficient and smooth operation of parallel and dual mode hybrid vehicles. The design of these components is relatively straightforward and not much different than that for similar components for conventional engine-powered vehicles. In a parallel hybrid (Figure 1b), the engine is connected to the drive shaft through a clutch that opens and closes as the engine power is needed. The speed ratio between the engine and the wheels is determined by the gear ratios in the transmission. Mechanical design of the engine clutch so that it has a long life and smooth operation is one of the critical tasks in the development of the parallel hybrid vehicle. Many hybrid vehicles are built with manual transmissions, because automatic transmissions with a torque converter have unacceptably high losses. A recent development is the use of a continuously variable transmission (CVT) in a parallel hybrid driveline. The operation of the CVT is like an automatic transmission from the driver's point-of-view with the advantages of lower losses and a wider range of continuous gear ratios, which result in efficient driveline operation in both the electric and hybrid modes. The disadvantages of the CVT are that control of the system is more difficult than with a manual transmission and the steel belt used in the CVT is much less tolerant of abuse (sudden changes in speed and torque). Nevertheless, it appears that in the future years CVTs will have application in parallel hybrid drivelines.

There are several arrangements of the dual mode hybrid driveline. In the simplest arrangement (Figure 2), the generator can be used as a motor to start the engine and/or supplemental torque of the traction motor to drive the vehicle. In this dual mode configuration, the batteries can be recharged either by the generator or by the traction motor acting as a generator. This simple dual mode system does not require a transmission. A second dual mode system utilizes a planetary gear set to couple the engine, generator, and main driveshaft. In this arrangement, the speed ratio between the engine and the main driveshaft depends on the fraction of the engine power that is applied to the generator. This second arrangement is used by Toyota in the Prius hybrid car. This system is less flexible and less efficient than the first system in which the engine and generator are directly connected on the same drive shaft, but it does not require a clutch, which must be opened/closed smoothly and reliably under computer control. Operation of the Toyota dual mode hybrid driveline has proven to be smooth and reasonably efficient.

Fuel Cells

Fuel cells can be utilized in electric hybrid vehicles as the means of converting chemical fuel to electricity. Rapid progress has been made in the development of fuel cells, especially proton exchange membrane (PEM) fuel cells, for transportation applications. This progress has resulted in a large reduction in the size and weight of the fuel cell stack and as a result, there is now little doubt that the fuel cell of the required power (20–50kW) can be packaged under the hood of a passenger car. The primary question regarding fuel cells in light duty vehicles is how they will be fueled. The simplest approach is to use high pressure hydrogen as has been done in the most successful bus demonstration to date. This approach is satisfactory for small test and demonstration programs, but the development of the infrastructure for using hydrogen as a fuel in transportation will take many years. Considerable work is underway to develop fuel processors (reformers) to generate hydrogen onboard the vehicle from various chemical fuels (e.g., methanol or hydrocarbon distillates). Most of the hydrogen used for industrial and transportation applications is presently generated by reforming natural gas using well-developed technology. A promising approach to fuel processing to hydrogen (H^+ and electrons) onboard the vehicle is direct oxi-

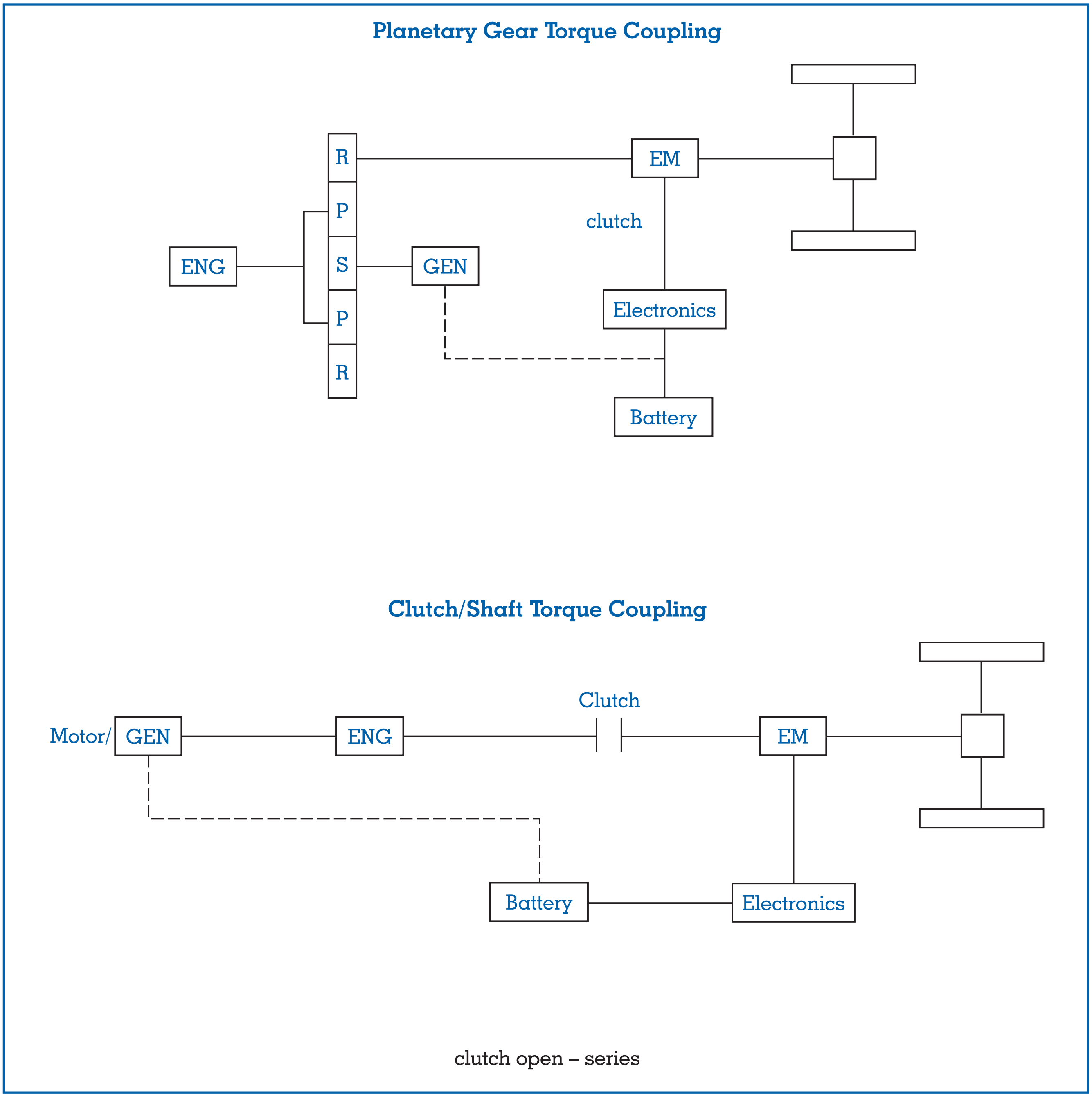


Figure 2
Dual Mode Hybrid Vehicle Driveline Schematics

dation of methanol within the fuel cell stack. When technology for the efficient, direct conversion of a liquid fuel to hydrogen within the PEM fuel cell is developed, the commercialization of fuel cells in light duty vehicles will occur rapidly.

CONTROL STRATEGIES

Control Strategies for Series Hybrid Vehicles

The intent of the control strategy is to maintain the state-of-charge of the energy storage unit within a prescribed range regardless of the driving cycle and the resultant power demand on the driveline. This should be done so that the onboard electrical generator (engine/generator or fuel cell) is operated at high efficiency and low emissions. This is done more easily when the energy storage capacity is reasonably large as with a battery than when it is small as using ultracapacitors. The strategy used for vehicles having a significant all-electric range is to discharge the battery to a prescribed state-of-charge (20 to 30%) and then to turn on the engine to maintain the battery within 10 to 20 percent of that condition. Electrical energy is generated at a rate slightly greater than the average power demand of the vehicle to account for losses associated with storing the energy. In the case of an engine/generator, a minimum power level is set so that the engine is never operated below it. Proper selection of this minimum power can have an important effect on fuel economy. When the battery charge reaches the maximum permitted, the engine is turned off and it remains off until the battery state-of-charge falls to the engine turn-on state-of-charge. When the series hybrid is operated so that the battery is permitted to discharge to a relatively low state-of-charge, it is termed a charge depleting hybrid. If the battery is maintained at a high state-of-charge (60 to 70%), it is termed a charge sustaining hybrid and the battery is seldom, if ever, recharged from the wall-plug. A significant fraction of the energy used by charge depleting hybrid vehicles is from the wall-plug and their average annual emissions and energy consumption are dependent on the use-pattern (miles of travel per day) of the vehicle and how the electricity used to recharge the batteries is generated.

Control Strategies for Parallel Hybrid Vehicles

The control strategies for parallel hybrid vehicles are more complicated than those for series hybrids primarily because they are dependent on both vehicle speed and state-of-charge of the energy storage unit and should include a criteria for splitting the driveline torque between the engine and the electric motor. The intent of the strategy is to permit the electric motor to provide the torque if it can at vehicle speeds below a prescribed value and permit the engine to provide the torque at higher speeds. If the vehicle is operating in the all-electric mode, the motor provides the torque and the engine is not turned on regardless of the torque demand or vehicle speed. Since the all-electric range of a hybrid vehicle is usually less than 80 km, operation of the vehicle should change automatically to the hybrid mode when the all-electric range is exceeded. The control strategy in the hybrid mode can be either charge sustaining or charge depleting. In the case of charge sustaining, the battery state-of-charge is maintained at a near constant value by a control strategy using electrical energy produced by the engine and the motor acting as a generator and consequently little electrical energy is used from the wall-plug. For the charge depleting case, the control strategy permits the battery state-of-charge to decrease as the vehicle is driven and the battery is then recharged from the wall-plug at night. Parallel hybrids usually have a multi-speed transmission so the control strategy must also include a gear shifting algorithm that depends on whether the motor or engine or both are producing torque. A continuously variable transmission (CVT) would be particularly attractive for use in a parallel hybrid driveline.

In order to achieve high fuel economy with a parallel hybrid, it is necessary to avoid engine operation below some minimum engine torque (or effective throttle setting) where the engine brake specific fuel consumption (gm/kWh) is relatively high and to manage engine turn on and off carefully to minimize emissions and wasted fuel. In urban driving, the control strategies for parallel hybrids often result in the engine being turned on and off frequently because the vehicle speed and power demands vary rapidly in stop-and-go driving. The effects of this on-off engine operation on fuel usage and emissions for the parallel hybrids are neglected in most simulations at the present time, so further analysis and vehicle testing is needed to determine whether the high fuel economy and low emissions projected for parallel hybrids can be attained. The control strategies for parallel hybrids are necessarily more complex than those for series hybrids and the uncertainty in the simulation results for parallel hybrids are greater.

Control Strategies for Dual Mode Hybrid Vehicles

The control strategies for dual mode hybrid vehicles are a combination of those used for series and parallel hybrids. There are so many possible hardware arrangements and associated control strategies, it is not possible to summarize them in a simple manner as was the case for series and parallel hybrids. The objective of the dual mode operation is to use the possibility of battery charging simultaneously with the use of the engine and electric motor to power the vehicle as a means of maintaining engine operation at high efficiency at all times. At highway speeds, the engine can be used directly to power the vehicle with the engine operating at high efficiency. This mode of operation is essentially that of a parallel hybrid. At low vehicle speeds, when the battery does not need charging (state-of-charge greater than a specified value), the driveline would operate in an electric-only mode if the electric motor can provide the power required by the vehicle. If the power demand is greater than that available from the electric motor, the engine is turned on to assist the electric motor. At low speeds when the battery requires charging, the engine output is split between powering the generator and the vehicle. The possibility of splitting the engine output in this way at low vehicle speeds is the distinguishing feature of the dual mode hybrid configuration. This permits the engine to be operated near its maximum efficiency at all times and the battery to be recharged, when needed, regardless of the vehicle speed and power demand. The dual mode arrangement also reduces the need for on-off engine operation as required in the series and parallel control strategies. Dual mode hybrids are operated with the battery state-of-charge maintained in a narrow range (charge sustaining) and thus require no recharging of the battery from the wall-plug. The Toyota Prius hybrid vehicle uses this dual mode operating strategy.

PERFORMANCE OF HYBRID VEHICLES

The added complexity of the various hybrid vehicle designs relative to battery-powered electric vehicles and conventional engine-powered vehicles is evident. In order to justify this added complexity, hybrid vehicles must be more marketable than pure electric vehicles and have higher fuel economy and lower emissions than conventional engine-powered vehicles. All the vehicle types (electric, hybrid, and conventional) can be designed to have the same acceleration performance and top speed by the proper selection of the driveline components. Acceleration times of 0–96 km/hr (60 mph) in less than 8 sec and top speeds in excess of 120 km/hr for passenger cars have been demonstrated for both pure electric and hybrid vehicles. The primary advantage of the hybrid vehicle compared to the electric vehicle is that its range and refueling time can be the same or better than the conventional vehicle because it is refueled in the same manner—at the fuel pump. The key comparisons of interest are the fuel economy, emissions, and costs of hybrid and conventional vehicles.

Computer simulations have been performed for both midsize and compact, lightweight vehicle designs for driving on the Federal Urban Driving Schedule (FUDS) and the Federal Highway Driving Schedule (FHWDS). These driving cycles (speed vs. time) are intended to simulate vehicle operation in city and highway driving. For each vehicle type, computer simulations were run using gasoline fuel injected engines, diesel engines, and Stirling engines. Electricity was generated on board the hybrid vehicles by coupling the engines to a generator or by utilizing a fuel cell fueled using compressed hydrogen. In all cases, electrical energy was recovered into the energy storage unit during braking using the traction motor as a generator. The control strategies used in the simulations were essentially those previously discussed in the section on control strategies. In all cases, the engines and fuel cell were operated in an on-off mode to maintain the energy storage unit in the state-of-charge range specified by the control strategy. The minimum power setting of the engines and fuel cell when they were "on" were set so their efficiency was not outside the high efficiency portion of their operating maps.

Fuel Economy

Fuel economy simulation results for various engines in series hybrids are compared in Table 2 for the FUDS and FHWDS driving cycles. For both the midsize and compact cars, fuel economy depends significantly on the technology used in the driveline. The use of diesel engines results in the highest fuel economy (miles per gallon of diesel fuel); however, from the energy consumption (kJ/mi) and CO_2 emission (gm CO_2/mi) points-of-view, the advantage of diesel engine relative to gasoline-fueled engines should be discounted to reflect the higher energy and the carbon content per gallon of diesel fuel compared to gasoline. These discount factors are 15 to 20 percent. The simulation results also indicate that for the same type of engine, the fuel economy can be 10 to 20 percent higher using ultracapacitors in place of batteries as the energy storage device. The highest fuel economics are projected for vehicles using fuel cells. The fuel economies (gasoline equivalent) of the fuel cell vehicles using compressed hydrogen are

Vehicle	*Engine*	*Energy Storage*	*Miles per Gallon* *FUDS*	*Highway*
Midsize	Honda Gasoline	Ni. Mt. Hy. Bat.	36.1	45.4
	Direct Injection Gasoline	Ni. Mt. Hy. Bat.	47.3	56.0
	Sw. Ch. Diesel	Ni. Mt. Hy. Bat.	49.7	56.8
	Direct Injection Diesel	Ni. Mt. Hy. Bat.	60.5	71.1
	Stirling	Ni. Mt. Hy. Bat.	50.0	57.2
	Honda Gasoline	Capacitor	44.3	47.3
	Sw. Ch. Diesel	Capacitor	62.3	65.7
	Fuel Cell (H_2)	Ni. Mt. Hy. Bat.	89.2	105.0
Lightweight Compact	Honda Gasoline	Ni. Mt. Hy. Bat.	69.6	71.4
	Direct Injection Gasoline	Ni. Mt. Hy. Bat.	82.9	84.9
	Sw. Ch. Diesel	Ni. Mt. Hy. Bat.	98.2	95.6
	Direct Injection Diesel	Ni. Mt. Hy. Bat.	107.3	110.4
	Stirling	Ni. Mt. Hy. Bat.	89.5	92.7
	Honda Gasoline	Capacitor	81.4	75.5
	Sw. Ch. Diesel	Capacitor	109.8	104.0
	Fuel Cell (H_2)	Ni. Mt. Hy. Bat.	16.3	17.9

Table 2.
Summary of Hybrid Vehicle Fuel Economy Results on the FUDS and Highway Driving Cycles using Various Engines and a Fuel Cell
(1) mpg diesel fuel for diesel engine and mpg gasoline equivalent for fuel cell powered vehicles

about twice those of hybrid vehicles with direct injected gasoline engines and about 80 percent higher than vehicles with diesel engines. All the fuel cell vehicle designs utilized a fuel cell load-leveled with a nickel metal hydride battery permitting it to operate at high efficiency at all times.

In comparisons between the fuel economies of conventional passenger cars and those using series hybrid divelines, the hybrid vehicles have the same weight and road load as the conventional cars. Still, the utilization of the hybrid driveline resulted in about a 50 percent improvement in fuel economy for the FUDS cycle and about a 10 percent improvement on the FHWDS (highway cycle). The fuel economy of the conventional cars was taken from the EPA Fuel Economy Guide corrected by 10 percent for the FUDS and 22 percent for the highway cycle. These corrections were made, because the actual dynamometer fuel economy test data had been reduced by those factors so that the published fuel economies would be in better agreement with values experienced in the real world.

The fuel economy of series and parallel hybrid vehicles are compared in Table 3 for both the compact, lightweight, and midsize cars. The series hybrids are assumed to operate only in the charge sustaining mode (no battery recharging from the wall plug), but the parallel hybrids can operate in either the charge sustaining or charge depleting mode. In the case of the parallel hybrid in the charge depleting mode, the fuel economy is given for gasoline alone and at the powerplant (pp) including energy needed to recharge the batteries from the wall plug. For hybrid vehicles using gasoline engines (port injected), the fuel economy of the parallel hybrid vehicles in the charge sustaining mode (batteries charged from the engine—not from the wall plug) is 9 to 12 percent higher than that of the series hybrids. For the powerplant efficiency (33%) assumed in the calculations, the parallel hybrids operating in the charge depleting mode (battery charged only from the wall plug) had only 1 to 4 percent higher equivalent fuel economy than the same vehicle operating in the charge sustaining mode. If the batteries were recharged using electricity from a higher efficiency powerplant, the fuel economy advantage of the parallel hybrid in the charge depleting mode would be lighter.

Vehicle	*Fuel Economy (mpg) Gasoline Engine* *FUDS*	*Highway*	*Fuel Economy (mpg) Swirl Chamber Diesel* *FUDS*	*Highway*
Small, lightweight				
Series Hybrid				
Charge Sustaining	62.4	71.1	75.7	85.2
Parallel Hybrid				
Charge Sustaining	68.2	79.5	75.9	88.8
Charge Depleting				
gasoline alone	90.1	86.1	95.6	94.0
Including power plant	71.1	80.5	75.8	88.1
Midsize (1995 materials)				
Series Hybrid				
Charge Sustaining	39.2	48.2	47.9	58.5
Parallel Hybrid				
Charge Sustaining	42.8	54.1	48.0	60.6
Charge Depleting				
gasoline alone	55.3	56.6	59.0	62.3
Including power plant	45.4	54.7	49.1	60.6

Table 3.
Comparisons of the Fuel Economy for Series and Parallel Hybrid Vehicles
(1) Fuel economy shown for diesel engines is gasoline equivalent

Full Fuel Cycle Emissions

The full fuel cycle emissions of the hybrid vehicles are the total of all the emissions associated with the operation of the vehicle and the production, distribution, and dispensing of the fuel and electricity to the vehicle. The total emissions can be calculated for all the vehicle designs utilizing as inputs the vehicle simulation results for the electricity consumption, fuel economy, and exhaust emissions in the all-electric and hybrid modes and the upstream refueling, evaporative, and fuel production emissions based on energy usage—bothfuel and electricity. Both regulated emissions (nonmethane organic gases [NMOG], CO, NO_x) and CO_2 emissions can be calculated.

Regulated Emissions for Hybrid, Electric, and Conventional Cars

Hybrid vehicles operated in the charge depleting mode (battery charged from the wall plug) have total emissions comparable to those of electric vehicles if their all-electric range is 50 mi or greater. Hybrid vehicles operating in the charge sustaining mode have much greater NMOG emissions than electric vehicles when the refueling and evaporative emissions are included. The calculated total emissions of the electric vehicles are close to the equivalent zero emission vehicle (EZEV) emissions when the battery charging is done in the Los Angeles (LA) basin. These comparisons are based on the total NMOG, CO, and NO_x emissions for the FUDS and highway driving cycles for electric vehicles and conventional ICE vehicles as well as hybrid vehicles. A baseline use pattern of 7,500 miles per year random, city travel, and a round trip to work of 15 miles was assumed.

Total CO_2 Emissions

The difference in the CO_2 emissions between operating a hybrid vehicle in charge depleting and charge sustaining modes, regardless of its all-electric range, is not large using nickel metal hydride batteries. The CO_2 emissions of the gasoline and diesel engine powered hybrids vary only about 25%—not as much as might be expected based on the differences in their fuel economies—because of the higher energy content and the higher carbon-to-hydrogen ratio of the diesel fuel. The fuel cell powered, hydrogen fueled vehicles are projected to have the lowest CO_2 emissions by 25 to 30 percent when compared to the most efficient of the engine powered vehicles even when the hydrogen is produced by reforming natural gas. The CO_2 emissions of the conventional ICE vehicles are directly proportional to their fuel economy, which is projected to be significantly less than the hybrid vehicles. ICE powered vehicles will have low CO_2 emissions only when their fuel economy is greatly increased. CO_2 emissions of vehicles are highly dependent on the technologies used to power them and can vary by a factor of at least two for the same size and weight of vehicle.

TESTING OF HYBRID VEHICLES

There have been relatively few tests of hybrid vehicles in recent years. One example of such tests is that of the Toyota Prius by the EPA. Special care was taken in those tests to account for changes in the net state-of-charge of the batteries on the vehicle. The simulation results were obtained using the same hybrid vehicle simulation program used to obtain the fuel economy projections given in Table 2. There is good agreement between the measured and calculated fuel economies for the Prius. The EPA emissions data indicate that the CO and NO_x emissions of the Prius are well below the California ULEV standards and that the NMOG emissions are only slightly higher than the 0.04 gm/mi ULEV standard. The fuel economy of the Prius on the FUDS cycle (in city driving) is 56 percent higher than a 1998 Corolla (equipped with a 4-speed lockup automatic transmission) and 11 percent higher for highway driving. These two improvements in fuel economy for a hybrid/electric car compared to a conventional ICE car of the same size are consistent with those discussed earlier in the section on fuel economy.

PROSPECTS FOR MARKETING HYBRID CARS

The societal advantages of the hybrid-electric vehicles will come to fruition only when a significant fraction of vehicle purchasers decide to buy one of them. This will occur if the purchase of the hybrid vehicles makes economic sense to them and the vehicle meets their needs. Otherwise vehicle buyers will continue to purchase conventional ICE-powered vehicles. The key to any workable marketing strategy is the availability of hybrid driveline technologies that make the transition from engine-based to electric-based drivelines manageable and attractive to the consumer with only modest financial incentives. The state of development of the new driveline technologies at the time of introduction must be such that vehicles meet the needs of the first owners and they find vehicles to be reliable and cost-effective to operate. Otherwise the market for the new technologies will not increase and the introduction of the new technologies at that time will be counterproductive. Even after the technical and economic feasibility of a new technology is shown in prototype vehicles, a large financial commitment is needed to perform the preproduction engineering and testing of the vehicles before the vehicles can be introduced for sale.

Starting in the fall of 1997, Toyota offered for sale in Japan the Prius Hybrid at a price close to that of the comparable conventional car. The initial response of the public was enthusiastic and the production rate quickly rose to more than 1,000 vehicles per month. Toyota is planning to introduce a redesigned Prius in the United States in the fall of 2000. Honda began selling a subcompact hybrid/electric car, the Insight, in the United States in the fall of 1999 at a price of less than $20,000, which is about $5,000 higher than the Honda Civic. According the EPA tests, the corporate average fuel economy (CAFÉ) (combined city and highway cycles) of the Honda Insight is 76 mpg.

Recent advances in exhaust emission technologies have resulted in the certification by several auto manufacturers of conventional gasoline fueled cars that can meet the California ULEV and SULEV standards. It seems important that the driving force for the eventual introduction of the hybrid-electric and fuel cell cars will be improved fuel economy and lower CO_2 emissions and not lower regulated emission standards. It appears likely that a significant increase in the price of energy (either because of scarcity or higher taxes), regulation (for example, the CAFÉ standard), or financial incentives to purchase and license hybrid vehicles will be necessary before advanced technology hybrid vehicles become popular.

Andrew Burke

See also: Capacitors and Ultracapacitors; Drivetrains; Electric Motor Systems; Electric Vehicles; Engines; Fuel Cell Vehicles; Fuel Cells.

BIBLIOGRAPHY

Ahmed, S. (1997). "Partial Oxidation Reformer Development for Fuel Cell Vehicles." Proceedings of the 32nd Intersociety Energy Conversion Engineering Conference. Paper 97081 (August).

Burke, A. F. (1996). "Hybrid/Electric Vehicle Design Options and Evaluations." Paper 920447 (February). Warrendale, PA: Society of Automotive Engineers.

Burke, A. F., and Miller, M. (1997). "Assessment of the Greenhouse Gas Emission Reduction Potential of Ultra-clean Hybrid-Electric Vehicles." Institute of Trans-portation Studies. Report No. UCD-ITS-RR-97-24 (December). Davis: University of California.

Conway, B. E. (1999). *Electrochemical Capacitors: Scientific Fundamentals and Technological Applications*. New York. Kluwer Academic/Plenum Publishers.

Craig, P. (1997). The "Capstone Turbogenerator as an Alternative Power Source." Paper 970292 (February). Pittsburgh, PA: Society of Automotive Engineers.

U.S. Department of Energy. (1997). "Fuel Economy Guide—Model Year 1998." Publication DOE/EE-0102. Washington, DC: U.S. Government Printing Office.

Gottesfeld, S., and Zawodzinski, T. A. (1998). *Polymer Electrolyte Fuel Cells, Advances in Electrochemical Science and Engineering*, ed. R. Alkire et al. New York: Wiley.

Helman, K. H.; Peralta, M. R.; and Piotrowski, G. K. (1998). "Evaluation of a Toyota Prius Hybrid System (THS)." U.S. Environmental Protection Agency, Report No. EPA 420-R-98-006(August). Washington, DC: U.S. Government Printing Office.

Hermance, D. (1999). "The Toyota Hybrid System." Pres. At SAE TOPTEC, Hybrid Electric Vehicles: Here and Now (27 May).

Honda Multi-Matic News Release. (1997). "Technical Description of the System and its Operation." Torrance, CA: Author.

Honda Unveils World's Highest Mileage Hybrid Vehicle. (1999). *Honda Environmental News* (March). Torrance, CA: Author.

Howard, P. F. (1996). "The Ballard Zero-Emission Fuel Cell Engine." Prese. At Commercializing Fuel Cell Vehicles, InterTech Conference (September).

Johansson, L., and Rose, T. (1997). "A Stirling Power Unit for Series Hybrid Vehicles." Stirling Thermal Motors, Poster at the 1997 Customer Co-ordination Meeting (October).

Kluger, M. A., and Fussner, D. R. (1997). "An Overview of Current CVT Mechanisms, Forces, and Efficiencies." Paper 970688. Pittsburgh, PA: Society of Automotive Engineers.

Michel, S., and Frank, A. (1998). "Design of a Control Strategy for the Continuously Variable Transmission for the UC Davis Future Car (Parallel Hybrid Vehicle)." Department of Mechanical Engineering Report (September). Davis, University of California.

Moore, R. M.; Gottesfeld, S.; and Zelenay, P. (1999). "A Comparison Between Direct-Methanol and Direct Hydrogen Fuel Cell Vehicles." SAE Future Transportation Technologies Conference. Paper 99FTT-48 (August).

NECAR 4—DaimlerChrysler's Sixth Fuel Cell Vehicle in Five Years. (1999). Press Release. March 17.

Pepply, B. A., et al. (1997). "Hydrogen Generation for Fuel Cell Power Systems by High Pressure Catalytic Methanol Steam Reforming." Proceedings of the 32nd Intersociety Energy Conversion Engineering Conference. Paper 97093 (August).

Prater, K. B. (1996). "Solid Polymer Fuel Cells for Transport and Stationary Applications." *Journal of Power Sources* 61:105–109.

Society of Automotive Engineers. (1998). "Recommended Practice for Measuring the Exhaust Emissions and Fuel Economy of Hybrid-Electric Vehicles." Document J1711 (September). Pittsburgh, PA: Author.

Ren, X.; Springer, T. E.; and Gottesfeld, S. (1998). "Direct Methanol Fuel Cell: Transport Properties of the Polymer Electrolyte Membrane and Cell Performance." Vol. 98-27. Proc. 2nd International Symposium on Proton Conducting Membrane Fuel Cells. Pennington, NJ: Electrochemical Society.

Takaoka, T., et al. (1998). "A High-Expansion—Ratio Gasoline Engine for the Toyota Hybrid System." *Toyota Review* 47(2).

Toyota. (1997). "Toyota's Green Machine." *Business Week* (December 15):108–110.

Vaughn, M. (1997). "Reinventing the Wheel." *Autoweek* (March 3).

HYDROELECTRIC ENERGY

Hydroelectric energy—electric power created by the kinetic energy of moving water—plays an important role in supplying the world's electricity. In 1996, nearly 13 trillion kilowatt-hours of electricity were generated worldwide; almost one-fifth of this electricity was produced with hydroelectricity. On average hydropower provides about ten percent of the U.S. electricity supply, although the annual amount of electricity generated by hydroelectric resources varies due to fluctuations in precipitation. In many parts of the world, reliance on hydropower is much higher than in the United States. This is particularly true for countries in South America where abundant hydroelectric resources exist. In Brazil, for example, 92 percent of the 287 billion kilowatt-hours of electricity generated in 1996 were generated by hydroelectricity.

There are many benefits for using hydro resources to produce electricity. First, hydropower is a renewable resource; oil, natural gas, and coal reserves may be depleted over time. Second, hydro resources are indigenous. A country that has developed its hydroelectric resources does not have to depend on other nations for its electricity; hydroelectricity secures a country's access to energy supplies. Third, hydroelectricity is environmentally friendly. It does not emit greenhouse gases, and hydroelectric dams can be used to control floods, divert water for irrigation purposes, and improve navigation on a river.

There are, however, disadvantages to developing hydroelectric power. Hydroelectric dams typically require a great deal of land resources. In conventional hydroelectric projects, a dam typically is built to create a reservoir that will hold the large amounts of water needed to produce power. Further, constructing a hydroelectric dam may harm the ecosystem and affect the population surrounding a hydro project. Environmentalists often are concerned about the adverse impact of disrupting the flow of a river for fish populations and other animal and plant species. People often must be relocated so that a dam's reservoir may be created. Large-scale dams cause the greatest environmental changes and can be very controversial. China's 18.2 gigawatt Three Gorges Dam project—the world's largest hydroelectric—will require the relocation of an estimated 1.2 million people so that a 412-mile reservoir can be built to serve the dam.

Another potential problem for hydroelectricity is the possibility of electricity supply disruptions. A severe drought can mean that there will not be enough water to operate a hydroelectric facility. Communities with very high dependence on the hydroelectric resources may find themselves struggling with electricity shortages in the form of brownouts and black-outs.

This article begins with a description of how hydroelectricity works, from the beginning of the hydrological cycle to the point at which electricity is transmitted to homes and businesses. The history of the dam is outlined and how dams evolved from structures used for providing a fresh water supply to irrigation and finally to providing electricity. The history of hydropower is considered and the different hydroelectric systems (i.e., conventional, run-of-river, and pumped storage hydroelectricity) currently in use are discussed.

HOW DOES HYDROELECTRICITY WORK?

Hydroelectricity depends on nature's hydrologic cycle (Figure 1). Water is provided in the form of rain, which fills the reservoirs that fuel the hydroelectric plant. Most of this water comes from oceans, but rivers and lakes and other, smaller bodies of water also contribute. The heat of the sun causes the water from these sources to evaporate (that is, to change the water from its liquid state into a gaseous one). The water remains in the air as an invisible vapor until it condenses and changes first into clouds and eventually into rain. Condensation is the opposite of evaporation. It occurs when the water vapor changes from its gaseous state back into its liquid state.

Condensation occurs when air temperatures cool. The cooling occurs in one of two ways. Either the air vapor cools as it rises and expands or as it comes into contact with a cool object such as a cold landmass or an ice-covered area. Air rises for several reasons. It can be forced up as it encounters a cooler, denser body of air, or when it meets mountains or other raised land masses. It can rise as it meets a very warm surface, like a desert, and become more buoyant than the surrounding air. Air also can be forced to rise by storms—during tornadoes particles of air circling to the center of a cyclone collide and are forced up. When the water vapor collides with a cold object, it can become fog, dew, or frost as it condenses. The vapor cools as it rises into the atmosphere and condenses to form clouds and, sometimes, rain.

In order for rain to form, there must be particles in the air (i.e., dust or salt) around which the raindrop can form and which are at temperatures above freezing. When the particles are cooled to temperatures below the freezing point water condenses around them in layers. The particles grow heavy enough that they eventually fall through the clouds in the form of raindrops or— if the air temperature is below the freezing point all the way to the ground — as snow, sleet, or hail.

Much of the rain that reaches the ground runs off the surface of the land and flows into streams, rivers, ponds and lakes. Small streams lead to bigger ones, then to rivers, and eventually back to the oceans where the evaporation process begins all over again.

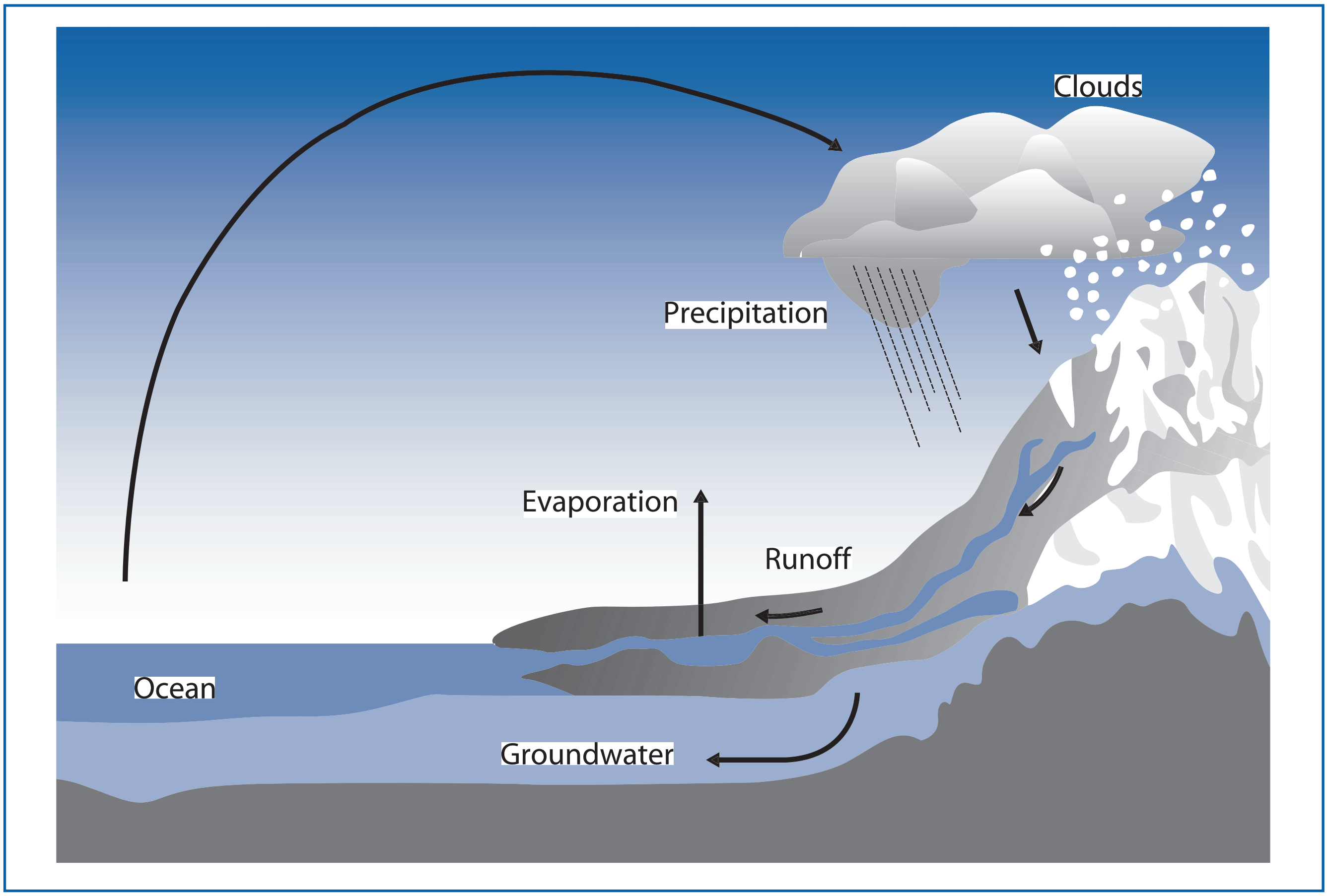


Figure 1.
The Hydrologic Cycle.
SOURCE: U.S. Department of Energy Hydropower Program.

Although water continuously changes from solid to liquid to gas, the amount of water on the earth remains the same; there is as much water today as there was hundreds of millions of years ago.

This water cycle—the process of moving water from oceans to streams and back again— is essential to the generation of hydroelectricity. Moving water can be used to perform work and, in particular, hydroelectric power plants employ water to produce electricity. The combination of abundant rainfall and the right geographical conditions is essential for hydroelectric generation.

Hydroelectric power is generated by flowing water driving a turbine connected to an electric generator. The two basic types of hydroelectric systems are those based on falling water and those based on natural river current, both of which rely on gravitational energy. Gravitational forces pull the water down either from a height or through the natural current of a river. The gravitational energy is converted to kinetic energy. Some of this kinetic energy is converted to mechanical (or rotational) energy by propelling turbine blades that activate a generator and create electricity as they spin.

The amount of energy created by a hydroelectric project depends largely upon two factors: the pressure of the water acting on the turbine and the volume of water available. Water that falls 1,000 feet generates about twice as much electric power as the same volume of water falling only 500 feet. In addition, if the amount of water available doubles, so does the amount of energy.

The falling water hydrosystem is comprised of a dam, a reservoir, and a power generating unit (Figure

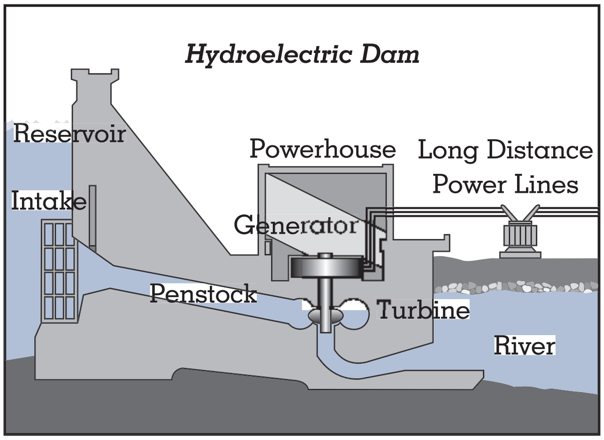


Figure 2.
Conventional, Falling Water Hydroelectric System.
SOURCE: U.S. Department of Energy Hydropower Program.

2). The dam is constructed so that a reservoir is created within which water accumulates and may be stored until it is needed. Water is released as required to meet electricity demands of customers. At the bottom of the dam wall is the water intake. The water intake controls when and how much water is moved into a steel called the "penstock." Gravity causes the water to fall through the penstock. This pipe delivers the running water to a turbine—a propeller-like machine with blades like a large fan. The water pushes against the turbine blades and the blades turn. Because the turbine is connected to an electric generator, as the turbine gains speed, it powers the generator and electricity is produced. The largest falling water facility in the United States is the Grand Coulee hydroelectric project on the Columbia River, in Washington State. Indeed, the largest power production facilities in North America are found at the Grand Coulee Dam project where an average 21 billion kilowatt-hours of electricity are produced each year.

The second type of hydroelectric plant is called a run-of-the-river system. In this case, the force of the river current applies pressure to the turbine blades to produce electricity. Run-of-the-river systems do not usually have reservoirs and cannot store substantial quantities of water. As a result, power production from this type of system depends on the river flow—the electricity supply is highly dependent upon seasonal fluctuations in output. Run-of-river projects are most successful when there are large flows in flat rivers or when a high natural geological drop is present, and when the required electricity output is below the maximum potential of the site.

Hydropower systems, as in all electricity-producing systems, require a generator to create the electricity. An electric generator is a device that converts mechanical energy into electric energy. The process is based on the relationship between magnetism and electricity. When a wire or any other electrically conductive material moves across a magnetic field, an electric current occurs in the wire. In a power plant, a strong electromagnet, called a rotor, is attached to the end of a shaft. The shaft is used to spin the rotor inside a cylindrical iron shell with slots—called a stator. Conducting wires are wound through the slots of the stator. When the rotor spins at a high rate of speed, an electric current flows through the conducting wires.

In a hydroelectric system, flowing water is used to propel a turbine that spins the shaft connected to the generator and creates the electric current. The kinetic energy of the moving water is changed into rotational energy and thereby causes the turbine blades to rotate. The turbine is attached to an electric generator and the rotational energy of the turbine is then converted to electric energy. The electricity then leaves the generator and is carried to the transformers where the electricity can travel through electric power lines and is supplied to residential, commercial, and industrial consumers. After the water has fallen through the turbine, it continues to flow downriver and to the ocean where the water cycle begins all over again.

Pumped storage hydroelectricity is an extended version of the falling water hydroelectric system. In a pumped storage system, two water sources are required—a reservoir located at the top of the dam structure and another water source at the bottom. Water released at one level is turned into kinetic energy by its discharge through high-pressure shafts that direct the downflow through the turbines connected to the generator. The water flows through the hydroelectric generating system and is collected in a lower reservoir. The water is pumped back to the upper reservoir once the initial generation process is complete. Generally this is done using reversible turbines—that is, turbines that can operate when the direction of spinning is reversed. The pump motors are powered by conventional electricity from the national grid. The pumping process usually occurs overnight when electricity demand is at its lowest. Although the pumped storage sites are not net energy producers—pumped storage sites use more energy pumping the water up to the higher reservoir than is recovered when it is released—they are still a valuable addition to electricity supply systems. They offer a valuable reserve of electricity when consumer demand rises unexpectedly or under exceptional weather conditions. Pumped storage systems are normally used as "peaking" units.

HISTORY AND EVOLUTION OF HYDROELECTRIC DAMS

Dams have existed for thousands of years. The oldest known dam, the Sadd el-Kafara (Arabic for "Dam of the Pagans"), was constructed over 4,500 years ago twenty miles south of Cairo, Egypt. The 348-feet wide, 37-feet high dam was constructed to create a reservoir in the Wadi el-Garawi. The dam was built with limestone blocks, set in rows of steps about eleven inches high. It appears that the dam was supposed to be used to create a reservoir that would supply drinking water for people and animals working in a nearby quarry. Scientists and archaeologists believe that the Sadd el-Kafara failed after only a few years of use because there is no evidence of siltation at the remains of the dam. When water flows into the reservoir created by a dam, the silt—sand and other debris carried with the stream—is allowed to settle, rather than be borne further downstream by the force of the flow of a stream or river. In the still water behind a dam this sediment is deposited on the bottom of the reservoir.

The earliest dams were built for utilitarian purposes, to create reservoirs for drinking supplies—as in the Sadd el-Kafara—or to prevent flooding or for irrigation purposes. Early dams were even used to create lakes for recreational purposes. In the middle of the first century, the Roman emperor, Nero, constructed three dams to create three lakes to add to the aesthetic beauty of his villa.

Today a variety of dam structures are utilized. These can be classified as either embankment dams or concrete dams. Embankment dams are constructed with locally available natural resources. There are several different types of embankment dams, including earth dams—which are constructed primarily of compacted earth; tailings dams—constructed from mine wastes; and rockfill dams—which are constructed with dumped or compacted rock. The shape of these dams is usually dependent on the natural settling angle of the materials used to build them. Concrete, bitumen, or clay is often used to prevent water from seeping through the dam. This can be in the form of a thin layer of concrete or bitumen facing which acts as a seal, or in the form of a central core wall of clay or other fine materials constructed within the dam, allowing water to penetrate the upstream side of the structure, but preventing it from moving beyond the clay core.

Concrete dams are more permanent structures than embankment dams. Concrete dams can be categorized as gravity, arch, or buttress. Because concrete is a fairly expensive material, different construction techniques were developed to reduce the quantity of concrete needed. This is highly dependent upon geological considerations. In particular, the rock foundations must be able to support the forces imposed by the dam, and the seismic effects of potential earthquakes. Gravity dams work by holding water back by way of their own weight. A gravity dam can be described as a long series of heavy vertical, trapezoidal structural elements firmly anchored at the base. It is generally a straight wall of masonry that resists the applied water-pressure by its sheer weight. The strength of a gravity dam ultimately depends on its weight and the strength of its base.

An arch dam, on the other hand, relies on its shape to withstand the pressure of the water behind it. The arch curves back upstream and the force exerted by the water is transferred through the dam into the river valley walls and to the river floor. They are normally constructed in deep gorges where the geological foundations are very sound. The United States's Hoover Dam is an example of a concrete arch dam.

The buttress dam uses much less concrete than the gravity dam, and also relies on its shape to transfer the water load. Buttress dams were developed in areas where materials were scarce or expensive, but labor was available and cheap. They have been built primarily for purposes of irrigation. The dams are particularly suited for wide valleys. They have a thin facing supported at an incline by a series of buttresses. The buttresses themselves come in a variety of shapes, including the multiple arch and the simple slab deck (see Figure 3). The weight of the concrete is transferred to bedrock through the downstream legs or "buttresses" of the structure. The Coolidge Dam near Globe, Arizona is an example of a buttress dam, constructed of three huge domes of reinforced concrete.

Using water as a source of power actually dates back more than 2,000 years when the Greeks used water to turn wheels to grind wheat into flour. The first recorded use of water power was a clock, built around 250 B.C.E. and since that time, falling water has provided power to grind corn, wheat, and sugar cane, as well as to saw mills. In the tenth century, water wheels were used extensively in the Middle East for milling and irrigation. One of these dams, built at Dizful—in what is now Iran—raised the water 190 feet and supplied the residents with water to grind corn and sugar cane.

The discoveries associated with electromagnetism in the early nineteenth century had a major impact on the development of hydropower. The development of electric power generators and the fact that electric power was the only form of energy in a "ready to use" state which can be transmitted over long distances has been particularly significant to hydroelectricity. Dams located away from population centers could be useful generators if there was a way to supply the consumers. Water turbines were also developed during the nineteenth century as a natural successor to the water wheel. The high performance and small size of the turbine relative to the water wheel were important advancements. Combining the technology of electric generators, turbines, and dams resulted in the development of hydroelectric power.

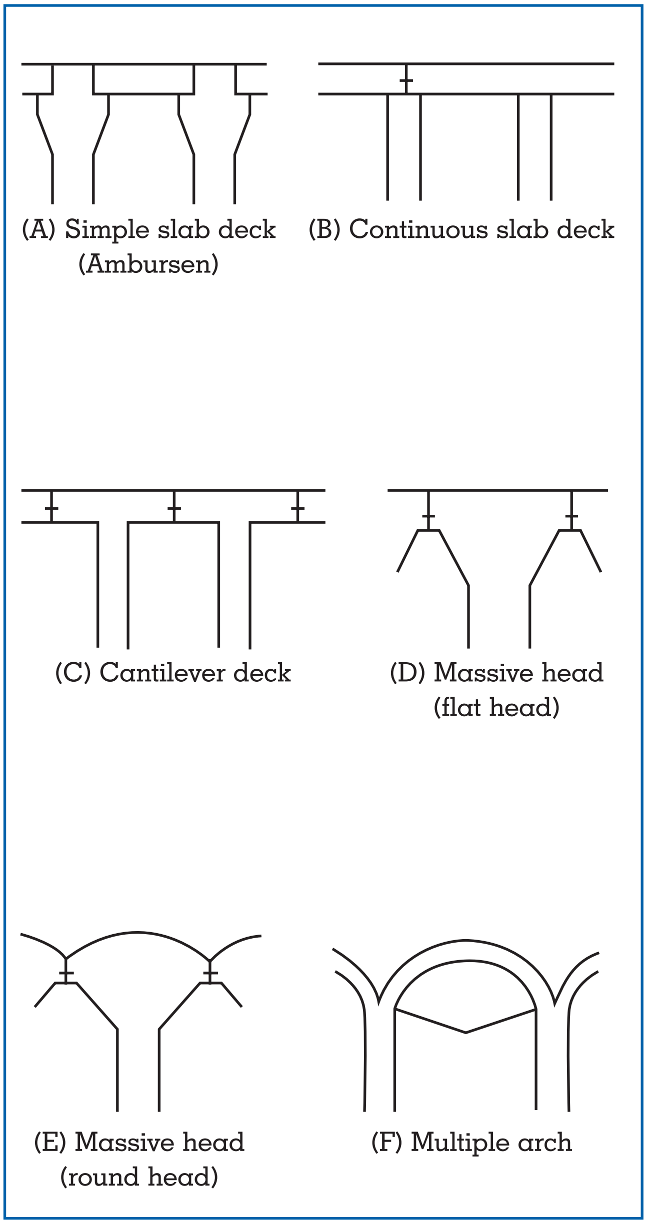


Figure 3.
Types of buttress dams.
SOURCE: University of Colorado, Department of Civil, Environmental, and Architectural Engineering.

Water was first used to generate electricity in 1880 in Grand Rapids, Michigan when a water turbine was used to provide storefront lighting to the city. In 1882—only two years after Thomas Edison demonstrated the incandescent light bulb—the first hydroelectric station to use Edison's system was installed on the Fox River at Appleton, Wisconsin. In 1881, construction began on the first hydroelectric generating station on the Niagara River in Niagara Falls on the New York-Canadian border. For twenty years, this project met the small electricity needs of the city. Water from the upper Niagara River fell 86 feet down a flume (a narrow gorge with a stream running through it) onto spinning water wheels on the lower river to generate electricity. The electricity was used to run the equipment of a paper company, other small factories, and sixteen open arc-lights on village streets. By 1896, the first long-distance transmission of electricity allowed the Niagara Falls project to provide electricity to Buffalo, New York—some 26 miles away.

By the early twentieth century, hydroelectric power was providing more than 40 percent of electricity generation in the United States. In 1940, hydropower supplied about three-fourths of all the electricity consumed in the West and Pacific Northwest, and still supplied about one-third of the total U.S. electricity supply. Although hydroelectricity's share of total electricity generation has since fallen to about 10 percent in the United States, hydroelectricity provides almost one-fifth of the world's total electricity generation today.

Hydroengineering has evolved over the past century so that it has become possible to build larger and larger hydroelectric projects. In 2000, the largest hydroelectric project in the United States was the Grand Coulee power plant on the Columbia river in Washington State. Grand Coulee is also the third largest currently-operating hydroelectric project in the world. The Grand Coulee project began operating in 1941 with 20 megawatts of installed capacity. The project has been expanded so that in 2000 it operated with an installed capacity of 6,180 megawatts.

The world's largest hydroelectric plant is the Itaipú power plant, located on the Paraná river separating Brazil and Paraguay. This 12,600-megawatt hydroelectric project is jointly-owned by Brazil and Paraguay. It is comprised of eighteen generating units, each with an installed capacity of 700 megawatts. The plant produces an estimated 75 billion kilowatt-hours each year. In 1994, Itaipú supplied 28 percent of all the electricity consumed in Brazil's south, southeast, and central-west regions, and 72 percent of Paraguay's total energy. Construction on this mammoth project began in 1975 and was not completed until 1991 at a cost of about $18 billion (U.S.). It required fifteen times more concrete than the Channel Tunnel that connects France and the United Kingdom; and the amount of steel and iron used in its construction would have built 380 Eiffel towers. The main dam of Itaipú is a hollow gravity type dam, but a concrete buttress dam and two embankment dams were also incorporated into the system design. At its highest point, the main dam is 643 feet—more than twice as high as the Statue of Liberty; it extends nearly five miles across the Paraná River.

The large-scale disruptions to the environment of projects like the Itaipú hydroelectric system are profound and this is one of the major disadvantages to constructing large-scale hydroelectric facilities. As a result of construction on Itaipú, over 270 square miles of forest land has been negatively impacted, mostly on the Paraguayan side of the Paraná River where an estimated 85 percent of the forest was destroyed during the early years of construction. Despite programs to minimize the environmental damage through migration to reserves of the wildlife and plants facing extinction, several plant species became extinct mostly because they did not survive the transplant process.

Itaipú no longer represents the upper limit of large-scale hydroelectric expansion. At the end of 1994, the Chinese government announced the official launching of construction on the Three Gorges Dam hydroelectric project. When completed in 2009, this 18,200-megawatt project will provide the same amount of electricity as thirty 600-megawatt coal-fired plants. Project advocates expect the dam to produce as much as 85 billion kilowatt-hours of electricity per year, which is still only about 10 percent of the 881 billion kilowatt-hours of electricity consumed by China in 1995. It will also be used to control flooding along the Yangtze River and will improve navigational capacity, allowing vessels as large as 10,000 tons to sail upstream on the Yangtze as far as Chongqing, 1,500 miles inland from Shanghai. It is a very controversial project that will require the relocation of an estimated 1.2 million people, and that will submerge thirteen cities, 140 towns, 1,352 villages, and some 650 factories in its 412-mile reservoir that is to be created to support the dam. Disrupting the flow of the river will place several rare plant and animals at risk, including the endangered Yangtze River dolphin. Some environmentalists believe the reservoir created for the Three Gorges Dam project will become a huge pollution problem by slowing the flow of the Yangtze River and allowing silt to build up, possibly clogging the planned harbor at Chongqing within a few decades. The World Bank and the U.S. Export-Import Bank have refused to help finance the project, primarily because of the adverse environmental effects this massive hydroelectric project might have. Opponents to the Three Gorges Dam project have stated that smaller dams built on Yangtze River tributaries could produce the same amount of electricity and control flooding along the river, without adversely impacting the environment and at a substantially lower price than this large-scale project.

Hydroelectric projects can have an enormous impact on international relations between surrounding countries. For instance, the controversy between Hungary and Slovakia over the Gabcĺkovo dam began at the dam's opening in 1992 and as of 2000 had not been completely resolved, despite a ruling by the International Court of Justice at The Hague in 1997. In its first major environmental case, the Court ruled that both countries were in breach of the 1977 treaty to construct two hydroelectric dams on the Danube River—the Slovakian Gabcĺkovo and, 80 miles to the south, the Hungarian Nagymaros. Hungary suspended work on its portion in 1989 due to protests from the populace and international environmental groups. In 1992, Czechoslovakia decided to complete Gabcĺkovo without Hungarian cooperation. Today, the $500 million, 180-megawatt project supplies an estimated 12 percent of the electricity consumed in Slovakia.

The Court stated that Hungary was wrong to withdraw from the treaty, but that Slovakia also acted unlawfully by completing its part of the project on its own. After almost a year of talks to resolve the differences between the two countries, Hungary and Slovakia agreed to construct a dam either at Nagymaros—the original site of the Hungarian portion of the project—or at Pilismarot. Unfortunately, soon after the agreement was signed, environmental protests began anew and little progress has been made to resolve the situation to everyone's satisfaction.

Another illustration of the political ramifications often associated with constructing dam projects concerns Turkey's plans to construct the large-scale Southeast Anatolia (the so-called GAP) hydroelectric and irrigation project. When completed, GAP will include twenty-one dams, nineteen hydroelectric plants (generating 27 billion kilowatthours of electricity), and a network of tunnels and irrigation canals. Neighboring countries, Syria and Iraq, have voiced concerns about the large scale of the hydroelectric scheme. Both countries argue that they consider the flow of the historic rivers that are to be affected by GAP to be sacrosanct. Scarce water resources in many countries of the Middle East make the disputes over the resource and the potential impact one country may have on the water supplies of another country particularly sensitive.

In the United States, in 2000 there were still over 5,600 undeveloped hydropower sites with a potential combined capacity of around 30,000 megawatts, according to estimates by the U.S. Department of Energy. It is, however, unlikely that a substantial amount of this capacity will ever be developed. Indeed, there is presently a stronger movement in this country to dismantle dams and to restore the natural flow of the rivers in the hopes that ecosystems damaged by the dams (such as fish populations that declined when dams obstructed migratory patterns) may be repaired. The U.S. Department of Interior (DOI) has been working to decommission many hydroelectric dams in the country and to restore rivers to their pre-dam states. In 1997, the Federal Energy Regulatory Commission (FERC) ordered the removal of the 160-year old Edwards Dam on the Kennebec River in Augusta, Maine and the demolition of the dam began in July 1999. Although Edwards Dam has only a 3.5-megawatt installed generating capacity and provided less than one-tenth of one percent of Maine's annual energy consumption, the event is significant in that it represented the first time that FERC had used its dam-removal authority and imposed an involuntary removal order.

In 1998, DOI announced an agreement to remove the 12-megawatt Elwha Dam near Port Angeles, Washington, but removal has been delayed indefinitely because Congress withheld the funds needed to finance the project. Some small dams in the United States have been successfully removed—such as the 8-foot Jackson Street Dam used to divert water for irrigation in Medford, Oregon and Roy's Dam on the San Geronimo Creek outside San Francisco, California. However, efforts to dismantle many of the larger dams slated for removal in the United States have been delayed by Congressional action, including four Lower Snake River dams and a partially built Elk Creek Dam in Oregon.

Linda E. Doman

See also: Electric Power, Generation of; Electric Power Transmission and Distribution Systems; Magnetohydrodynamics.

BIBLIOGRAPHY

The American University, School of International Service (1997). <http://www.gurukul.ucc.american.edu/salla/sis-home.htm>.

Baird, S. (1993). *Energy Fact Sheet: Hydro-Electric Power.* Ontario: Energy Educators of Ontario.

Biolchini dos Santos, M., and de Assis, R. (1998). *Information System of the Brazilian Embassy in London (INFOLONDRES)* "Brazil in the School." <http://www.brazil.org.uk>.

Blue Bridge Enterprises, Inc. (1996). "The Three Gorges Dam—II: The Super Dam Along the Yangtze River." New York: Author.

California State Government. (2000). *Energy Education from California Energy Commission.* "The Energy Story — Hydro." <http://www.energy.ca.gov>.

First Hydro Company. (1998). "About Pumped Storage." <http://www.fhc.co.uk>.

Hydro Tasmania. (1999). *Water Power: An Introduction to Hydro-electricity in Tasmania.* <http://www.hydro.com.au>.

Mufson, S. (1997). "The Yangtze Dam: Feat or Folly?" *The Washington Post.* (November 7).

National Hydropower Association. (2000). *Hydro Facts.* <http://www.hydro.org>.

New York Power Authority. (1999). *NYPA: Our History: The Niagara Power Project.* <http://www.nypa.gov>.

Powercor Australia, Ltd. (1997). *At School.* <http://www.powercor.com.au.>.

Smith, Norman (1972). *A History of Dams.* Seacaucus, NJ: Citadel Press.

Tyson, J. L. (1997). "Ardent Foe Takes on China Dam." *The Christian Science Monitor.* (November 12).

U.S. Bureau of Reclamation, Power Program. (1999). *The Hydropower Industry.* <http://www.usbr.gov>.

U.S. Department of Energy, Energy Information Administration. (1999). *International Energy Annual 1997.* Washington, DC: U.S. Government Printing Office.

U.S. Department of Energy, Energy Information Administration. (1998). *Electric Power Annual 1997: Volume I.* Washington, DC: U.S. Government Printing Office.

U.S. Department of Energy, Energy Information Administration. (1997). *Inventory of Power Plants in the United States as of January 1, 1998.* Washington, DC: U.S. Government Printing Office.

U.S. Department of Energy, Energy Information Administration. (1998). *International Energy Outlook 1998.* Washington, DC: U.S. Government Printing Office.

U.S. Department of Energy, Idaho National Engineering and Environment Laboratory (2000) *Hydropower Program* <http://www.inel.gov>.

U.S. Department of Interior. (2000). *Hoover Dam: National Historic Landmark.* <http://www.hooverdam.com>.

U. S. Geological Survey. (1998). *Water Science for Schools.* "Hydroelectric Power: How it Works." <http://ga.water.usgs.gov/edu>.

University of Colorado at Boulder, Department of Civil, Environmental, and Architectural Engineering. (1999). *Intro to Civil and Architectural Engineering.* "Geo Applications: Dams." <http://bechtel.colorado.edu/courseware/mos-cven/ce/3.2/texts/dams.html>.

HYDROGEN

Hydrogen is a high-quality energy carrier that can be employed with high conversion efficiency and essentially zero emissions at the point of use. Hydrogen can be made from a variety of widely available primary energy sources, including natural gas, coal, biomass (agricultural or forestry residues or energy crops), wastes, solar, wind, or nuclear power. Technologies for production, storage, and transmission of hydrogen are well established in the chemical industries. Hydrogen transportation, heating, and power generation systems have been technically demonstrated, and in principle, hydrogen could replace current fuels in all their present uses. If hydrogen is made from renewable or decarbonized fossil sources, it would be possible to produce and use energy on a large scale system with essentially no emissions of air pollutants (nitrogen oxides, carbon monoxide, sulfur oxides, volatile hydrocarbons or particulates) or greenhouse gases during fuel production, transmission, or use. Because of hydrogen's desirable environmental characteristics, its use is being proposed to reduce emissions of air pollutants and greenhouse gases. However, technical and economic challenges remain in implementing a hydrogen energy system.

CHARACTERISTICS OF HYDROGEN AS A FUEL

Table 1 summarizes some physical characteristics of hydrogen relevant to its use as a fuel. Specifically,

- Hydrogen is a low-density gas at ambient conditions. It liquefies at –253°C. For storage hydrogen must be compressed or liquefied.
- Hydrogen has the highest heating value per kilogram of any fuel (which makes it attractive as rocket fuel), although it has the lowest molecular weight. The specific heat of hydrogen is higher than that of other fuels.
- Hydrogen has a wider range of flammability limits and a lower ignition energy than other fuels.

- Hydrogen can be used with very little or no pollution for energy applications. When hydrogen is burned in air, the main combustion product is water, with traces of nitrogen oxides. When hydrogen is used to produce electricity in a fuel cell, the only emission is water vapor.

HISTORICAL PERSPECTIVE

Although it is not considered a commercial fuel today, hydrogen has been used for energy since the 1800s. Hydrogen is a major component (up to 50% by volume) of manufactured fuel gases ("town gas") derived from gasification of coal, wood, or wastes. Town gas was widely used in urban homes for heating and cooking in the United States from the mid-1800s until the 1940s, and is still used in many locations around the world (including parts of Europe, South America, and China) where natural gas is unavailable or costly. Hydrogen-rich synthetic gases also have been used for electric generation. Hydrogen is an important feedstock for oil refining, and indirectly contributes to the energy content of petroleum-derived fuels such as gasoline. Liquid hydrogen also is a rocket fuel, and it has been proposed as a fuel for supersonic aircraft. Despite hydrogen's many applications, primary energy use of hydrogen for energy applications (including oil refining) is perhaps 1 percent of the world total.

The idea of boosting hydrogen's role—a "hydrogen economy" or large-scale hydrogen energy system—has been explored several times, first in the 1950s and 1960s as a complement to a largely nuclear electric energy system (where hydrogen was produced electrolytically from off-peak nuclear power), and later, in the 1970s and 1980s, as a storage mechanism for intermittent renewable electricity such as photovoltaics, hydroelectric and wind power. More recently, the idea of a hydrogen energy system based on production of hydrogen from fossil fuels with separation and sequestration (e.g., secure storage underground in depleted gas wells or deep saline aquifers) of by-product CO_2 has been proposed.

Concerns about global climate change have motivated new interest in low-carbon or noncarbon fuels. Recent rapid progress and industrial interest in low-temperature fuel cells (which prefer hydrogen as a fuel) for transportation and power applications have also led to a reexamination of hydrogen as a fuel.

Molecular weight of H_2 (g/mole)	2.016
Mass density of H_2 gas (kg/Nm^3) at standard conditions (P=1 atm=0.101 MPa, T=0°C)	0.09
Higher Heating Value (MJ/kg)	141.9
Lower Heating Value (MJ/kg)	120.0
Gas Constant (kJ/kg/°K)	4.125
Specific heat (c_p) (kJ/kg/°K) at 20°C	14.27
Flammability Limits in air (% volume)	4.0-75.0
Detonability Limits in air (% volume)	18.3-59.0
Diffusion velocity in air (meter/sec)	2.0
Buoyant velocity in air (meter/sec)	1.2-9.0
Ignition energy at stoichiometric mixture (milliJoules)	0.02
Ignition energy at lower flammability limit (milliJoules)	10
Temperature of liquefaction (°C)	-253
Mass density of liquid H_2 at -253°C (kg/Nm^3)	70.9
Toxicity	non-toxic

Table 1
Physical Properties of Hydrogen

HYDROGEN ENERGY TECHNOLOGIES

Hydrogen Production

Hydrogen is the most abundant element in the universe and is found in a variety of compounds, including hydrocarbons (e.g., fossil fuels or biomass) and water. Since free hydrogen does not occur naturally on earth in large quantities, it must be produced from hydrogen-containing compounds.

More than 90 percent of hydrogen today is made thermochemically by processing hydrocarbons (such as natural gas, coal, biomass, or wastes) in high-temperature chemical reactors to make a synthetic gas or "syngas," comprised of hydrogen, CO, CO_2, H_2O and CH_4. The syngas is further processed to increase the hydrogen content, and pure hydrogen is separated out of the mixture. An example of making hydrogen thermochemically from natural gas is shown in Figure 1.

Where low-cost electricity is available, water electrolysis is used to produce hydrogen. In water electrolysis

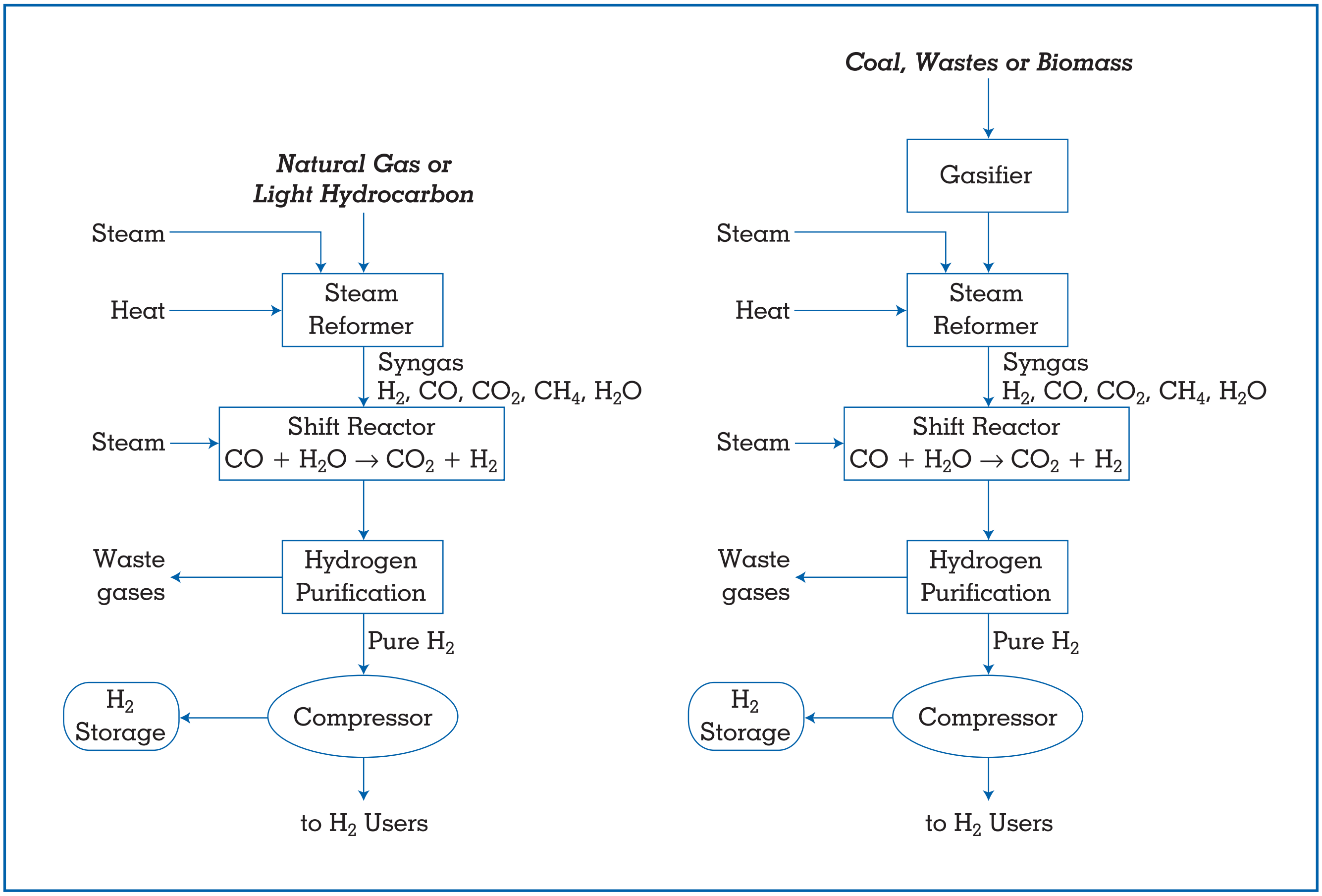


Figure 1.
Thermochemical Hydrogen Production Methods.

electricity is passed through a conducting aqueous electrolyte, breaking down water into its constituent elements, hydrogen and oxygen (see Figure 2). The hydrogen can be compressed and stored for later use. Any source of electricity can be used, including intermittent (time-varying) sources such as off-peak power and solar or wind energies. Fundamental research is being conducted on experimental methods of hydrogen production, including direct conversion of sunlight to hydrogen in electrochemical cells, and hydrogen production by biological systems such as algae or bacteria.

Hydrogen Transmission and Distribution

The technologies for routine handling of large quantities of hydrogen have been developed in the chemical industry. Hydrogen can be liquefied at low temperature (-253°C) and delivered by cryogenic tank truck or compressed to high pressure and delivered by truck or gas pipelines.

If hydrogen were widely used as an energy carrier, it would be technically feasible to build a hydrogen pipeline network similar to today's natural-gas pipeline system. It has been suggested that the existing natural-gas pipeline system might be converted to transmit hydrogen. With modifications of seals, meters, and end-use equipment, this could be done if pipeline materials were found to be compatible. However, hydrogen embrittlement (hydrogen-induced crack growth in pipeline steels that are subject to changes in pressure) could be an issue, especially for long-distance gas pipelines. Rather than retrofitting existing pipelines, new hydrogen pipelines might be built utilizing existing rights-of-way.

Hydrogen Storage

Unlike gasoline or alcohol fuels, which are easily handled liquids at ambient conditions, hydrogen is a

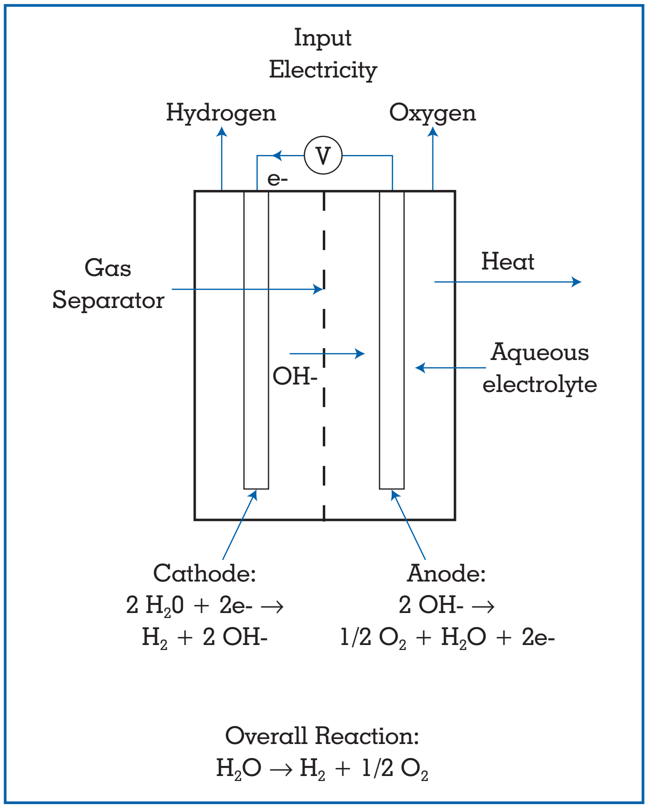


Figure 2.
Hydrogen Production Via Water Electrolysis.

lightweight gas and has the lowest volumetric energy density of any fuel at normal temperature and pressure. Thus hydrogen must be stored as a compressed gas (in high-pressure gas cylinders), as a cryogenic liquid at –253°C (in a special insulated dewar), or in a hydrogen compound where the hydrogen is easily removed by applying heat (such as a metal hydride). All these storage methods for hydrogen are well known in the chemical industry. Innovative storage methods such as hydrogen adsorption in carbon nanostructures are being researched.

Hydrogen onboard storage systems for vehicles are bulkier, heavier, and costlier than those for liquid fuels or compressed natural gas, but are less bulky and less heavy than presently envisaged electric batteries. Even with these constraints, it appears that hydrogen could be stored at acceptable cost, weight, and volume for vehicle applications. This is true because hydrogen can be used so efficiently that relatively little fuel is needed onboard to travel a long distance.

HYDROGEN FOR TRANSPORTATION AND POWER APPLICATIONS

Hydrogen Engines

Hydrogen engines resemble those using more familiar fuels, such as natural gas, gasoline, or diesel fuel. There are several key differences: (1) the emissions from a hydrogen engine are primarily water vapor with traces of nitrogen oxides, (2) the engine can be made more energy-efficient than with other fuels, and (3) there is much less need for postcombustion clean-up systems such as catalytic converters. Use of hydrogen in engines would substantially reduce pollutant emissions and would improve efficiency somewhat.

While hydrogen engines have some advantages over natural-gas engines, the hydrogen fuel cell offers a true “quantum leap” in both emissions and efficiency.

Hydrogen Fuel Cells

A fuel cell is shown in Figure 3. A fuel cell is an electrochemical device that converts the chemical energy in a fuel (hydrogen) and an oxidant (oxygen in air or pure oxygen) directly to electricity, water, and heat. The emissions are water vapor. The electrical conversion efficiency can be quite high. Up to 60 percent of the energy in the hydrogen is converted to electricity, even at small scale. This compares to 25 to 30 percent conversion efficiency for small fossil-fuel-powered engines.

The fuel cell works as follows. Hydrogen and oxygen have “chemical potential,” an attraction driving them to combine chemically to produce water. In a fuel cell hydrogen is introduced at one electrode (the anode) and oxygen at the other electrode (cathode). The two reactants (hydrogen and oxygen) are physically separated by an electrolyte, which can conduct hydrogen ions (protons) but not electrons. The electrodes are impregnated with a catalyst, usually platinum, which allows the hydrogen to dissociate into a proton and an electron. To reach the oxygen, the proton travels across the electrolyte, and the electron goes through an external circuit, doing work. The proton, electron and oxygen combine at the cathode to produce water.

Although the principle of fuel cells has been known since 1838, practical applications are fairly recent. The first applications were in the space program, where fuel cells powered the Gemini and Apollo spacecraft. In the 1960s and 1970s, fuel cells

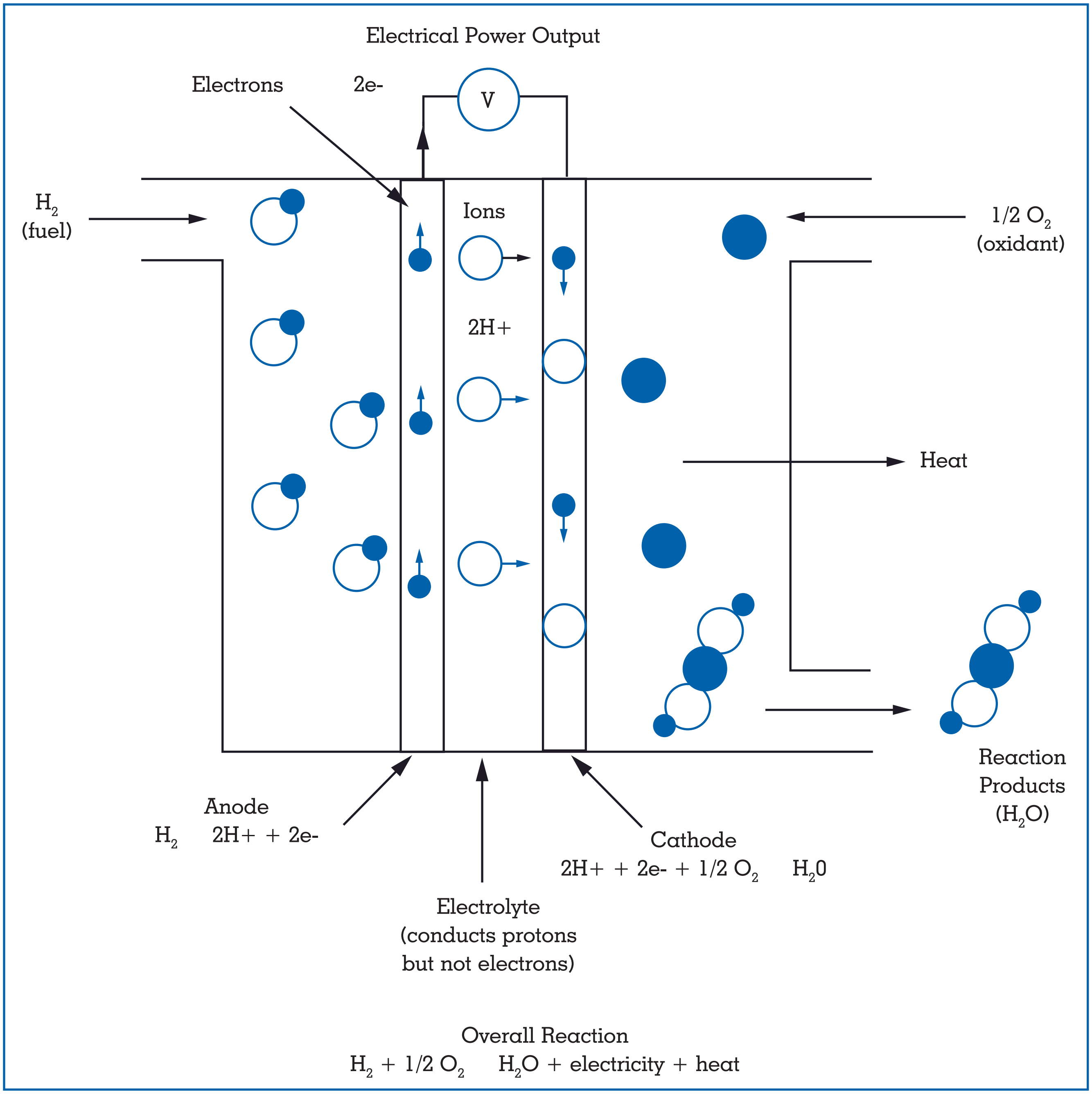


Figure 3.
Diagram of a Hydrogen-Oxygen Fuel Cell.

were used in space and military applications such as submarines. More recently fuel cells have been developed for low-polluting cogeneration of heat and power in buildings. In the past few years there has been a large worldwide effort to commercialize fuel cells for use in zero-emission vehicles.

Hydrogen for Transportation

A number of experimental hydrogen-powered vehicles have been built, dating back to the 1930s. Beginning in the early 1990s, "zero-emission-vehicle" regulations (enacted first in California and later in Massachusetts and New York) and government programs encouraging the development of high-efficiency automobiles (notably the Partnership for a New Generation of Vehicles) led to increased levels of R&D on fuel-cell and electric vehicles. Progress toward a commercial fuel cell vehicle is proceeding at a rapid pace. As of 1999, eight major automobile manufacturers were developing fuel-cell vehicles, and Ford, Daimler Chrysler, GM, and Toyota have announced their intent to commercialize a fuel-cell vehicle by 2005. Several oil companies (Mobil, Shell, Exxon, Texaco, and ARCO) are partners in demonstrations of fuel-cell-vehicle technologies. Shell has recently started a new business unit, Shell Hydrogen.

Hydrogen for Power Generation and Heating

Hydrogen also can be used to heat buildings and for power production. If low-cost fuel cells are commercialized, this could open the way for efficient combined heat and electric power systems for use in commercial and residential buildings. Because of ongoing deregulation, distributed, smaller-scale production of power may play an increasing role in future electric utilities. Several companies are developing fuel cell systems in a range of sizes. The economics are particularly attractive for buildings far from the existing electrical transmission grid, where resources for hydrogen production are present, such as remote sites or islands.

ENVIRONMENTAL AND SAFETY CONSIDERATIONS

Emissions from a Hydrogen Energy System

Hydrogen can be used with zero or near-zero emissions at the point of use. When hydrogen is burned in air, the main combustion product is H_2O, and there are traces of nitrogen oxides, which can be controlled to very low levels. No particulates, carbon monoxide, unburned hydrocarbons, or sulfur oxides are emitted. With hydrogen fuel cells, water vapor is the sole emission. Moreover, the total fuel cycle emissions of pollutants and greenhouse gases (such as CO_2, which could contribute to global climate change) can be much reduced compared to conventional energy systems, which vent all of the CO_2 in the flue gas.

Fuel cycle emissions are all the emissions involved in producing, transmitting, and using an alternative fuel. For example, for hydrogen made from natural gas, there would be emissions of CO_2 and nitrogen oxides at the hydrogen production plant, emissions associated with producing electricity to run hydrogen pipeline compressors (the nature of these emissions would depend on the source of electricity), and zero local emissions if the hydrogen is used in a fuel cell. The more efficient the end-use device (e.g., a fuel-cell vehicle), the lower the fuel cycle emissions per unit of energy service (e.g., emissions per mile traveled).

If hydrogen is made from decarbonized fossil fuels, fuel-cycle emissions can be cut by up to 80 percent. With renewable energy sources such as biomass, solar, or wind, the fuel cycle greenhouse gas emissions are virtually eliminated. It is possible to envision a future energy system based on hydrogen and fuel cells with little or no emissions of pollutants or greenhouse gases in fuel production, distribution, or use.

Resource Issues

In contrast to fossil energy resources such as oil, natural gas, and coal, which are unevenly distributed geographically, primary sources for hydrogen production are available virtually everywhere in the world. The choice of a primary source for hydrogen production can be made based on the best local resource.

Can hydrogen be produced sustainably? Over the next few decades, and probably well into the twenty-first century, fossil sources such as natural gas or coal may offer the lowest costs in many locations, with small contributions from electrolysis powered by low-cost hydropower.

In the longer term (or where locally preferred), renewable resources such as wastes, biomass, solar, or wind might be brought into use. It has been estimated that hydrogen derived from biomass produced on about two-thirds of currently idle cropland in the United States would be sufficient to supply transportation fuel to all the cars in the United States, if they used fuel cells (Ogden and Nitsch, 1993). Municipal solid waste could be gasified to produce transportation fuel for perhaps 25 to 50 percent of the cars in U.S. metropolitan areas (Larson, Worrell, and Chen, 1996). Solar power and wind power are potentially huge resources for electrolytic hydrogen production, which could meet projected global demands for fuels, although the delivered cost is projected to be about two to three times that for hydrogen from natural gas (Ogden and Nitsch, 1993).

Hydrogen Safety

When hydrogen is proposed as a future fuel, the average person may ask about the *Hindenburg*, the *Challenger*, or even the hydrogen bomb. Clearly, consumers will not accept hydrogen or any new fuel unless it is as safe as our current fuels.

Table 2 shows some safety-related physical properties of hydrogen as compared to two commonly accepted fuels, natural gas and gasoline.

In some respects hydrogen is clearly safer than gasoline. For example, hydrogen is very buoyant and disperses quickly from a leak. Experiments have shown that is it difficult to build up a flammable concentration of hydrogen except in an enclosed space, because the hydrogen disperses too rapidly. This contrasts with gasoline, which puddles rather than dispersing, and where fumes can build up and persist. Hydrogen is nontoxic, which also is an advantage.

Other safety concerns are hydrogen's wide flammability limits and low ignition energy. Hydrogen has a wide range of flammability and detonability limits (a wide range of mixtures of hydrogen in air will support a flame or an explosion). In practice, however, it is the lower flammability limit that is of most concern. For example, if the hydrogen concentration builds up in a closed space through a leak, problems might be expected when the lower flammability limit is reached. Here the value is comparable to that for natural gas.

The ignition energy (e.g., energy required in a spark or thermal source to ignite a flammable mixture of fuel in air) is low for all three fuels (hydrogen, gasoline, and natural gas) compared to commonly encountered sources such as electrostatic sparks from people. The ignition energy for hydrogen is about an order of magnitude lower for hydrogen than for methane or gasoline at stoichiometric conditions (at the mixture needed for complete combustion). But at the lower flammability limit, the point where problems are likely to begin, the ignition energy is about the same for methane and hydrogen.

Although safe handling of large quantities of hydrogen is routine in the chemical industries, people question whether the same safety can be achieved for hydrogen vehicle and refueling systems. According to a 1994 hydrogen vehicle safety study by researchers at Sandia National Laboratories, "There is abundant evidence that hydrogen can be handled safely, if its unique properties—sometimes better, sometimes worse and sometimes just different from other fuels—are respected." A 1997 report on hydrogen safety by Ford Motor Company (Ford Motor Company, 1997) concluded that the safety of a hydrogen fuel-cell vehicle would be potentially better than that of a gasoline or propane vehicle, with proper engineering. To assure that safe practices for using hydrogen fuel are employed and standardized, there has been a considerable effort by industry and government groups (within the United States and internationally) in recent years to develop codes and standards for hydrogen and fuel-cell systems.

ECONOMICS OF HYDROGEN PRODUCTION AND USE

If hydrogen technologies are available now and the environmental case is so compelling, why aren't we using hydrogen today? Part of the answer lies in the economics of hydrogen use. There are substantial capital and energy costs involved in hydrogen production. Comparing the delivered cost of hydrogen transporation fuel on an energy cost basis, we find that it is more costly than natural gas or gasoline. However, hydrogen can be used more efficiently than either natural gas or gasoline, so the cost per unit of energy service is comparable. Studies of the projected cost of hydrogen-fueled transportation have shown that if fuel-cell vehicles reach projected costs in mass production, the total life-cycle cost of transportation (accounting for vehicle capital costs, operation and maintenance costs, and fuel) will be similar to that for today's gasoline vehicles.

SCENARIOS FOR DEVELOPING A HYDROGEN ENERGY SYSTEM

The technical building blocks for a future hydrogen energy system already exist. The technologies for producing, storing, and distributing hydrogen are well known and widely used in the chemical industries today. Hydrogen end-use technologies—fuel cells, hydrogen vehicles, and power and heating systems—are undergoing rapid development. Still, the costs and the logistics of changing our current energy system mean that building a large-scale hydrogen energy system probably would take many decades.

Because hydrogen can be made from many different sources, a future hydrogen energy system could evolve in many ways. In industrialized countries,

	Hydrogen	*Methane*	*Gasoline*
Flammability Limits (% volume)	4.0-75.0	5.3-15.0	1.0-7.6
Detonability Limits (% volume)	18.3-59.0	6.3-13.5	1.1-3.3
Diffusion velocity in air (meter/sec)	2.0	0.51	0.17
Buoyant velocity in air (meter/sec)	1.2-9.0	0.8-6.0	non-buoyant
Ignition energy at stoichiometric mixture (milliJoules)	0.02	0.29	0.24
Ignition energy at lower flammability limit (milliJoules)	10	20	n.a.
Toxicity	non-toxic	non-toxic	toxic in concentrations > 500 ppm

Table 2
Safety-Related Properties of Hydrogen, Methane, and Gasoline

hydrogen might get started by "piggybacking" on the existing energy infrastructure. Initially, hydrogen could be made where it was needed from more widely available energy carriers, avoiding the need to build an extensive hydrogen pipeline distribution system. For example, in the United States, where low-cost natural gas is widely distributed, hydrogen probably will be made initially from natural gas, in small reformers located near the hydrogen demand (e.g., at refueling stations). As demand increased, centralized production with local pipeline distribution would become more economically attractive. Eventually hydrogen might be produced centrally and distributed in local gas pipelines to users, as natural gas is today. A variety of sources of hydrogen might be brought in at this time. In developing countries, where relatively little energy infrastructure currently exists, hydrogen from local resources such as biomass may be more important from the beginning.

Joan M. Ogden

See also: Fuel Cells; Nuclear Energy; Refineries; Reserves and Resources.

BIBLIOGRAPHY

Ford Motor Company. (1997). *Direct Hydrogen-Fueled Proton Exchange Membrane Fuel Cell System for Transportation Applications: Hydrogen Vehicle Safety Report*. Contract No. DE-AC02-94CE50389. Washington, DC: United States Department of Energy.

Herzog, H.; Drake, E.; and Adams, E. (1997). *CO_2 Capture, Reuse, and Storage Technologies for Mitigating Global Climate Change*. Washington, DC: United States Department of Energy.

Hord, J. (1976). "Is Hydrogen a Safe Fuel?" *International Journal of Hydrogen Energy* 3:157–176.

James, B. D.; Baum, G. N.; Lomax, F. D.; Thomas, C. E.; Kuhn, I. F. (1996). *Comparison of Onboard Hydrogen Storage for Fuel Cell Vehicles*. Washington, DC: United States Department of Energy.

National Research Council. (1998). *Review of the Research Program of the Partnership for a New Generation of Vehicles: Fourth Report*. Washington DC: National Academy Press.

Ogden, J. (1999). "Prospects for Building a Hydrogen Energy Infrastructure," In *Annual Reviews of Energy and the Environment*, ed. R. Socolow. Palo Alto, CA: Annual Reviews.

Ogden J., and Nitsch, J. (1993). "Solar Hydrogen." In *Renewable Energy Sources of Electricity and Fuels*, ed. T. Johannsson, H. Kelly, A. K. N. Reddy, and R. H. Williams. Washington DC: Island Press.

Ogden, J.; Steinbugler, M.; and Kreutz, T. (1999). "A Comparison of Hydrogen, Methanol, and Gasoline as Fuels for Fuel Cell Vehicles." *Journal of Power Sources* 79:143–168.

Socolow, R., ed. (1997). *Fuels Decarbonization and Carbon Sequestration: Report on a Workshop*. Princeton, NJ: Princeton University Center for Energy and Environmental Studies.

Taylor, J. B.; Alderson, J. E. A.; Kalyanam, K. M.; Lyle, A. B.; and Phillips, L. A. (1986). "Technical and Economic Assessment of Methods for the Storage of Large

Quantities of Hydrogen." *International Journal of Hydrogen Energy* 11:5.

Thomas, C. E.; Kuhn, I. F.; James, B. D.; Lomax, F. D.; and Baum, G. N. (1998). "Affordable Hydrogen Supply Pathways for Fuel Cell Vehicles." *International Journal of Hydrogen Energy* 23(6).

Williams, R. H. (1998). "Fuel Decarbonization for Fuel Cell Applications and Sequestration of the Separated CO_2." In *Eco-restructuring: Implications for Sustainable Development*, ed. W. Ayres. Tokyo: UN University Press.

Winter, C.-J., and Nitsch, J. (1998). *Hydrogen as an Energy Carrier*. New York: Springer-Verlag.

HYDROGEN BOMB

See: Nuclear Fusion

I

IMPORT/EXPORT MARKET FOR ENERGY

The import and export of energy on the international market helps create world economic growth. Countries that need energy because they do not have very large domestic reserves must import energy to sustain their economies. Countries that have large reserves export energy to create growth in their energy industries. The export and import of energy also contributes to specialization and comparative advantage in the world economy, lowering costs of production, increasing productivity, and expanding world economic output.

Comparative advantage in energy trade creates more robust economic growth, similar to any economic trade. For example, an oil worker has a comparative advantage over the farmer in producing oil as a farmer has a comparative advantage over the oil worker in producing food. An oil worker who specializes in pumping oil out of the ground is better off trading the oil for food from a farmer. The oil worker could grow his own food, but he could better utilize his time by simply pumping oil and trading for food. He specializes in oil production because he can obtain more food that way, or more of other goods and services. The farmer likewise could try to produce his own oil, but would spend an enormous amount of time and resources exploring for and developing an oil field. Both the farmer and the oil producer gain from trade. The medium of exchange for this trading is money, and both the farmer and the oil worker make more money by specializing and trading than by being totally self-sufficient.

The dynamics of specialization and comparative advantage that apply to the individuals work similarly for nations as well. On the world market, energy exporters specialize in producing energy, and trade their energy to other countries for other goods and services. Energy importers specialize in producing other goods and services, and sell those goods and services for energy. Saudi Arabia, for example, obtains high-tech computers from the United States, automobiles from Japan, and manufactured goods from Europe in exchange for its oil. If Saudi Arabia had to produce all its own computers, their computers would cost a lot more and probably not be nearly as good. If Europe, Japan and the United States had to produce all their own oil by, for example, converting coal into oil, oil would be much more costly, and that cost would inhibit economic growth. The trade in energy therefore benefits both sides.

The largest international energy trading on a Btu basis is with oil, partly because oil is the world's most valuable and versatile existing energy resource. The international natural gas trade is also increasing but because of the portability problems with natural gas (far less energy content given volume making it more expensive and more difficult to transport) its greater potential is limited. But within geographical regions around the world. Electricity is traded internationally more frequently now that many countries have deregulated their utility markets. The Middle East is the most prolific energy exporter, exporting about five and a half billion barrels of oil a year in 1998, which was over 50 percent of all world oil exports, and one-third of all world energy exports. That share looks to increase in the future as many older oil-producing regions, including the United States and Russia, have declining production.

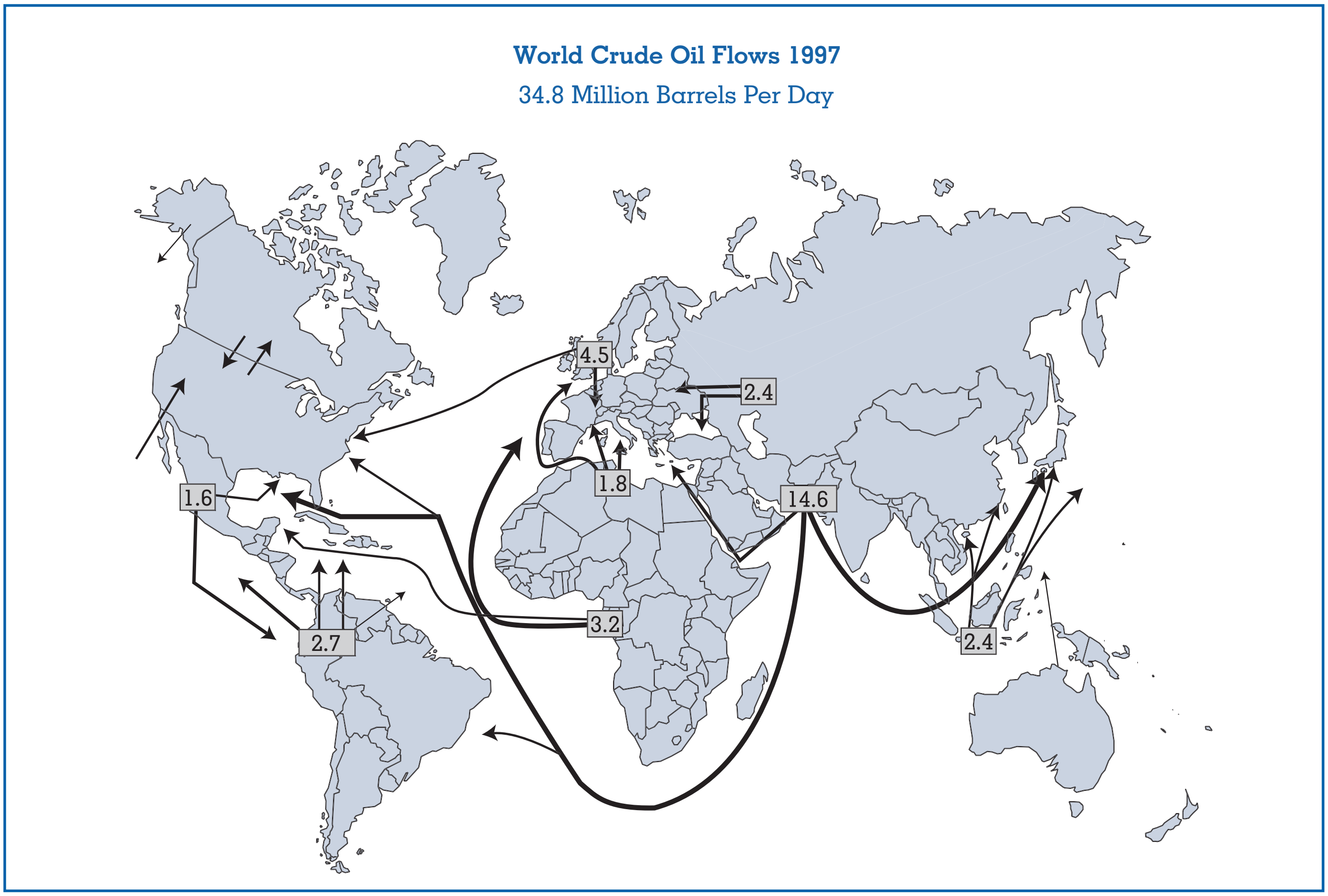


Figure 1.

SOURCE: Energy Information Agency, <http://www.eia.doe.gov/emeu/security/oilflow2.gif>.

MARKET POWER IN ENERGY TRADE

Energy exports not only depend on geological reserves of energy and internal consumption requirements of a country, but also on the market power of that country in energy markets. For example, in geological terms, 65 percent of world oil reserves reside in only five countries in the Middle East—Saudi Arabia, Iran, Iraq, Kuwait, and the United Arab Emirates—also called the big five Middle Eastern oil producers. The big five had proven oil reserves of more than 660 billion barrels of oil in 1998, with a theoretical potential to produce and export as much as 180 million barrels of oil a day (mbd), yet they produced and exported only 18 mbd on average in 1998. By restricting their oil exploration, production, and exports, the big five cause oil prices to rise on the world market. Therefore, it is not enough to look at world oil reserves and the theoretical potential oil production to determine how long oil will last or when an oil shortage will occur; market power must also be taken into consideration. World oil production will undoubtedly last for hundreds of years, but at low levels. Production will be restricted by the use of market power so that the price of oil on the world market will be exceptionally high, decreasing the demand for oil.

The percent of world oil production from the Middle East has changed over the years. In 1950, the Middle East produced about 16 percent of world production, but by 1975 it had a 35 percent market share of all production. That share declined to 25 percent by 1990 due to a decline in world demand, and stayed at that level through 1998. Eventually that share will rise. The big five Middle Eastern oil producers have the largest oil reserves in the world and therefore have the greatest potential to supply increasing world demand, giving them an increasing market share of world production, which in turn will give them enough market power to substantially increase prices by limiting their production. It is the share of total world production rather than the share of the export market itself that determines market power. Middle Eastern oil market power may be a good thing, however, since by raising oil prices and limiting production the world's oil reserves will last longer, albeit at a higher price.

THE HISTORY OF U.S. ENERGY IMPORTS

The United States became the world's first producer of deep crude oil from an oil well when in 1859 Colonel Edwin Drake successfully used a pipe drilled into the ground to obtain oil. From then until about 1970, the United States was virtually energy-independent with only some oil and gas imports from Mexico and Canada. While U.S. reserves of coal, natural gas and uranium continue to be large enough to supply internal demand with enough left over to export, the supply of oil took a sharp turn downward. After 1970, even while U.S. demand continued to increase at a steep 6.5 percent per year, the supply of U.S. oil began to decline, necessitating sharp increases in U.S. oil imports.

Suddenly there was a significant change in the import/export dynamics of the world. From 1970 to 1973 the United States increased its proportion of imported oil from 10 percent to 23 percent of domestic demand, a substantial 2 mbd increase in imports. At the same time, Saudi Arabia increased its exports by 120 percent, or about 4 mbd, to become the world's most important oil exporter. This change in the balance of exports on the international market gave Saudi Arabia and the rest of the Organization of Petroleum Exporting Countries (OPEC) tremendous market power. It was only a matter of time before they would use it.

The 1973 Yom Kippur war caused Arab oil producers to boycott oil exports to the United States and some allies. While the boycott may look like the culprit in the 1973 oil price shock, the cuts in exports actually did not last long, with the oil price increases continuing long after the Arab/Israeli war was over. The real reason why oil prices surged up and stayed there was because world oil demand was high, while Saudi Arabia and OPEC kept supplies low. OPEC used its market power to curtail its own oil exports in the face of surging world demand. The result was an incredible quadrupling of the international price for oil in only one year. The sharp increase in oil prices created an immediate call for energy independence.

After the first oil shock, world demand for oil began to subside somewhat, but then increased again. Exports by OPEC members increased simultaneously as imports by the United States and other countries also increased. In 1979 Iran had a revolution that cut exports of Iranian oil to virtually nothing. The Iranian supply cut pushed oil prices into another shock. Again oil prices jumped high, but this time the extremely high price pushed oil demand down. After 1980, world oil demand declined for awhile and then stayed low due to the structural changes in the world's economies. Firms began using less oil and more coal, natural gas and nuclear power. The resulting softening of demand cut oil exports by OPEC members, particularly Saudi Arabia.

By 1986, world demand for oil was so low, and non-OPEC oil producers were exporting oil so copiously, that the price of oil took a tremendous dive. The combination of low demand and more abundant supplies could not sustain a high world oil price. U.S. demand, for example, had reached 19 mbd in 1979, but fell to 15 mbd by 1984. At the same time, imports of oil fell from 8 mbd to 4.25 mbd (i.e., from 42 percent to 28 percent of demand). This same trend occurred in all the developed countries resulting in the price plunge. However, since the crash in oil prices, oil demand has turned up again. After 1984, U.S. demand increased at a rate of 1.5 percent per year to 22 mbd by 1998. Yet, U.S. oil supplies continued to fall, even taking into account the great Alaska North Slope oil fields. By 1996, the United States was importing 50 percent of its oil demand, or 8.5 mbd, which has created calls for energy independence for America.

ENERGY INDEPENDENCE

The idea behind energy independence is that if all energy production occurs within a country's borders, then that country's economy will be insulated from any energy supply disruptions. The country would then have less unemployment and less economic decline if and when the world's energy exporters, especially OPEC members, cut their supplies. France, has developed a strong nuclear power industry so that it would not have to import as much oil from the Middle East. A close policy alternative for energy independence is to have all imported energy come from friendly countries. In 1998, the United States imported much of its oil from Mexico and Venezuela rather than from Iraq, Libya, or Iran.

One of the problems with attaining energy independence has been the cost of alternative energies. In the late 1970s and early 1980s, the United States tried to produce oil from oil shale in order to become energy-independent. However oil from oil shale cost twice as much as oil from OPEC. The problem was that if the United States were to rely on shale oil, then the economy would have been worse off, with even more unemployment than if it imported relatively cheaper OPEC oil. The expansion of the oil shale industry itself would have caused an increase in employment in that particular industry, but at a drastically higher cost of energy to the economy as a whole. The higher cost of domestically produced energy would have caused the rest of the U.S. economy to go into more of a recession than it already was in. This is why a policy of energy independence has not been successful.

Another U.S. policy to attain energy independence was to force all Alaskan North Slope crude oil to be consumed inside the United States and not be allowed to be exported. The problem was that North Slope crude oil is relatively heavy and not suitable for west coast fuel needs. The mismatch of supply and demand caused California refineries to sell heavy distillate fuels abroad and import lighter fuel additives. Furthermore, the forced selling of Alaska crude oil on a very saturated west coast market caused Alaska crude prices to be $1 to $5 per barrel less than the international price, resulting in less oil exploration and development in Alaska. The upshot of all this was lower tax revenue, a loss of jobs in the oil fields, and less oil exploration and development on the North Slope. The United States actually exported heavy bunker fuel oil at a loss, as opposed to the profit that could have been attained by simply exporting crude oil directly.

FRIENDLY COUNTRY IMPORTS

According to some experts, the next best solution for energy independence is to make sure that all imported energy comes from friendly countries. The idea is to ensure a guaranteed source of energy in case a major energy exporter were to suddenly cut off its exports. The United States, for example, can buy all its oil from Mexico or Venezuela and be independent of crises in the Middle East. The problem with this strategy is that markets have a tendency to be independent of political expediency.

In a world market, oil is often bought by and sold to the highest bidder. Since all oil is fungible, the oil exporter usually does not care where the oil is sold as long as he receives top dollar for it. Oil import competitors such as Japan will always be ready to bid higher prices for oil no matter where it comes from. For example, the United States may import oil from Mexico. If a supply disruption occurred in Iran so that Japan could not get enough oil, Japan would immediately offer a higher price to Mexico for its oil. Even though Mexico is friendly to the United States it is not responsible for the U.S. economy. Mexico wants to get as much money as it can for its oil, and so will sell its oil to the highest bidder regardless of politics. In this case, Mexico would sell its oil to Japan if Japan were the highest bidder. This is how events in the Middle East can affect U.S. oil imports even if the U.S. imports its oil from somewhere other than the Middle East.

If Mexico did sell its oil to Japan rather than the United States, then the United States would immediately offer a high price for oil from another country, for example, Libya. Even though Libya has been a political foe of Americas, Libya would be willing to sell its oil to the United States if the United States paid more than European countries. In fact, events outside the United States have more to do with the price of oil than where the United States gets its oil. Whether U.S. oil is obtained from Alaska, Mexico, or Libya, it is the supply and demand for oil on the world market that determines the price of oil, not the location of supplies. Oil demand in East Asia, political tension in the Middle East, and lower oil production in the former Soviet republics, are much bigger determinants of U.S. energy prices than the source of imported oil. The world price of oil is independent of secure supplies from friendly countries, and as such it is the world oil market that affects the price of oil and brings about the economic instability caused by changing oil prices. The only way to ensure a stable low price for oil is to make sure that the international oil market stays competitive and free from supply disruptions caused by war or terrorist acts. This is why the United States, Europe, and the rest of the world defended Kuwait and Saudi Arabia in the Gulf War in 1991. However, as long as the United States or any other country is willing to pay the world price for oil, then it can purchase as much oil as it wants.

THE AMAZING INCREASE IN U.S. COKE IMPORTS

Coal is a very useful energy resource in the production of steel. However, before coal can be used in steel mills, it must be converted into another form called coke, which is the source of the carbon monoxide used in steel-making. The coal is heated up in a low-oxygen chamber and burned slightly to produce coke, which is similar to wood charcoal. The United States has a 200-year supply of coal, and therefore vast potential supplies of coke. Nevertheless, the United States imported more coke than it exported in 1998, mostly from Japan and China. Even though imported coke represents only eight percent of U.S. consumption, it is still surprising that the United States, which has the largest reserves of coal in the world, would import coke at all.

That imports of coke have increased in the United States is mostly due to the high cost of transportation within the United States and the lower cost of shipping by foreign firms. It is cheaper to extract coal in China, coke it and ship it to California than to buy coke from U.S. producers and transport it within the United States. The trade in coke doesn't hurt the U.S. economy any more than the trade in jet aircrafts to China hurts China. Mutual trade is always beneficial to trading partners.

HYDRO-QUEBEC CANADIAN ELECTRICITY

Quebec, Canada, has great hydroelectric generating potential. Hydroelectricity is a very cheap source of electric power and relatively environmentally benign. As the U.S. continues to deregulate its power industry, Quebec hydroelectricity can be marketed for export to the U.S. northeast where electricity is very costly. It is even possible for Quebec to obtain market power in the U.S. electric power market. However, electric power was much less regional in 1999 than it was in the 1980s and before. In 1999 the technology existed to transport electric power over high voltage power lines as far as two thousand miles. This means the lowest priced electric producer could compete in the power market from very far away. If Quebec begins exporting electric power to the U.S. northeast, it will compete with electric producers as far away as Texas and Oklahoma that can produce cheap electricity from natural gas.

The overriding concern for energy imports and exports should not be where energy comes from or whether an industrialized country should be completely self reliant for its energy needs. The overriding concern is simply the price of energy. Whether energy is produced inside a country or not, world supply and demand for energy, energy market power, and the costs of energy production are the biggest factors in determining energy prices. The price of energy then affects how well the world's economy grows. This is the reason the United States and the rest of the world are concerned about all economic and energy events around the world, including energy supply cutoffs, recessions, and a generally increasing world demand for energy.

Douglas B. Reynolds

See also: Hydroelectric Energy.

BIBLIOGRAPHY

Adelman, M. A. (1993). *The Economics of Petroleum Supply.* Cambridge MA: MIT Press.

Brantly, J. E. (1960). *History of Oil Drilling.* Houston, TX: Gulf Publishing Co.

BP Amoco. (1996). *Statistical Review of World Energy 1996.* London, England: Author.

Cuff, D. J., and Young, W. J. (1986). *The United States Energy Atlas,* 2nd ed. New York: Macmillan.

Energy Information Administration. (1997). *International Energy Annual 1997.* Washington DC: Author.

Energy Information Administration. (1999). *International Energy Data.* Washington, DC: Author.

Exxon. (1984). *Middle East Oil and Gas.* Exxon Background Series, December 1984. New York: Author.

Federal Energy Regulatory Commission. (1996). *Promoting Wholesale Competition Through Open Access Non-Discriminatory Transmission Services by Public Utilities.* Final Order No. 888, 75-FERC 61, 080, Docket No. RM95-8-000, April, 1996. Washington DC: Author.

IMPULSE TURBINE

See: Turbines, Steam

INCANDESCENT LAMP

See: Lighting

INCINERATION

See: Waste-to-Energy Technology; Cogeneration; Materials

INDUCTION MOTOR

See: Electric Motor Systems

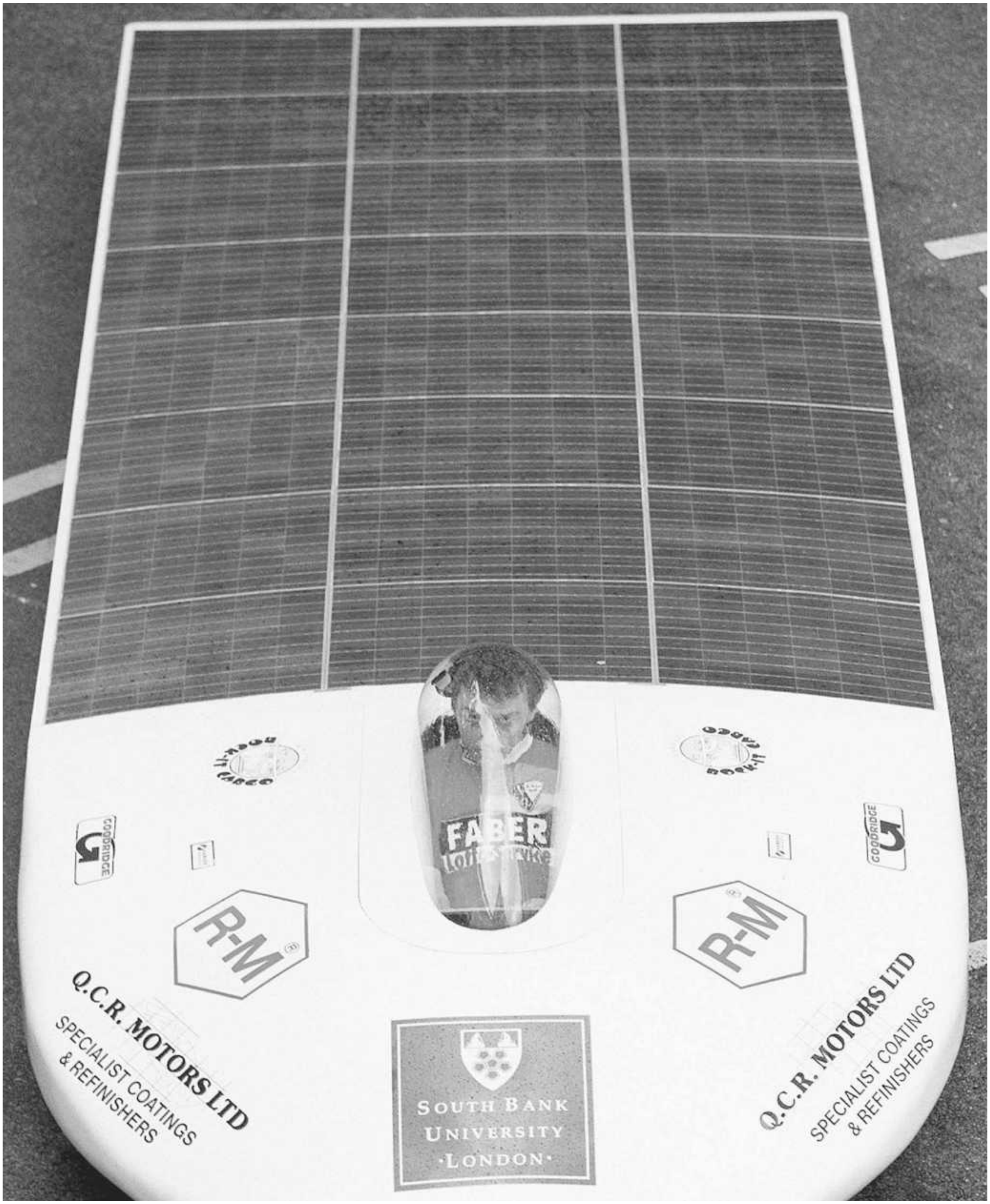
INDUSTRY AND BUSINESS, ENERGY AS A FACTOR OF PRODUCTION IN

The production of goods and services requires energy as an input, which is called a factor of production. Energy sources vary in their effectiveness as a factor of production, depending on their energy characteristics. The energy characteristics are measured in energy grades, which indicate the levels of usefulness of any given energy input. Low-grade energy resources are less useful to an economy than high-grade resources, because any given process will be able to produce more economic value from a high-grade energy resource than from a low-grade energy resource. For example, it is easier to fly a passenger jet aircraft using jet fuel rather than coal because jet fuel has more suitable energy characteristics.

ENERGY GRADES

There are four energy grades used to measure energy characteristics: weight, volume, area, and state. The weight grade is British thermal units (Btus) per pound of the energy resource. For example, coal has 10,000 Btus per pound, while oil has 20,000 Btus per pound, making oil the higher-grade resource. The volume grade is Btus per cubic foot of the energy resource. Oil has 1 million Btus per cubic foot while compressed natural gas, at 3,000 pounds per square inch, has 177,000 Btus per cubic foot, which makes oil the higher-grade resource. The weight and volume grades are important determinants for how easy energy is to transport. Light, compact energy sources are much easier to store and use than heavy, voluminous energy sources. The area grade is Btus per acre where the energy resource is found in its original state. Wood has 1 billion to 5 billion Btus per acre as a forest, whereas coal has 10 billion to 1 trillion Btus per acre in a mine. The area grade generally determines how costly it is to extract or produce energy. Energy that is diffuse over an area, such as trees in a forest, tends to require more capital and labor to extract each Btu of energy than concentrated sources. The state grade is the original physical state of the energy resource, such as a liquid, gas, or solid as measured at standard atmospheric temperature and pressure. The highest state grade is the liquid state, followed by the gas state, the solid state, and the field state. The liquid state is the highest state grade because liquids are easier to use than gases and solids. The field state is the lowest state grade, since energy from energy fields such as solar energy is difficult to store. The field state is any kind of energy field such as a magnetic field, an electric field, or a radiation field. Nuclear energy is a field state grade, since it derives its energy from a radiation field. The state grades are fundamental in determining how well various energy resources can produce economically valuable outputs.

High-grade energy resources can create higher-valued, lower-cost outputs. For example, oil is one of the highest-grade energy resources there is. It is a liquid that is a very high state grade. It has a high weight and volume grade, better than any other energy resource except nuclear fuels, making it cost-effective to carry with a mobile machine. For example, chain saws and automobiles work better with a light-weight fuel than with a heavy fuel, because there is less fuel weight to carry. Oil also has a high area grade—100



The solar-powered car Mad Dog III is put through its paces before it is packed up on August 26, 1999, to travel to Australia to compete in the World Solar Challenge. Its 760 solar cells generate 1,200 watts of power and give it a cruising speed of 75km/hr. (Corbis Corporation) billion to 1 trillion Btus per acre in oil fields—which makes it easier and cheaper to produce. Because oil is a liquid, it is easy to extract with no mining, and it is easy to convert into a refined liquid fuel. Finally, liquid fuels are the easiest of all energy resources to store and transport. Liquid fuels can be used in internal-combustion engines, which are lighter in weight and have more power per pound than external-combustion engines (steam engines) that operate on solid fuel (coal). And although internal-combustion engines can operate more cleanly on natural gas, the infrastructure needed to compress and store gas is much more complex and expensive.

THE ENERGY UTILIZATION CHAIN (EUC)

For any energy to produce goods and services, it goes through a sequence of usage called an energy utilization chain (EUC). The EUC determines how energy will be used to produce goods and services. Link 1 of the EUC is simply obtaining the energy source. This includes exploration and extraction of the energy or in some way producing it. Link 2 is energy conversion. More often than not, energy must be refined or converted into a more useful form of energy for consumption to occur. Link 3 is energy transportation and storage. All energy must be brought to the consumer or firm for use and if necessary be stored for later use. Link 4 is energy consumption. This is where energy is burned or used up. Conservation of energy resources is also a part of link 4—that is, consuming less energy resources in link 4 is conservation of energy. Link 5 is the energy service. The ultimate end of using and consuming any energy resource is to provide some sort of service to society. The service can be used directly by consumers or be an input into the production of other goods and services.

The use of oil for transportation, such as in automobiles and in aircraft, for example, is generally called the oil EUC. The chain of EUC links are the following: the exploration and production of oil; the transportation of the oil by pipeline or tanker; the refining of the oil in a refinery; the transportation of the gasoline to filling stations; the consumption of the gasoline in automobiles, which is sometimes conserved by the use of high-mileage cars; and finally the transportation service from driving the automobile. The EUC then explains the system for using energy. If an alternative energy resource is used to replace oil, then some or all of the EUC links must change. For example, replacing the oil EUC with coal converted into oil, requires only changing the first two links of the oil EUC. Using solar energy to replace the oil EUC may require all of the links to change. For example, gasoline-driven automobiles would be replaced with electric vehicles, which give different services than normal automobiles. Electric vehicles based on solar energy may have less range of operation, carry less cargo, and take longer to refuel than ordinary automobiles.

The oil EUC is one of the highest-value EUCs in the economy. Alternative EUCs for transportation are the natural gas EUC, the solar/electric EUC, the synthetic fuels EUC, and so on. The reason it is so costly to use these alternative EUCs to replace the oil EUC is because of the low energy grades of the energy sources for the alternative EUCs. For example, to replace the oil EUC with an ethanol alcohol fuel, the alcohol EUC would have to use industrial distilleries to convert grain into alcohol. However, grain has a low area grade of about 40 million Btus per acre, as compared to oil's 1 trillion Btus per acre. Oil has about five magnitudes greater energy per acre in its original state, making it much cheaper to produce. It takes about 100,000 acres of farmland planted in grain to equal one acre of an oil field. In terms of supply, it takes about three times as much capital and 10 times as much labor to extract a Btu of ethanol from farmland as to extract a Btu of gasoline from an oil field. If all U.S. farmland were used to make ethanol, it could replace only 35 percent of U.S. oil needs. In addition, the grain needs to be converted from a solid-state grade energy resource into a liquid. Because oil is already a liquid, this transition is much easier and cheaper.

ALTERNATIVE EUCS

In the future, as oil and other energy resources begin to deplete, the economy will need to use alternative EUCs. One alternative EUC is the solar EUC. Solar energy is a field grade—that is, it is and energy radiated in light waves similar to an electric field. Because it is a field, it cannot be stored easily. An acre of solar collectors can catch about 65 million Btus per hour. One hour of an oil field operation on one acre of land can produce 2 billion Btus. The oil requires much less capital to extract its energy than does solar energy. This is why it is cheaper to obtain energy from oil than from solar energy; in addition, it is easier to convert, store, and transport the oil energy, allowing it to produce cheaper and more useful energy services than solar energy can. Once oil supplies decline substantially, the economy may be forced to use the solar EUC, but at a substantially higher cost than for the oil EUC.

Alternative EUCs provide the economy with energy services such as transportation at a certain cost. Usually, low-cost, high-value EUCs are used wherever possible and are used before higher-cost, lower-value EUCs. However, as a high-grade energy resource declines, lower-grade energy resources must be used, meaning the economy must begin to use higher-cost, lower-valued EUCs—that is, the cost of energy as a factor of production may rise, at least in the short run. In the longer run, breakthroughs in advanced energy sources, such as from hydrogen, may actually lower energy costs. The cost structure of alternative EUCs, though, depends on the cost of inputs. Ironically, one of the inputs that goes into every EUC is energy itself. It takes energy to produce energy. For example in the case of the synthetic fuels EUC, oil shale is used as an energy source to replace crude oil. To produce oil from oil shale, the oil shale must be converted from a solid-state energy grade into a liquid energy grade, which requires much more capital and labor inputs than does converting crude oil into fuels. However, the capital and labor require energy to produce oil from oil shale. When the price of oil goes up, the costs of the labor and capital inputs also rise causing the price of the shale oil to go up. In 1970, before the first oil shock, oil from oil shale cost about $3 per barrel to produce, while oil cost $1.50 per barrel. However, by 1982, when oil was $30 per barrel, shale oil cost $60 per barrel to produce. The high cost of oil made other inputs into the synthetic fuels EUC cost more, creating a higher price for the synthetic fuel, which in turn resulted in an inflation cost spiral. The nature of energy grades and the inflation cost spirals of inputs into the EUCs tend to make it difficult to pin down the cost of energy.

In general and in the short run, alternative EUCs have higher costs and lower-valued services than currently used EUCs. As the economy begins to use alternative EUCs, they tend to cost more, possibly creating an inflationary cost spiral. Technology can help to make alternative EUCs cost less and create more value, but is not likely to change the physical characteristics of alternative energy resources in the short run, making alternative EUCs overall less valuable than currently used high-grade-energy EUCs. One way to deal with higher-cost, lower-value alternative EUCs is for society to change its lifestyle. The value of alternative lifestyles must be evaluated in comparison to the cost of alternative EUCs. Higher-cost EUCs may force society into changing lifestyles to be able to afford energy services.

In the longer run, perhaps new alternative EUCs, even superior to the current high-grade-energy EUCs, can be developed. This possibility offers the opportunity of defeating any energy-induced inflationary cost spiral that might have developed, and opening up new possibilities of economic expansion driven by inexpensive energy as a factor of production. Given the critical role of energy to the production process, our economic output in the future is dependent on further advances in energy technology.

Douglas B. Reynolds

See also: Auditing of Energy Use; Capital Investment Decisions; Economically Efficient Energy Choices; Energy Economics; Industry and Business, Productivity and Energy Efficiency in.

BIBLIOGRAPHY

Cuff, D. J., and Young, W. J. (1986). *The United States Energy Atlas*, 2nd ed. New York: Macmillian.

Graham, S. (1983). “U.S. Pumps $2 Billion into States Oil Shale.” *Rocky Mountain News*, July 31, pp. 1, 22, 25.

Katell, S., and Wellman, P. (1971). *Mining and Conversion of Oil Shale in a Gas Combustion Retort*, Bureau of Mines Oil Shale Program Technical Progress Report 44. Washington, DC: U.S. Department of the Interior.

Reynolds, D. B. (1994). “Energy Grades and Economic Growth.” *Journal of Energy and Development* 19(2):245–264

Reynolds, D. B. (1998). “Entropy Subsidies.” *Energy Policy* 26(2):113–118.

Ricci, L. (1982). *Synfuels Engineering*. New York: McGraw-Hill.

INDUSTRY AND BUSINESS, PRODUCTIVITY AND ENERGY EFFICIENCY IN

Improving productivity is one of the central concerns of businesses. In buildings, energy-efficient technologies and design strategies can improve labor productivity (output of goods and services per hour worked) far in excess of the improvement in energy productivity (output per unit energy consumed). Similarly, in manufacturing, energy efficiency can improve total factor productivity (product output as a function of all labor, capital, energy, and materials consumed in its production) far in excess of the improvement in energy productivity.

OFFICE PRODUCTIVITY AND ENERGY EFFICIENCY

Offices and buildings are not typically designed to minimize either energy use or labor costs (by maximizing worker productivity). Almost everyone involved in building construction—such as the developer, architects, and engineers—is rewarded by the ability to minimize the initial cost of a building, as opposed to its life-cycle cost. Moreover, the designers are rarely the ones who will be paying the energy bill or the salaries of the people working in the building. The missed opportunity is revealed by the total life-cycle costs of a building (Table 1).

Initial cost (including land and construction)	2 percent
Operation & Maintenance (including energy)	6 percent
People Costs	92 percent

Table 1.
30-Year Life-Cycle Costs of a Building

A systematic approach to energy-efficient design can cost-effectively cut energy costs by 25 percent to 50 percent, as has been documented in both new construction and retrofit. The evidence on how energy-efficient design can increase worker productivity has been limited, so the issue has been a controversial one.

A growing body of international research suggests that specific design approaches can simultaneous save energy and increase productivity. The fundamental goal in productivity-enhancing design is to focus on the end users—workers—giving them the lighting, heating, and cooling they need for the job. This end-use approach maximizes productivity by ensuring, for instance, that workers don't have too much light or inadequate light quality for the job. It maximizes energy savings because, in most cases, this end-use approach eliminates excess and/or inefficient lighting, heating, and cooling, and because new technologies that provide higher quality services (such as better light quality) typically use far less energy than the technologies they replace.

Daylighting—use of natural light—is a key strategy because it is both the highest quality lighting and the most energy-saving, when it is systematically integrated into a design. In Costa Mesa, California, VeriFone achieved both large energy savings and productivity gains when it renovated a 76,000-square-foot building containing offices, a warehouse, and light manufacturing. The upgrade included energy-efficient air handlers, high-performance windows, 60 percent more insulation than is required by code, a natural-gas-fired cooling system, occupancy sensors, and a comprehensive daylighting strategy, including a series of skylights.

On sunny days, workers in the remanufacturing area construct circuit boards with only natural light and small task lighting. In the office area, on the other hand, the design minimizes direct solar glare on computers, while providing enough daylight to allow workers there to see changes associated with the sun's daily and seasonal variation.

The building beat California's strict building code by 60 percent, with a 7.5-year payback on energy-efficient technologies based on energy savings alone. Workers in the building experienced an increase in productivity of more than 5 percent and a drop in absenteeism of 45 percent, which brought the payback to under a year—a return on investment of more than 100 percent.

INDIVIDUAL CONTROL OVER THE WORKPLACE ENVIRONMENT

An increasingly popular strategy is to give individuals control over their workplace conditions. The benefits of this approach were documented at West Bend Mutual Insurance Company's 150,000-square-foot building headquarters in West Bend, Wisconsin. The design used a host of energy-saving design features, including efficient lighting, windows, shell insulation, and HVAC (heating, ventilation, and air conditioning).

In the new building, all enclosed offices have individual temperature control. A key feature is the

Environmentally Responsive Work-stations (ERWs). Workers in open-office areas have direct, individual control over both the temperature and air-flow. Radiant heaters and vents are built directly into their furniture and are controlled by a panel on their desks, which also provides direct control of task lighting and of white noise levels (to mask out nearby noises). A motion sensor in each ERW turns it off when the worker leaves the space, and brings it back on when he or she returns.

The ERWs give workers direct control over their environment, so that individuals working near each other can and often do have very different temperatures in their spaces. No longer is the entire HVAC system driven by a manager, or by a few vocal employees who want it hotter or colder than everyone else. The motion sensors save even more energy. The lighting in the old building had been provided by overhead fluorescent lamps, not task lamps. The workers in the new building all have task lights and they can adjust them with controls according to their preference for brightness. The annual electricity costs of $2.16 per square foot for the old building dropped to $1.32 per square foot for the new building, a 40 percent reduction.

The Center for Architectural Research and the Center for Services Research and Education at the Rensselaer Polytechnic Institute (RPI) in Troy, New York conducted a detailed study of productivity in the old building in the 26 weeks before the move, and in the new building for 24 weeks after the move. To learn just how much of the productivity gain was due to the ERWs, the units were turned off randomly during a two-week period for a fraction of the workers. The researchers concluded, “Our best estimate is that ERWs were responsible for an increase in productivity of about 2.8 percent relative to productivity levels in the old building.” The company’s annual salary base is $13 million, so a 2.8 percent gain in productivity is worth about $364,000—three times the energy savings.

FACTORY PRODUCTIVITY AND ENERGY EFFICIENCY

Just as in offices, energy efficiency improvements in manufacturing can generate increases in overall productivity far in excess of the gains in output per unit energy. This occurs in two principal ways: improved process control and systemic process redesign.

Process Control

Many energy-efficient technologies bring with them advanced controls that provide unique and largely untapped opportunities for productivity gains. In factories, probably the biggest opportunity is available in the area of variable-speed drives for motors. Variable- or adjustable-speed drives are electronic controls that let motors run more efficiently at partial loads. These drives not only save a great deal of energy, but also improve control over the entire production process. Microprocessors keep these drives at precise flow rates. Moreover, when the production process needs to be redesigned, adjustable drives run the motor at any required new speed without losing significant energy efficiency. Two examples are illustrative.

In Long Beach, Toyota Auto Body of California (TABC) manufactures and paints the rear deck of Toyota pickup trucks. In 1994, the company installed variable-speed motor drives for controlling the air flow in the paint booths. Applying paint properly to truck beds requires control over the temperature, air flow/balance, and humidity in the paint booths. Before the upgrade, manually-positioned dampers regulated airflow into the booths. Since the upgrade, the dampers are left wide open, while the fan motor speed changes automatically and precisely with touch screen controls, which also provide continuous monitoring of the airflow.

The improvements to the motor systems reduced the energy consumed in painting truck beds by 50 percent. In addition, before the upgrade, TABC had a production defect ratio of 3 out of every 100 units. After the upgrade, the ratio dropped to 0.1 per hundred. In 1997, the plant received a special award for achieving zero defects. The value of the improvement in quality is hard to put a price tag on, but TABC’s senior electrical engineer Petar Reskusic says, “In terms of customer satisfaction, it’s worth even more than the energy savings.”

The Department of Energy documented another typical case at the Greenville Tube plant in Clarksville, Arkansas, which produces one million feet of customized stainless steel tubing per month for automotive, aerospace, and other high-tech businesses. Greenville’s production process involves pulling, or “drawing,” stainless steel tubing through dies to reduce their diameter and/or wall thickness. The power distribution system and motor drive were inefficient and antiquated, leading to overheating, overloading, and poor control of the motor at low speed. A larger, but more efficient, motor (200 hp) was installed along with a computerized control system for $37,000. Electricity consumption dropped 34 percent, saving $7,000 a year, which would have meant slightly more than a five-year payback.

The greater horsepower meant that many of the tubes needed fewer draws: On average, one draw was eliminated from half the tubes processed. Each draw has a number of costly ancillary operations, including degreasing, cutting off the piece that the motor system latches on to, and annealing. Reducing the number of draws provided total labor cost savings of $24,000 a year, savings in stainless steel scrap of $41,000, and additional direct savings of $5,000. Thus, total annual savings from this single motor system upgrade was $77,000, yielding a simple payback of just over five months, or a return on investment in excess of 200 percent.

Process Redesign

The second principal way that energy efficiency can bring about an increase in industrial productivity is by achieving efficiency through process redesign. That process redesign could simultaneously increase productivity, eliminate wasted resources, and save energy has been understood as far back as Henry Ford. Perhaps the most successful modern realization and explication of the connection between redesigning processes to eliminate waste and boost productivity was achieved in the post-war development of the "lean production" system.

To understand lean production it is important to distinguish between improving processes and improving operations:

- An automated warehouse is an *operations* improvement: It speeds up and makes the operation of storing items more efficient. Eliminating all or part of the need for the warehouse by tuning production better to the market is a *process* improvement.
- Conveyor belts, cranes, and forklift trucks are *operations* improvements: They speed and aid the act of transporting goods. Elimination of the need for transport in the first place is a *process* improvement.
- Finding faster and easier ways to remove glue, paint, oil, burrs and other undesirables from products are *operations* improvements; finding ways not to put them there in the first place is a *process* improvement.

These examples show that when one improves the process, one does not merely cut out unnecessary operations, critical though that is to increasing productivity; one invariably reduces energy consumption as well as environmental impact. Reducing warehouse space reduces the need for energy to heat, cool, and light it. Reducing transportation reduces fuel use and exhaust fumes. Eliminating "undesirables" means no glue, paint, oil or scrap; it also avoids the resulting clean-up and disposal.

Thus, there is an intimate connection between redesigning processes to increase productivity (by eliminating wasted time, which is the essence of lean production) and eliminating wasted energy and resources. The most-time consuming steps in any process also tend to produce the most pollution and use the most energy. An integrated redesign can minimize everything.

For instance, in 1996, 3M company announced a breakthrough in the process for making medical adhesive tapes, a process that reduces energy consumption by 77 percent. The new process also cuts solvent use by 2.4 million pounds, lowers manufacturing costs, and cuts manufacturing cycle time by 25 percent. The proprietary process took researchers nine years from conception through final implementation.

The Sealtest ice cream plant in Framingham, Massachusetts, modified its refrigeration and air-handling system, cutting energy use by one third. The new system blew more air and colder air, and it defrosted the air handler faster. As a result, the time required to harden the ice cream was cut in half. The overall result was a 10 percent across-the-board increase in productivity, which is worth more to the company than the energy savings.

Process redesign drives a company toward a number of practices associated with productivity gain, including cross-functional teams, prevention-oriented design, and continuous improvement. One of the best programs for continuously capturing both energy and productivity gains was developed by Dow Chemical's Louisiana Division.

In 1982, the Division began a contest in which workers were invited to propose energy-saving projects with a high return on investment. The first year's result—27 winners requiring a capital investment of $1.7 million providing an average return on investment of 173 percent—was somewhat surprising. What was astonishing to everyone was that, year after year, Dow's workers kept finding such projects. Even

Reducing warehouse space reduces the need for energy to heat, cool, and light it. (Corbis-Bettmann)

as fuel prices declined, the savings kept growing. Contest winners increasingly achieved their economic gains through process redesign to improve production yield and capacity. By 1988, these productivity gains exceeded the energy and environmental gains.

Even after 10 years and nearly 700 projects, the two thousand employees continued to identify high-return projects. The contests in 1991, 1992, and 1993 each had in excess of 100 winners, with an average return on investment of 300 percent. Total energy savings to Dow from the projects of those three years exceeded $10 million, while productivity gains came to about $50 million.

CONCLUSION

Not all energy-efficiency improvements have a significant impact on productivity in offices or factories. Nonetheless, productivity gains are then possible from a systematic approach to design. The key features that simultaneously save energy and enhance productivity in office and building design are (1) a focus on the end user, (2) improved workplace environment, especially daylighting, and (3) individual control over the workplace environment. The key features that save energy and enhance productivity in factories are (1) improved process control, and (2) systematic process redesign.

Joseph Romm

See also: Building Design, Commercial; Building Design, Energy Codes and.

BIBLIOGRAPHY

Brill, M.; Margulis, S.; and Konar, E. (1984). *Using Office Design to Increase Productivity,* Volume 1. Buffalo, NY: Workplace Design and Productivity, Inc.

Ford, H. (1988). *Today and Tomorrow*, reprint. Cambridge, MA: Productivity Press.

Kroner, W.; Stark-Martin, J. A.; and Willemain, T. (1982). *Using Advanced Office Technology to Increase Productivity*. Troy, NY: The Center for Architectural Research.

Loftness, Vivian, et al. (1998). *The Intelligent Workplace Advantage*, CD-ROM. Pittsburgh: Center for Building Performance and Diagnostics, School of Architecture, Carnegie Mellon University.

Romm, J. (1999). *Cool Companies: How the Best Businesses Boost Profits and Productivity By Cutting Greenhouse Gas Emissions*. Washington, DC: Island Press.

Shingo, S. (1990). *Modern Approaches to Manufacturing Improvement: The Shingo System*, ed. Alan Robinson. Cambridge, MA: Productivity Press.

U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy. (1997). "The Challenge: Improving the Efficiency of a Tube Drawing Bench." Washington, DC: Author.

Womack, J., and Jones, D. (1996). *Lean Thinking: Banish Waste and Create Wealth in Your Corporation.* New York: Simon & Schuster.

INSULATION

The term "thermal insulation" refers to a material or combination of materials that slows the transfer of heat from high temperature (hot) regions to low temperature (cold) regions. Thermal insulation is placed between regions or surfaces having different temperatures to reduce heat flux—heat flow rate per unit area. In general, the heat flow increases as the temperature difference increases. The heat flux also depends on the type of material between the hot and cold surfaces. If the material between the surfaces is a thermal insulation—a material with relatively low values of a property called apparent thermal conductivity—then the heat flux will be small. As an added bonus, most building thermal insulation also function as effective sound or acoustical insulation.

The development of better thermal insulation is important since space heating and cooling account for the majority of energy consumption in the residential sector, and are second only to lighting in the commercial sector. Because of advances in insulation and more efficient heating systems, the U.S. Department of Energy projects that the energy used for space heating will drop at least 25 percent per household by 2020 relative to 1997 usage.

HISTORICAL DEVELOPMENT

The use of thermal insulation dates back to ancient times, when primitive man used animal skins for clothing and built structures for protection from the elements. Primitive insulation included fibrous materials such as animal fur or wool, feathers, straw, or woven goods. Bricks and stone, while not highly efficient thermal insulation, provided protection from the elements, reduced the loss of heat from fires, and provided large masses that moderate temperature changes and store heat.

Origins of the science associated with thermal insulations coincide with the development of thermodynamics and the physics associated with heat transfer. These technical subjects date to the eighteenth century. Early observations that a particular material was useful as thermal insulation were not likely guided by formal theory but rather by trial and error. Sawdust was used, for example, in the nineteenth century to insulate ice storage buildings.

MODERN DEVELOPMENT

Both building and industrial insulation experienced widespread use in the twentieth century. Because the use of thermal insulation in buildings significantly reduces the energy required to heat and cool living or working space, the use of thermal insulation has grown as expectations for comfort and the cost of energy have increased. The amount of insulation used in a particular application is economically justified by balancing the cost of the insulation against the value of energy saved during the lifetime of the insulation. For buildings, it depends on geographic location; for industry, it depends on the process temperature and safety considerations. The outside surface of insulation that is installed around a steam pipe, for example, will be cooler than the metal surface of the pipe.

Thermal insulation is available over a wide range of temperatures, from near absolute zero (–273°C) (–459.4°F) to perhaps 3,000°C (5,432°F). Applications include residential and commercial buildings, high- or low-temperature industrial processes, ground and air vehicles, and shipping containers. The materials and systems in use can be broadly characterized as air-filled fibrous or porous, cellular solids, closed-cell polymer foams containing a gas other than air, evacuated powder-filled panels, or reflective foil systems.

Thermal insulation in use today generally affects the flow of heat by conduction, convection, or radiation. The extent to which a given type of insulation affects each mechanism varies. In many cases an insulation provides resistance to heat flow because it contains air, a relatively low thermal conductivity gas. In general, solids conduct heat the best, liquids are less conductive, and gases are relatively poor heat conductors. Heat can move across an evacuated space by radiation but not by convection or conduction.

The characterization of insulation is simplified by the use of the term "apparent thermal conductivity" (k_a). Apparent thermal conductivity is used to signify that all three modes of heat transfer have been included in the evaluation of k_a. The k_a of an insulation can be determined by a standard laboratory test without specifying the heat transfer mechanism. The thermal resistivity (R^*) of an insulation is the reciprocal of k_a. R^* is the R-value per unit thickness of insulation, which is one inch in the inch-pound (IP) system. The ratio of insulation thickness over k_a is called the thermal resistance or R-value of the insulation. The property k_a is often referred to as the k-factor. In some cases the Greek letter λ is used. These terms are appropriate for fibrous or porous insulation containing air or some other low-thermal-conductivity gas or vacuum. Insulation such as cellulose, cork, fiberglass, foamed rubber, rock wool, perlite or vermiculite powder, and cellular plastics or glasses fall into this category of insulation sometimes called mass insulation. The thermal resistivity is useful for discussing insulations since the R-value is the product of thickness (in appropriate units) and k_a. Insulation in the building industry is commonly marketed on the basis of R-value while industrial users often prefer k_a, or k-factor. The greater the R-value or the smaller the k_a, the greater the insulation's effectiveness. The physical units commonly associated with the terms k_a, R^*, and R are listed in Table 1. The R-values of insulation of the type mentioned above are directly proportional to thickness as long as the thickness exceeds a minimum value, which is a few inches in most conditions.

In the case of thermal insulation that primarily reduces thermal radiation across air spaces, the term k_a is not used. This type of insulation is called reflective insulation, and R is not always directly proportional to thickness. The R-value of a reflective system is the temperature difference across the system divided by the heat flux.

Term	***Inch-Pound System (IP)***	***Scientific-International (SI)***
k_a	Btu·in./ft^2·h·°F	W/m·k
R^*	ft^2·h·°F/Btu in.	m·K/W
R	ft^2·h·°F/Btu	m^2·K/W
Thickness	in.	m
Heat flow rate	Btu/h	W
Heat flux	Btu/h·ft^2	W/m^2

Table 1.
Units Associated with Thermal Insulation

Fibrous, open-cell, and particulate insulations reduce heat flow because of the low thermal conductivity of the air in the insulation, which occupies a large fraction of the material's volume. The conduction across the material is limited because the air in the void fraction has low thermal conductivity. The structure provided by the solid restricts the movement of the air (convection) and limits radiation across the insulated space by intercepting and scattering thermal radiation. The type of "mass" insulation eliminates internal convection, reduces radiation, and increases conduction slightly because a small fraction of air volume is replaced by solid volume. The R-values of this type of insulation generally decrease with increasing temperature because the thermal conductivity of air increases with temperature. The R-value of mass-type insulation also depends on density, which is directly related to the amount of solid present. In fiberglass insulation at low density and 37.7°C (100°F), for example, R^* increases as density increases primarily because of reduced radiation. At high densities, R^* decreases because of increased solid conduction.

Closed-cell plastic foams such as polystyrene, polyurethane, and polyisocyanurate initially containing a gas other than air represent an important class of insulations. The gases that are used in the manufacturing process (blowing agents) usually have thermal conductivites less than that of air. Gases such as propane, carbon dioxide, and fluorinated hydrocarbons are examples of gases used as blowing agents. Closed-cell foams containing these or similar low-thermal-conductivity gases have R^* values up to about two times those of air-filled products. These products retain their relatively high R^* values as long as the blowing agent is not lost or unduly diluted by air. The lifetime average R-values for these products are less than the R-values of fresh products because air diffuses into the cells with time and the blowing agent diffuses out, thus changing the chemical composition and thermal conductivity of the gas mixture in the cells. Insulation board stock made of closed-cell polymeric foam are often faced with a gas-impermeable barrier material such as metal foil to reduce or prevent gas diffusion from taking place and thus maintain their R-values with time.

Fiberglass and rock wool insulations are produced by spinning and cooling to produce fine fibers. Fiberglass is produced from molten glass, binder or adhesive is added to produce batt-type insulation, and dye can be added for color. Rock wool insulation is produced from molten slag from iron or copper production. Cellulosic insulation is manufactured from recycled newsprint or cardboard, using equipment that produces small particles or flakes. Fire-retardant chemicals such as boric acid, ammonium sulfate, borax, and ammoniated polyphosphoric acid are added to reduce the smoldering potential and surface burning. Perlite and vermiculite particulate insulations are produced from ores passed through high-temperature furnaces. The high-temperature processing results in a lightweight particle with low bulk density. These particulate insulations are not flammable.

Reflective insulation and reflective insulation systems reduce the heat flux across enclosed air spaces using surfaces that are poor thermal radiators. These insulations have little effect on conduction across the air space, usually decrease the natural convection, and significantly reduce radiation. The R-values for reflective insulations like mass insulations depend on temperature. The terms k_a and R*, however, are not appropriate. The R-values for reflective systems depend on the number and positioning of the reflective surfaces, the heat flow direction, and the temperature difference across the system. The unique feature of reflective insulations is that they include low-emittance, high-reflectance surfaces. The low-emittance material commonly used is aluminum foil, with emittance in the range of 0.03 to 0.05. The emittance scale goes from 0 to 1, with 0 representing no radiation being emitted. The aluminum foil is commonly bonded to paper, cardboard, plastic film, or polyethylene bubble pack. The backing materials provide mechanical strength and support needed for handling and attachment. R-values for reflective insulations are generally measured for heat flow directions up, horizontal, and down.

Thermal insulations used in industrial applications include high-density fiberglass, high- density rock wool, particle insulations such as perlite and vermiculite, solids formed from perlite or calcium silicate, and occasionally reflective systems. The temperature range for industrial applications is much greater than that for buildings, so there is a wider variety of materials available. Fine perlite powder, for example, is used in low-temperature applications, while formed perlite or calcium silicate is used for high-temperature applications. Calcium silicate, perlite, high-density fiberglass, high-density rock wool, and foam glass are used for applications such as pipe insulation and furnace insulation. Each material has a range of temperatures in which it can be used, so it is important to match the material with the application. Reflective insulation produced from polished aluminum or stainless steel is also used for high-temperature applications.

Selected references that include thermal insulation data are listed in the bibliography at the end of this article. Table 2 contains nominal thermal resistivities of ten commonly used insulations. The thermal resistivities in Table 2 are at 23.9°C (75°F) and include the effects of aging and settling.

To obtain thermal resistance from R* multiply R* by the thickness of the insulation in the appropriate units, inches or meters. In the case of R* in IP units ($ft^2 \cdot h \cdot °F/Btu \cdot in.$), the thickness is in inches to give R_{IP} in $ft^2 \cdot h \cdot °F/Btu$, the units commonly used in the United States. For R_{SI}, the thickness is in meters to obtain units $m^2 \cdot K/W$. Conversely, the thickness required for a specific R-value is the ratio R_{IP}/R^*_{IP} or R_{SI}/R^*_{SI} . To obtain $R_{IP} = 30$, for example, a thickness of 10.3 inches of a fiberglass batt insulation with $R^*_{IP} = 2.9$ is required.

The R*s of a fibrous or cellular insulation like those in Table 2 generally decrease as the temperature increases. In the case of closed-cell polymeric foams like polyurethane or polyisocyanurate board, the R* may decrease if the insulation temperature drops below the condensation temperature of the blowing agent in the cells. This is because of changes in the gas- phase composition and therefore the gas-phase thermal conductivity. The R* of insulations also depends on density when all other factors are constant. The relationship between R* and density

Insulation Type	*R*(ft^2·h·°F/Btu·in.)$^{(a)}$*	*R*(m·K/W)$^{(b)}$*
Fiberglass batts (standard)	2.9 - 3.8	20.1 - 26.3
Fiberglass batts (high performance)	3.7 - 4.3	25.7 - 29.8
Loose-fill fiberglass	2.3 - 2.7	15.9 - 18.7
Loose-fill rock wool	2.7 - 3.0	18.7 - 20.8
Loose-fill cellulose	3.4 - 3.8	23.6 - 26.3
Perlite or vermiculite	2.4 - 3.7	16.6 - 25.7
Expanded polystyrene board	3.6 - 4.0	25.5 - 27.7
Extruded polystyrene board	4.5 - 5.0	31.2 - 34.7
Polyisocyanurate board, unfaced	5.6 - 6.3	38.8 - 43.7
Plyurethane foam	5.6 - 6.3	38.8 - 43.7

Table 2.
Nominal Values for R* at 23.9°C (75°F)
Note: (a) The values listed are for one inch of thickness.
(b) The values listed are for one meter of thickness.

depends on the type of insulation in question. In general, the compression or settling of an insulation results in a decrease in the R value.

Reflective insulations consist of enclosed air spaces bounded by low-emittance surfaces. The distance across the air spaces is set by the insulation designer. The spacing is generally in the range of 0.5 to 2.0 in, with one to seven air spaces provided by different products. The low-emittance (high-reflectance) surfaces serve as radiation barriers and, as a result, thermal radiation across the enclosed air spaces is dramatically reduced. The open air spaces, however, can have natural convection (air movement) occurring so the R-values depend on the heat-flow direction and the magnitude of the temperature difference across each air space in the system. The R-value of a reflective system is greatest when the heat-flow direction is downward, and least when the heat-flow direction is upward.

TRENDS

The design and manufacturing of cellulose, fiberglass, and rock wool are constantly being improved. The resulting increases in R*, however, are relatively small. The fiberglass and rock wool producers seek optimum fiber diameters and improved radiation attenuation. The cellulosic insulation producers have changed from hammer mills producing relatively high-density products to fiberizers that produce relatively low-density products. The result in most cases is improved R* with less material required. In some cases adhesives are added to building insulations to improve mechanical stability so that long-term settling of the insulation is reduced. Adhesives also have been added to loose-fill fiberglass insulation to produce a blow-in-blanket insulation (BIBS) widely used as insulation in cavity walls.

More advanced insulations are also under development. These insulations, sometimes called superinsulations, have R* that exceed 20 ft^2·h·°F/Btu·in. This can be accomplished with encapsulated fine powders in an evacuated space. Superinsulations have been used commercially in the walls of refrigerators and freezers. The encapsulating film, which is usually plastic film, metallized film, or a combination, provides a barrier to the inward diffusion of air and water that would result in loss of the vacuum. The effective life of such insulations depends on the effectiveness of the encapsulating material. A number of powders, including silica, milled perlite, and calcium silicate powder, have been used as filler in evacuated superinsulations. In general, the smaller the particle size, the more effective and durable the insulation packet. Evacuated multilayer reflective insulations have been used in space applications in past years.

Silica aerogels, a newly developing type of material, also have been produced as thermal insulations with superinsulation characteristics. The nanometer-size cells limit the gas phase conduction that can take place. The aerogels are transparent to visible light, so they have potential as window insulation. The use of superinsulations at present is limited by cost and the need to have a design that protects the evacuated packets or aerogels from mechanical damage.

David W. Yarbrough

See also: Building Design, Commercial; Building Design, Energy Codes and; Building Design, Residential; Domestic Energy Use; Economically Efficient Energy Choices; Efficiency of Energy Use.

BIBLIOGRAPHY

Goss, W. P., and Miller, R. G. (1989). "Literature Review of Measurement and Predictions of Reflective Building Insulation System Performance." *ASHRAE Transactions* 95(2):651–664.

Govan, F. A.; Greason, D. M.; and McAllister, J. D., eds. (1981). *Thermal Insulation, Materials, and Systems for energy Conservation in the 80s* (STP 789). West Conshohocken, PA: American Society for Testing and Materials.

Graves, R. S., and Wysocki, D. C., eds. (1991). *Insulation Materials: Testing and Applications*, vol. 2 (STP 1116). West Conshohocken, PA: American Society for Testing and Materials.

Graves, R. S., and Zarr, R. P., eds. (1997). *Insulation Materials: Testing and Applications*, vol. 3 (STP 1320). West Conshohocken, PA: American Society for Testing and Materials.

Industrial Insulation (ORNL/M-4678). (1995). Oak Ridge, TN: Oak Ridge National Laboratory.

Kreith, F., ed. (1999). *The CRC Handbook of Thermal Engineering*. Boca Raton, FL: CRC Press.

Kreith, F., and Bohn, M. S. (1997). *Principles of Heat Transfer*. Boston: PWS.

McElroy, D. L., and Kimpflen, J. F., eds. (1990). *Insulation Materials: Testing and Applications*, vol. 1 (STP 1030). West Conshocken, PA: American Society for Testing and Materials.

McElroy, D. L., and Tye, R. P., eds. (1978). *Thermal Insulation Performance* (STP 718). West Conshohocken, PA: American Society for Testing and Materials.

Powell, F. J., and Matthews, S. L., eds. (1984). *Thermal Insulation: Materials and Systems* (STP 922). West Conshohocken, PA: American Society for Testing and Materials.

Pratt, A. W. (1981). *Heat Transmission in Buildings*. New York: John Wiley and Sons.

Rohsenow, W. M.; Hartnett, J. P.; and Cho, Y. I. (1998). *Handbook of Heat Transfer*, 3rd ed. New York: McGraw-Hill.

Turner, W. C., and Malloy, J. F. (1998). *Thermal Insulation Handbook*. New York: McGraw-Hill.

Tye, R. P., ed. (1977). *Thermal Transmission Measurements of Insualtion* (STP 660). West Conchohocken, PA: American Society for Testing and Materials.

INSULATOR, ELECTRIC

See: Electricity

INTELLIGENT TRANSPORTATION SYSTEMS

See: Traffic Flow Management

INTERNAL COMBUSTION ENGINES

See: Engines

IPATIEFF, VLADIMIR NIKOLAEVITCH (1967–1952)

Vladimir Nikolaevitch Ipatieff, one of the founding fathers of high-pressure catalysis, was the innovating force behind some of this country's most important petroleum processing technologies in the years leading up to World War II.

THE EARLY YEARS

Ipatieff was born in Russia in 1867. As a member of the privileged (i.e., noble) class, Ipatieff prepared for a military career. Early on in his education, Ipatieff gravitated toward the sciences, and in particular chemistry.

Ipatieff's formal training in chemistry began in earnest when, at the age of twenty-two, he entered the Mikhail Artillery Academy in St. Petersburg. The Academy was founded to give technical training to officers who were to serve as engineers and in other positions within the Russian military.

Chemical training at the Academy was particularly strong. The school taught Ipatieff how to apply textbook concepts to actual plant conditions. As a student, and then later as an instructor at the Academy, he regularly visited local metallurgical and chemical plants to examine first-hand industrial production.

While still a student at the Academy, Ipatieff began to make a name for himself in the Russian chemical community as he began to publish some of his laboratory findings. His first professional milestone as a chemist came in 1890 when he joined Russia's Physical-Chemical Society. Here he came into close contact with Russia's most famous chemists, including Dimitri Mendeleev, discoverer of the periodic table and one of the founders of the Society. In 1891, upon graduating from the school, he was appointed lecturer in chemistry at the Academy where he also continued to undertake original chemical research for his doctoral dissertation. In 1895, he was made assistant professor and, upon completion and acceptance of his dissertation in 1899, he became a full professor of chemistry.

Right up to World War I, Ipatieff, as part of his responsibilities at the Academy, traveled extensively to conferences and seminars throughout Europe and the United States. As a result, he made the acquaintance of some of the major chemists of the day.

During these years he gained an international reputation in the area of experimental chemistry in general and high pressure catalysis in particular. At the Academy he designed, built, and directed one of the first permanent high-pressure laboratories in Europe. Ipatieff investigated the composition and synthesis of numerous aliphatic compounds, including isoprene and the complex alcohols. He demonstrated for the first time the catalytic effect of the metal reactor walls on reactions, a result which would have important commercial implications. Ipatieff also undertook the first sophisticated experiments in catalytic polymerization, isomerization, and dehydrogenation of organic materials, including alcohols and petroleum.

He became intimately familiar with a wide range of catalytic materials—including aluminum oxide, silica, and clay, as well as nickel, platinum, zinc, and copper—and their role individually and as mixtures in effecting chemical transformation. One of Ipatieff's most important lines of research was his breakthrough work on the nature and mechanisms of catalytic promoters on organic reactions.

Although Ipatieff was dedicated to research, his practical training at the Academy prepared him well for applying his research to commercial problems. Ipatieff consulted to industry and government on a regular basis. He was often called upon to make commercial grade catalysts for a variety of applications. One of his most important early assignments, obtaining gasoline from Caspian sea petroleum, introduced Ipatieff to industrial research in fuels. He also used his expertise to undertake research for the government and helped expand Russia's explosives capacity during World War I.

THE POSTREVOLUTIONARY YEARS

Ipatieff's life changed dramatically with the Russian Revolution in 1917. The Bolsheviks were hesitant to retain in official positions those who were too closely allied with the old regime.

Ipatieff, however, remained loyal to Russia, even in its new form. He was also a pragmatist who understood that he must find a way to accommodate the new way of life in Russia, and gain the confidence of those currently holding power if he was to continue his chemical research.

Over the next few years, Ipatieff called upon the wide contacts he had made within Russian scientific and administrative circles before the revolution, and his growing international reputation, to negotiate with the Bolsheviks a favorable place for his research activity. An important element in his success was his ability to convince the government of the strategic importance of his work, which (he pointed out) had just been demonstrated during the war. Under Lenin, who knew of Ipatieff and his work, the scientist received government support. He remained at the Academy, which was retained to train future Bolshevik military officers. Most importantly, Ipatieff was given an official and prestigious research position as Director of the newly created Institute for High Pressures.

In this favorable environment, Ipatieff undertook some of his most important work in high pressure reactions. He studied the effect of catalysts and high pressures on reactions taking place in hydrogen atmospheres, and the effect of such metals as platinum and nickel on forming cyclic compounds from olefin-based materials. This was to become very important later in the 1930s and 1940s in the United States with the development of the platforming process.

When Stalin came to power, the political climate in Russia turned against the creative individuals—scientists, musicians, and writers—especially those who had strong roots in Czarist Russia. During the 1920s Ipatieff saw his colleagues and coworkers arrested, sent to labor camps, or executed.

THE LATER YEARS: AMERICAN PETROLEUM TECHNOLOGY

The opportunity for Ipatieff to emigrate to the United States came through a job offer by Dr. Gustav Egloff, Director of Research at the Universal Oil Products Co. (UOP). UOP, a Chicago-based firm, was a petroleum process development firm known in particular for its commercialization of the first continuous thermal cracking technology (i.e., the Dubbs Process). In the late 1920s, UOP wanted to go into catalytic cracking research. Having met Ipatieff in Germany while attending an international chemical conference, Egloff, who was familiar with Ipatieff's reputation, suggested that Ipatieff think of relocating to Chicago to direct UOP in these efforts.

Reluctantly, in 1929 Ipatieff left his homeland for the United States. He remained at UOP for the remainder of his life. Northwestern University, which had close ties to UOP, appointed Ipatieff to a professorship and directorship of the university's high-pressure research laboratory.

At UOP, Ipatieff had the opportunity to apply his former research in catalytic promoters and high-pressure technique to develop important catalytic petroleum processing technologies. In contrast to the way he conducted science, Ipatieff's technical efforts were conducted in teams comprised of a wide assortment of specialists.

Through the 1930s, Ipatieff led UOP in its effort to develop two catalytic processes for the production of high-octane fuel: alkylation and polymerization—the first, a reaction of a hydrocarbon with an olefin (double-bonded compound); the second, the formation of long molecules from smaller ones. Both processes produce high-octane blending compounds that increase the quality of cracked gasoline.

Ipatieff's work on promoted phosphoric acid and hydrogen fluoride catalysts, which extended his earlier research in Russia, provided the key to commercializing alkylation and polymerization technologies. Shell and other refiners quickly established industrial scale operations. These technologies, generally operating together and in tandem with cracking operations, created vast amounts of high quality aviation fuel. They were a vital part of the Allied war effort during World War II.

Ipatieff's final contribution to catalytic technology was more indirect but essential. His guidance and suggestions to Vladimir Haensel, a fellow Russian émigré who worked on catalytic reforming at UOP and studied under Ipatieff at Northwestern, were a significant contribution to Haensel's development of the high-pressure reforming technology known as platforming. An extension of Ipatieff's previous work involving platinum catalysts, high pressure, and hydrogen environments, platforming represents the first continuous catalytic reforming process. It produces large tonnages of ultra-high octane gasoline materials and critical organic intermediates (benzene, toluene, and xylene), previous obtained only in limited quantities from coal tar. Platforming, along with fluid catalytic cracking, is generally considered one of the great petrochemical innovations of the century.

Ipatieff never received the honor he coveted the most, the Nobel Prize. However, he continued to publish voluminously and, ever the practical scientist, he obtained numerous patents. He continued to receive honors in the United States and internationally. He became a member of the National Academy of Sciences and received the prestigious Gibbs medal for his many achievements. Ipatieff lived long enough to see the petroleum industry transformed with process technologies that he created and that were rooted in his early scientific research. The first platforming plant, the culmination of his life's work in catalytic research, came on line just shortly before his death in 1952.

Sanford L. Moskowitz

See also: Catalysts; Refining, History of.

BIBLIOGRAPHY

Enos, J. (1962). *Petroleum, Progress and Profits: A History of Process Innovation*. Cambridge, MA: MIT Press.

Frankenburg, W. G., et. al., eds. (1954). *Advances in Catalysis and Related Subjects*. New York: Academic Press.

Haensel, V. (1935). "Polymerization of Butylene." M.S., MIT.

Haensel, V. (1983). "The Development of the Platforming Process—Some Personal and Catalytic Recollections." In *Heterogeneous Catalysis: Selected American Histories*, eds. B. H. Davis and W. P. Hettinger. American Chemical Society Symposium Series No. 222. Washington, DC: American Chemical Society.

Ipatieff, V. N. (1936). *Catalytic Reactions at High Pressures.* New York: Macmillan.

Ipatieff, V. N. (1946). *The Life of a Chemist: Memoirs of V. N. Ipatieff*, trans. V. Haensel and Mrs. R. H. Lusher, eds. X. J. Eudin, H. D. Fisher, and H. H. Fisher. Stanford, CA: Stanford University Press.

Remsberg, C., and Higden, M. (1994). *Ideas for Rent: The UOP Story*. Chicago, IL: Universal Oil Products Co.

Rideal, E. K., and Taylor, H. S. (1926). *Catalysis in Theory and Practice*. London: Macmillan.

Spitz, P. (1988). *Petrochemicals: The Rise of an Industry*. New York: John Wiley and Sons.

Williamson, H. F. (1963). *The American Petroleum Industry, Vol. II: The Age of Energy, 1899-1959*. Evanston, IL: Northwestern University Press.

J

JET ENGINE

See: Aircraft

JET FUEL

See: Aviation Fuel

JOULE

See: Units of Energy

JOULE, JAMES PRESCOTT (1818–1889)

James Joule was born in Salford, near Manchester, England, on December 24, 1818. He was the second son of a wealthy brewery owner and was educated at home by private tutors. For three years he was fortunate enough to have the eminent British chemist, John Dalton as his chemistry teacher. He never attended a university; as a consequence, while he was bright enough to learn a great deal of physics on his own, he remained, like Michael Faraday, unskilled in advanced mathematics.

Joule had the means to devote his time to what became the passion of his life — obtaining highly accurate experimental results in physics, for which he displayed a precocious aptitude. His genius showed itself in his ability to devise new methods, whenever needed, to improve on the accuracy of his quantitative results.

Joule had no real profession except as an amateur scientist, and no job except for some involvement in running the family brewery. Since his father was ill and forced to retire in 1833, his son had to become more involved in the affairs of the brewery from 1833 to 1854, when the brewery was sold by his family. While Joule was working at the brewery, he carried out his experiments before 9:00 A.M., when the factory opened, and after 6:00 P.M., when it closed. Because his father built a laboratory for him in his home, in 1854 he had the time and means to devote himself completely to physics research. Later in life, he suffered severe financial misfortune, but the Royal Society and Queen Victoria in 1878 each provided a £200 subsidy for Joule to continue his important researches.

In 1847 Joule married Amelia Grimes, and they had two children who survived them. Another son was born on June 8, 1854, but died later that month. This was followed by an even greater tragedy—within a few months Joule's wife also passed away. Joule never remarried, but spent the rest of his life with his two children in a variety of residences near Manchester.

Joule died in Sale, Cheshire, England, on October 11, 1889. He always remained a modest, unassuming man, and a sincerely religious one (even though he

James Prescott Joule. (Library of Congress)

was in the habit of falling asleep during sermons). Two years before his death he said to his brother, "I have done two or three little things, but nothing to make a fuss about." Those "two or three little things" were so important for the advancement of science that Joule was elected in 1850 as a fellow of the Royal Society of London, received the Copley Medal (its highest award) in 1866, and was elected president of the British Association for the Advancement of Science in 1872 and again in 1887. Joule is memorialized by a tablet in Westminster Abbey, and constantly comes to the attention of physicists whenever they use the unit of energy now officially called the joule (J).

JOULE'S CONTRIBUTIONS TO THE PHYSICS OF ENERGY

Joule's interest in the conservation of energy developed as a consequence of some work he did in his teens on electric motors. In 1841 he proposed, on the basis of his experiments, that the rate at which heat Q is generated by a constant electric current i passing through a wire of electrical resistance R is:

$$DQ/Dt = i^2R$$

now called Joule's Law.

From 1841 to 1847 Joule worked steadily on measuring the heat produced by electrical processes (Joule's Law), mechanical processes (rotating paddles churning water or mercury), and frictional processes (the rubbing of materials together, as Count Rumford had done in 1798). In each case he compared the amount of energy entering the system with the heat produced. He proved his mettle as a physicist by spending endless days ferreting out the causes of errors in his experiments and then modifying his experimental set-up to eliminate them. In this way he produced a remarkably precise and accurate value for the constant that relates the energy entering the system (in joules) with the heat produced (in calories). This constant is now called Joule's Equivalent, or the mechanical equivalent of heat.

In 1847 Joule published a paper that contained an overwhelming amount of experimental data. All his results averaged out to a value of 4.15 J/cal (in modern units), with a spread about this mean of only five percent. The best modern value of Joule's Equivalent is 4.184 J/cal, and so his results were accurate to better than one percent. This was truly amazing, for the heat measurements Joule performed were the most difficult in all of physics at that time.

At the British Association meeting at Oxford in June 1847, at which Joule presented his results, his audience's reaction was much more subdued and uninterested than he had expected. Joule fully believed that his paper would have passed unnoticed had not the 23-year-old William Thomson (later Lord Kelvin) asked a number of penetrating questions. These awakened his colleagues to the significance of Joule's work as a proof of the conservation-of-energy principle (now commonly called the first law of thermodynamics) under a variety of experimental conditions and involving many different types of energy.

This event marked the turning point in Joule's career. From 1847 on, when Joule spoke, scientists listened. His research results were one of the two major contributions to the establishment of the first law of thermodynamics, the other being that of the

German physician Julius Robert Mayer. Mayer's work, although historically important for its insights into the conservation-of-energy principle, was however tainted by errors in physics and an unacceptable reliance on philosophical arguments.

In addition to his work on the conservation of energy, Joule made a number of other important contributions to physics. In 1846 he discovered the phenomenon of magnetostriction, in which an iron rod was found to change its length slightly when magnetized. In 1852, together with William Thomson, he showed that when a gas is allowed to expand into a vacuum, its temperature drops slightly. This "Joule-Thomson effect" is still very useful in the production of low temperatures.

Joule believed that nature was ultimately simple, and strove to find the simple relationships (like Joule's law in electricity), which he was convinced must exist between important physical quantities. His phenomenal success in finding such relationships in the laboratory made a crucial contribution to the understanding of energy and its conservation in all physical, chemical and biological processes.

Joseph F. Mulligan

BIBLIOGRAPHY

Cardwell, D. S. L. (1971). *From Watt to Clausius.* Ithaca, New York: Cornell University Press.

Cardwell, D. S. L. (1989). *James Joule: A Biography.* Manchester, England: Manchester University Press.

Crowther, J. G. (1936). "James Prescott Joule." In *Men of Science.* New York: W.W. Norton.

Joule, J. P. (1963). *The Scientific Papers of James Prescott Joule*, 2 vols. London: Dawson's.

Rosenfeld, L. (1973). "Joule, James Prescott." In *Dictionary of Scientific Biography,* ed. Charles Coulston Gillispie, Vol. 7, pp. 180–182. New York: Scribner.

Steffens, H. J. (1979). *James Prescott Joule and the Development of the Concept of Energy.* New York: Science History Publications.

Wood, A. (1925). *Joule and the Study of Energy.* London: G. Bell and Sons, Ltd.

K

KAMERLINGH ONNES, HEIKE (1853–1926)

Heike Kamerlingh Onnes was born on September 21, 1853, in Groningen, the Netherlands, and died on February 21, 1926, in Leiden. His father owned a roof-tile factory. Heike Kamerlingh Onnes entered the University of Groningen to study physics. When the government threatened to permanently shut down the university for monetary reasons, he led a delegation to the seat of the government in the Hague as president of the student government. During his studies he won several prizes in physics; his Ph.D. thesis demonstrates superior mathematical abilities. Despite delicate health through much of his life, he showed an enormous capacity for work.

Kamerlingh Onnes does not fit the description of a loner; on the contrary, he created one of the first laboratories to be set up as if it were a factory. Per Dahl in his comparison between Kamerlingh Onnes and his British counterpart, James Dewar, states that Onnes, to be sure was paternalistic, opinionated, and a man of strong principles—traits not uncommon among the moguls of late nineteenth-century science—but that he proved to be a benevolent leader, kind and scrupulously fair in his relations with friends and pupils alike—behavior that was certainly within the norms of his time.

After Kamerlingh Onnes was appointed professor in experimental physics at the University of Leiden in 1882, he stated in his inaugural lecture that physics is capable of improving the well-being of society and proclaimed that this should be accomplished primarily through quantitative measurements. He laid out what he was going to do and, to the surprise of onlookers, that is what he did. He was an excellent organizer, and set out to equip his laboratory on a grand scale. The city of Leiden did not yet provide electricity, so he acquired a gas motor and generator and made his own electricity. Pumps and compressors were barely available, so he had them made in the machine shops of the laboratory. There was a need for measuring instruments, so he created an instrument makers' school. There was an even greater demand for glassware (Dewars, McLeod gauges and connecting tubes, etc.), so he created a glassblowers' school that became famous in its own right.

All currents that had to be measured were sent to a central "measurement room" in which many mirror galvanometers were situated on top of vibration-free columns that were separated from the foundations of the building. One should realize that the many announcements in the early literature of the liquefaction of specific gases pertained to not much more than a mist or a few drops; Kamerlingh Onnes planned to make liquid gases by the gallon. A separate hydrogen liquefaction plant was located in a special room with a roof that could be blown off easily.

The availability of large quantities of liquid helium as well as an excellent support staff led to the undertaking of many experiments at 8 K (the boiling temperature of helium) as well as the lower temperatures obtained by pumping. One subset was the measurement of the resistivity (conductivity) of metals, since this property was useful as a secondary thermometer. Although a linear decline was observed, various speculations were made as to what the result would be when zero absolute temperature was reached. In April 1911 came the surprising discovery that the resistivity in mercury disappeared. At first the sur-

Heike Kamerlingh Onnes. (Library of Congress)

prise was a sharp jump to what was thought to be a small value. The first reaction to this baffling result was to suspect a measurement error. All electrical machines in the laboratory were shut down to be sure that there were no unforeseen current leaks and the experiment was repeated several times. Very careful verifications finally showed that the effect was real and that the resistance was indeed unmeasurably small ("sinks below 10^{-4} of the resistance at 0°C"), and moreover it was found that the effect existed also in other metals, even those that could not be purified as well as mercury.

Initially there was speculation about building an "iron-less" magnet, but this hope was dashed when it was discovered that a small field destroyed the superconductivity. Not until 1960 when materials where found that could sustain high fields, did superconductivity show promise for building strong magnets. The other obstacle was the need for a low-temperature environment. Raising the critical temperature had been a goal for many years, and a spectacular breakthrough was made in 1986.

The result was called by Dutch physicist H. A. Lorentz "perhaps the most beautiful pearl of all [of Kamerlingh Onnes's discoveries]." However, as H. B. G. Casimir describes in his memoirs, he refused to give any credit to the graduate student who observed the phenomenon and who realized its importance.

Although the discovery can be called accidental, one may ask what led up to it. The decline in resistivity of metals when the temperature was lowered clearly invited further study. This program (J. van den Handel calls it Kamerlingh Onnes's second major field of research; liquefaction being the first) was both intrinsically interesting as well as relevant to the construction of a good secondary thermometer. It was known that the resistivity was a linear function of the temperature but was noticed to level off at lower temperatures. The height of this plateau was found to depend on the amount of impurities, using a series of experiments with gold, since the amount of admixture in this metal can be easily controlled. To lower this plateau, a metal of very high purity was needed. Since zone melting, in the modern sense, did not exist, the choice fell on mercury, because this metal could be purified by distillation, and the purified liquid was then placed in glass capillaries. When the liquid in these capillaries was frozen, it formed a "wire." Moreover, it became clear that these wires were free of dislocations, to use a modern term, because it was found that pulling of the wires resulted in increased values of the residual resistance. Hence mercury was the best bet to see how far the linear part of the resistivity curve could be extended to lower temperatures. The hope to have a linear resistor at very low temperatures was certainly a driving factor for the research that led to the discovery of superconductivity.

Heike Kamerlingh Onnes was awarded the Nobel Prize in physics in 1913.

Paul H. E. Meijer

See also: Electricity; Energy Intensity Trends; Heat and Heating; Heat Transfer; Magnetism and Magnets; Molecular Energy; Refrigerators and Freezers; Thermal Energy.

BIBLIOGRAPHY

Bruyn Ouboter, R. de. (1997). "Heike Kamerlingh Onnes's Discovery of Superconductivity." *Scientific American* 276:98–103.

Casimir, H. B. G. (1983). *Haphazard Reality: Half a Century of Science.* New York: Harper and Row.

Dahl, P. F. (1992). *Superconductivity: Its Historical Roots and Development from Mercury to the Ceramic Oxides.* New York: American Institute of Physics.

Handel, J. van den. (1973). "Kamerlingh Onnes, Heike." In *Dictionary of Scientific Biography*, Vol. 7, ed. C.C. Gillispie. New York: Charles Scribner's Sons.

Meijer, P. H. E. (1994). "Kamerlingh Onnes and the Discovery of Superconductivity." *American Journal of Physics* 62:1105–1108.

Nobel, J. de. (1996). "The Discovery of Superconductivity." *Physics Today* 49(9):40–42.

KEROSENE

Kerosene is that fraction of crude petroleum that boils between about 330° to 550° F (183° to 305° C) (i.e. the distillate cut between gasoline and heavy fuel oil). What is usually thought of as kerosene, the domestic fuel or "range oil" commonly used for space heaters, is only a small part of this fraction. Diesel fuel, aviation jet fuel, and No. 1 and No. 2 heating oils also come from this general boiling range.

HISTORY

Petroleum was discovered by Drake in Pennsylvania in 1859. It is a mixture of hydrocarbon compounds ranging from highly flammable volatile materials like propane and gasoline to heavy fuel oil, waxes, and asphalt. It was soon discovered that crude oil could be separated by distillatoin in to various factions according to their boiling point (molecular weight). Kerosene, a middle distillate, was the first product made from petroleum that had a substantial commercial market. The Pennsylvania crudes happened to be low in sulfurous and aromatic components, so the straight run (unrefined) distillate was relatively odorless and clean-burning. The distillate rapidly replaced whale oil in lamps, and coal in home room heaters, and by 1880 kerosene accounted for 75 percent by volume of the crude oil produced. However, after the turn of the century, the advent of electric lights soon made kerosene lamps obsolete. Later, central home heating based on heating oil and natural gas largely replaced space heating based on kerosene, wood, and coal.

By 1980 production of "range oil" type kerosene was less than one percent of the crude oil refined. By the end of the twentiteth century, kerosene's use for space heating, water heating, cooking, and lighting was largely limited to camps, cabins, and other facilities remote from centralized energy sources. Kerosene is a relatively safe fuel, but to prevent fires care must be taken not to tip over operating kerosene appliances. While burner designs have been improved to minimize accidental upset, carbon monoxide formation is still a potential hazard. This deadly gas can result from use of dirty, poorly maintained burners or if the heater is used in a space with insufficient ventilation.

Kerosene is also used as a solvent for herbicides and insect sprays. However, most of the kerosene fraction in crude oils is used to make Diesel engine fuel and aviation jet fuel.

PROPERTIES

Kerosene is an oily liquid with a characteristic odor and taste. It is insoluble in water, but is miscible with most organic solvents. Structurally, it is composed mostly of saturated hydrocarbon molecules containing twelve to fifteen carbon atoms. When a sweet (low sulfur content) paraffinic crude oil is cut to the proper boiling range in a refinery's atmospheric pipe still, the resulting kerosene may only require a drying step before use. However, if the crude oil contains aromatic ring compounds like xylenes, they must be eliminated by solvent extraction because they burn with a smoky flame. Olefinic (unsaturated) molecules in this boiling range must also be removed by refining, usually by hydrogenation. These compounds tend to form color bodies and polymerize, thus imparting poor storage stability. Sulfur compounds are undesirable because of their foul odor and formation of air polluting compounds. They can also be removed by hydrogenation.

Other important properties include flash point, volatility, viscosity, specific gravity, cloud point, pour point, and smoke point. Most of these properties are related directly to the boiling range of the kerosene and are not independently variable. The flash point, an index of fire hazard, measures the readiness of a fuel to ignite when exposed to a flame. It is usually mandated by law or government regulation to be 120° or 130° F (48° or 72° C). Volatility, as measured

	Kerosene	*Diesel Fuel*	*#2 Heating Oil*
Gravity, A.P.I.	40	37	34
Boiling Range, °F.	325 - 500	350 - 650	325 - 645
Viscosity, SSU @ 100° F.	33	35	35
Flash Point, °F.	130	140	150
Sulfer, Weight Percent	0.05 - 0.12	0.30	0.40

Table 1.
Comparison of Heating Fuels

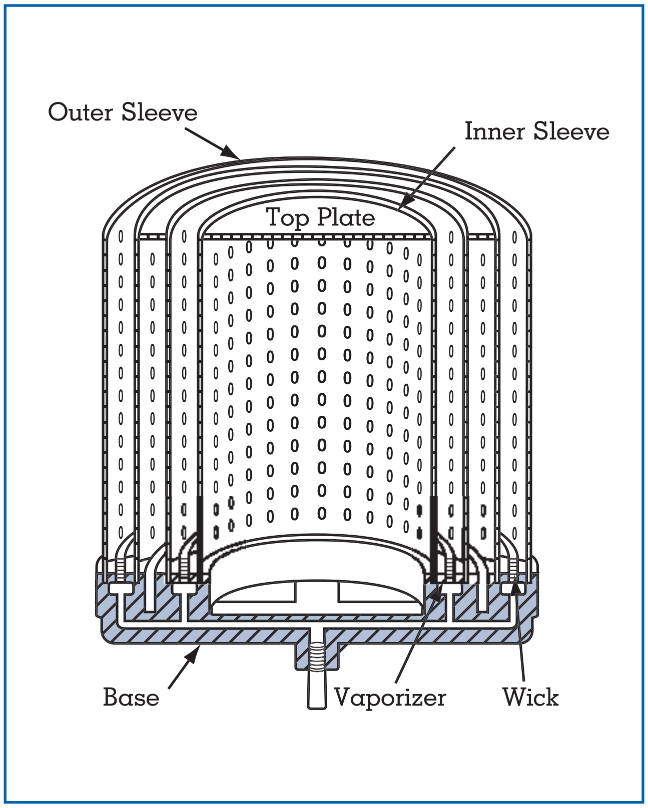


Figure 1.
Cross-section of a range burner.

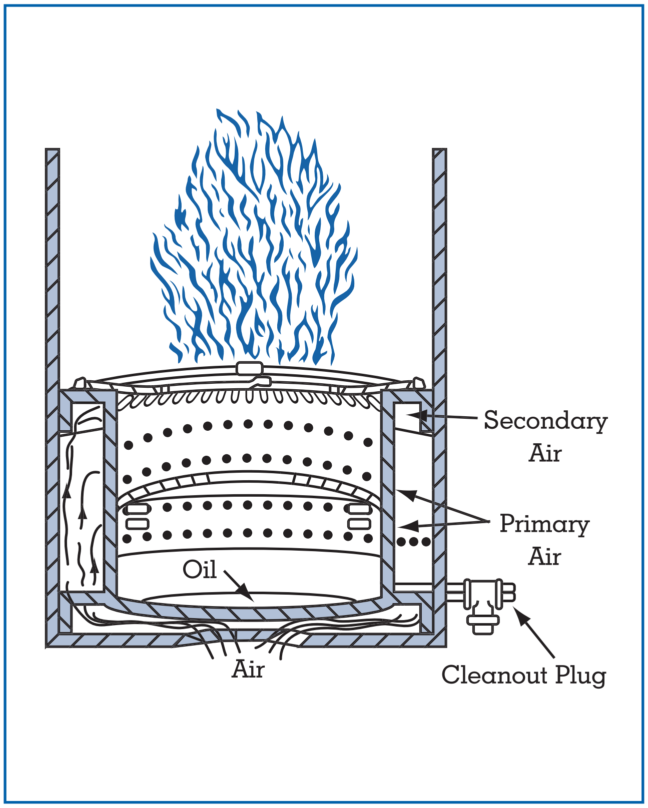


Figure 2.
Cross-section of a vaporizing pot burner.

by the boiling range, determines the ease with which the fuel can evaporate in a burner or lamp. Viscosity measures the fuel's resistance to flow, and thus determines the ease with which it can be pumped or atomized in a burner nozzle. Specific gravity is the ratio of the weight of a given volume of fuel to the same volume of water. Denser, higher gravity fuels have higher heating values. Typical kerosenes have heating values of about 139,000 BTU per gallon. Cloud and pour points indicate the temperature at which the precipitation of waxy constituents affect performance in cold climates. As wax begins to form (cloud point) there is danger of plugging fuel nozzles and filters. At an even lower temperature (the pour point), enough wax has formed so that the fuel sets up and can no longer flow as a liquid. The smoke point is the flame height in millimeters at which a kerosene lamp begins to smoke.

Other properties of interest are carbon residue, sediment, and acidity or neutralization number. These measure respectively the tendency of a fuel to foul combustors with soot deposits, to foul filters with dirt and rust, and to corrode metal equipment. Cetane number measures the ability of a fuel to ignite spontaneously under high temperature and pressure, and it only applies to fuel used in Diesel engines. Typical properties of fuels in the kerosene boiling range are given in Table 1.

Because of its clean burning characteristics, kerosene commands a higher price than other fuels in its boiling range.

EQUIPMENT

The technology of kerosene burners is quite mature. The most popular kerosene heater is the perforated sleeve vaporizing burner or range burner (Figure 1). It consists of a pressed steel base with concentric, interconnected grooves and perforated metal sleeves, between which combustion takes place. Kerosene is maintained at a depth of about 1/4 inch in the grooves. As the base heats up, oil vaporizes from the surface, and the flame lights from asbestos wicks. Combustion air is induced by natural draft. The flame is blue, and the burner is essentially silent, odorless, and smokeless.

For larger capacity space heaters vaporizing pot burners are used (Figure 2). They consist of a metal pot perforated with holes for combustion air. Oil flows into the bottom of the pot by gravity, and is vaporized from the hot surface. Fuel vapors mix with primary air from the lower holes, then with additional secondary air in the upper section, and burn at the top of the vessel. Between periods of high demand, pot burners idle at a low fire mode. The flames burn in the bottom section of the pot, at a fraction of the high fire fuel feed rate, using only primary air. Thus no extraneous pilot light or ignition source is required for automatic operation.

Kerosene lamps have a flat cloth wick. Flame height is determined by the height of the wick, which is controlled by a ratchet knob. A glass chimney ensures both safety and a stable, draft-free flame.

Herman Bieber

BIBLIOGRAPHY

Gary, J. H., and Handwerk, G. E. (1993). *Petroleum Refining Technology and Economics*, 3rd ed. New York: M. Dekker.

Ramose, D. (1983). *Kerosene Heaters*. Blue Ridge Summit, PA: Tab Books.

U.S. Consumer Product Safety Commission. (1988). *What You Should Know About Kerosene Heaters*. Washington, DC: U.S. Government Printing Office.

Wood, Q. E. (1986). *Quaker State Roots Go Deep in World's First Oil Field*. New York: Newcomen Society of the United States.

Yergin, D. (1991). *The Prize: the Epic Quest For Oil, Money and Power*. New York: Simon & Schuster.

KILOWATT

See: Units of Energy

KINETIC ENERGY

Energy is “capacity (or ability) to do work,” and work is “the result of a force acting through some distance.” A car running into the rear of a stalled car exerts a force on it, pushing it some distance, doing work in the process. The capacity of the moving car to do work is termed its kinetic energy. The greater a car's speed and/or mass, the greater its capacity to do work—that is, the greater its kinetic energy.

Formally, the kinetic energy (K) of a mass (m) moving with speed (v) is defined as $K=1/2\ mv^2$. Kinetic energy is measured in joules (J) when m and v are expressed in kilograms (kg) and meters per second (m/s). A 1,000-kg car traveling 15 m/s (about 30 miles per hour) has 112,500 J of kinetic energy. Kinetic energy depends much more on speed than on mass. That is because doubling the mass of an object doubles the kinetic energy, but doubling the speed quadruples the kinetic energy. A 4,000-kg tractor trailer traveling at 30 m/s has the same kinetic energy as a 1,000-kg car traveling at 60 m/s.

Invariably, energy of use to a society is kinetic. A car is useful when it is in motion. Water in motion is useful for driving turbines in a hydroelectric plant. Electricity, the most versatile of all forms of energy, involves electric charges (electrons) in motion. And thermal energy, which provides energy for a steam turbine, is associated with the kinetic energy of molecules.

Most energy converters convert sources of potential energy to forms of kinetic energy that are useful. Gasoline in the tank of an automobile has potential energy. When burned, potential energy is converted to heat, a form of kinetic energy. Uranium in the core of a nuclear reactor has nuclear potential energy. Conversion of nuclear potential energy through nuclear fission reactions produces nuclei and neutrons with kinetic energy. This kinetic energy is the source for the electric energy produced by the nuclear power plant. Water atop a dam has gravitational potential energy. Flowing toward the bottom of the dam, the water continually loses potential energy but gains kinetic energy (and speed). The potential energy the water had at the top of the dam is converted entirely to kinetic energy at the bottom of the dam.

Every molecule in the air around you has kinetic energy because each has mass and is in incessant motion. According to the kinetic theory of gases, each of the molecules has average kinetic energy given by $E=3/2\ kT$ where k is the Boltzmann constant, having a value of 1.38×10^{-23} joules per Kelvin, and T is the Kelvin temperature. Interestingly, the average kinetic energy is independent of the mass of the molecule. Thus, nitrogen molecules and oxygen molecules in the air we breathe have the same average kinetic energy. On the other hand, their average speeds differ because their masses differ. When considering the fundamental meaning of temperature, it is appropriate to think of temperature as measuring the average kinetic energy of a molecule in a gas.

Flywheels of reasonable size and speed can store energy comparable to that of batteries and have promise for storing energy for electric vehicles. When a car is brought to rest by braking, the kinetic energy of the car is converted into heat that dissipates into the environment. Conceivably, the car could be brought to rest by transferring the linear kinetic energy of the car to rotational energy in a flywheel, with very little lost as heat during the conversion. The rotational energy then could be recovered by allowing the flywheel to operate an electric generator.

Joseph Priest

See also: Conservation of Energy; Flywheels; Gravitational Energy; Nuclear Energy; Potential Energy.

BIBLIOGRAPHY

Hobson, A. (1995). *Physics: Concepts and Connections.* Englewood Cliffs, NJ: Prentice-Hall.

Priest, J. (2000). *Energy: Principles, Problems, Alternatives*, 5th ed. Dubuque, IA: Kendall/Hunt Publishing Company.

Serway, R. A. (1998). *Principles of Physics*, 2nd ed. Fort Worth: Saunders College Publishing.

KINETIC ENERGY, HISTORICAL EVOLUTION OF THE USE OF

Historically humans have used three natural sources of kinetic energy: wind, water, and tides.

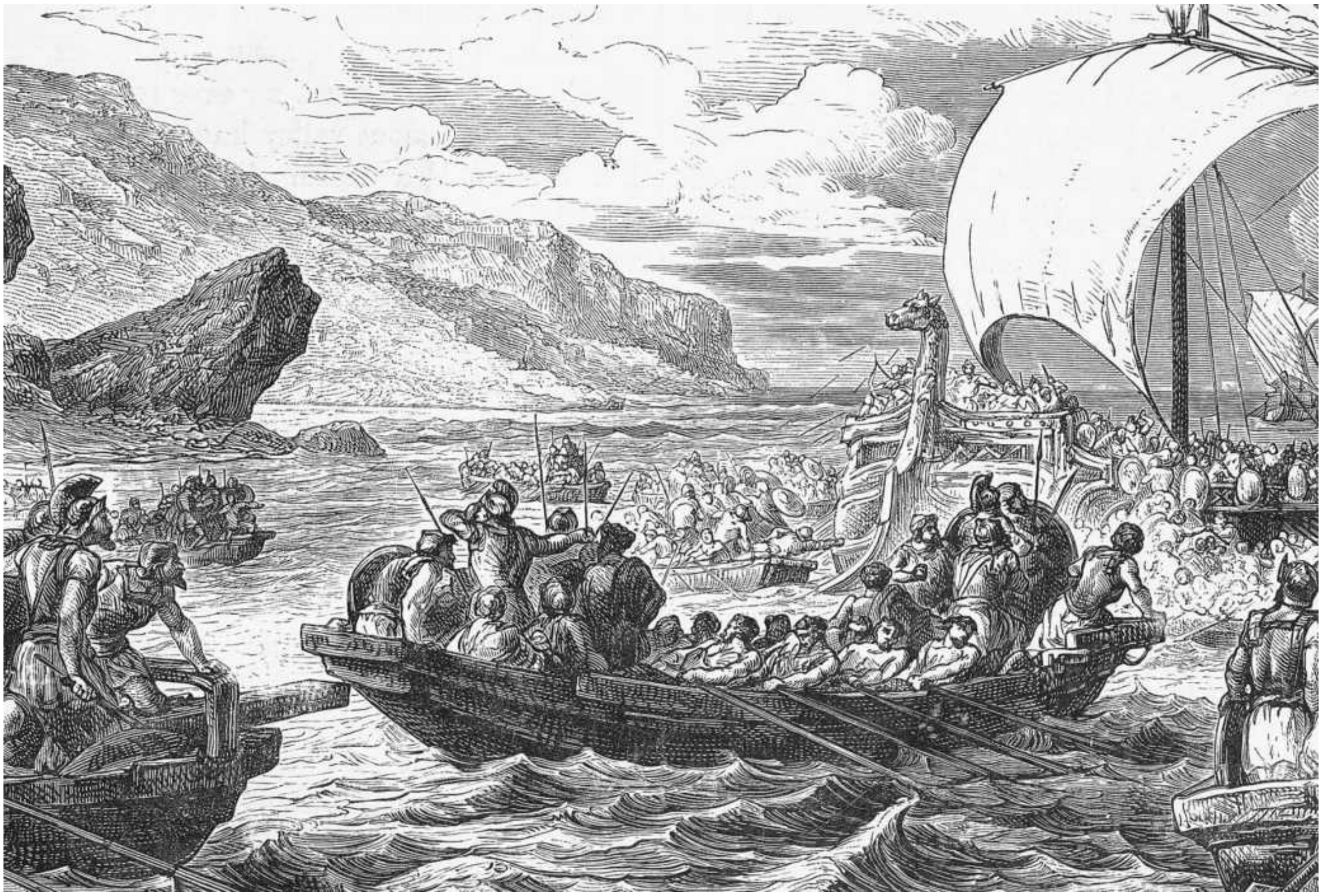
THE ANCIENT WORLD (TO 500 C.E.)

Although early humans often inadvertently tapped into the kinetic energy of moving air or water to do things such as separate grain from chaff or float downstream, the deliberate use of kinetic energy to power machinery came only in the historical era.

Only one machine in classical antiquity made deliberate use of the kinetic energy of wind: the sailing vessel. As early as 3000 B.C.E., paintings illustrated Egyptian vessels using sails. By the first millenium B.C.E., the use of sails was common for long-distance, water-borne trade. However, ancient sails worked poorly. The standard sail was square, mounted on a mast at right angles to the ship's long axis. It was effective only if the wind was dead astern. It was barely adequate if the wind was abeam, and totally inadequate in head winds. As a result, ancient mariners usually timed sailings to correspond with favorable wind direction and often averaged only 1 to 1.5 knots.

Very late in the ancient period the square sail was challenged by a more effective design: the triangular lateen sail, aligned with the vessel's long axis (i.e., fore-and-aft). The origin of this rigging is uncertain; unlike square sails, lateen sails operated as fabric aerofoils and permitted vessels to sail more closely into headwinds.

The only evidence for the use of wind in antiquity to power other machinery occurs in the *Pneumatica*



The sails on these Phoenician ships exhibit an early use of naturally occuring kinetic energy. (Corbis Corporation)

of Hero of Alexandria dating from the first century C.E. Hero described a toy-like device with four small sails, or blades, attached perpendicularly to one end of a horizontal axle. Wind struck the blades and turned the device. Pegs mounted on the opposite end of the axle provided reciprocating motion to a small air pump feeding an organ. There is no evidence that the idea was expanded to a larger scale, and some experts are suspicious of the authenticity of this portion of the *Pneumatica*.

Water, the other source of kinetic energy used in antiquity, saw wider application to machinery than did wind. The first evidence of the use of waterpower comes from the first century B.C.E., simultaneously in both China and the Mediterranean region. In China the preferred method of tapping the power of falling water was the horizontal water wheel, named after the plane of rotation of the wheel. Around the Mediterranean, the preferred form was the vertical water wheel. The vertical wheel came in two major forms: undershot and overshot. An undershot wheel had flat blades. Water struck the blades beneath the wheel and turned it by impact. The overshot wheel emerged later—the first evidence dates from the third century C.E. The overshot wheel's periphery consisted of containers, called buckets. Water, led over the top of the wheel by a trough, was deposited in the buckets; weight rather than impact turned the wheel.

The diffusion of waterpower was initially slow—perhaps due to its relatively high capital costs, its geographical inflexibility, and the abundance of manual labor in both the classical Mediterranean world and in China. Only in the declining days of the Roman Empire, for example, did watermills become the standard means of grinding grain in some areas, displacing animal- and human-powered mills.

THE MEDIEVAL WORLD (500–1500 C.E.)

Knowledge of how to tap the energy of wind and water was passed by both the Romans and Han Chinese to their successors. In the conservative eastern realm of the old Roman Empire (the Byzantine Empire), little was done to develop either wind power or water power, although Byzantine vessels did make increased use of lateen sails after the eighth century. On the other hand, the western part of the old Roman Empire, perhaps due to labor shortages, saw a very significant increase in the use of both wind and water power, especially between 900 and 1300.

The expansion of medieval Europe's use of natural sources of kinetic energy was most significant in waterpower. Although many Roman watermills were destroyed during the collapse of centralized authority in the fifth through seventh centuries, recovery and expansion beyond Roman levels was relatively rapid. By the eleventh century, sparsely populated England had over 5,600 watermills, a level of dependence on nonhuman energy then unparalleled in human history. Nor was England unusual. Similar concentrations could be found elsewhere. The Paris basin, for example, had around 1,000 watermills in the same era, and canal networks around medieval European cities were often designed to benefit mills.

In coastal estuaries, European craftsman by the twelfth century or earlier had also begun to use water wheels to make use of the movement of the tides. The usual arrangement required impounding incoming tides with a barrier to create a reservoir. During low tide, this water was released through a gate against the blades of a watermill.

With few exceptions, waterpower had been used in classical antiquity for only one purpose—grinding grain. By the tenth, and definitely by the eleventh century, European technicians had begun significantly to expand the applications of waterpower. By 1500 waterpower was used to grind not only wheat, but mustard seed, gunpowder, and flint for use as a glaze. It was used to crush ore, olives, and rags (to make paper). It was used to bore pipes, saw wood, draw wire, full (shrink and thicken) wool, pump water, lift rocks, shape metal, and pump bellows.

Nor was Europe the only civilization to make increased use of waterpower in this era. Although not blessed with the same abundance of stable, easy-to-tap streams, the Islamic world also increased its use of waterpower. Medieval travelers mention numerous mills and water-lifting wheels (norias) along rivers near Islamic cities such as Baghdad, Damascus, Antioch, and Nishapur. China, too, saw the expanded use of waterpower, for powering metallurgical bellows, grinding grain, spinning hemp, driving fans, crushing minerals, and winnowing rice.

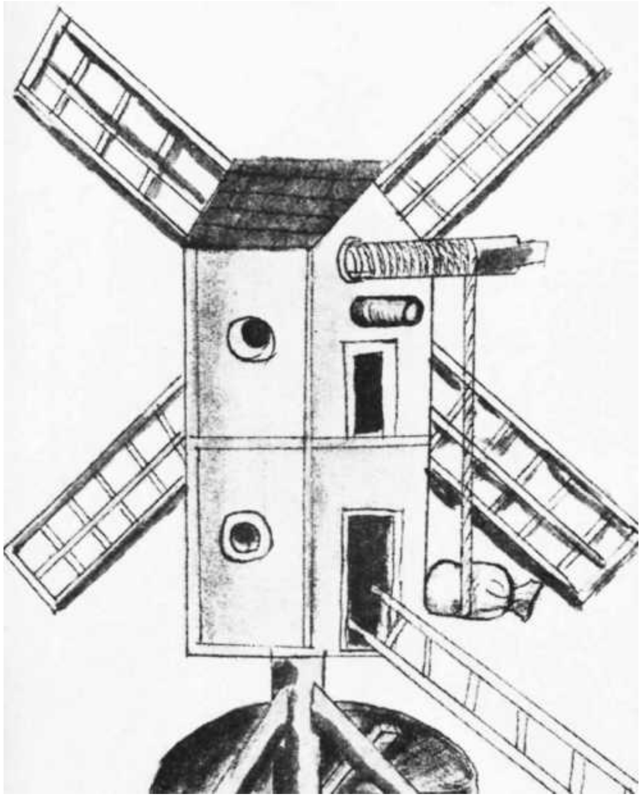
The wood water wheels used in the medieval period were, by modern standards, inefficient. Medieval undershot and horizontal wheels probably had an efficiency of about 15 percent to 25 percent, medieval overshot wheels about 50 percent to 60 percent. Commonly, their power output was only about 2–5 hp. But relative to the alternatives available at the time—human or animal power—they offered a very substantial gain in power.

Water provided a more reliable source of energy than the winds, but labor shortages in medieval Europe also encouraged the further development of wind power. The sailing vessel continued to be the most important wind-powered machine, and it saw significant improvement. Between 1000 and 1500, Chinese and European shipbuilders began to use multiple masts, stern post rudders, and deeper keels to tap larger volumes of air more effectively, and Europeans by 1500 were making effective use of combined square and lateen riggings.

The first solid evidence of the use of air to provide mechanical power to something other than sailing vessels comes from tenth-century Islamic travel accounts that describe wind-powered mills for grinding grain in eastern Persia. These were horizontal windmills, so named because the plane of rotation of their sails, or blades, was horizontal. The rotors were two-story devices, surrounded by walls with openings facing the prevalent wind direction. Wind passed through these orifices, struck the exposed blades, and exited through other orifices in the rear. In heavy winds they might produce as much as 15 hp, but more typically their output was about 2 to 4 hp.

Diffusion was very slow. The earliest clear Chinese reference to a windmill occurs only in 1219. The Chinese adopted the horizontal rotor used in Central Asia, but equipped it with pivoted blades, like venetian blinds. These windmills did not require shielding walls like Central Asian mills, and could tap wind from any quarter.

European technicians departed much more radically from Central Asian designs. Drawing from close experience with the vertical watermill, by 1100–1150



An early wind mill (c. 1430), with an automatic elevator for lifting flour bags. The post was designed to turn in accordance with the direction of the wind. (Corbis Corporation)

they had developed a vertical windmill called a post mill. The blades, gearing, and millstones/machinery of the post mill were all placed on or in a structure that pivoted on a large post so operators could keep the mill's blades pointed into the wind. European windmills typically had four blades, or sails, consisting of cloth-covered wooden frames. They were mounted on a horizontal axle and set at a small angle with respect to their vertical plane of rotation.

The European windmill diffused rapidly, especially along the Baltic and North Sea coasts. By the fourteenth century they had become a major source of power. Eventually, England had as many as 10,000 windmills, with comparable numbers in Holland, France, Germany, and Finland. In some areas of Holland one could find several hundred windmills in a few square miles.

Around 1300, Europeans improved on the post mill by devising the tower mill. In the tower mill, only a cap (containing the rotor axle and a brake wheel), mounted atop a large, stationary tower, had to be turned into the wind. The tower design allowed the construction of larger windmills containing multiple pairs of millstones, living quarters for the miller and his family, and sometimes machinery for sawing wood or crushing materials.

By the end of the medieval period, growing European reliance on wind and waterpower had created the world's first society with a substantial dependence on inanimate power sources.

THE EARLY MODERN ERA (1500–1880)

Dependence on wind or waterpower does not seem to have grown substantially in other civilizations after the medieval era, but this was not the case in Europe. By 1700, most feasible sites along streams convenient to some mercantile centers were occupied by water-powered mills, making it difficult for newer industries, such as cotton spinning, to find good locations. Moreover, European colonists carried waterpower technology with them to North and South America. By 1840 the United States, for example, had nearly 40,000 water-powered mills.

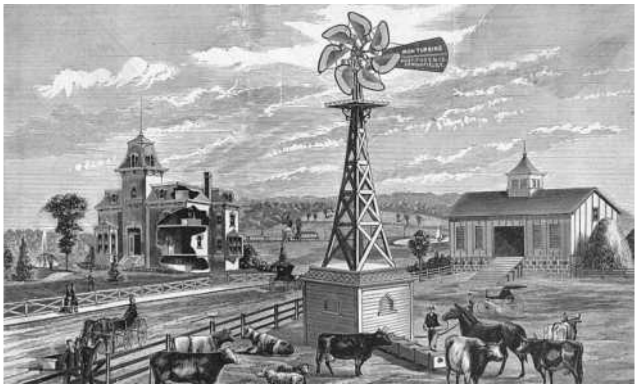
Pressure on available, easily tapped streams pushed the development of waterpower in several different directions simultaneously between 1700 and 1850. One direction involved the application of quantitative techniques to analyze wheel performance. In 1704 Antoine Parent carried out the first sophisticated theoretical analysis of water wheels. Erroneously believing that all wheels operated like undershot wheels, he calculated the maximum possible efficiency of water wheels at an astoundingly low 4/27 (15%). When John Smeaton in the 1750s systematically tested model undershot and overshot wheels, he discovered Parent's error. Smeaton found the optimum efficiency of an undershot vertical wheel was 50 percent (his model wheels actually achieved around 33 percent), and that of a weight-driven overshot wheel was much higher, approaching 100 percent (his model wheels achieved about 67–70%). Smeaton thereafter made it his practice to install overshot wheels where possible. Where impossible, he installed an intermediate type of vertical wheel called the breast wheel. With breast wheels, water was led onto the wheel at or near axle level. A close-fitting masonry or wooden casing held the water on the wheel's blades or buckets so it acted by weight (as in an overshot wheel) rather than by impact.

The mechanization of various elements of textile production, especially carding, spinning, and weaving, after 1770 created important new applications for waterpower. As textile factories grew larger, engineers modified the traditional wooden water wheel to enable it to better tap the kinetic energy of falling water. In addition to following Smeaton's example and using weight-driven wheels as much as possible, they replaced wooden buckets, with thinner sheet-iron buckets and used wrought iron tie-rods and cast iron axles to replace the massive wooden timbers used to support traditional water wheels. By 1850 iron industrial water wheels, with efficiencies of between 60 percent and 80 percent, had an average output of perhaps 15–20 hp, three to five times higher than traditional wooden wheels. Moreover, iron industrial wheels developing over 100 hp were not uncommon, and a rare one even exceeded 200 hp. The iron-wood hybrid wheel erected in 1851 for the Burden Iron works near Troy, New York, was 62 feet (18.9 m) in diameter, by 22 feet (6.7 m) wide, and generated around 280 hp.

Despite these improvements, in the more industrialized parts of the world the steam engine began to displace the waterwheel as the leading industrial prime mover beginning around 1810–1830. Steam was not able, however, to completely displace waterpower, in part because waterpower technology continued to evolve. Increased use of mathematical tools provided one means of improvement. In the 1760s the French engineer Jean Charles Borda demonstrated, theoretically, that for a water wheel to tap all of the kinetic energy of flowing water, it was necessary for the water to enter the wheel without impact and leave it without velocity.

In the 1820s, another French engineer, Jean Victor Poncelet, working from Borda's theory, designed an undershot vertical wheel with curved blades. Water entered the wheel from below without impact by gently flowing up the curved blades. It then reversed itself, flowed back down the curved blades, and departed the wheel with no velocity relative to the wheel itself. Theoretically the wheel had an efficiency of 100 percent; practically, it developed 60 percent to 80 percent, far higher than a traditional undershot wheel.

In the late 1820s another French engineer, Benoit Fourneyron, applied the ideas of Borda and Poncelet to horizontal water wheels. Fourneyron led water into a stationary inner wheel equipped with fixed, curved guide vanes. These vanes directed the water against a mobile outer wheel, also equipped with curved blades. These curved vanes and blades ensured that the water entered the wheel with minimum impact and left with no velocity relative to the wheel itself. Moreover, because water was applied to the entire periphery of the outer wheel at once, instead of to only a portion of the blades, Fourneyron's wheel developed much more power for its size than a comparable vertical wheel could have. This new type of water wheel was called a water turbine. Its high velocity and the central importance of the motion of the water on the wheel (water pressure) to its operation distinguished it from the traditional water wheel.



The "Iron Turbine" windmill is a more modern example of windpower. (Corbis Corporation)

By 1837 Fourneyron had water turbines operating successfully on both small falls and large ones. At St. Blasien in Germany, a Fourneyron turbine fed by a pipe, or penstock, used a fall of 354 feet (107.9 m), far more than any conventional water wheel could hope to. It developed 60 hp with a wheel only 1.5 foot (0.46 m) in diameter that weighed less than 40 pounds (18.2 kg).

Engineers quickly recognized that turbines could be arranged in a variety of ways. For example, water could be fed to the wheel internally (as Fourneyron's machine did), externally, axially, or by a combination. Between 1830 and 1850 a host of European and American engineers experimented with almost every conceivable arrangement. The turbines that resulted—the most popular being the mixed-flow 'Francis' turbine—quickly demonstrated their superiority to traditional vertical wheels in most respects. Turbines were much smaller per unit of energy produced, and cheaper. They could operate submerged when traditional wheels could not. They turned much faster, and they did all of this while operating at high efficiencies (typically 75% to 85%). By 1860 most new water-powered wheels were turbines rather than vertical wheels.

The continued growth and concentration of industry in urban centers, however, most of which had very limited waterpower resources, meant that steam power continued to displace water power in importance, even if the development of the water turbine delayed the process.

European engineers and inventors also made significant improvements in windmills between 1500 and 1850, but windmills fared worse in competition with steam than water wheels and water turbines.

The sails of the windmills that emerged from Europe's medieval period were inclined to the plane of rotation at a uniform angle of about 20°. By the seventeenth century, millwrights in Holland had improved the efficiency of some mills by moving away from fixed angles and giving the windmill blades a twist from root to tip (i.e., varying the inclination continuously along the blade's length) and by putting a camber on the leading edge, features found on modern propeller blades. Another advance was the fantail. Developed by Edmund Lee in 1745, the fantail was a small vertical windmill set perpendicular to the plane of rotation of the main sail assembly and geared so that a change in wind direction would power the fantail and rotate the main sails back into the wind. In 1772 Andrew Meikle developed the spring sail, which used hinged wooden shutters mounted on the blades, operating like venetian blinds, to secure better speed control. Other inventors linked governors to spring sails to automatically control wind speeds.

As with water wheels, engineers also began to carry out systematic, quantified experiments on windmills to test designs. In 1759 John Smeaton published a set of well-designed experiments on a model windmill. Since he had no means of developing a steady flow of air, he mounted 21-inch (53-cm) long windmill blades on a 5.5-feet (1.54-m) long arm that was pulled in a circle in still air to simulate the flow of air against a windmill. His experiments refuted prevalent theory, which recommended that the blades of windmills be inclined at a 35° angle. They confirmed that the 18–20° angle generally used in practice was far better. Smeaton's work also confirmed the even greater effectiveness of giving the blades a twist, that is, inclining sails at a variable angle like a propeller. By 1800 a large, well-designed Dutch windmill might develop as much as 50 hp at its axle in exceptional winds (normally 10–12 hp), although gearing inefficiencies probably reduced the net output at the machinery during such winds to around 10–12 hp.

Despite the empirical improvements in windmill design and attempts to apply quantitative methods to its design, the energy that could be tapped from winds was too erratic and had too low a density to compete with newly developed thermal energy sources, notably the steam engine. Following the development of an effective rotary steam engine by James Watt, and the introduction of mobile, high-pressure steam engines (locomotives) in the early-nineteenth century, that brought the products of steam-powered factories cheaply into the countryside, windmills began a rapid decline.

Only in limited arenas did the windmill survive. It continued to produce flour for local markets in unindustrialized portions of the world. It also found a niche—in a much-modified form—as a small power producer in isolated agricultural settings such as the Great Plains of North America. The American farm windmill of the nineteenth century was a vertical windmill, like the traditional European windmill, but it had a much smaller output (0.2-1 hp in normal breezes). Consisting of a small annular rotor usually no more than 6 feet (2 m) in diameter, with a very large number of blades to produce good starting torque, American windmills were mounted atop a high tower and usually used to pump water. By the early twentieth century several million were in use in America, and the technology had been exported to other flat, arid regions around the world, from Australia to Argentina to India.

THE MODERN ERA (1880–2000)

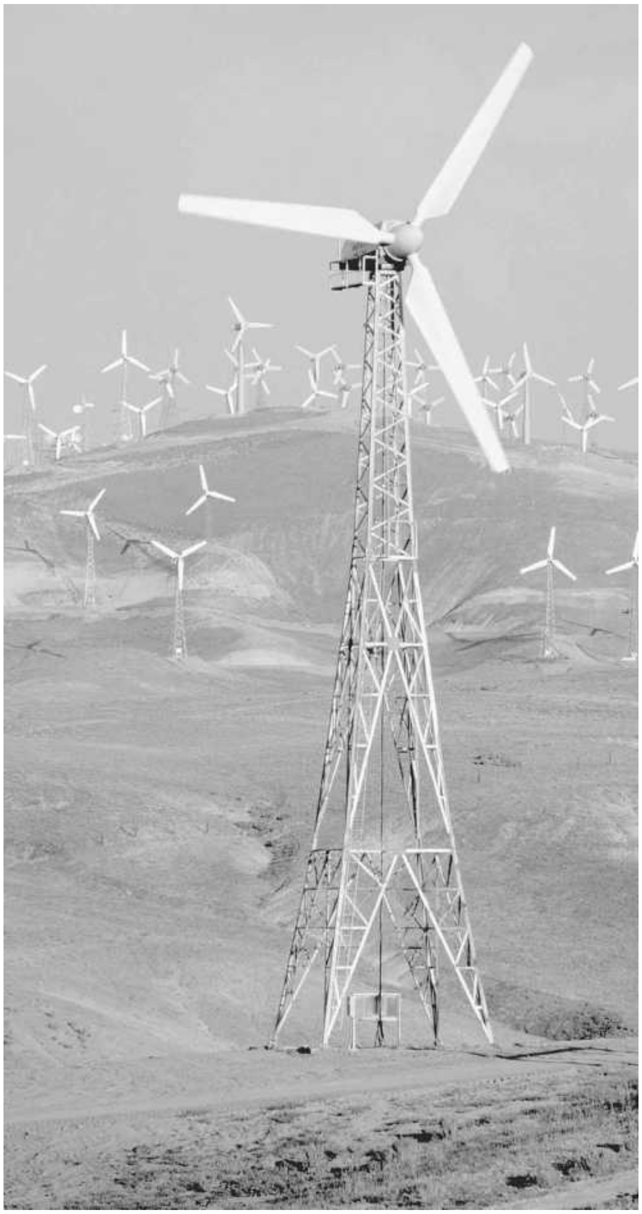
By the late nineteenth century, both wind and water were in decline as power producers almost everywhere. Steam-powered vessels had, in the late nineteenth century, begun to rapidly replace sailing vessels for hauling bulk merchandise. Only in isolated regions did sailing vessels continue to have a mercantile role. Increasingly, sailing vessels became objects of recreation, not commerce. In areas such as Holland, traditional windmills continued to pump away here and there, but they were on their way to becoming historical relics more than anything else. In the twentieth century even the American windmill went into decline as electric power lines belatedly reached rural areas and electric pumps replaced wind-powered pumps. Meanwhile, growing dependence on steam engines steadily reduced the importance of waterpower in all but a few favorable locations.

The emergence of electricity as a means of transmitting power late in the nineteenth century offered new life to devices that tapped the kinetic power of wind and water, since it offset one of the greatest advantages of steam power: locational flexibility. With electricity, power could be generated far from the point at which it was used.

Around 1890 Charles Brush erected a wind-powered wheel over 56 feet (17 m) in diameter to test the possibilities of wind as an economical generator of dc electricity. Coal- and waterpower-poor Denmark provided much of the early leadership in attempts to link wind machinery with the newly emerging electrical technology. The Danish engineer Poul LaCour, for example, combined the new engineering science of aerodynamics with wind tunnel experiments and new materials such as steel to try to develop a wind engine that would be a reliable and efficient generator of electricity. By 1910 he had several hundred small wind engines operating in Denmark generating dc electricity for charging batteries, but he had few imitators.

The development of the airplane stimulated intensive research on propeller design. In the 1920s and 1930s, a handful of engineers began to experiment with fast-moving, propeller-shaped windmill rotors (wind turbines), abandoning traditional rotor designs. For example, Soviet engineers in 1931 erected a 100 kW (about 130 hp) wind turbine on the Black Sea. In 1942 the American engineer Coslett Palmer Putnam erected the first megawatt-scale wind turbine designed to feed ac into a commercial electrical grid—a very large, two-blade rotor—at Grandpa's Knob in Vermont. It failed after 1,100 hours of operation, and cheap fuel prices after World War II limited further interest in wind-generated electricity.

The fuel crises and environmental concerns of the 1970s revived interest in the use of wind to produce electricity. Governmental subsidies, especially in the United States, supported both the construction of wind turbines and research, and led to the erection of large arrays of wind generators (wind farms) in a few favorable locations. By the 1990s, over 15,000 wind turbines were in operation in California in three mountain passes, generating a total of about 2.4 million hp (1,800 mW), but wind's contribution to total electrical energy supplies remains miniscule (<1%).



Modern wind turbine on a hill. (Corbis Corporation)

Waterpower fared better. When electrical engineers demonstrated the possibility of long-distance transmission of electrical power between 1875 and 1885, waterpower suddenly assumed new importance. Many of the best waterpower sites had been untapped because they were in remote regions far from manufacturing centers. As long as power had to be directly transmitted by mechanical shafts and belts, the locational flexibility of steam power made it the overwhelming first choice of industry. Electric power transmission made low-cost hydropower more competitive.

The first experiments with commercial water-generated electricity took place in the 1880s, but applications were limited since the predominant form of electricity generated, low voltage dc suffered large power losses in transmission. The introduction of high voltage ac around 1890, with its sharply reduced transmission losses, led to the first hydroelectric plants of significant scale. The spectacular success of the first hydroelectric plant at Niagara Falls in 1895, which used multiple turbine generator units, each developing 5,000 hp (3,750 kW) under a 136-foot head, and transmitting that power via high voltage ac some 22 miles to Buffalo, attracted enormous capital into the field. Between 1895 and 1930 large hydroelectric plants sprang up all across the United States and Europe. In some regions such as Norway and California, hydroelectricity became the dominant form of energy used.

The move toward gigantic hydroelectric developments accelerated in the 1930s, especially in the United States and the Soviet Union, as government-sponsored projects replaced privately funded projects. The Dnieprostroy hydroelectric plant in the Soviet Union, completed in 1932, was equipped with nine 85,000 hp turbine-generator units. The seventeen turbine-generators installed at Hoover Dam in the mid-1930s were even more powerful: rated at 115,000 hp (~86 MW each). More recent hydroelectric plants have grown even larger. For example, Itaipú (1984–1991) on the Paraná River in South America has eighteen turbine-generator units, each with a capacity of around 940,000 hp (700 MW), for a total capacity of around 17 million hp (12,600 MW).

Although the continued growth of energy consumption and the fixed amount of energy available from waterpower has prevented it from returning to its position as humanity's primary source of inanimate power, waterpower, unlike wind power, continues to be an important producer of energy. Today, waterpower is responsible for nearly 20 percent of the world's electrical energy. In some regions where hydrological and terrain conditions are appropriate, for example in Canada, Norway, and New Zealand it supplies a much higher proportion.

The technologies that enabled waterpower to remain important in the twentieth century—turbines, high voltage ac, and large dams—were also applied to harness the tides. In a few select estuaries where tidal variations were significant and construction conditions appropriate, large dams with embedded turbine-generators impounded high tides and used the outflow during low tides to generate electricity. The largest of these was placed in operation in 1966 on the Rance River on the coast of Brittany in France, where an average tidal range of around 28 feet was available. At Rance, a 0.4 mile long dam equipped with twenty-four turbines, generates around 320 MW of electric power. The high capital cost of tidal plants, the very limited number of sites with sufficient tidal variation and suitable construction conditions, and environmental concerns sharply limit the use of this form of kinetic energy. Thus, tides do not produce a significant portion of the world's energy (much less than 1%).

Terry S. Reynolds

See also: Hydroelectric Energy; Smeaton, John; Turbines, Wind; Watt, James.

BIBLIOGRAPHY

Baker, T. L. (1985). *A Field Guide to American Windmills.* Norman, OK: University of Oklahoma Press.

Hay, D. (1991). *Hydroelectric Development in the United States, 1880–1940.* Washington, DC: Edison Electric Institute.

Heymann, M. (1998). "Signs of Hubris: The Shaping of Wind Technology Styles in Germany, Denmark, and the United States, 1940–1990." *Technology and Culture* 39: 641–70.41

Hills, R. L. (1994). *Power from Wind: A History of Windmill Technology.* Cambridge: Cambridge University Press.

Hunter, L. C. (1979). *A History of Industrial Power in the United States, Vol. 1: Waterpower in the Century of the Steam Engine.* Charlottesville, VA: University Press of Virginia.

Hunter, L. C., and Bryant, L. (1991). *A History of Industrial Power in the United States, Vol. 3: The Transmission of Power.* Cambridge, MA: The MIT Press.

Reynolds, T. S. (1983). *Stronger Than a Hundred Men: A History of the Vertical Water Wheel.* Baltimore, MD: Johns Hopkins University Press.

Smil, V. (1994). *Energy in World History.* Boulder, CO: Westview Press.

Smith, N. (1976). *Man and Water: A History of Hydro-Technology.* London: P. Davies.

L

LAMPS

See: Lighting

LANDFILL GAS

See: Methane

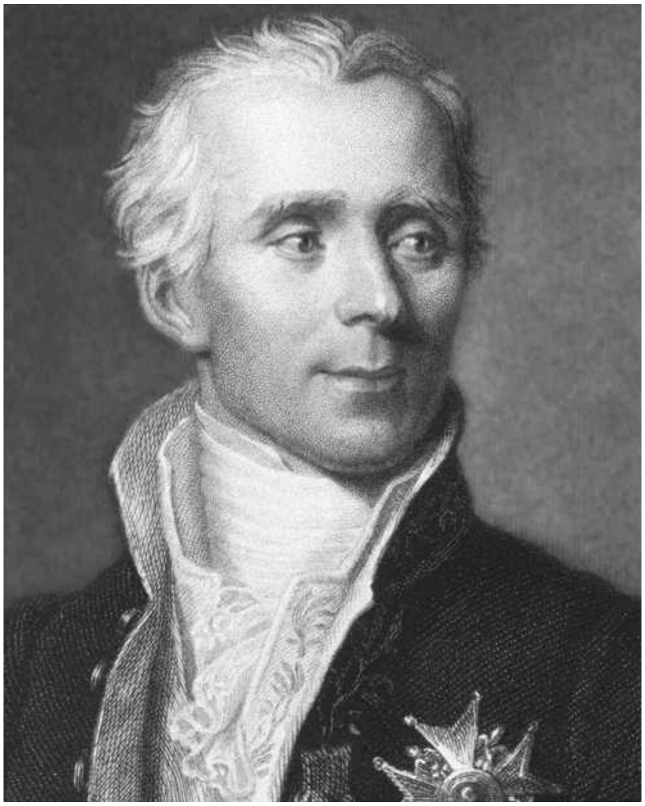
LAND RIGHTS

See: Property Rights

LAPLACE, PIERRE SIMON (1749–1827)

Pierre Simon Laplace, the most influential of the French mathematician-scientists of his time, made many important contributions to celestial mechanics, the theory of heat, the mathematical theory of probability, and other branches of pure and applied mathematics. He was born into a Normandy family of well-to-do peasant farmers and merchants. Intended for an ecclesiastical career, he matriculated at the University of Caen in theology but, discovering his aptitude for mathematics, departed for Paris at age nineteen bearing a letter from his instructors to the mathematician d'Alembert, a leading intellectual figure in prerevolutionary France.

D'Alembert, impressed by Laplace, took him on as his protégé and obtained an appointment for him as a professor of mathematics at the École Militaire, where he taught mathematics to teenage cadets. In this position, Laplace was also expected to present his work to the Académie Royal des Sciences, which was under royal patronage. He presented thirteen papers to the Academy in his first three years in Paris. In 1773, his fifth year in Paris, he was elected an adjunct member of the Academy. By the 1780s he was regarded as one of its leading figures and was appointed to important royal committees, largely based on his published papers on probability, astronomy, and geophysics. In 1785 he was promoted to the rank of senior pensioner in the Academy. However, Laplace was far from a popular figure. He rightly considered himself to be the leading mathematician in France and presumed upon it, pronouncing judgment on matters in other sciences. Even when he was correct, his abrasive manner created enemies. Nevertheless, he played a preeminent role in the Academy committee to reform the system of weights and measures. During this period (1782–1793) he also collaborated with Lavoisier in a series of investigations on the nature of heat and the phenomena of combustion and respiration. The collaboration terminated during the Reign of Terror, when the Academy was suppressed. Toward the end of 1793, probably in fear for his safety, Laplace left Paris until after the fall of Robespierre in the following year.



Pierre Simon Laplace. (Library of Congress)

The functions of the Académie Royal des Sciences were assumed in 1795 by a branch of the newly formed National Institute. Laplace was elected vice president of this reincarnated Academy and then elected president a few months later, in 1796. The duties of this position put him in contact with Napoleon Bonaparte. Three weeks after Napoleon seized power in 1799, Laplace presented him with copies of his work on celestial mechanics. Bonaparte quipped that he would read it "in the first six weeks I have free" and invited Laplace and his wife to dinner. Three weeks later, Napoleon named Laplace his minister of the interior. After six weeks, however, he was replaced; Napoleon thought him a complete failure as an administrator. However, Napoleon continued to heap honors and rewards upon him, regarding him as a decoration of the state. He made Laplace a chancellor of the Senate with a salary that made him wealthy, named him to the Legion of Honor, and raised him to the rank of count of the empire. Laplace's wife was appointed a lady-in-waiting to the Italian court of Napoleon's sister. Laplace responded with adulatory dedications of his works to Napoleon. When Napoleon fell from power, however, Laplace carefully dissociated himself from the emperor. In the Senate, Laplace voted for the return of the Bourbon monarchy and absented himself from Paris in 1815 during Napoleon's brief hundred-day return from Elba. In 1817 Louis XVIII raised Laplace to the rank of marquis. Laplace remained loyal to the Bourbons for the rest of his life, and his 1826 refusal to sign a petition supporting freedom of the press condemned him as far as the liberals in the Academy were concerned.

Laplace's lifelong work was the successful application of Newtonian gravitation to the entire solar system, accounting for all the observed deviations of the planets and satellites from their theoretical orbits. He began this work in 1773. His five-volume work on celestial mechanics (1798–1827) provided a complete mechanical description of the solar system. During the course of his investigations, Laplace discovered that the force on a body in a gravitational field could always be derived as the gradient of a potential function, a discovery that had profound implications in other branches of physics.

In mathematics, Laplace's name is most often associated with the "Laplace transform," a technique for solving differential equations. Laplace transforms are an often-used mathematical tool of engineers and scientists. In probability theory he invented many techniques for calculating the probabilities of events, and he applied them not only to the usual problems of games but also to problems of civic interest such as population statistics, mortality, and annuities, as well as testimony and verdicts.

In the first and second decades of the nineteenth century, Laplace exercised a powerful influence on physics in France, not only through his publications but also through the power to direct the research of others by virtue of his prestige and his position in the Academy. The genius of Laplace lay in his skill in surmounting mathematical difficulties. However, unlike other great scientists, he never attempted to go beyond the view of the world that existed when his career began. Laplace was an apostle of Newtonian mechanics. He believed, for example, that phenomena such as the behavior of light in material media could be understood on the basis of Newton's corpuscular theory of light and short-range molecular forces acting on the corpuscles of light; he directed research to strengthen this approach. As the weight of evidence in favor of the wave nature of light grew larger, Laplace's influence and power waned. In his last years he developed an elaborate caloric theory of heat, again based on a theory of short-range mechanical forces, but the theory was greeted with indifference and generated no further research.

Leonard S. Taylor

See also: Gravitational Energy; Heat and Heating.

BIBLIOGRAPHY

Arago, F. (1972). *Biographies of Distinguished Scientific Men*, tr. W H. Smyth, B. Powell, and R. Grant. Freeport, NY: Books for Libraries Press.

Gillespie, C. C. (1997). *Pierre-Simon Laplace*. Princeton NJ: Princeton University Press.

LASERS

It is said that necessity is the mother of invention. This adage says volumes about the early development of the laser. During World War II, U.S. military and civilian scientists searched frantically for improved radar. While these researchers met with only mixed success, their efforts spurred basic research. After the war, using knowledge gained from this line of inquiry, the first successful laser was developed in 1960.

Applications of the laser were found in numerous fields, including science, medicine, industry, and entertainment. By the 1980s, with the widespread use of lasers in industry and in commercial devices such as compact disk players and retail store price scanners, lasers were affecting the daily lives of nearly everyone in the developed world, whether they realized it or not. In a few decades, the laser has gone from a new cutting-edge technology to one that is so pervasive it is difficult to imagine many fields even existing without it.

WHAT IS A LASER?

The word "laser" is an acronym for "light amplification by the stimulated emission of radiation." Lasers of all kinds consist of several basic components: an active medium, an outside energy source, and an optical cavity with carefully designed mirrors on both ends. One of the mirrors is 100 percent reflective while the other is somewhat less reflective, so a beam can be emitted. The active medium is inside the optical cavity and is excited with an external energy source (typically electricity). For the cavity to emit laser radiation, the active medium has to achieve an unusual energy state called "population inversion."

All atoms consist of electrons orbiting around a nucleus. Normally these electrons prefer to remain in their lowest energy state, or orbit, which is known as ground level. When excited, however, electrons will jump to a higher orbit and then drop back down by radiating energy in the form of light at a discrete wavelength. This release of energy is called spontaneous emission. The energy or wavelength of the emitted light determines its color. Spontaneous emission occurs all by itself and is the source of virtually all light we see from natural sources (stars) and man-made ones (television sets, fluorescent lamps.)

Unlike these light sources, laser technology relies on a concept known as stimulated emission. When an excited atom is stimulated by a photon, light is emitted at precisely the same wavelength and precisely in phase with the light wave that stimulated it.

As energy is added and photons collide with other atoms, more and more electrons gain energy and jump orbits. The excited atoms all emit photons at the exact same wavelength. At some point there are more atoms with electrons in excited states than at ground level. This is population inversion.

As the emitted radiation bounces back and forth between the two mirrors, it becomes coherent. Some of the energy traveling back and forth through the optical cavity is transmitted though the less reflective mirror and becomes a laser beam.

The coherence of the beam makes a laser extremely powerful. The beam is light if a single frequency or color in which all the components are in step with one another. Normal radiation (such as the light from the sun) consists of many different wavelengths of light traveling in different directions. A laser's energy is concentrated, traveling in one direction, and traveling at the same wavelength, so its effect on other matter is extraordinary. A 100-watt lightbulb, for example, can barely light a living room, but a pulsed, 100-watt laser can be used to cut or drill holes in metal.

Another feature of many (though not all) lasers is a cooling system to dissipate energy wasted in the form of heat. Although some are much more efficient than others, all lasers waste significant amounts of energy. For example, a CO_2 laser (the most popular laser for industrial uses) operates at approximately 20 percent efficiency. To produce 100 watts of coherent light, these lasers require 500 watts of input power. A ruby laser is even less efficient, at about 0.1 percent; it would require 100,000 watts of electricity to produce that same 100 watts of laser light. Conversely, one company introduced a semiconductor laser in 1998 that operates at 56 percent, the most efficient up to that time.

DEVELOPMENT OF THE LASER

The first known laser was made by Theodore Maiman at Hughes Research Laboratories in Malibu, California, in 1960, but the seeds of this breakthrough were planted years before. In 1917 Albert Einstein, through his work on the quantum theory of light, theorized that stimulated emission of light radiation could occur. The idea was forgotten, though, until the middle of the century.

In the 1940s, researchers were working on better military radar. Radar is used to locate and track objects using pulses of long-wavelength microwave radiation created according to similar principles as laser radiation. Scientists knew that if they could achieve shorter wavelengths, more accurate images can be obtained using more compact equipment. The race was on to reach the infrared and visible portions of the spectrum. Although technical stumbling blocks prevented researchers from achieving this goal during the war, the pursuit continued during peacetime. The war had taught U.S. military and government officials the importance of technological superiority. Worldwide political tensions kept research grants flowing into America's scientific laboratories.

In 1948 Charles Townes was a professor of physics at Columbia University, working with microwaves. During the early 1950s he speculated that stimulated emission could generate microwaves, but he also knew that a population inversion was necessary. By combining the work done by researchers at Harvard University on this concept as well as his own radar-related work on amplifying a signal, Townes came up with the idea for a closed cavity in which radiation would bounce back and forth, exciting more and more molecules.

By 1954 Townes, with the help of graduate students Herbert Zeiger and James Gordon, developed the maser, an acronym for microwave amplification by stimulated emission of radiation. The maser had all the components of a laser—the resonator cavity and an active medium (Townes used ammonia)—but the wavelengths produced were much longer. In 1957 Townes developed the equation showing that much smaller wavelengths (in the infrared and visible-light range) were possible. Townes collaborated with Arthur Schawlow, a physicist at AT&T Bell Laboratories, and the two worked together to achieve a laser.

Theodore H. Maiman with a ruby that was used in early laser studies in the 1960s. (Corbis Corporation)

Others had similar ideas. R. Gordon Gould, a graduate student at Columbia, had come to the same conclusions. In November 1957 he wrote up his notes on a possible laser. In 1958 Townes and Schawlow patented a similar idea. Theodore Maiman learned of this research in September 1959 at a conference on quantum electronics that Townes had organized. Afterward he began his own research on a laser, using a pink ruby as the active medium. Ironically, Schawlow had rejected ruby because he felt it would require too much energy. By pumping the ruby with the light from a photographer's flash lamp, however, Maiman created the world's first laser, in June 1960.

The floodgates were now open. Lasers using calcium fluoride, helium-neon, glass, and cesium vapor as active media followed within two years. Others substances followed throughout the early 1960s. By 1962 about 400 companies had some kind of laser research. U.S. government spending on laser development soared, from $1.5 million in 1960 to $20 million in 1963. Scientific journals became inundated with papers, and meeting rooms at conferences overflowed with inquisitive attendees.

Lasers were still tools of elite scientists, though. Moving the technology from the lab and into the marketplace took years. The first devices produced in significant quantities were military systems in the early 1970s, followed by a small but growing number of industrial and medical lasers. The laser's time had finally come.

LASERS AND THEIR APPLICATIONS

Different lasers use different materials as the active medium. The medium can be either solid, liquid, or gas, and there are advantages for each in the amount of energy that can be stored, ease of handling and storage, secondary safety hazards, cooling properties, and physical characteristics of the laser output.

Solid-State Lasers

The term "solid-state laser" refers to lasers that use solids as their active medium. However, two kinds of materials are required: a "host" crystal and an impurity "dopant." The dopant is selected for its ability to form a population inversion. The Nd:YAG laser, for example, uses a small number of neodymium ions as a dopant in the solid YAG (yttrium-aluminum-garnet) crystal. Solid-state lasers are pumped with an outside source such as a flash lamp, arc lamp, or another laser. This energy is then absorbed by the dopant, raising the atoms to an excited state. Solid-state lasers are sought after because the active medium is relatively easy to handle and store. Also, because the wavelength they produce is within the transmission range of glass, they can be used with fiber optics.

Diode (Semiconductor) Lasers

Also using solids but considered separately because of their unique characteristics, diode lasers are the most common lasers in use. Compact size and reliability are the chief benefits of this kind of laser. The two common families of diode lasers contain active mediums composed of GaAlAs (gallium-aluminum-arsenite) or InGaAsP (indium/phosphorus). These media emit radiation in the infrared range. Much like those working with radar in the 1940s and '50s, researchers in the 1980s and '90s found ways to shorten wavelengths of lasers produced by diodes to reach into the range of blue visible light.

Liquid (Dye) Lasers

The common liquid lasers utilize a flowing dye as the active medium and are pumped by a flash lamp or another laser. These are typically more complex systems requiring more maintenance. They can be operated as either CW (continuous wave) or pulsed. One advantage liquid lasers have is they can be tuned for different wavelengths over a 100-nm range.

Gas Lasers

Gas lasers are not unlike fluorescent light bulbs and neon signs. Gas is confined to a hollow tube, and electricity passing through it excites the atoms. The most common gas lasers use carbon dioxide, argon, and helium-neon. Gas lasers are relatively inexpensive and can produce very high-powered beams.

Excimer Lasers

Excimer lasers use gases, but because of their special properties are usually considered as a class of their own. Excimer is short for "excited dimer," which consists of two elements, such as argon and fluorine, that can be chemically combined in an excited state only. These lasers typically emit radiation with very small wavelengths, in the ultraviolet region of the electromagnetic spectrum. This shorter wavelength is an enormous advantage for many applications.

There are literally thousands of uses for lasers. One of the largest applications is telecommunications—sending a signal through fiber optic cables, for example. This application grew rapidly in the 1990s with the phenomenal increase in traffic on the Internet. Optical data storage, such as on compact disks, CD-ROMs, and DVDs, is another important use for lasers. The information age was obviously a boon to this application, and as researchers obtained smaller wavelengths with diode lasers, they were able to fit more information on smaller storage devices.

Another group of applications is collectively known as materials processing. This includes the processes used in manufacturing. Production facilities use lasers to cut, weld, drill, mark, and heat-treat numerous materials such as metals, plastics, wood, ceramics, and even diamonds. Lasers are much more precise than other mechanical means used to process materials, and lasers make it possible to build devices with tiny, even microscopic, dimensions. A subgroup of this category, medical device manufacturing, relies on lasers to machine stents and other devices for implantation into the human body.

Lasers are used for medical procedures as well. The laser has been called a bloodless scalpel because it can be used to simultaneously cut and cauterize tissue. Photodynamic therapy is a cancer treatment that relies on lasers, and laser eye surgery can correct vision, eliminating the need for glasses and contact lenses. Plastic surgeons use lasers for hair removal, skin resurfacing, tattoo removal, and many other applications. New medical imaging and diagnostic methods have been devised thanks to the laser.

Military uses for lasers are abundant, from rangefinding to guided munitions to laser aiming devices on firearms. Warfare has been revolutionized by the laser. Law enforcement uses lasers to lift hard-to-recover fingerprints and in laser radar speed guns.

Laser printers, bar code readers, unmanned freeway tollbooths, laser pointers—none of these very common devices would be possible without laser technology. This is just a minor sampling; the list of laser applications goes on and on.

The laser has revolutionized many aspects of science and other disciplines, as well as the daily lives of millions of people. When it was first invented, the laser was referred to by some as a solution looking for a problem because it came about mostly from basic research rather than the active solution to a particular concern. At the time, no one could have predicted the far-reaching effects it would have in the second half of the twentieth century, or that it would come to be considered by many as one of the most influential technological achievements of that time.

Karl J. Hejlik

See also: Industry and Business, Productivity and Energy Efficiency in; Lighting; Matter and Energy; Townes, Charles Hard.

BIBLIOGRAPHY

Bromberg, J. L. (1991). *The Laser in America, 1950–1970.* Cambridge, MA: MIT Press.

Bromberg, J. L. (1992). "Amazing Light." *Invention and Technology*. Spring:18–26.

Charschan, S.(1993). *Guide to Laser Materials Processing.* Orlando, FL: Laser Institute of America.

Hecht, J. (1992). *Laser Pioneers*. Boston: Academic Press.

Muncheryan, H. (1983). *Principles and Practices of Laser Technology*. Blue Ridge Summit, PA: TAB Books.

National Academy of Sciences. (1987). *Lasers: Invention to Application*. Washington, DC: National Academy Press.

National Academy of Sciences. (1998). *Harnessing Light: Optical Science and Engineering for the 21st Century*. Washington, DC: National Academy Press.

Svelto, O., and Hanna, D. (1998). *Principles of Lasers*, 4th ed. New York: Plenum Press.

LAUNCH VEHICLES

See: Spacecraft Energy Systems

LAWRENCE BERKELEY LABORATORY

See: National Energy Laboratories

LAWRENCE LIVERMORE LABORATORY

See: National Energy Laboratories

LEVER

See: Mechanical Transmission of Energy

LEWIS, WARREN K. (1882–1975)

INTRODUCTION

As a central figure in twentieth century petrochemical technology, Warren K. Lewis is widely viewed as the father of American chemical engineering. His work opened up an entirely new and powerful engineering discipline applicable across a broad range of manufacturing industries, including chemical synthesis, steel production, and power generation.

Lewis had an enormous influence on the energy industries. Along with the Russian, Vladimir Ipatieff, Lewis was pivotal in advancing petroleum refining technology. As an instructor and mentor, Lewis left his mark on three generations of chemical engineering students, many of whom entered and influenced the U.S. petroleum and petrochemical industry. As a consultant and technological innovator, Lewis directly applied his principles of chemical engineering, an essentially American development, to revamping and extending the production capability of U.S. catalytic cracking facilities. Lewis' greatest single engineering achievement was his central role during World War II in the design of the fluid catalytic cracking process, an innovation of strategic and economic importance.

THE EARLY YEARS

Warren K. Lewis was born in 1882 on a small farm in southern Delaware. Farm life exerted a strong early influence on the future engineer. Believing that after college he would return to manage the family farm, his academic goal was to learn the fundamentals of agricultural science in order to apply them in a rational manner to improve the farm's productivity. To Lewis, this meant employing a thorough understanding of the mechanical operations of advanced (i.e., efficient) equipment and tools. In 1901 Lewis entered MIT as an undergraduate with a major in mechanical engineering.

THE MIT YEARS

Soon after entering MIT, Lewis came under the influence of Professor William H. Walker, who was instituting a new and innovative program in chemical engineering. In his sophomore year, Lewis transferred to the chemical engineering program, which at that time was part of the chemistry department.

After graduation, and at the urging of Walker, Lewis won a two-year fellowship for study in Germany. Studying under the great physical chemist Abegg at the University of Breslau, Lewis obtained his doctorate in 1908 while developing a thorough grounding in the theories of and mathematical structures associated with phenomena underlying numerous industrial engineering processes.

Spending only a brief period in industry, Walker urged Lewis back to MIT in 1910 to become an assistant professor of chemical engineering. Both Walker and Lewis sought to form closer ties between the chemical engineering program and industry. To this end, and with the cooperation of the great industrial chemist A. D. Little, they established in 1916, MIT's School of Chemical Engineering Practice. The Practice School, a successor to Walker's Research Laboratory of Applied Chemistry (established at MIT in 1908), centered around practical chemical engineering instruction within operating plants in and around the New England area.

In addition to being (by all accounts) a brilliant and riveting teacher, Lewis was an innovative administrator in the cause of furthering chemical engineering at MIT. In 1920, largely due to the efforts of Walker and Lewis, MIT's chemical engineering program became a separate department, with the Practice School becoming part of the newly-formed department. In that year, Lewis became the first head of the chemical engineering department. During his nine years as department head, Lewis fought successfully for a strong, independent chemical engineering program. In their path breaking text, The Principles of Chemical Engineering (1923), Lewis and Walker for the first time systematized the engineering study of unit operations and provided a model of instruction for burgeoning chemical engineering departments. Eventually, Lewis secured for the department approval by MIT to develop a curriculum leading to a B.S. in chemical engineering, one of the first such programs in the country.

ENGINEERING, CONSULTING AND PETROLEUM PROCESSING

Lewis' influence on the U.S. energy sector evolved from his career as a consultant to the petroleum refining industry, and in particular Standard Oil of New Jersey (currently Exxon). Lewis' success as a consultant rested in large part on his ability to apply to practical problems his knowledge of and experience in industrial chemical engineering. Between 1919 and 1927, Lewis virtually restructured Jersey Standard's manufacturing operations. His work spanned both oil production and refining.

One of Lewis' first assignments was to make the distillation process more precise and continuous. By the early 1920s, Lewis introduced to Jersey Standard the use of vacuum stills. These were able to operate at lower temperatures that limited coking and fouling of equipment. Thus production engineers did not have to periodically clean out and repair equipment, which in turn facilitated the transformation of distillation from batch to continuous operations.

Lewis directed subterranean reservoir studies to improve the efficiency of extracting oil from the ground. He designed the first bubble tower to effect more precise and efficient fractionation operations and provided important assistance in developing one of the first continuous thermal cracking process.

In 1927, Lewis and Frank A. Howard, head of the Development Department at Jersey Standard, established a research center at Jersey's Baton Rouge plant to research and commercialize advanced petrochemical technology. This development marks the beginning of the Southwest as a center of petroleum-based technology. Lewis populated the facility with colleagues and graduates from MIT's Chemical Engineering Department and the Practice School. These included Robert Haslam, the head of the Practice School, who became director of Baton Rouge R&D and Eger Murphree, a protégé of Lewis, who was made Manager of Development and Research for Jersey Standard's Development Department.

In the years leading to World War II, Lewis and Haslam hired fifteen additional MIT-affiliated chemical engineers to work at the Baton Rouge plant. These men felt great loyalty to Lewis who they considered their mentor. They were instrumental in championing Lewis and his style of chemical engineering at Jersey Standard. Through the 1930s and during World War II, Lewis and Jersey's R&D people at Bayway and Baton rouge worked closely together to develop some of the most important innovations in the history of petrochemical technology. The most significant of these was fluid catalytic cracking.

WORLD WAR II, FLUID CRACKING, AND THE LATER YEARS

Fluid cracking, which began operations in 1942 and was responsible for providing the Allies with sufficient supplies of aviation fuel and synthetic rubber, was Lewis' last and greatest engineering achievement. Its successful development depended on Lewis' engineering genius, on the close relationship established between Jersey Standard and MIT by Lewis, and the ability of his protégés both within MIT and Jersey Standard to transform his core ideas and designs into commercial reality.

In addition to his work on fluid cracking, Lewis served as Vice Chairman of the Chemistry Division of the National Defense Research Committee during World War II. In this capacity, he directed chemical engineering research on problems associated with the development of the atomic bomb. Lewis officially retired from teaching in 1948, although he continued to meet with students at MIT well into his eighties. Lewis died at the age of 92 in 1975. The American Institute of Chemical Engineers, in honor of his achievements in the academic arena, established an annual award in his name for those who have made significant contributions to education in the field.

Sanford L. Moskowitz

BIBLIOGRAPHY

Enos, J. (1962). *Petroleum, Progress and Profits: A History of Process Innovation*. Cambridge, MA: MIT Press.

Frankenburg, W. G., et al., eds. (1954). *Advances in Catalysis and Related Subjects, Vol. 6*. New York: Academic Press.

Grace, J. R., and Matsen, J. M., eds. (1980). *Fluidization*. New York: Plenum Publishing.

Landau, R., and Rosenberg, N. (1990). "America's High-Tech Triumph." *Invention & Technology*. Fall:58–63.

Lewis, H. C. (1980). "W. K. Lewis, Teacher." In *History of Chemical Engineering*, edited by W. F. Furter. Washington, DC: American Chemical Society.

Larson, H. M., et al. (1971). *History of Standard Oil Company (New Jersey): New Horizons, 1927–1950*. New York: Harper and Row.

Mattill, J. (1991). *The Flagship: The MIT School of Chemical Engineering Practice, 1916–1991*. Cambridge, MA: David H. Koch School of Chemical Engineering Practice, MIT.

Moskowitz, S. L. (1999). "Science, Engineering and the American Technological Climate: The Extent and Limits of Technological Momentum in the Development of the U.S. Vapor-Phase Catalytic Reactor, 1916–1950." Diss. Columbia University.

Murphree, E. V., et al. (1945). "Improved Fluid Process for Catalytic Cracking." *Transactions of the American Institute of Chemical Engineers* 41:19–20.

Othmer, D. F. (1956). *Fluidization.* New York: Reinhold Publishing Corp.

Pigford, R. L. (1976). "Chemical Technology: The Past 100 Years." *Chemical and Engineering News* (April 6):190–195.

Popple, C. S. (1952). *Standard Oil Company (New Jersey) in World War II.* New York: Standard Oil Company (New Jersey).

"Reminiscences of William H. Walker." (1952). *Chemical Engineering* (July):158–178.

Russell, R. P. (1944). "The Genesis of a Giant." *Petroleum Refiner* 23:92–93.

Spitz, P. (1988). *Petrochemicals: The Rise of an Industry.* New York: John Wiley and Sons.

Squires, A. M. (1986). "The Story of Fluid Catalytic Cracking: The First Circulating Fluid Bed." *Proceedings of the First International Conference on Circulating Fluid Beds, November 18–20, 1985.* New York: Pergamon Press.

Weber, H. C. (1980). "The Improbable Achievement: Chemical Engineering at MIT." In *History of Chemical Engineering,* edited by W. F. Furter. Washington, DC: American Chemical Society.

Williams, G. C., and Vivian, J. E. (1980). "Pioneers in Chemical Engineering at MIT." In *History of Chemical Engineering,* edited by W. F. Furter. Washington, DC: American Chemical Society.

Williamson, H. F., et al. (1963). *The American Petroleum Industry, Vol. II: The Age of Energy, 1899–1959.* Evanston, IL: Northwestern University Press.

LIGHTING

Light is essential for human life. Modern societies have created homes, schools, and workplaces that rely on electric light sources. Some of the electricity used to generate the light in these spaces is wasted, largely owing to ignorance. The efficient application of electric, as well as natural, light sources to the human condition is a sophisticated effort, but one that is essential to a sustainable and enjoyable future.

WHAT IS LIGHT?

Humans are a diurnal species, which means that we are active in the day and asleep at night. Indeed, daylight is the primary stimulus to the photobiological system that regulates our sleep-awake cycle. Of course, while we are awake, we see, and we depend a great deal on seeing. Approximately 80 percent of the human brain devoted to sensing the environment is devoted to vision. It is not surprising, then, that from the beginning of human history we have strived to produce and control light.

Until very recent human history, the Sun had been our primary source of light for both seeing and waking. Over the past two millennia, and particularly over the past two centuries, our direct reliance on the Sun for light has diminished. Today it can be argued that large segments of affluent human societies are exposed to light from the Sun only rarely. Many people spend virtually all of their active lives under manufactured light sources of various types.

These manufactured light sources are, perhaps ironically, largely dependent on the Sun. The radiant energy from the Sun has been stored in the fossilized remains of billions of creatures over millions of years and is used to power the electric light sources created by modern humans. The power generated by hydroelectric sources also is a result of solar evaporation and subsequent rainfall. Only nuclear reactors provide power independent of the Sun, which is, of course, the largest nuclear reactor in the solar system.

Ironically, too, without humans there cannot be "light." In other words, we formally define light in human terms. Radiation from the Sun and from other sources varies in frequency. These different frequencies have been categorized for convenience into different bands. The highest frequencies are in a band known as cosmic rays, and the lowest are in the radio frequency band. Between these two extremes is a very small band of radiant energy known as visible radiation, or light. Only radiation in the narrow region of the electromagnetic spectrum visible to humans can be called light. All other radiation, *even that seen by other species* outside this band, cannot technically be referred to as light. We measure the frequency of radiation within the visible band in terms of wavelength. By convention, the visible band ranges from 380 to 780 nanometers; one nanometer is one billionth (10^{-9}) of a meter.

Photoreceptors in the eyes convert radiation in the visible band into neural signals that reach the brain. Photoreceptors are located throughout the retina, a sensory membrane that covers the entire back of the

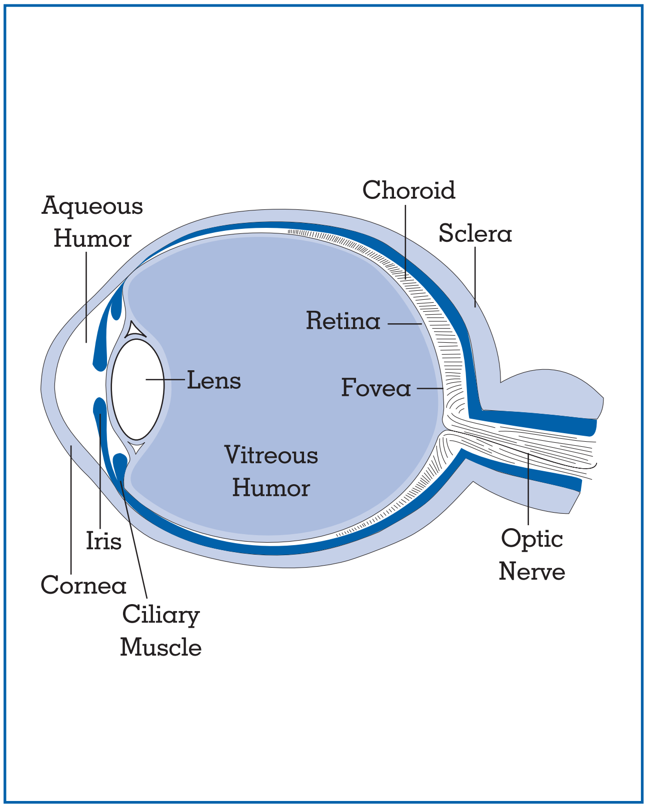


Figure 1.
Cross section of the human eye.

human eye, as shown in Figure 1. One eye contains approximately 130 million photoreceptors. We have, however, two distinct classes of photoreceptors, rods and cones. Rods are more sensitive to light and are used primarily for nighttime vision, whereas cones, of which there are three types, are used for daytime vision. These two classes of photoreceptors enable human vision to comfortably span light levels from a sunny, snow-capped mountain to faint starlight, a billion-to-one range. The three cone types provide us with the ability to convert light into color. Indeed, color is not an inherent property of light. Rather, the human brain "calculates" color from the neural signals generated by the three cone types.

MEASUREMENT OF LIGHT AND COLOR

Light is measured in several ways. Until about 1960, most light measurements were obtained by visual comparison. A standard light source of known, but adjustable, brightness was compared by a trained technician to another light source of unknown, but fixed, brightness. When the two lamps were seen to be equally bright, they were said to produce the same amount of light. Naturally, this technique was fraught with problems of inconsistency and inconvenience. Today photosensitive electronic detectors are used to measure light accurately and reliably. They convert radiant energy into a measurable electrical signal that can be used to quantify the amount of light generated by a source.

The measurement of light is known as photometry. (Color measurement is known as colorimetry; see below.) Photometry can be performed in several different ways, depending on the geometric relationship between a light source and a detector. Most light measurements are based on the flux (photon) density on a detector. This quantity is known as illuminance and represents the rate of photon absorption by a detector of known area. From an illuminance measurement other photometric quantities can be derived. Flux, measured in lumens, is simply the total amount of light generated by a source. Intensity, measured in candelas (cd), is the amount of light projected in a given direction and within a two-dimensional (solid) angle. Luminance, measured in candelas per square meter, is comparable to the human perception of brightness because it accounts for both the amount of light reaching a surface and the amount of light reflected from that surface back to the eye. For many years luminance was known as photometric brightness, but this term is no longer used.

Photometric measurements do not weight all wavelengths in the visible band equally. Rather, a specific weighting function for the electromagnetic spectrum is employed to define "light" (see Figure 2). This function, known as the photopic luminous efficiency function, is not based on the responses of all of the photoreceptors in the eyes. Rather, it is based on the spectral sensitivity of only two cone types found only in the small region of the retina known as the fovea. Responses by the photoreceptors in the fovea enable us to read and see fine detail, but they represent only about 4 percent of all the photoreceptors in the retina. The photopic luminous efficiency function was established by international agreement in 1924 and has been used as the standard weighting function for light ever since.

Light also can be defined in other ways, even though these definitions are rarely used. For exam-

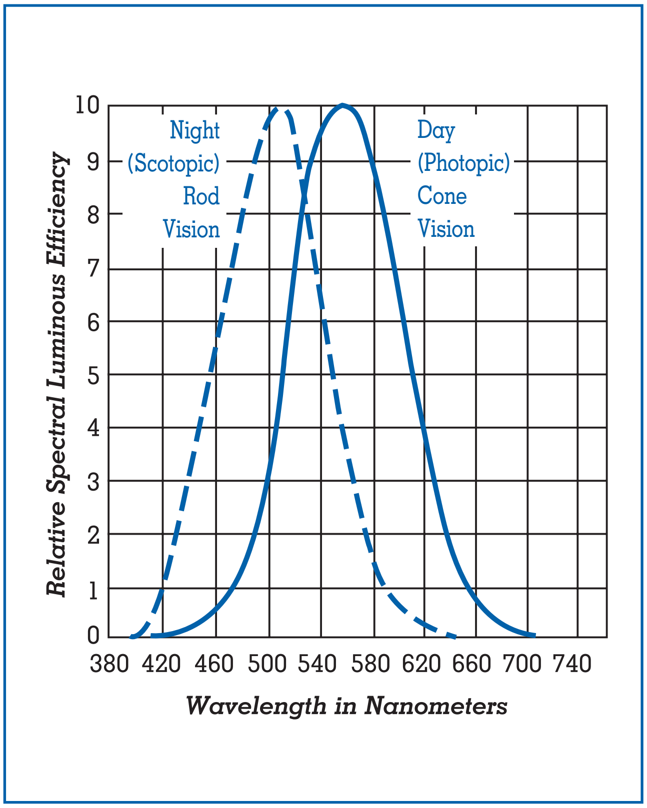


Figure 2.
Photopic (right) and scotopic (left) luminous efficiency functions.

ple, the scotopic luminous efficiency function was established in 1951 by international agreement to represent the spectral sensitivity of rods (see Figure 2). This function is rarely if ever used in photometry because the presence of almost any light source, including moonlight, will raise light levels to a point where some cones also function for vision.

Light sources vary in their ability to produce light—that is, to produce radiation within the photopic luminous efficiency function. In general the efficiency of extracting energy from fossil fuels or other energy sources and converting it into light is very low. For example, the efficiency of light extraction by an open gas flame is only about 0.04 percent. Not until very recently, with the advent of electric light sources, was it possible to substantially increase extraction efficiency. The conversion efficiency of the most efficacious electric light source presently manufactured (low-pressure sodium) is, however, only about 30 percent.

All electric light source efficacies are measured in terms of their ability to generate light per unit electric power, measured in lumens per watt (lm/W). However, efficacy is only one consideration in selecting a light source. For example, the most efficacious light source, low-pressure sodium, essentially produces a single wavelength of light (589 nanometers), making every object color appear as different shades of yellow. For some human activities the absence of color information is not important, but for others it is not only unpleasant but also dangerous (e.g., in surgery). It is beyond the scope of this article to compare light source cost, maintenance, safety, flexibility, and operating conditions, but all of these factors, not just efficacy, are extremely important in selecting the right light source for a particular application.

However, two points about energy efficiency need to be stressed. First, power and energy are two different quantities, and second, high light source efficacy (lm/W) does not always indicate good energy efficiency.

Power is the rate at which a source generates energy, measured in ergs per second. Obviously, then, because energy is the product of power and time, energy conservation techniques can either reduce the power required to generate light or the total time that the power is being supplied to the light source. Many electric light sources can be used with associated electrical dimming circuits to reduce light output by reducing the power supplied to that source. Photosensors, for example, can be used to dim electric lighting levels when daylight enters a room through a window or a skylight. Time clocks, occupant sensors, and even manually operated switches are effective techniques for turning lights off when a room or a building is unoccupied. Both strategies reduce the energy used for lighting; dimming matches the light source intensity with the needs of the people, and switching extinguishes light when no one needs it. Dimming and switching should be used to reduce only wasted lighting energy. It should be reemphasized that light is necessary for human activity. Our societal goal should not be simply to reduce energy consumption but rather to reduce energy wasted by too much light and by lighting that meets no human need.

Light sources are only rarely used without a fixture to house the light source. A fixture makes handling and operating the light source safer and helps provide

Common Lamp Types	*Typical CCT*	*Typical CRI*
Incandescent		
household A-lamp	2800	100
tungsten-halogen	3190	100
photoflood	3400	100
Fluorescent		
warm-white	3000	51
cool-white	4150	64
delux cool-white	4160	89
lite-white	4200	40+
daylight	6380	76
RE70 (rare earth phosphor, 70+ CRI)	2700, 3500, 4100, or 5000	70+
RE80 (rare earth phosphor, 80+ CRI)	2700, 3500, 4100, or 5000	80+
C75	7500	90+
full-spectrum	5500	90+
Low pressure		
sodium	1740	-44
HID		
metal halide	4220	67
clear mercury	6410	18
high pressure sodium	2100	24
xenon	5920	94

Table 1.
Typical values of correlated color temperature (CCT) and color rendering index (CRI) for some common electric light sources.

light where it is needed by controlling the direction of light emitted by the source. As already noted, electric lamp efficacy is measured in lumens per watt. However, the total flux generated by a source is measured without regard to direction. Because light should be directed toward an object or an area to achieve a purpose, flux emitted from a fixture is not necessarily the most useful photometric quantity to assess efficacy in a given application. An automobile headlamp, for example, should direct as much of the light generated by the lamp as possible onto the road in front of the automobile. The light directed into the night sky cannot be used, and thereby this lost light reduces the fixture (headlamp) efficacy. Therefore, efficacy measured in terms of lumens per watt does not provide a true measure of fixture efficacy because all the flux from the fixture are not always useful for a given application. The photometric quantity, intensity, is the amount of light flowing within an imaginary cone connecting the illuminated object and the light source. Intensity, measured in candelas, is a measure of how much light generated by the light source actually arrives where it is needed. The amount of light arriving at a specific location is critical for headlamp design, for example. In general, intensity per watt is a valuable measure of efficacy for any lighting system because, ultimately, light must arrive at a specified location to be useful. Other light is wasted, and this wasted light should be reflected in a measure of low fixture efficacy. Different fixtures are produced by manufacturers for different applications using a variety of light sources. The most efficacious light fixture for a given application should be used, but the most efficacious light fixture (cd_{max}/W)

does not always employ the most efficacious light source (lm/W).

The term "color" can be used in two ways to describe the light generated by a source or reflected from an object. As already noted, color appearance is a perceptual phenomenon that, despite much scientific investigation, defies precise quantification. For example, the same physical light may look brown or orange depending on its brightness. Further, this same light may look more or less red, yellow, or brown depending on the apparent color adjacent to it. At present we have only a qualitative understanding of these color appearance phenomena. Color can, however, be quantified very precisely in terms of color matching. This system of quantifying color is known as colorimetry. Any light can be matched in appearance with the right combination of idealized, but quantifiable, red, green, and blue lights. These three lights are known as primary colors or, simply, primaries. Color televisions are practical examples of color matching. They can produce a wide gamut of colors through adjustments of three color pixels: red, green, and blue. Thus, although we can predict precisely what colors will match in terms of these three primaries, we have no precise way to predict whether the pixels will be seen as, for example, brown or orange. Through manual adjustments of the color television pixels, however, each person can reach an acceptable appearance of a televised image.

The color of light generated by a source can be precisely described by colorimetry. From a colorimetric description of the light, other color measures can be derived. Two derived measures—correlated color temperature (CCT) and color rendering index (CRI)—are commonly used to characterize light generated by manufactured sources. It is very important to emphasize that both CCT and CRI are derived from the science of color matching but are used, incorrectly, to describe different aspects of color appearance. Despite the technical error, both CCT and CRI have been found to be practically useful by the lighting industry for color appearance.

CCT refers to the appearance of the light generated by a very hot (i.e., incandescent) object, the temperature of which is measured in kelvins (K). As a body is heated, it begins to produce a reddish-yellow, and then a yellow-white light. As temperature increases, the apparent color of the light changes to blue-white. In astronomy, for example, older, cooler stars appear yellow or red, whereas younger, hotter stars look blue. Paradoxically, electric light sources with CCTs between 2,700 K and 3,200 K generate a yellowish-white light and are termed "warm"; those with CCTs between 4,000 K and 7,500 K produce a bluish-white light and are termed "cool." This paradox seems to have originated from the association between yellow light and a hot fire. The association between apparent color and tactile temperature seems to have been reinforced by the cold feel of glass admitting blue light from the sky on a clear day.

CRI is a measure of how "true" or "natural" colors will appear when illuminated by a light source. Light sources that generate light evenly throughout the visible spectrum, such as daylight, have high values of CRI (maximum CRI = 100); those that have gaps in the visible spectrum (e.g., clear mercury) have low CRI. Electric light sources, particularly fluorescent lamps, have undergone a great deal of development in recent years and now have much higher CRI values than were available in 1970. Although real improvements have been made to the color-rendering properties of these lamps, to some extent these lamp developments are only a game of numbers. Given the technical flaws inherent in CRI for describing color appearance, this measure should not be expected to precisely characterize the color-rendering properties of these lamps. For example, a ten-point difference in CRI values is probably unimportant.

TYPES OF LIGHTING

Much of the history of electric light sources (see Figure 3), fixtures (see Figure 4), and control technologies (dimming and switching) centers around improving energy efficiency. It is often assumed that incandescent lamps were the first electric light source. Actually, carbon arc lamps were the first practical electric light source, preceding the incandescent lamp by almost half a century. Carbon arc lamps employ two carbon electrodes separated by a gap. When current is supplied to the electrodes, the lamp produces a very bright, blue-white arc. These lamps literally burn the electrodes while the lamp is operating, so a clock device is required to continuously feed carbon into the arc to keep the gap width constant.

Carbon arc lamps were first developed in the 1840s, and sometimes elaborate towers were created to provide illumination to streets in a few European

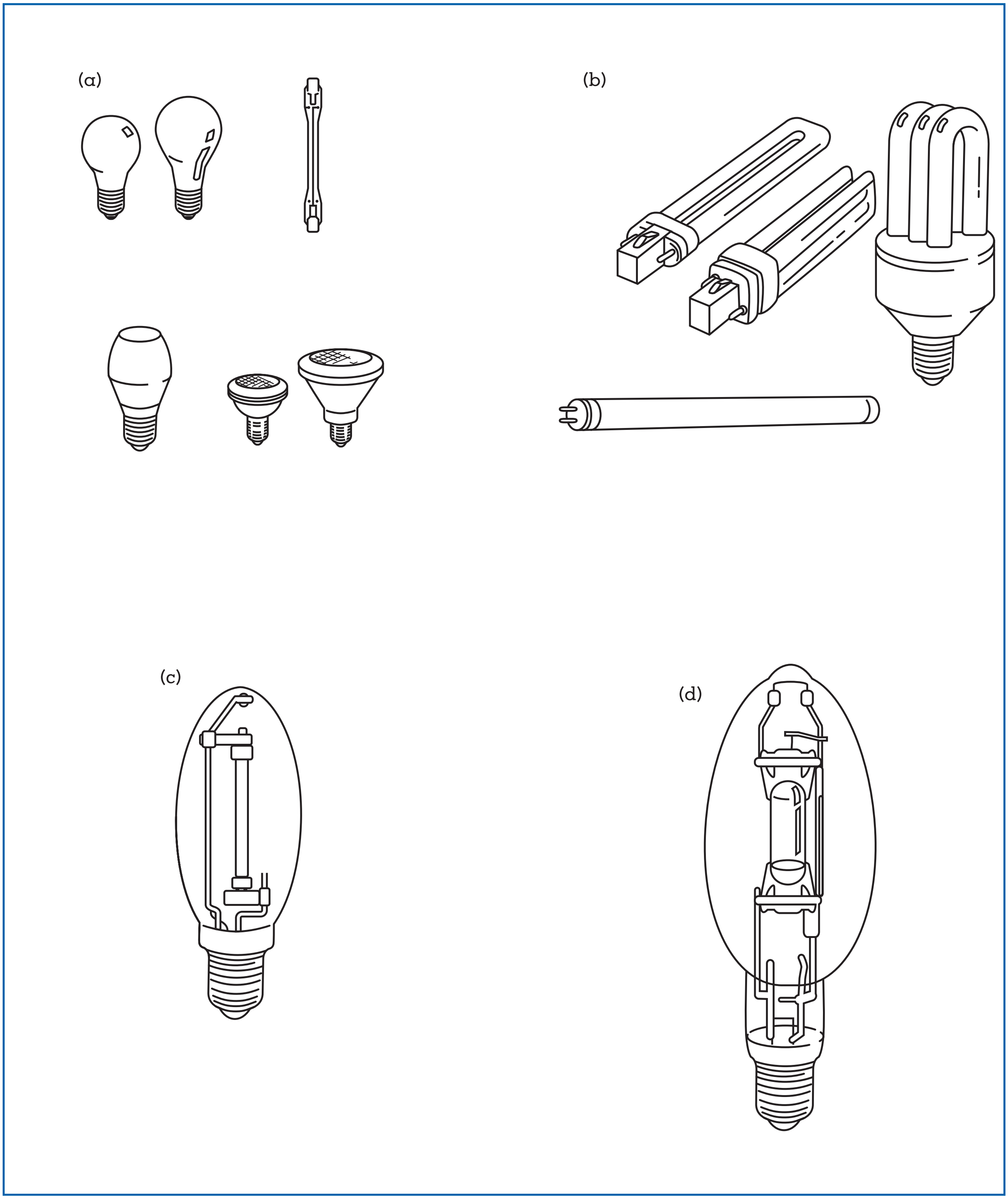


Figure 3.
Lamp types: (a) incandescent and tungsten-halogen lamp shapes; (b) fluorescent and compact fluorescent lamp shapes; (c) typical high-pressure sodium lamp; (d) typical metal halide lamp.

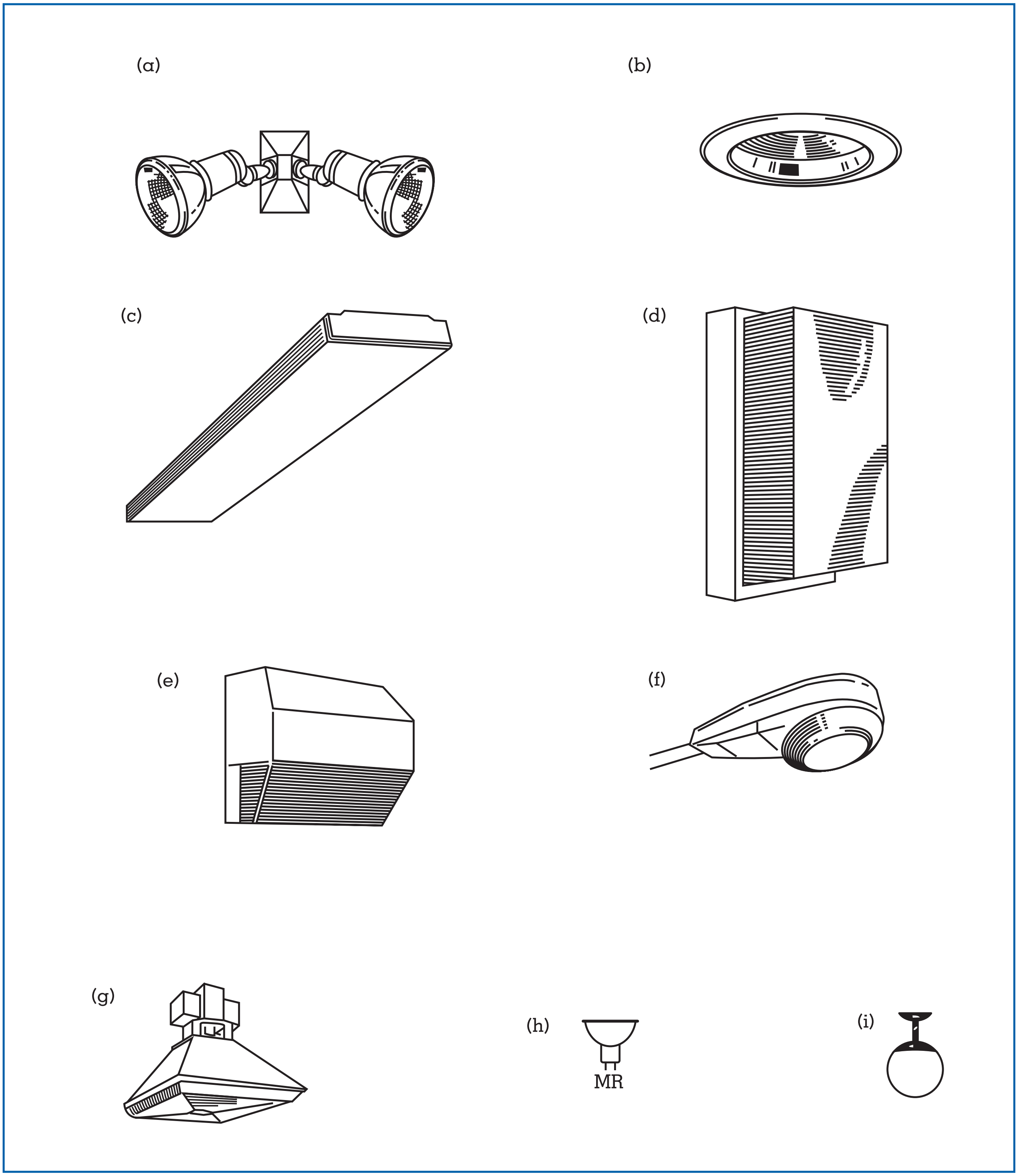


Figure 4.
Luminaire types: (a) parabolic aluminized reflector (PAR) lamp holder; (b) incandescent, compact fluorescent or high-intensity discharge lamp downlight; (c) lensed fluorescent lamp luminaire; (d) wall-mounted compact fluorescent lamp luminaire; (e) wall-mounted high-intensity discharge lamp luminaire; (f) roadway luminaire; (g) low bay industrial luminaire; (h) multi-faceted reflector (MR) lamp; (i) pendant sphere.

cities. Carbon arc lamps are still used today for searchlights because they produce the very bright, concentrated point of light needed for a high-intensity beam. These lamps are impractical for other purposes, however, because it is difficult to segregate and distribute this concentrated light into many small packages useful for human activities indoors. Also, cost and the pollution generated by these lamps make them unacceptable by modern standards.

Incandescence means to heat an object to the point of producing light. The incandescent lamp, then, is a lamp with a filament heated to the point of glowing. The trick accomplished by Thomas Edison and his team at Menlo Park, New Jersey, in 1879 was to produce an inexpensive lamp with a carbon filament that would glow for several hundred hours. After more than a year of experimenting, Edison was able to demonstrate an incandescent lamp producing approximately 2 lumens per watt and lasting several hundred hours. Incremental improvements in that basic design, especially the use of the metal tungsten in filaments, have increased the efficacy of commercially available incandescent lamps to between 10 and 15 lumens per watt. More significantly, perhaps, these lamps are easy to install and cost as little as fifty cents each.

It has been argued that Edison's greatest vision was not the invention of the incandescent lamp but rather his insight two years earlier into the possibilities of providing small "packages" of light where people needed them. A high-voltage power supply with a high resistance light source was required to meet his vision; the incandescent lamp was, therefore, a logical outcome of that insight. The incandescent lamp and the associated electrical distribution system changed human history forever. An inexpensive, manageable light source not only illuminated the night but also made it possible to build the large windowless buildings so prevalent today.

One of the major innovations in incandescent lamps has been the tungsten-halogen lamp, developed in 1959. As the tungsten filament of an incandescent lamp glows, some of it evaporates and is deposited on the inside surface of the bulb. This evaporation not only shortens lamp life, but also the deposited tungsten reduces the light output from the lamp. Iodine, bromine, and chlorine are within the family of elements called halogens. These molecules will combine with the evaporating tungsten and, rather than be deposited on the bulb wall, will deposit the combined molecules back onto the filament. At this point the halogen molecule disassociates from the tungsten molecule, and the cycle begins again. The halogen cycle prolongs incandescent lamp life and keeps the lamp burning brightly. Special bulb shapes and materials are needed for tungsten-halogen lamps because very high temperatures are required for their operation. The high temperatures increase efficacy so that most tungsten-halogen lamps available to the market produce approximately 15 to 30 lumens per watt.

Probably every reader has heard of the mythical incandescent lightbulb that has burned for more than fifty years. Because we are all susceptible to believing conspiracy theories, we are suspicious that manufacturers could make all lightbulbs last that long, if only they would. Most inexpensive incandescent lamps available in stores are rated for 750 hours of operation. In reality, an incandescent lamp can be made to last a very long time, certainly longer than 750 hours, but there is no "free lunch." Incandescent lamp life can be prolonged considerably if the lamp is operated at low temperatures. Lamps operated at low temperatures, however, produce relatively more heat than light, so the efficacy drops below the rated 10 to 12 lumens per watt. Lamp life could also be improved if the lamp were more expensive. Incandescent lamps used in traffic signals are rated at about 8,000 hours of operation but are approximately five to ten times as expensive as the conventional household incandescent lamp, because of their more durable filament construction. An incandescent lamp can have long life, high efficacy, or low cost, but not all three at the same time. So although the mythical lamp may exist, and indeed it can be made, physics demands that it must either produce light expensively or inefficiently, and probably both.

Fluorescent lamps were the next major innovation in lamp technology. Introduced commercially in 1938, they were a radical improvement in lighting energy efficiency. At nearly 50 lumens per watt, the first fluorescent lighting systems immediately replaced those using incandescent lamps in large industrial and commercial applications. Even today, fluorescent lamp technologies are arguably the most cost-effective, energy-efficient, and reliable source of illumination for interior applications.

Fluorescent lamps are termed a low-pressure discharge lamp. An electric current passes through mercury contained at normal vapor pressure within the bulb. The current vaporizes the mercury and liberates electrons from the molecules. The liberation of electrons from mercury molecules produces radiation, much of which cannot be seen by humans. Phosphors that coat the inside of the bulb absorb the nonvisible radiation. The irradiated phosphor molecules are themselves excited to liberate electrons that *do* emit radiation within the visible band of the electromagnetic spectrum. This multistage process sounds inefficient, but fluorescent lamps can be made to operate at nearly 100 lumens per watt, with lamp life greater than 20,000 hours. New lamp designs as well as improvements to the electrical device needed to start and operate the fluorescent lamp, known as a ballast, have improved system efficacy by more than 40 percent since 1980. Not only has efficacy been improved, but also the color characteristics are better, and the audible noise generated by the ballast has been reduced. Moreover, smaller, compact fluorescent lamps have been introduced and are beginning to replace some of the less efficacious incandescent lamps used most commonly in homes.

Many consumers continue to complain about fluorescent lamps, however, arguing that they "buzz," distort the color of natural objects, and cause headaches. These attitudes are barriers to societal goals for energy efficiency, particularly in the home. As stated above, however, most of the technical issues leading to these complaints have been resolved. Nevertheless, these negative attitudes toward fluorescent lamps persist and are, in fact, reinforced by exposure to the older technologies still in operation and by low-cost, inefficient products being introduced by manufacturers from developing countries. A major barrier to widespread introduction of fluorescent lamps is initial price. Fluorescent lighting systems are much more expensive to purchase than incandescent systems. In the long run, however, the energy savings associated with the fluorescent lighting system would more than pay for its higher initial cost.

The flip side of consumer bias is that some people argue that only fluorescent lamps should be used in buildings. This attitude, while well intentioned, is based on an unsophisticated knowledge of the performance of lighting systems. As previously discussed, the efficacy of a lighting system cannot be characterized by lamp efficacy alone. For example, although a fluorescent lamp and ballast may produce 80 lumens per watt compared to a tungsten-halogen lamp at 20 lumens per watt, the system efficacy of a recessed open downlight ceiling fixture may, in fact, be better with a tungsten-halogen source than with a fluorescent source. This perhaps surprising result is due to the fact that the tungsten-halogen filament is very compact. The light from the small tungsten-halogen filament can be optically controlled much easier than it can from the relatively large fluorescent lamp. Where little or no optical control is necessary, as with general illumination from a standing floor-lamp fixture, fluorescent lighting systems are much more efficacious than tungsten-halogen lighting systems in producing the same visual effect. Again, the application is important for selecting the most efficacious lighting system.

Low-pressure sodium lamps, another low-pressure discharge lamp, produce very bright, monochromatic light at relatively high wattage. These lamps are used exclusively for outdoor applications where a high-light-output lamp can distribute light over a relatively large area. Low pressure sodium lamps have the highest efficacy (180 lumens per watt), but as discussed above, provide people with no color perception. Despite its high efficacy, this monochromatic source has made little inroads into the outdoor lighting market, except in the United Kingdom, which consumes approximately half of all low-pressure sodium lamps manufactured.

High pressure can be used to expand the spectral emission of the sodium gas. Although expanding the sodium spectrum reduces the lamp efficacy to between 60 and 120 lumens per watt, it significantly improves the color characteristics of the lamp. High-pressure sodium lamps were introduced in 1965, and the yellow-orange light produced by these lamps is now found throughout modern societies in roadway, security, and parking lot applications. Newer, color-corrected high-pressure sodium lamps can be found in some indoor applications but have still lower efficacies.

High pressure can also be used to expand the spectral emission of mercury gas. High-pressure, clear mercury lamps are no longer widely used because, even under high pressure, they are relatively inefficient and do not provide good color perception. By mixing halides with the mercury under high pressure, however, good color can be produced at relatively high efficacies ranging from 60 to 110 lumens per watt. These metal halide lamps have been gaining steadily in popularity since they were introduced in 1964, particularly as an outdoor light source associated with retail spaces (e.g., shopping centers, gas stations, and facade lighting). These lamps also are being used in many indoor applications, such as warehouses and shopping malls. They are even being used in modern automobile headlamps as a replacement for tungsten-halogen headlamps.

Perhaps the most radical new development in light sources has come from the electronics industry rather than the traditional lighting industry, established in the nineteenth century. Light-emitting diodes (LEDs) were invented in the 1960s, but only since 1997 have these sources become an important light source for widespread applications. Before then, LEDs were used as indicator lights on electronic equipment and were largely restricted to a few colors, the first being the red LED. Today many colors, including white, can be produced economically with efficacies ranging from 5 to 25 lumens per watt. It has been projected that efficacies of some colored LEDs may approach 100 lumens per watt in the near future.

LEDs are a highly directional light source, ideally suited to applications such as automobile taillights and traffic signals. In architectural applications, LEDs have already proven successful in exit sign applications. The high visibility necessary for transportation applications is also a positive attribute in emergency egress fixtures. In addition, the low power requirements of LEDs in exit signs (fewer than 5 or 6 watts per face) compared to other technologies provide substantial energy savings when multiplied by hours of operation typical of exit signs (usually twenty-four hours per day every day). The long life of LEDs also provides the reliability needed for transportation and egress applications.

In the future, LED fixtures will be produced for architectural applications by clustering the individual LEDs. Coupled with electronic controls, the LED fixtures will enable the color and intensity distribution of light to be customized and easily changed. These systems will provide cost-effective and energy-efficient lighting solutions to future lighting artists, designers, and engineers. Indeed, this technology may change architecture in ways not seen since the time of Edison.

The efficient application of lighting technologies to a residential, commercial, or industrial space is much more than simply picking the lamp with the highest lumens per watt. As described above, it is important to have light where you need it and when you need it. Color, cost, ease of maintenance, heat, and durability—as well as how people will use, operate, and maintain the space—are among the other factors to consider in selecting the right lighting equipment for an application.

FUTURE TRENDS

The world population continues to expand. Fuel reserves continue to be depleted. Pollution associated with power generation continues to increase. Technical advances in light sources, fixtures, and controls as well as those in power generation provide a modestly optimistic picture of the future. In 1850 the extraction efficiency of light from carbon-based fuels such as open gas flames was about 0.04 percent. Today we have increased that efficiency more than a hundredfold. Estimates from the U.S. Department of Energy in 1999 suggest that efficiency improvements in lighting technologies will reduce the required energy for lighting in commercial buildings, although the amount of commercial floor space will continue to grow in the United States. Even if this projection is correct for the United States, the need for light by people around the world will outpace the increased extraction efficiency offered by new lighting technologies in the decade 2000 to 2010.

Our societal ambition to reduce energy consumption is, without any doubt, both correct and urgent. It must always be remembered, however, that humans always will need light. Our goal, then, should be to reduce wasted lighting energy, energy expended on lighting that meets no real purpose. The technological advances in light source efficacy must also be coupled with technological advances in controls—both optical control to deliver light where it is needed, and power control to deliver the right amount of light when it is needed. Lighting control systems are slowly becoming more sophisticated, utilizing both automatic and manual controls, to tune lights to occupant needs. Estimates made in the late 1990s suggest that lighting energy used in offices can be reduced by as much as 80 percent using controls relative to static lighting systems (Maniccia 1999). We must also strive to better understand exactly what humans need. Recommended illumination levels in North America, for example, have been reduced by roughly two-thirds since the oil embargo of 1972, with no noticeable loss in human productivity or satisfaction.

Arguably, a reason why lighting energy is still being wasted is ignorance of policymakers, building owners, and developers. Well-intentioned policymakers legislate power but not energy, failing to consider the importance of lighting controls in meeting our societal goals. They also regulate lamp efficacy rather than lighting system efficacy, implicitly failing to recognize that light should be directed to a location where it can be used. Another reason lighting energy is being wasted is the emphasis on the purchase price of lighting equipment. Often the cheapest lighting products are the most energy-wasteful. A lighting system will be operated for many years, and the cost of energy, even at current low prices, far exceeds the initial cost of even the most expensive lighting equipment.

Mark S. Rea

See also: Conservation of Energy; Consumption, Culture and Energy Usage; Economically Efficient Energy; Edison, Thomas Alva; Elctricity; Electricity, History of; Electric Power, Generation of; Power.

BIBLIOGRAPHY

Bierman, A. (1999). "LEDs: From Indicators to Illumination." *Lighting Futures* 3(4):1–5.

Bowers, B. (1998). *Lengthening the Day*. New York: Oxford University Press.

Bright, A. A. (1949). *The Electric-Lamp Industry*. New York: Macmillan.

Cox, J. A. (1979). *A Century of Light*. New York: Benjamin.

Drucker, H. (1997). "1972–1997: Twenty-five Years of Energy and Environmental History: Lessons Learned." In *Twenty-Five Years of Energy and Environmental Policy: Proceedings of the 25th Annual Illinois Energy Conference*. Chicago: University of Illinois at Chicago, Energy Resources Center.

Howell, J. W., and Schroeder, H. (1927). *History of the Incandescent Lamp*. Schenectady, NY: Maqua.

Illuminating Engineering Society. (1947). *IES Lighting Handbook*, 1st ed. New York: Author.

Illuminating Engineering Society. (1952). *IES Lighting Handbook*, 2nd ed. New York: Author.

Illuminating Engineering Society. (1993). *IES Lighting Handbook*, 8th ed. New York: Author.

International Lighting Review. (1979). "1879–1979 [Edison Lamp Centenary Issue]" 30(1):1–13.

Leslie, R. P., and Conway, K. M. (1996). *The Lighting Pattern Book for Homes*, 2nd ed. New York: McGraw-Hill.

Leslie, R. P., and Rogers, P. (1996). *Outdoor Lighting Pattern Book*. New York: McGraw-Hill.

Lighting Research Center. (1994). *A&P Food Market, Old Lyme, Connecticut*, DELTA Portfolio 1(1). Troy, NY: Rensselaer Polytechnic Institute.

Lighting Research Center. (1994). *Linens 'n Things, Patchogue, New York*, DELTA Portfolio 1(2). Troy, NY: Rensselaer Polytechnic Institute.

Lighting Research Center. (1995). *450 South Salina Street, Syracuse, New York*, DELTA Portfolio. 1(3). Troy, NY: Rensselaer Polytechnic Institute.

Lighting Research Center. (1996). *DeGraff Street Industrial Center, Amsterdam, New York*, DELTA Portfolio 1(4). Troy, NY: Rensselaer Polytechnic Institute.

Lighting Research Center. (1996). *Prudential HealthCare, Albany, New York*, DELTA Portfolio 1(5). Troy, NY: Rensselaer Polytechnic Institute.

Lighting Research Center. (1997). *Sacramento Municipal Utility District Customer Service Center, Sacramento, California*. DELTA Portfolio, 2(2). Troy, NY: Rensselaer Polytechnic Institute.

Lighting Research Center. (1997). *Sony Disc Manufacturing, Springfield, Oregon*. DELTA Portfolio 1(6), 2(1). Troy, NY: Rensselaer Polytechnic Institute.

Lighting Research Center. (1998). *Mary McLeod Bethune Elementary School, Rochester City School District, Rochester, New York*. DELTA Portfolio, 2(3). Troy, NY: Rensselaer Polytechnic Institute.

Lighting Research Center. (1999). *South Mall Towers Apartments, Albany, New York*. DELTA Portfolio, 2(4). Troy, NY: Rensselaer Polytechnic Institute.

Lighting Research Center. (1999). *Staples Distribution Center, Killingly, CT*, DELTA Portfolio 2(5). Troy, NY: Rensselaer Polytechnic Institute.

Lighting Research Center. <www.lrc.rpi.edu>.

Lightolier. (1994). *Journey: Lightolier and Lighting in the Twentieth Century*. New York: Lightolier.

Luckiesh, M. (1920). *Artificial Light*. New York: Century.

Maniccia, D.; Ratledge, B.; Rea, M.; and Morrow, W. (1999). "Occupant Use of Manual Lighting Controls in Private Offices." *Journal of the Illuminating Engineering Society* 28(2):42–56.

O'Dea, W. T. (1958). *A Short History of Lighting*. London: HMSO, Ministry of Education, Science Museum.

Rea, M. S., ed. (1991). *Selected Papers on Architectural Lighting*. Bellingham, WA: SPIE—The International Society for Optical Engineering.

LIQUEFIED PETROLEUM GAS

Propane is the most widely used commercial term to describe a family of liquefied petroleum gases (LP-gas or LPG) that also includes ethane and butane.

Name and Formula	*Ethane* C_2H_6	*Propane* C_3H_8	*N-butane* C_4H_{10}
Vapor pressure @ 100 d.F			
lbs/sq. in. absolute	780.0	190.0	51.6
Boiling point of liquid at			
atmospheric pressure-F	-127.5	-43.7	31.1
Weight of liquid @ 60 d.F			
Pounds per gallon	3.11	4.23	4.86
Specific gravity	0.374	0.508	0.584
Gross heat of combustion			
Btu per pound	22,329	21,670	21,315
Btu per cu. ft. @ 60 d.F	1,783	2,558	3,368
Btu per gallon @ 60 d.F	69,433	91,044	103,047
Flammability limits			
Lower % in air	3.0	2.2	1.9
Upper % in air	12.5	9.5	8.5
Freezing point of liquid at			
atmospheric pressure F	-297.8	-305.9	-216.9
Gallons per pound mol @ 60 d.F	9.64	10.41	11.94
Molecular weight	30.07	44.09	58.12

Table 1.
Physical Constants of Selected Hydrocarbons Found in LP Gas.
SOURCE: Handbook/Butane-Propane Gases

Propane is a nontoxic, colorless, odorless hydrocarbon that occurs naturally in natural gas streams and crude oil. At normal atmospheric pressure and temperature, it is a gas; under moderate pressure propane becomes liquid. The ratio of liquid to gas is 270—one unit of liquid expands to 270 units of vapor.

HISTORY

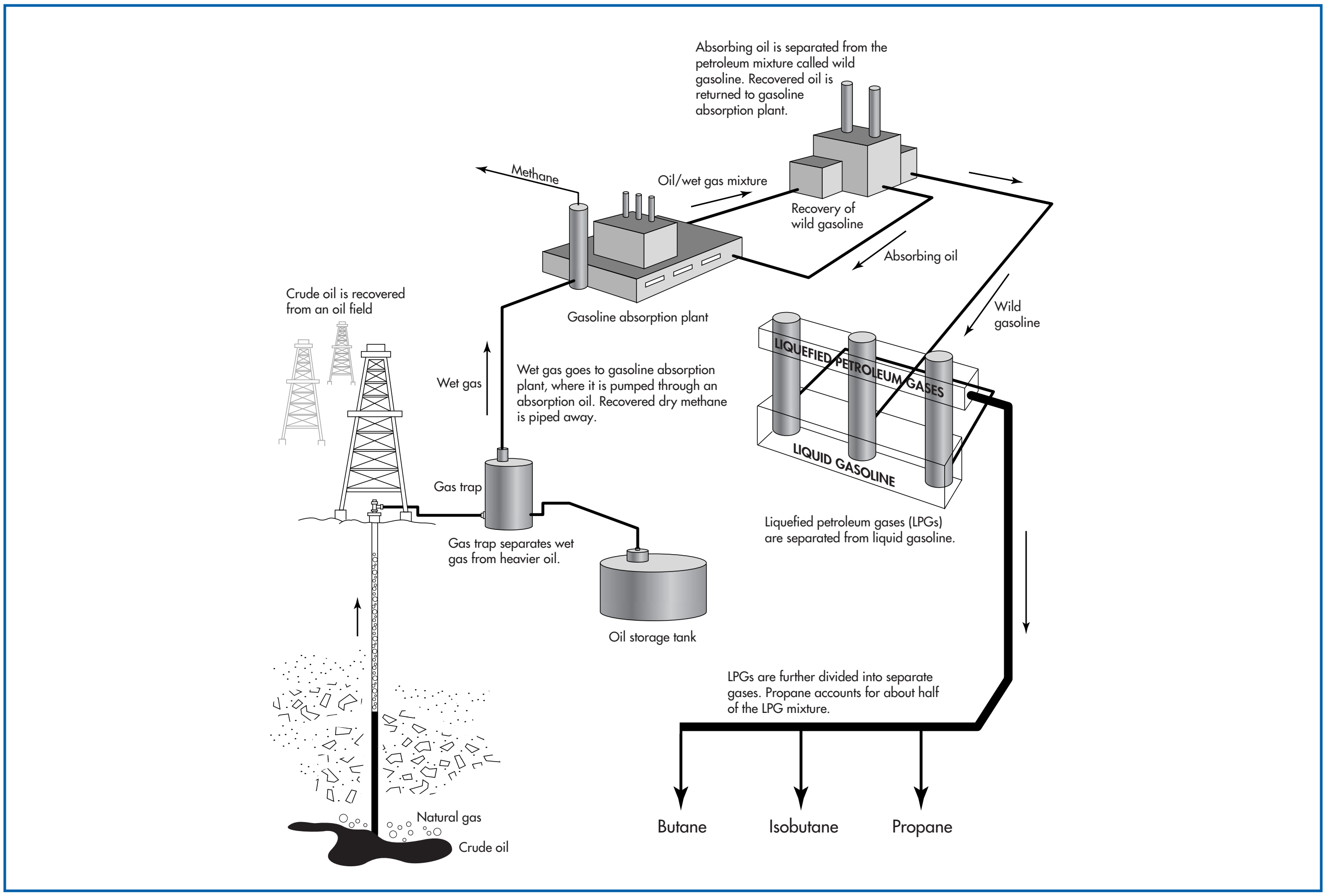
The process of separating "liquefied gases" from natural gasoline was developed in 1912 by Walter O. Snelling, a chemist for the U.S. Federal Bureau of Mines. He wanted to develop better reading light for rural people than the then commonly used candles and kerosene. The first houses were piped for use of this "gasol" in Pennsylvania in 1912. Later, propane was used for cooking and refrigeration. Space heating use soon followed.

Propane became popular as a cutting fuel replacing more expensive acetylene. Another of its first uses was motor fuel. City buses began using it in the 1930s, new farm uses were added in the 1940s including flame weeding, crop drying, hog farrowing, and chicken brooding, and even a locomotive was powered with propane. Recreational uses for barbecues and campers became popular. In the 1950s LP-gas farm tractors and trucks began production. Chicago Transit began converting their city buses to propane.

In the 1960s, use extended to hot air balloons and industrial burners. Residential and commercial use expanded because of new construction without access to natural gas pipelines and the growing popularity of camping. In the 1970s, propane became a popular supplement for natural gas utility systems that were experiencing shortages of natural gas. Refrigerated import terminals were built. When government price controls were imposed in response to the oil embargo, massive vehicle conversions took place to take advantage of the lower-priced and price-controlled propane. Gasoline prices were approaching 100 percent higher and were expected to keep rising.

PHYSICAL PROPERTIES

Propane (C_3H_8) is one of the saturated open-chain hydrocarbons that form the paraffin or alkane series



The process of making propane. (Gale Group)

(C_nH_{2n+2}). The parent compound of this family is methane. The propane molecule, with three carbon and eight hydrogen atoms, is third in the series after ethane (C_2H_6). The specific gravity is 0.508–0.510 at 60°F (15.6°C). The melting point is –309.8°F (–189.9°C). (See Table 1 for other selected properties of LP-gases.)

When liquefied, propane has a Btu content of 91,044 per gallon (higher heating value). At ordinary temperatures, propane is relatively unreactive with other chemicals such as acids, alkalis, or oxidizers.

COMMERCIAL PROPERTIES

Upon release from a pressurized container, propane vaporizes immediately. This property makes liquid propane economical to store and to transport by pipeline, rail, barge, or truck. Because they are relatively light, propane tanks and canisters can be carried by hikers or wheeled to outdoor sites for barbecuing, lighting, and campfires. Propane has been nicknamed “the portable gas.”

Although propane is nontoxic, exposure to liquefied propane can cause skin burns. In gas form, propane is a simple asphyxiant that can displace oxygen. The gas is highly flammable but has a relatively narrow flammability range of 2.2 to 9.5 percent concentration in air. Numerous safety features are incorporated into propane equipment. Storage tanks are constructed to withstand pressures a minimum of four times actual normal operating pressures. Relief valves are integrated into the tanks to relieve extreme pressures that might occur when the tank is subjected directly to fire. To aid in detection of a leak an odorant such as mercaptan is added to the fuel. Although heavier than air, a propane leak readily dissipates into the atmosphere as it reverts from the liquid phase inside the tank to its normal vapor state.

Propane is desirable as a fuel because of its clean-burning properties, its portability, closed-to-atmosphere storage, and availability even in remote locations. It is an approved alternative fuel listed in the 1990 Clean Air Act Amendments and the National Energy Policy Act of 1992. Current factory-produced propane vehicles meet the U.S. Environmental Protection Agency (EPA) ultralow emission vehicle (ULEV) standards and prototype vehicles meet the even more stringent super ultralow emission (SULEV) standards. Because propane does not contaminate soil or water supplies it is exempt from EPA's leaking underground storage regulations.

Compared with other alternative motor fuel options (reformulated gasoline, compressed or liquefied natural gas, ethanol from corn or coal, methanol and electricity), propane has the lowest greenhouse gas emissions except for natural gas. According to a 1998 study by the Institute of Transportation Studies, greenhouse emissions from propane vehicles are 21.8 percent less than from gasoline or diesel.

PRODUCTION AND DELIVERY

There are two primary sources of commercial propane: natural gas processing (55 percent) and crude oil refining (45 percent). A typical natural gas raw stream is comprised of about 90 percent methane and 3 percent propane, and a remainder of other gases such as ethane, butane and pentane. Ninety percent of the propane used in the United States is produced domestically. Ten percent is imported with 50 percent of imports originating in Canada and Mexico. Domestic production comes from several hundred refineries and gas processing plants. From the production source, propane is transported to bulk storage plants in pipelines, railroad tank cars, trucks and barges. Combinations of several types of transportation may be used depending on distance and economics of each transportation type. From the bulk plant, deliveries are made to end users in delivery trucks of 2,500-3,000 gallon capacity. Imports are by pipeline and refrigerated ships.

CONSUMPTION

Worldwide consumption of propane in 1998 was about 96 billion gallons, of which 15 billion gallons were used in the United States by some 60 million consumers. This represents about 4 percent of the total United States energy market. (Each year an additional 8-22 billion gallons of propane are used by the petrochemical industry where it is reformulated into basic building blocks for the production of a variety of polymeric products and chemicals.)

Homes and commercial establishments are the largest users of propane as a fuel where it is used for space heating, water heating, cooking and clothes drying. Fuel storage tanks can be located underground or above ground. Residential, commercial and industrial appliances and equipment can generally operate on either propane or natural gas with only slight modifications to accommodate the different heating values and fuel/air ratios of the fuels.

Industries use propane for space heating, concrete drying, steel cutting, process heat, and asphalt laying, and to power forklifts. Farms use it for crop drying, irrigation engines, animal barn heating and flame cultivation.

Roy W. Willis
Robert Myers

See also: Import/Export Market for Energy.

BIBLIOGRAPHY

Cannon, R. E. (1998). *The Gas Processing Industry: Origins and Evolution,* 2nd ed. Tulsa: Gas Processors Association.

Delucchi, M. A. (1998). *LPG for Motor Vehicles: A Total Fuel Cycle Analysis of Emissions of Urban Air Pollutants and Greenhouse Gases,* prepared for the Propane Vehicle Council and the Propane Education & Research Council, Washington, DC.

Hayduk, W., ed. (1996). *Propane, Butane and 2-Methylpropane.* New York: Pergamon Press.

National Propane Gas Association. *Propane Gas Facts.* <http://www.propanegas.com/consumer>.

Propane Research & Education Council. <http://www.propanecouncil.org>.

Propane Research & Education Council. *The NEED Project. What Is Propane?* <http://www.propanecouncil.org/need/secondary/htm>.

U.S. Geological Survey. (1995). *The Future of Energy Gases.* Reston, VA: Author.

World LP Gas Association. *LP Gas Uses.* <http://worldlpgas.com/uses/d_uses.htm>.

LOAD MANAGEMENT

See: Electric Power, Generation of

LOCOMOTIVE TECHNOLOGY

A machine designed to convert the potential energy stored in fuel to tractive effort needed for pulling unpowered wheeled vehicles (usually in a train on railway tracks) is called a "locomotive." The complexity of this definition reflects the fact that "railroads" are a composite technology made up of tracks, locomotives, and trains. There were "trains" of wheeled vehicles pulled by animals or gravity on "tracks" laid to reduce rolling resistance before there were locomotives, but there were no "railroads" until all three components were brought together. The final link was the locomotive—one of the most useful inventions ever made. Within decades of its development locomotives had multiplied overland travel speeds and freight tonnages more than tenfold—in the process revolutionizing land transportation for all time.

HISTORY

Historians usually credit the invention of the locomotive (and hence the innovation of the railroad) to Richard Trevithick, a Cornish mining engineer. In 1804 he assembled a steam-powered locomotive employing used cylinder steam to increase stack draft for a hotter boiler fire, with the resultant puffing smoke and sound characteristic of steam locomotives. Trevithick then used his effective little engine to pull some ten tons of iron and about seventy persons in railed wagons for a distance of nine miles on the Penydarren Iron Works tramroad in South Wales. The Frenchman Nicholas Cugnot had devised a steam-powered, self-propelled street wagon as early as 1769, but before Trevithick's demonstrations there was much disbelief that a locomotive could pull heavy loads without its wheels spinning uselessly on the iron rails.

The word "locomotive" means, literally, "power for moving from one place to another." Early locomotives were given the nickname "iron horse" because of their material and function. Most European languages use similar nomenclature for "railroads," as in the French *chemin de fer*, Italian *ferrovie*, Spanish *ferrocarril* and German *eisenbahn*, all of which mean "iron way" in literal English.

Work is movement of mass through distance, and *power* is work accomplished or energy transferred at a given rate or in a given time. Horsepower is a measure of power—first used, so it is said, by one of the pioneers of steam engine development, James Watt. One horsepower is, approximately, the work a good draft horse could accomplish in a minute—now given as 33,000 foot pounds/minute or 746 watts. A locomotive's power is often rated by the horsepower generated by its prime mover (the engine for converting fuel to power and motion), and the pulling force it can give to the first car in its train (drawbar horsepower). The difference between engine horsepower and drawbar horsepower is power used or lost in the locomotive's transmission, wheel slippage, and auxiliary functions such as producing air pressure for braking or electric current for lighting and other onboard train devices, plus the power needed to move the locomotive itself. A locomotive's tractive effort (work performed by wheel torque at the rails) is estimated in pounds. Actual tractive effort can be directly measured with the use of a wheel dynamometer. What is most important to a railroad, however, is drawbar pull, which equals tractive effort minus the force necessary to move the locomotive.

POWER GENERATION

Locomotives are often classified into three basic types derived from the kind of prime mover and the fuel or energy conversion system employed. These three main types are steam, diesel-electric, and electric (in which the prime mover is actually an off-board power generating station). By historical usage, steam locomotives are often called "engines" and electric locomotives or powered passenger cars are often referred to as "motors" or "traction." Locomotives with internal combustion or turbine prime movers are usually identified further by the type of transmission apparatus, thus "diesel-electric," "LNG diesel," "gas-electric," "turbine-electric," "diesel-hydraulic," or "diesel-mechanical," etc. Innovations in each generic type of locomotive gradually improved its energy efficiency, but inherent features of each type limited technology improvements and ultimately determined (along with other capital and operating cost factors) its success in the marketplace. *Dual mode* or *hybrid* locomotives refer to designs enabling operation with more than one prime mover, such as diesel-electric plus third-rail electrification (see Table 1).

	Steam	*Electric*	*Diesel - DC*	*Diesel - AC*
OUTPUT SIDE:				
Power Generation-Availability per Unit	High	Very High*	Limited	High
Starting Ability	Poor	Excellent	Excellent	Outstanding
Acceleration	Good	Outstanding	Good	Excellent
Operational Control	Poor	Excellent	Good	Excellent
INPUT SIDE:				
Capital Cost - Equipment	Low	Very High	Average	High
- Fuel (and water) Infrastructure	High	Very High	Average	Average
Operating Cost - Labor	High	Low	Average	Average
- Fuel Cost (North America)	High / Varies	High / Varies	Low / Varies	Low / Varies
Maintenance Cost - Equipment	Very High	Average/High	Average	Average/Low
- Shops and Line-side Infrastructure	Very High	High	Low	Low

Table 1.
Technical and Economic Characteristics of Locomotive Types.

Wood was the easiest fuel to use in early steam locomotives, but it was soon realized that the logistics of wood fuel were limiting. Steam engines were developed that could burn coal, peat, or (later) oil where those fuels were more abundant. For intercity railroads (especially in the Americas, Asia, Australia, and Africa), coal remained the fuel of choice for one hundred years. Despite impressive technology development, steam locomotives never could achieve thermal efficiencies greater than about 6 to 8 percent.

As electric and internal combustion motors were developed and improved in the second half of the nineteenth century, railroads began to look at alternatives to steam locomotion. In most urban transit applications, electric traction overtook horse-drawn and cable-drawn cars, as well as any remaining coal burners. Electrification offered smoke-free locomotive operation, the ability to tap hydroelectric energy supplies (important in Europe and Japan), and the benefit of drawing peak power levels when needed for climbing steep mountain grades or rapid acceleration of passenger trains. With regard to intercity rail applications as opposed to urban transit, however, more often than not initial construction costs for electric power distribution systems were prohibitive. The most important extension of intercity rail electrification in many decades has been completed between New Haven, Connecticut and Boston.

Dual mode locomotives offer the ability to operate beyond electrified territory without changing locomotives, but at a cost of deploying additional on-board equipment and incurring consequent suboptimal performance in both modes. Hybrid locomotives have been used for many years in the approaches to New York City commuter terminals.

In America and elsewhere, coal powered steam locomotives gave way to diesel electric traction over the period 1925 to 1960. This three-decade-plus process of "dieselization"—from first innovation to universal application—became a textbook case in the study of the diffusion of new industrial technology. Professor Edwin Mansfield has demonstrated that the economics of dieselization were akin to those of stimulus and response in psychology; railroads that could benefit the most from diesel power implemented the new technology most expeditiously. Consequently (and like jet engines later replacing piston aircraft), dieselization followed a typical logistics curve—slow initial acceptance in the face of skepticism and uncertainty, then rapid deployment as benefits were understood, and finally tapering of demand as opportunities became saturated.

A diesel-electric locomotive uses as its prime mover a large, self-igniting, internal combustion engine of the type invented by Rudolf Diesel and first successfully demonstrated in 1897. Thermal efficiency of these engines exceeded 30 percent, compared

1712	**Thomas Newcomen**	**First recorded Newcomen atmospheric (vacuum) steam engine.**
1776	James Watt	Separately condensing steam engine.
1802	Trevithick & Vivian	High-pressure steam engine with feed water heater.
1804	**Richard Trevithick**	**Tramroad stack-blast steam locomotive and cars (first railroad).**
1812	William Chapman	Pivoting locomotive bogie.
1814-15	Hedley, *et al.*, Wylam Colliery RR	"Puffing Billy" with 4 and then 8 geared wheels.
1814-16	G. Stephenson, Dodd & Losh	Locomotives with connecting rods for transmission to wheels.
1821	**Michael Faraday**	**Demonstration of electro-magnetic rotation (invention of the electric motor).**
1825	**George Stephenson**	**First public railway, Stockton and Darlington; engine No. 1, "Locomotion."**
1827	William Chapman	Equalizing lever for axle bearings and locomotive weight distribution.
1827	Sequin (France); Booth (England)	Multi-tubular boiler with tubes surrounded by water.
1829	**Robert Stephenson**	**"The Rocket," drawing three times its weight, reached speed of 12.5 mph.**
1829	Delaware & Hudson Canal Co. RR	First locomotives in North America, imported from England, burned coal.
1830	George Stephenson	Locomotive with horizontal inside cylinders.
1830	**Peter Cooper**	**"Tom Thumb" runs on B&O from Baltimore to Ellicott Mills, MD.**
1831	Ross Winans	Improvements in the construction of locomotive axles with outside journals.
1831	Mohawk & Hudson River RR	Locomotive crew cab to protect operators from the elements.
1833	Leicester & Swannington Railway	Steam whistle, after train hit horse and cart at crossing.
1836	Henry Campbell	Locomotive of the 4-4-0 "American" type with front swivel truck.
1836	W.G. Whistler	First American steam whistle, on locomotive "Susquehanna."
1837	Thomas Rogers	Cast iron wheels with hollow spokes and rims.
1838	Rogers & Wakely	First successful oil-burning headlight, allowing nighttime operation.
1839	**R. & W. Hawthorn Co.**	**Superheater designed to make more efficient use of steam energy.**
1842	**Egide Walschaerts**	**Radial valve gear, enabled better control of steam locomotive operation.**
1849	Eugene Bourdon	First practical steam pressure gauge, enabling safe use of higher pressures.
1850	James Samuel	Compound cylinder locomotive (high pressure and low pressure steam cycles).
1857	George Griggs	Brick arch firebox for increased steam locomotive energy efficiency.
1862	Pacific Rail Act (USA)	Authorized transcontinental railroad and promoted standard national track gauge.
1869	Union Pacific and Central Pacific RRs	First transcontinental railroad; May 10 driving of Golden Spike at Promontory, UT.
1869	George Westinghouse	Compressed air brake enabled heavier and faster locomotives, longer trains.
1872	William Robinson	Closed-loop track circuits and wayside signals enabled faster, safer train speeds.
1872	Robert Davidson	First electric (battery powered) locomotive, pulled 6 tons at 4 mph.
1873	Eli H. Janney	Automatic coupler, for faster and safer train make-up.
1879	Werner von Siemens	Successful demonstration of the use of electric traction on railways.
1885	J. van Depoele	Single overhead wire system for electric railways.
1886	**Frank J. Sprague**	**Electric-powered operations from axle-mounted traction motors.**

Table 2.
Technological Development of the Locomotive.

1889	S.M. Vauclain	Compound low- and high-pressure cylinders.
1893	George Daniels, New York Central RR	First 100 mph record run (Engine #999)
1895	Frank Sprague	Multiple-unit control system.
1895	GE, Baltimore & Ohio RR	First major railroad electrification in U.S.
1897	**Rudolf Diesel**	**First compression-ignition internal combustion engine.**
1897		Steam-powered locomotive with electrical transmission.
1898	**Wilhelm Schmidt**	**Steam superheater for locomotives.**
1901		Mechanical coal stokers.
1906	Baldwin Locomotive Works	2-6-6-2 articulated compound steam locomotive.
1906	General Electric Co.	Gasoline engine equipped with electric drive.
1914	**Sulzers, Krupp, and the Prussian & Saxon State Railways**	**First large diesel-powered locomotive.**
1923-5	**Herman Lemp and GE**	**Separately excited DC generator for locomotive. Self-regulating loads in diesel engine, generator, and traction motor.**
1924	**Ingersoll-Rand and GE**	**First successful diesel-electric locomotive in USA.**
1920s	(end of decade)	Mechanically-driven pressure charger applied to diesel locomotive engine.
1930	Timken Roller Bearing Co.	Steam locomotive using roller bearings on the driving axles.
1934	Union Pacific RR	First distillate-electric streamliner in regular service, the M-10000.
1934	C, B & Q (Burlington) RR	First diesel-electric streamliner run, Denver to Chicago, *The Pioneer Zephyr*.
~1935	**Bucci (Italy)**	**Development of exhaust gas turbo-charger for diesel engines.**
1935	Pennsylvania RR	Electrified passenger-train service between New York and Washington.
1937	American Locomotive Co.	First diesel engine turbocharger applied to locomotive in America.
1938	Union Switch & Signal Co.	Communication system for trains.
1939	**GM Electro-Motive Co.**	**First successful diesel-electric locomotive for freight service, and first use of dynamic brakes in road freight locomotive: Demonstrator freight locomotive FT-103; 5400 hp in four units (A-B-B-A).**
1941	Union Pacific RR and ALCO	"Big Boy" 4-8-8-4 steam locomotives, largest ever built (540 long tons).
1941	EMD; A, T & SF (Santa Fe) RR	Diesel-electric road freight locomotives in regular service.
1948	GE & ALCO	First gas-turbine locomotive.
1952	Norfolk & Western Ry Roanoke Shops	Last steam locomotive delivered for use by Class I railroad.
1960	Canadian National RR, Norfolk & Western RR	Last steam locomotive runs in revenue service by US and Canadian trunk lines. Steam use continued longer in Mexico.
1960 -70	**GM-EMD and GE**	**SD40 and U30C class locomotives of 3000 hp and 6 powered axles. Over 5000 were purchased in U.S. alone.**
Late 1970s	ASEA - Brown Bovari (ABB)	Electric locomotive rectifier technology arrived from Europe for Amtrak's Northeast Corridor passenger trains.
1979	Federal Railroad Administration	Proposals for railroad electrification related to second OPEC energy crisis.
~1981	Canadian National	North American "Wide Cab" design to improve comfort and safety of crew.
1984	GE	First micro-processor controlled diesel-electric locomotive (Dash 8).
1984	AAR and Railway Association of Canada	Beginning of industry cooperative effort to develop open-standard Advanced Train Control System (ATCS).

Table 2 (continued).
Technological Development of the Locomotive.

1984 -86	GM-EMD	Testing of 4-axle radial trucks on BN and ATSF.
Mid '80s	Canadian Pacific RR, Bombardier	Experimentation with AC traction locomotives. (Conversion of M-640 to AC.)
1988	GM-EMD and ABB	F69PHAC alternating current locomotive for Amtrak.
1989	**GM-EMD and Siemens**	**Solid state Voltage Source Inverter (VSI) and Gate Turn-Off (GTO) thyristor technology applied to locomotive. Enabled changing DC to AC power of infinitely variable frequency and amplitude, facilitating high adhesion traction motor control.**
1989	**GM-EMD and Siemens**	**First successful three phase AC traction high horsepower freight / passenger locomotive (SD60MAC). Compact "brushless" induction rotor ("squirrel cage") traction motor is self-regulating and saves maintenance.**
1993	Burlington Northern RR & EMD	Largest single locomotive order in history, 460 SD70MAC 4000 hp locomotives; (order later increased to 680 units).
1994	Association of American Railroads	Publication of Locomotive Systems Integration (LSI) standards to improve inter-operability of on-board devices and components.
1996	**GE**	**First 6000 hp diesel-electric locomotives using a single diesel engine.**

Table 2 (continued).
Technological Development of the Locomotive.

with about 17 percent for contemporary low pressure oil engines and about 14 percent for a steam turbine. Locomotives use "medium speed" (typically 16 cylinders and 800 to 1200 rpm) diesel engines—in contrast to the "high speed" engines in trucks or "low speed" diesels in ships. Older locomotives powered an electric direct current (DC) generator, which in turn fed electric energy to DC traction motors mounted one to an axle, usually four or six in total, under the frame of the locomotive. Newer locomotives almost always generate alternating current (AC) in an alternator, and then rectify it to DC. If AC traction motors are to be used, the power is then inverted to AC of precisely variable frequency (see below).

Diesel engine output is controlled by throttle settings for the diesel engine ("notch eight" usually means full power), and is automatically balanced to the electrical load on the generator or alternator. Most large diesel-electric locomotives employ a turbo-supercharger, which uses exhaust from the engine to compress air for injection, along with additional fuel, to increase the combustion rate and thus the horsepower of the engine. This action is also automatic, in that the harder the engine runs, the more exhaust, the more power for compression of air, and hence the more oxygen injected for combustion of more fuel. Turbochargers are precision machines and require special maintenance, but they approximately double the engine's horsepower compared with naturally-aspirated diesels. Design improvements in diesel-electric locomotives continue to result in impressive gains in fuel efficiency. Revenue ton-miles per gallon of fuel have nearly doubled since 1975 (see Figure 2).

Top speeds achievable by locomotives depend on their mass, horsepower, gearing, and the quality of track on which they run. Steam and diesel-electric locomotives have been operated at more than 100

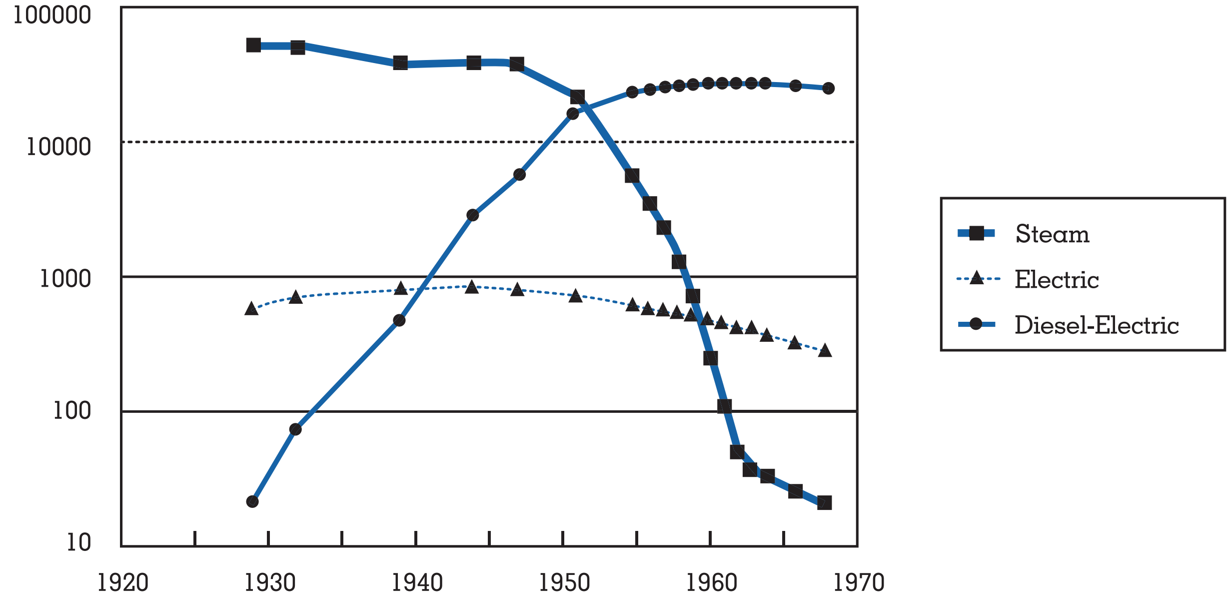


Figure 1.
Logarithmic Scale of Locomotives in Service—Class 1 Railroads.
SOURCE: Association of American Railroads. (1970). *Railroad Facts,* p. 68. Washington, DC: AAR.

mph, and electric locomotives in France first exceeded 200 mph in 1955. The Federal Railroad Administration is currently sponsoring development of a turbine-electric locomotive designed to operate at up to 150 mph.

TRACTIVE EFFORT

After the power rating of the prime mover, a second limitation to any locomotive is the tractive effort it can deliver to the unit's driving wheels without slipping on the rail. The proverbial "inherent advantage" of railroads due to low rolling resistance of steel wheels on steel rails now becomes a limitation, in that a positive coefficient of friction is needed to maintain pulling force at the face of the rail. In the old days, steam locomotive engineers developed great skill at adjusting steam admission to the cylinders and sanding of the rails to minimize slipping. To start a long train, the engineer would back the engine into the train in order to bunch the train's slack, then accelerate—causing each car in succession to jerk to a roll. This is unnecessary with an appropriately powered diesel locomotive consist.

In 1939 General Motor's Electro-Motive Division sent its famed demonstrator FT-103 diesel-electric units on a triumphant tour of America. One purpose was to show skeptical railroaders that this 5400 hp, four-unit diesel locomotive developed more low speed tractive effort than competitor steam engines, which meant smooth starts and excellent performance on long mountain grades. The effect was much like Trevithick's demonstrations 130 years earlier, and a new generation of railroad locomotive power was assured.

Modern locomotives have sensors that detect

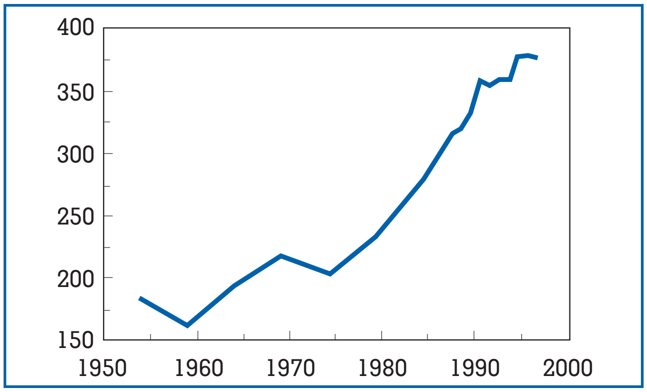


Figure 2.
Revenue Ton-Miles per Gallon of Diesel Fuel—USA Class 1 Railroads.

SOURCE: Association of American Railroads. (1998). *Railroad Facts*, p. 40. Washington, DC: AAR.

when the wheels of any powered axle are about to spin; the microprocessor-based slip control then reduces energy to that traction motor to enable smooth application of maximum pulling power from all wheels. Traditionally, locomotives could achieve (for dry, sanded rail) an adhesion factor of about 25 percent of their weight on driving wheels, but the refinement of wheel slip controls allowed increases in the adhesion factor to about 32 percent. Also to increase torque at the rail, DC locomotive traction motors are wired in series at low speeds, spreading generator voltage and maximizing tractive effort. When speed increases to a certain point, the motors are switched to a parallel connection permitting increased traction motor voltage; this allows the motors to run faster and cooler.

The newest AC locomotives use microprocessor controls and instrumentation that compares ground speed with traction motor rotation to control tractive effort and wheel slip with great precision. These locomotives have equipment that can change the phase angle of the three phase power supplied to the traction motors, and a device called a thyristor (a high speed solid state switch that can handle large ratios of controlled to controlling amperage) to adjust the frequency and amplitude of alternating current cycles. AC locomotives with single axle control will actually reallocate power from a slipping wheelset to those with greater adhesion.

The arrangement of diesel engine, on-board electric power generation, and axle-mounted traction motors also allows dynamic braking—the ability to reduce speed by converting train momentum back through torque developed between rail and wheel to the locomotive axles and traction motors. The motors, thus converted to electric power generators, resist the kinetic energy of the train and produce electric current, which must be carried away to prevent overheating. In electric locomotives (where it is known as regenerative braking) the current is returned to the catenary or third rail, while in diesel-electric engines the electric power is wasted as heat energy through ventilated grid resistors on the top of the locomotive.

OPERATIONAL CONTROL

The third limitation on performance of any locomotive is the ability, whether through the skill of the engineer or by good design of the technology, to control its overall operation effectively. Steam engines were tricky to operate, to the point of taking on almost humanoid characteristics in popular lore. Getting the boiler fire to burn just right, for example, required hard work and experience. Until the mechanical stoker was invented, the fireman's simple physical limitations constrained performance. An advantage of the diesel-electric power configuration is that loads balance nearly automatically between the engine, generator, and traction motors—making operation by the locomotive engineer or driver much simpler than on a steam engine. The "marriage" and self-correcting nature of the diesel engine, its generator (or alternator), and the traction motors was invented by Herman Lemp of General Electric in 1923–1925, and continues to be used today in all diesel-electric locomotives.

With use of electric and diesel-electric locomotives also came the ability to control power and tractive effort, not just for the unit in which a crew rides, but for multiple units (MU) coupled together in a locomotive consist—and this without need of additional crews in the following units. (A "consist" is the make-up of the train, in this case the full set of locomotives used.) With MU operation, locomotive managers had flexibility to size horsepower of the consist to operating requirements of a particular train and its territory or specific assignment.

In time, other aspects of operational control of the locomotive were made possible by electric and diesel-electric power. One important technology advance

	GM-EMD SD40-2 (1972)	GE U30C (1972)	GM-EMD SD90MAC (1997)	GE AC6000 CW (1997)
Traction Motors	Direct current	Direct current	Alternating current	Alternating current
Diesel Engine HP Rating	3000	3000	6000	6000
Continuous Tractive Effort	83,000 lbs.	90,600 lbs.	170,000 lbs.	166,000 lbs.
Adhesion Factor*	20%	20%	38%	40%
Maximum Braking Effort	61,000 lbs.	61,000 lbs.	115,000 lbs.	117,000 lbs.
Maximum Speed	70 mph	70 mph	75 mph	75 mph

Table 3.
Locomotive Performance Comparison: 1972 to 1997.

represents a logical extension of the principle of MU operation—remote control of additional locomotives placed deep in the train or at its end. This distributed power has proven highly effective in reducing drawbar forces in long, heavy trains, or indeed, in allowing safe handling of longer trains by the locomotive engineer. Experiments were made with distributed power in the 1960s and 1970s, but it was only with the development of reliable microprocessor controls and data radio communications in the 1990s that the concept proved widely effective.

Train control in a different sense of the term is a powerful potential breakthrough technology for railroads. In this concept, accurate knowledge of train position and dispatcher authorities interpreted by an on-board locomotive computer is used to prevent train collisions, overspeed accidents, and incursions into track territory reserved for maintenance crews. The Global Positioning System (GPS), digital data radio communications to and from locomotives, and computer display screens for crews are among the enabling technologies to be used in advanced train control systems, increasingly called positive train control (PTC). The on-board locomotive electronic networks and radio communications systems developed for applications such as power and train control will increasingly be used for locomotive health monitoring, event recording, train dispatching efficiency, on-time (scheduled) train movement, and more responsive customer service. Some of these other business functions are already being deployed on locomotives by railroads, and the Association of American Railroads has developed locomotive systems integration (LSI) standards to promote compatibility of devices from different suppliers.

EMISSIONS

Locomotive emissions have become an important aspect of performance in recent years. While overall fuel efficiency of railroad freight service has been shown by several detailed studies to be approximately three times better than highway motor carrier service, and while rail passenger advocates point out the energy use, congestion mitigation, and superior emissions performance of rail compared with other modes of transportation, environmental protection authorities are interested in reducing major sources of manmade pollution. The 1990 Clean Air Act Amendments specifically mandated extension of emission regulations to locomotives, which the U.S. Environmental Protection Agency (EPA) currently estimates to be responsible for nearly five percent of nationwide emissions of oxides of nitrogen.

In April of 1998, the EPA published a final rule for emission of oxides of nitrogen (NOx), hydrocarbons (HC), carbon monoxide (CO), particulate matter (PM), and smoke opacity for newly manufactured and remanufactured locomotives. The rulemaking took effect in 2000 and is estimated by the EPA to cost the railroads $80 million per year—about $163 per ton of NOx reduced, according to EPA figures. The emissions standards for the several pollutants will be implemented in three tiers—for locomotives manufactured or remanufactured from 1973 to 2001, from 2002 through 2004, and in or after 2005, respectively. The emission levels are based on achievability with reasonable technology and cost outlays. EPA has adopted fleet averaging and banking and trading provisions to provide flexibility in meeting goals more efficiently. The two largest railroads serving the Los Angeles Basin have entered into a separate agreement with California authorities to address the special air quality issues in that area.

LOOKING BACK AND AHEAD

After decades of economic decline and company bankruptcies from the end of World War II through the 1970s, American railroads were largely deregulated in the Staggers Act of 1980. Since then the industry has achieved an unprecedented renaissance in productivity, financial performance, and economic value to customers, due in part to significant technology advances facilitated by deregulation. Because of their labor and energy efficiency, favorable environmental characteristics relative to competing transportation modes, and remarkable ability to absorb and spawn new technologies, railroads have a bright future. Expected increases in traffic volumes in coming decades will require large investments in railroad capacity, including new locomotives. Continuing improvements in locomotive performance, together with the contributions of "intelligent systems" riding these powerful machines, will lead the procession into railroading's third century, as they did its first.

Robert E. Gallamore

See also: Diesel Fuel; Mass Transit; Railway Passenger Service; Steam Engines; Transportation, Freight Delivery and; Trevithick, Richard.

BIBLIOGRAPHY

Armstrong, J. H. (1993). *The Railroad: What It Is, What It Does*, 3rd ed. Omaha: Simmons-Boardman Books, Inc.

Bilz, F., and Holger, S. (1998). "A Decade of Three-phase Traction Technoogy for Diesel-Electric Locomotives in North America." Translated from *ZEV+DET Glassers Annalen* 121(9)/1997. Siemens AG Transportation Systems Group.

Bruce, A. W. (1952). *The Steam Locomotive in America: Its Development in the Twentieth Century*. New York: Bonanza Books.

Doherty, J. M. (1962). *Diesel Locomotive Practice*. London: Odhams Press, Ltd.

Draney, J. (1954). *Diesel Locomotives: Mechanical and Electrical Fundamentals*. Chicago: American Technical Society, Publishers.

Farrington, S. K. (1943). *Railroading from the Head End*. Garden City, NY: Doubleday, Doran & Co.

Grosser, M. (1978). *Diesel: The Man and The Engine*. New York: Atheneum.

Haine, E. A. (1990). *The Steam Locomotive*. New York: Cornwall Books.

Haut, F. J. G. (1970). *The Pictorial History of Electric Locomotives*. New York: A. S. Barnes & Co.

Illustrated Science and Invention Encyclopedia. (1983). Westport, CT: H. S. Stuttman.

Jackson, R., and Triest, M. (1999). "TTCI: Tough Tests for High Tech Power." *Railway Age*, January, pp. 55–58.

Kratville, W. W., ed. (1997). *The Car and Locomotive Cyclopedia,* 6th ed., Section 4:430–619. Omaha: Simmons-Boardman.

Marre, L. A. (1995). *The Contemporary Diesel Spotter's Guide*, 2nd ed. Waukesha, WI: Kalmbach Publishing Co.

Marshall, C. F. D. (1953). *A History of Railway Locomotives*. London: Locomotive Publishing Co.

Marshall, J. (1974). *Rail Facts and Feats*. New York: Two Continents Publishing Group.

Pinkepank, J. A. (1973). *The Second Diesel Spotter's Guide*. Milwaukee: Kalmbach Publishing Co.

Ransome-Wallis, P., ed. (1959). *The Concise Encyclopedia of World Railway Locomotives*. New York: Hawthorn Books.

Rolt, L. T. C. (1962). *The Railway Revolution: George and Robert Stephenson*. New York: St. Martin's Press.

Stretton, C. E. (1989). *The Development of the Locomotive*. London: Bracken Books.

Thomas, D. E., Jr. (1987). *Diesel: Technology and Society in Industrial Germany*. Tuscaloosa, AL: University of Alabama Press.

U.S. Environmental Protection Agency. <http://www.epa.gov/orcdizux/regs/nonroad/locomotv/frm/42097048.htm>.

White, J. H., Jr. (1997). *American Locomotives: An Engineering History, 1830–1880,* revised and expanded ed. Baltimore: Johns Hopkins University Press.

LOS ALAMOS NATIONAL LABORATORY

See: National Energy Laboratories

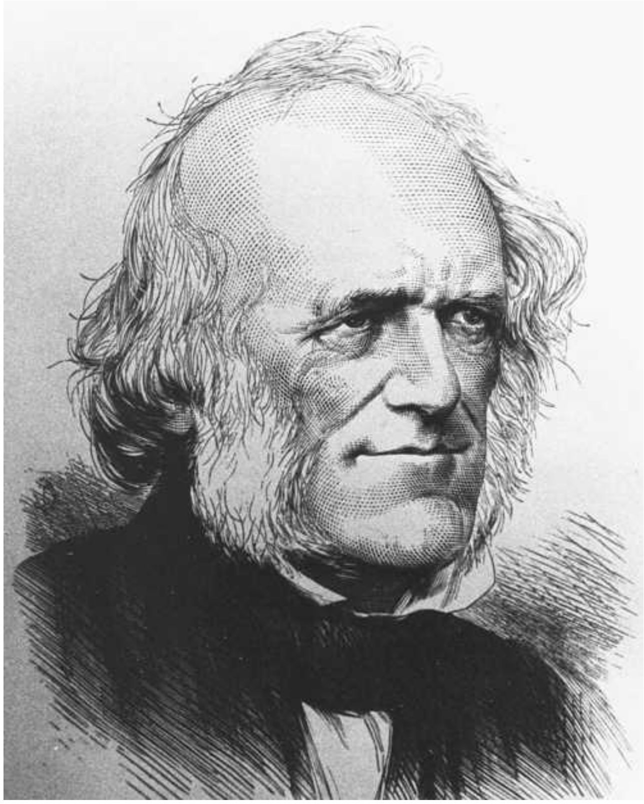
LUBRICANTS

See: Tribology

LYELL, CHARLES (1797–1875)

Charles Lyell was a founder of modern British geology. One of his most important contributions to science concerned the rates at which the earth's internal energy was released to affect the shape and form of its crust and thus to create the landscape we know today. Geological changes such as creation of valleys, mountain formation, deposition of sediments, and the like were not in his view caused by occasional "catastrophes" but rather the results of ordinary geological processes operating over an immense period of time. In other words the release of energy to produce geological change has occurred at a rate similar to that of the present time, and is largely uniform over the earth's long history. This doctrine of uniformitarianism has been rightly attributed to Lyell, though others before him (e.g., James Hutton) had expressed similar views.

Lyell was born on November 17, 1797, at Kinnordy near Forfar in Scotland. His father was a wealthy landowner with passions for both Italian literature and natural history. His mother came from Yorkshire and while he was still very young the family moved to England, taking a lease of a large house on the fringe of the New Forest in Hampshire. The boy was sent to schools in the locality, though he showed little evidence of scholastic promise. In 1816 he went to Exeter College, Oxford, to read classics, but he also attended lectures by William Buckland, a professor of geology. The effects of these lectures were reinforced by books from his father's ample library and by an encounter with Gideon Mantell, a doctor in the neighboring county of Sussex, renowned for his study of the fossils then being discovered in the chalkbeds and on the coast of southern England.



Sir Charles Lyell. (Library of Congress)

Lyell's 1819 Oxford degree was in classics, and he then began to study law at Lincoln's Inn, London. However, weak eyesight precluded much reading at that time, and this gave him a reason for further geological study in lectures, field studies, tours, and so on. He became a fellow of the Geological Society in 1819, and four years later was made one of its joint secretaries. He undertook geological tours in France, the west country of England, and the Scottish Highlands. A visit with Buckland to Glen Roy revealed to them the now famous Parallel Roads, three raised seabeaches now believed to be of glacial origin but even then posing great problems for conventional geology.

The central problems that Lyell was to address in the next few years involved detailed mechanisms of geological change, the age of the earth; and rates of energy release. The most popular view of geological history had proposed a series of great convulsions or "catastrophes" involving immense amounts of energy, alternating with longer periods of relative quiescence. Most famously, the biblical Flood of Noah was invoked to explain the present crust of the earth, including the regular strata of fossil beds. People who favored water as the chief agent of change ("diluvualists") had to contend with others who thought that fire was chiefly responsible ("vulcanists"). But each believed in a relatively short period of time for it all to happen, and sometimes drew confirmation from certain biblical data that *could* be interpreted to suggest a date of creation a mere few thousand years ago. Much support for this view was given by Buckland and by his Cambridge counterpart Adam Sedgwick, though both were later to change their minds. The opposite views of James Hutton were less popular, though Lyell was increasingly inclined toward them. One of several events that proved critical in his experience was a visit to Sicily in 1828. Before crossing the volcano Etna he had observed strata at the base of the mountain that seemed comparatively recent. Discovering similar deposits on the opposite side he concluded that the vast bulk of Etna rested on what were, in geological terms, recently recent rocks. Hence a vast age for the whole earth was indicated, and he embarked on a crusade for his uniformitarian ideas, publishing his monumental *Principles of Geology* in the early 1830s (and many subsequent editions thereafter). This work brought him lasting fame.

His legal practice largely forgotten, Lyell now devoted the rest of his life to promoting his doctrines and to traveling extensively. In 1832, at the new King's College, London, he became the first professor of geology, a post he occupied for only two years. He received from the Royal Society its Royal Medal in 1834, and a year later he became president of the Geological Society. He became much involved with the Great Exhibition in London in 1851 and in the British Association. He received a knighthood in 1848 and a baronetcy in 1864.

The influence of Charles Lyell on science was profound. Among the recipients of his *Principles* was a young naturalist embarking on a long sea voyage. This was Charles Darwin, who came to develop a strong interest in geology. He also accepted much of Lyell's arguments and, while his own theory of evolution was being formed, relied extensively on Lyellian arguments for an immensely old earth. He became a good friend as well as a disciple, though even in the 1860s Lyell was reluctant to give Darwin his public support, as he saw how thin the fossil evidence really was for transmutation of species. Not all scientists became enthusiastic Lyellians, however. William Thomson (Lord Kelvin) opposed him over the age of the earth, arguing on largely thermodynamic grounds for a shorter time span than Lyell or Darwin wanted. Only with the discovery of subterranean radioactivity were Kelvin's estimates shown to be erroneous and a Lyellian time scale rendered more credible. However, that of all the uniformitarians in the nineteenth century Lyell was the most extreme, and no one identified completely with a literal interpretation of uniformitarian change. There was too much evidence of catastrophic releases of energy in volcanic eruptions, flash flooding, and earthquakes for most people to deny their massive influence of earth history. Today a modified, and reduced, uniformitarianism seems more likely to fit the facts. Lyell's extreme views were probably related to a religious inclination to Unitarianism, which denies God's intervening activity in history through Christ, just as uniformitarianism cannot allow catastrophic interventions in geology. However, Lyell outwardly remained a member of the Church of England.

In 1842 he married Mary Horner, daughter of Leonard Horner, warden of the University of London. Lyell died in London on February 22, 1875.

Colin A. Russell

See also: Geography and Energy Use; Geothermal Energy; Thomson, William.

BIBLIOGRAPHY

Bailey, E. (1962). *Charles Lyell.* London: Thomas Nelson & Sons.

Lyell Centenary Issue. (1976). *British Journal for the History of Science* 9:9–242.

Rudwick, M. J. S. (1985). *The Great Devonian Controversy.* Chicago: University of Chicago Press.

Wilson, L. (1970). "Lyell, Charles." In *Dictionary of Scientific Biography*, ed. C. C. Gillispie. New York: Charles

M

MACHINES, SIMPLE

See: Mechanical Transmission of Energy

MAGNETIC LEVITATION

The term “maglev” was coined by Howard Coffey in the 1970s as a shortened form of “magnetic levitation” for transportation. One dictionary defines maglev as “having to do with a railroad system using magnets to float a swiftly moving train above its tracks.” This is a poor definition since there are no tracks, no need for a train of vehicles, and maglev can work at any speed. A better definition is “a transportation system in which the vehicles are suspended, guided and propelled by magnetic forces without any contact with the guideway.”

Maglev is a recent name for an old idea: support a moving vehicle with magnetic fields so there is no contact between the vehicle and a guideway. This apparently simple idea has long inspired inventors, but in spite of sustained efforts by hundreds of people, publication of thousands of technical papers and the expenditure of billions of dollars, there is no commercial maglev system in operation at the start of the Twenty-First Century. This is due to the formidable technical and nontechnical problems that must be solved, the inability to find a design and application that are well matched, and the complexity of the maglev puzzle. This article discusses the past, present, and probable future for this fascinating technology.

EARLY HISTORY: SIMPLE IDEA, A HARD REALITY

The dream of maglev may be as old as the discovery of permanent magnets, such as ones used in magnetic compasses by the Chinese in about 1250 C.E. Permanent magnets can attract or repel each other and it would seem that one might arrange magnets so that a body floats freely in space. The earliest history of maglev is not well known, but in 1839 Earnshaw proved an important theorem which, in simplified form, states that any arrangement of bodies that have attractive or repulsive forces that obey an inverse power law is unstable. For maglev this means that no configuration of permanent magnets, electromagnets with constant current excitation, and ferromagnetic material can be stable, with or without the presence of gravitational forces. If the magnetic field can change, such as via a feedback control mechanism, stability can be achieved. It is also possible to achieve stability by virtue of motion between interacting bodies or by using superconductors arranged so that their persistent currents change appropriately in response to changing position.

A major problem is the design of a propulsion system for a vehicle that has no contact with a guideway. The only reasonable choice is a linear motor that uses magnetic fields to propel the vehicle. The development of a linear motor and its control system is at least as formidable a challenge as the development of a maglev suspension system.

Magnetic scaling laws favor large systems. It takes strong and heavy magnets to suspend a vehicle with any significant spacing between a vehicle and a guideway, and this has made it difficult to build small maglev systems. We have many toy cars, planes, and trains, but no comparable maglev toys that can be used to explain the key ideas.

A train that operates by magnetic levitation, at a railroad research center in Japan. (Corbis-Bettmann)

TECHNOLOGICAL ALTERNATIVES FOR SUSPENSION AND PROPULSION

There are four competing maglev technologies: electromagnetic suspension, electrodynamic suspension, linear synchronous motor, and linear induction motor.

Electromagnetic Suspension (EMS)

Figure 1a shows an EMS system in which an array of magnets is attracted upwards to a steel rail. It is possible to design the magnets so that there is an upward force produced by magnetic attraction that cancels the downward gravitational force: the magnets are suspended in space! If steel beams were mounted on either side of a "guideway," then a vehicle with magnets on both sides could move along the guideway and be supported and guided by the steel rails. If permanent magnets, or electromagnets with constant current excitation, were used in the design, the system would be unstable: any disturbance from equilibrium would cause the magnets to move in the direction of the disturbance. The difference between stable and unstable equilibrium is like the difference between balancing a cone on its base and balancing it on its point. EMS was pioneered by B. Graeminger in 1911.

Earnshaw's theorem proves that some instability is inherent. Ideally there is neutral stability in the fore and aft direction, so that we have a choice: the vehicle can be stable laterally and unstable vertically, or vice versa. The usual approach is to use a design similar to Figure 1a and use electromagnets with power controllable current. Position and velocity sensors are used in conjunction with controllers to regulate the magnet currents and thereby achieve vertical stability, and lateral stability can be achieved passively or, possibly, with help from controlled magnetic forces. This approach is called "Electromagnetic Sus-pension" (EMS) and has been used in many designs.

EMS vehicles have operated at speeds from 0 to 440 km/h (270 mph).

EMS requires about 1 to 2 kW of magnet power for every ton of vehicle mass. At modest and high speeds this power loss is small compared with the power loss due to aerodynamic drag. EMS is the favored approach for urban maglev and is suitable for high speeds if the use of a minimal 10 mm air gap is acceptable.

Electrodynamic Suspension

Figure 1b shows an array of magnets moving over a conducting sheet and being pushed upwards by forces due to induced currents. This electrodynamic suspension that was pioneered by Gordon Danby and James Powell in the 1960s, can not work at zero speed, but at higher speeds it can be inherently stable.

The virtue of EDS designs is their ability to operate with larger air gaps than is feasible with EMS designs. Vehicles suspended in this way have operated at speeds from about 50 to 552 km/h (31 to 343 mph).

All EDS designs are highly underdamped and can even be negatively damped, the equivalent of the instability found in EMS designs. Many maglev designers have underestimated the importance of this damping problem, the equivalent of building a car with solid tires and no shock absorber.

A well designed EDS suspension will create 5 to 10 kW of power dissipation in the guideway for every ton of vehicle mass. This is almost an order of magnitude higher than the power required for EMS. Since the suspension power comes from the linear motor, at low speeds the motor must provide high thrust, and even at moderately high speeds, the suspension power is comparable to aerodynamic drag losses. The high losses can cause serious overheating of the guideway at low speeds. These factors, taken together, make EDS a questionable technology for urban maglev but quite appropriate for high speeds.

Linear Synchronous Motor

One propulsion alternative is to have a powered guideway in which a magnetic field is made to appear to move via electronic means, and it pushes on a magnetic field created by the vehicle. This is called a long stator linear synchronous motor, generally referred to as an LSM. It is possible to have magnets on the guideway and supply the moving field from the vehicle, but this is not a favorable design for maglev. The LSM creates forces by attractive and repulsive magnetic forces, much as an EMS design creates suspension via these forces. The rotary synchronous motor was invented more than 100 years ago and the linear version was pioneered by Henry Kolm and Richard Thornton at MIT in the 1970s.

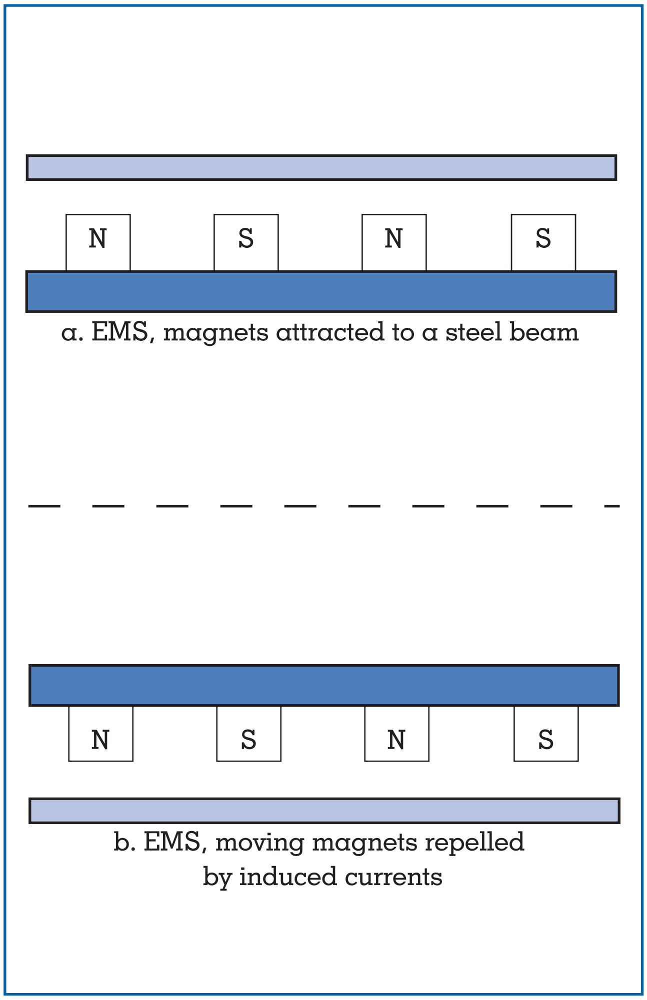


Figure 1.
Electromagnetic suspension: two views.

The advantage of this design is the simplified nature of the vehicle's portion of the propulsion system. In the simplest case the vehicle has a linear array of permanent magnets and is totally passive. This passive vehicle approach has been used successfully for material handling equipment and is probably the best design for small and moderate size vehicles. For larger vehicles the permanent magnets are usually replaced with electromagnets, that also provide the suspension force. This dual use of vehicle magnets for both suspension and propulsion has much merit for high speed transportation systems. The long stator design also has the major advantage that there is no need to transfer large amounts of power to the vehicle and that there is no bulky and heavy power conditioning equipment on the vehicle. Also, the guideway-based power conditioning can be sized according to terrain, with lower power controllers used in regions where high forces and velocity are not required.

A disadvantage of the long stator LSM is the higher cost of the guideway winding as compared with a simple conducting strip that can be used with a linear induction motor. If there are many closely spaced vehicles, each fairly small, then this cost is well justified, but it becomes a problem for some maglev designs.

A second disadvantage is the need to maintain synchronism. Unlike rotary synchronous machines, linear versions are difficult to operate "open loop" with the vehicle automatically maintaining synchronism with a moving field. In practice it is necessary to have a separate controller for each vehicle, and therefore not more than one vehicle can be in one block of guideway. For operation up to a few meters per second, it is possible to use a "stepping mode" with the vehicle following the moving field without feedback, but this method will not work at higher speeds because of inherent "mid frequency resonance" phenomenon. The solution is to use position sensors and electronically synchronize the vehicle with the moving magnetic field.

Linear Induction Motor

A propulsion alternative is to transfer power to the vehicle and then have a magnetic structure on the vehicle that can create a moving field that, in turn, interacts with a conducting guideway to provide propulsion. This is called a short stator linear induction motor (or linear asynchronous motor), generally referred to as a LIM. It is possible to put the powered winding on the guideway and induce currents in a conductor on the vehicle, in which case it would be a long stator design, but this turns out to be an expensive and inefficient approach to maglev propulsion. The LIM creates forces via induced current much as an EDS design creates suspension via induced currents. The rotary induction motor was invented by Nikola Tesla in 1886 and the linear version for transportation was pioneered by Eric Laithwaite in the 1950s.

The advantage of the LIM is the reduced cost of the guideway and the ability to operate closely spaced vehicles without the block limitation of the LSM. The LIM has been used for propelling both wheel based vehicles and magnetically suspended vehicles over a very wide range of speeds.

The principal disadvantage of the short stator design is the need to transfer power to the vehicle. If this is done with sliding contacts, as with conventional electric trains, then some of the advantages of maglev are lost. If it is done via inductive transfer, then the guideway cost is increased.

A second disadvantage is the need to have substantial amounts of power conditioning on the vehicle. This equipment must be sized to match the highest force and speed for the entire system, even if it rarely needs such high peak power. If the air gap is large this power conditioning equipment can be heavy and expensive.

Note that the popular press often misuses terms like "induction" and "synchronous" when applied to linear motors.

RECENT SYSTEMS

England

A low speed EMS maglev system was constructed in Birmingham in 1984. This design had an air gap of 10 mm and used short stator LIM propulsion with power delivered to the vehicle via sliding contacts and a third rail. It provided transportation between an airport and a train station and worked nearly flawlessly for more than twelve years. When problems did develop there was no one with interest and ability to repair it, so it was removed.

Japan

The Japanese constructed a high-speed maglev system in Yamanashi Prefecture using superconducting technology. This was the successor to several smaller and lower speed vehicles that had been tested in Miyazaki Prefecture starting in 1980. The Yamanashi system uses the null-flux version of EDS and a long stator LSM for propulsion. The latest test facility consists of a two-way, 42.8 km (27 mi.) long guideway, most of it in tunnels. In April 1999 a manned five-car train, MLX01, achieved a speed of 552 km/h (343 mph). In November 1999 two trains passed each other at a relative speed of 1,003 km/h (623 mph). This system shows the potential of EDS, but is relatively expensive to construct and operate and it is questionable whether it will ever become a commercially viable design.

Germany

The Germans constructed a 31.5-km (19.6-mile) test loop in Emsland, starting in 1974. This was the culmination of a research effort started in 1970 and involving experiments with both EDS and EMS and with various types of linear motors. The latest design, called Transrapid 08, has an EMS with 9 to 10 mm (3/8 in.) air gap and LSM propulsion using the same magnets for both suspension and propulsion. It achieved a top speed of 450 km/h (280 mph) in June 1993. Although the air gap is relatively small, the success of the Emsland tests makes it clear that this technology is usable for higher speeds than are ever likely to be commercially feasible with conventional railroad technology. A comparison of Transrapid with the Germany ICE high speed train indicates that maglev consumes 30 percent less energy and the differences are much larger when comparing maglev with automobiles and commercial aviation. There are plans to construct a 292-km (181-mi.) operational system between Hamburg and Berlin. The design calls for five stations and a travel time of less than one hour. The construction cost is estimated to be about the same as for a new ICE high speed rail line, but there is still some doubt whether this project will ever be completed. There are active proposals to construct Transrapid systems in the United States but it is too early to tell whether such a system will ever be built.

A second German effort, designated HSST, has been underway since 1974. This is a lower speed EMS design and uses a LIM with power transfer via sliding contacts. This system has been demonstrated on several occasions and there are pending plans for implementation, but past plans have never been carried out and the future is uncertain.

United States

In 1966 Danby and Powell at Brookhaven National Laboratory conceived of using vehicle-mounted superconducting magnets that induced current into their novel “null-flux” winding on the guideway. This concept helped precipitate maglev research efforts in the early 1970s, but these were soon terminated by political action. In 1987 Senator Patrick Moynihan created the U.S. National Maglev Initiative that was managed by the Federal Railway Administration. It led to four preliminary designs, but the effort was terminated by political action. In the late 1990s a maglev effort was reinitiated, but the combination of political and technical dispute makes the future uncertain.

There is an effort by the U.S. Federal Transit Administration to develop an Urban Maglev system that can be an alternative to urban rail-based transit systems. The objective is to achieve speeds up to 161 km/h (100 mph) with lower noise, lower energy consumption and lower operating cost that any other fixed guideway system. Based on the success of the Birmingham, England system it is likely that a viable urban system can be built. If the lower speed design was well conceived it could evolve into a design suitable for operation at higher speeds, just as the railroad system evolved.

Web Sites for More Information

The best and most up-to-date references are Internet websites. Most of these can be reached via links from the U.S. Department of Transportation site: <http://www.dot.gov>. This site has links to the Federal Railroad Administration and the Federal Transit Administration, each of which has links to other national and international maglev sites. Additional maglev sites can be reached from the *Innovative Transportation Technologies* site: <http://faculty.washington.edu/~jbs/itrans/maglevq.htm>. More details on EMS, LSM, and LIM can be found on the German *Transrapid* and Japanese *HSST* sites. More details on EDS and LSM can be found on the Japanese *Railroad Technical Research Institute* (RTRI) site.

A word of caution: Any publication that claims major advantages for a particular breed of suspension or propulsion should be viewed with skepticism.

PROS AND CONS OF MAGLEV

Advantages

Following are advantages that have been proven by actual tests on working system.

- *More efficient.* Maglev vehicles can operate at a given speed with less energy consumption than for almost any other transportation mode, particularly for short to medium distances. This is because the vehicles can be lighter and more streamlined than is possible for any wheel-based design and commercial aviation is very inefficient for short trips. It has frequently been proposed that the vehicles can operate at very high speeds in a partially evacuated tunnel with dramatic savings in energy and time, and several studies have shown that this is not as far-fetched as it might sound.

- *Less noisy.* Maglev vehicles are quiet. Transrapid test data shows 5 to 10 dB less noise than for a train at the same speed or the same noise for speeds that are 100 km/h (62 mph) higher. Some newer maglev designs with more streamlining have even lower noise.
- *Safer.* The system can be as safe or safer than any other system. If it is well designed with dedicated rights of way, a high level of automation and no physical contact, then the most common cause of accidents will have been eliminated. There have been no significant accidents in all of the maglev tests that have been run. If maglev attracts people away from other modes it could save many lives.
- *Environmentally friendly.* The reduced noise and lack of air pollution is environmentally desirable, and so also is the ability to climb steep grades and operate in smaller diameter tunnels. This combination is unbeatable by any other mode.
- *Faster.* The system can be designed to operate safely at very high speeds. But speed is relative so that 161 km/h (100 mph) operation in an urban area will be dramatically faster than alternatives. Some people believe that maglev designers have put too much emphasis on speed and this has made the designs expensive.
- *Reduced operating cost.* The elimination of bearings and wheel contact will reduce the maintenance time and cost. This is borne out by the high reliability of the Birmingham, England maglev shuttle and the reliability of the Transrapid test system in Emsland, Germany. The lower energy consumption and potential for complete automation creates additional savings.

Disadvantages and Problems

There are several disadvantages and problems that must be solved:

- There is a perception that maglev is too expensive. Part of the problem is a focus on very high speeds where any system would be expensive, if not impossible. For many new construction projects a well designed maglev system need be no more expensive than wheel-based designs, and lower maintenance and energy cost will provide additional savings. But this potential cost advantage must be proven before it is can be used to sell maglev.
- Transportation planners and managers can point to financial disasters in the early application of new technology. They feel, with justification, that no commitment for a commercial system should be made until there is a good demonstration system that proves the technology.
- There are many barriers to innovation in the transportation sector. Large corporations dominate the design and construction of guided systems and they have little incentive to change. They could lose by either failing to deliver a viable system or because another company is stronger in the new technology. There are many government regulations that pertain to any transportation system, and political involvement in transportation decision making tends to favor the status quo.
- There are too many inventors and little cooperation in finding the best combination of suspension, guidance, propulsion, and control. A national competition would be very helpful, such as the locomotive competition in 1829 when George Stephenson's *Rocket* beat several competing designs by hauling a coach of passengers at 39 km/h (24 mph).

PROGNOSIS FOR THE FUTURE

Future success will require a significant group to make a long-term commitment to develop an integrated maglev system. There should be a reduced emphasis on new inventions and increased attention to the myriad problems of integrating suspension, propulsion, guidance, and control. Some group must be willing to fund the development to the point that a potential buyer of a maglev system can clearly see that the advantages outweigh the disadvantages. The probability of this happening is good; the major question is "When, where, and how will maglev reach commercial fruition?"

Richard D. Thornton

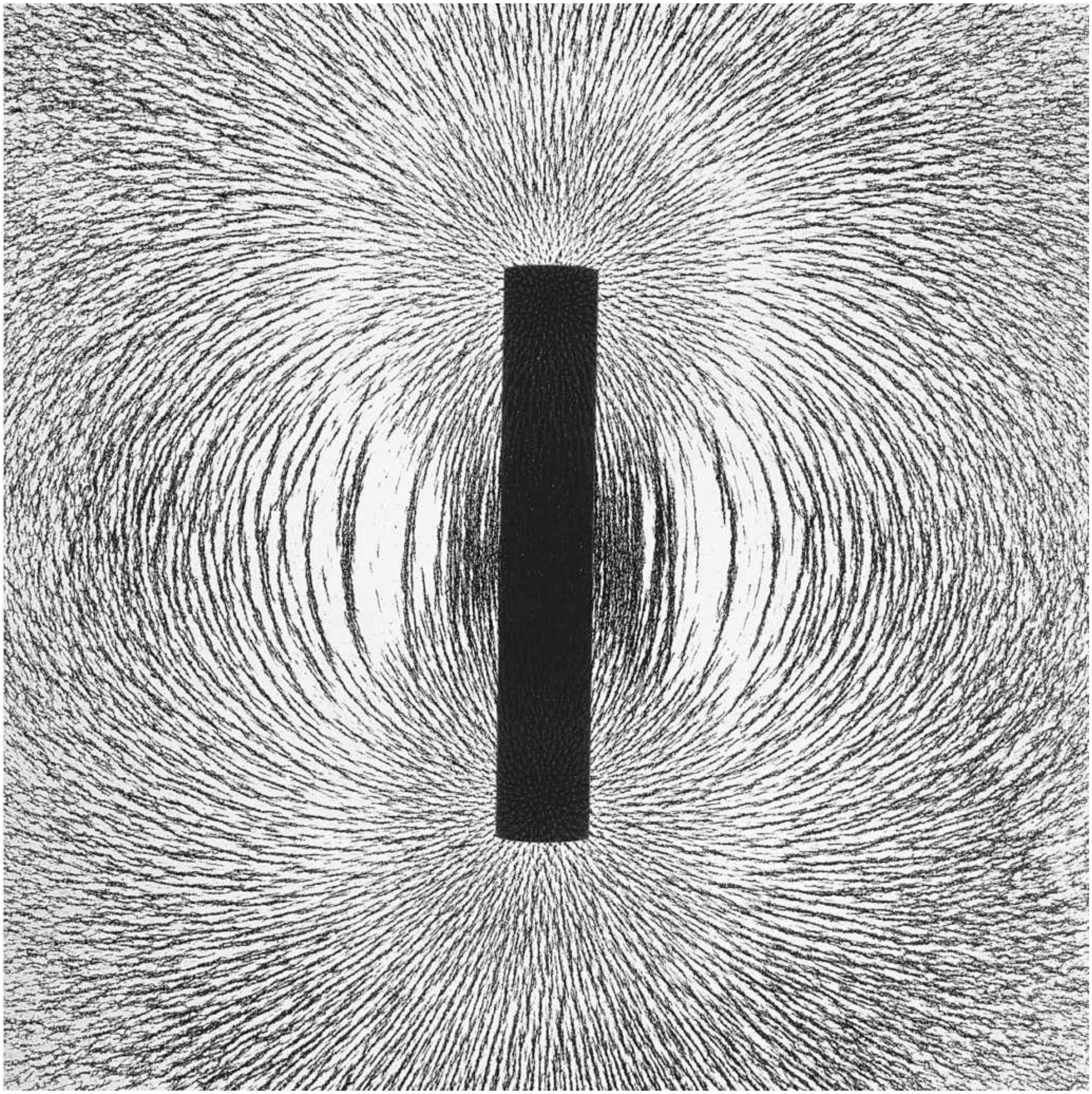
See also: Railway Passenger Service.

MAGNETIC FLIGHT

See: Magnetic Levitation

MAGNETISM AND MAGNETS

Magnetism is the phenomenon in which iron is attracted to a natural material called lodestone, the properties of which are similar to a magnet.



Metal filings form a magnetic field pattern around the poles of a magnet. (Corbis-Bettmann)

LODESTONE AND MAGNETS

Lodestone is a crystalline oxide of iron called magnetite. Until the 1800s, lodestones and the earth were the only sources of magnetism. Iron-based materials are attracted to a lodestone as well as to any other magnet. It is the attraction between a magnet and the iron in a refrigerator door that pins a photograph to the door. Bringing an iron-based material in contact with a magnet or lodestone will make a magnet of the material. Unfold a paperclip and rub it with a magnet. The clip becomes magnetized, with the magnetic properties concentrated near the ends of the clip. These end regions are called poles. Magnetize two unfolded paperclips and, in one parallel orientation, the ends or poles attract each other. Change the orientation of one clip and the poles repel each other. Attach a thread to the middle of one of the clips, suspend it, and one pole will point in the general direction of geographic north. The one pointing north is called an N-pole; the opposite pole is an S-pole. We observe, and say, “unlike poles attract, like poles repel.” It is an instructive experiment to cut the magnetized clip into two pieces in an effort to isolate a pole. Interestingly, each of the two pieces has an N-pole and an S-pole. Regardless of how many times a clip is cut in two, magnetic poles always occur in pairs of N-poles and S-poles.

MAGNETIC FIELD

A magnet does not materially change the space around it. Yet, if the magnet were not there, another magnet would not experience a force when brought into the space. Magnetically, the space around the magnet is altered and the modification is thought of as producing a magnetic field. When another magnet is brought into the magnetic field and experiences a force, the magnetic field is the mechanism for exerting the force. The concept of a field applies to gravitational and electric forces as well, and is an extremely important aspect of many energy applications.

Magnetic fields are measured in units of teslas, symbol T. A strong permanent magnet, such as might be found in a physics laboratory, produces a magnetic field of about 0.3 T. Such a magnet is capable of lifting several kilograms of iron. For comparison, the magnetic fields produced by other systems include

10^{8} T	neutron stars
10^{3} T	short bursts of electric current
10^{1} T	strong laboratory superconducting magne
10^{-5} T	Earth’s magnetic field
10^{-9} T	interplanetary magnetic fields
10^{-12} T	magnetic field associated with the human body

In 1819, Hans Christian Oersted, professor of physics at Copenhagen University, discovered that a magnet experiences a force when in the vicinity of a wire carrying an electric current. The fact that the magnet experiences a force is evidence that the electric current produces a magnetic field, which eventually led to the development of innumerable devices—electric motors, electric generators, speakers for hi-fidelity amplifiers, and electromagnets, to name a few—based on this principle.

ENERGY APPLICATIONS

Electric motors

Electric motors are found in nearly every room of a typical house. They power everything from the washing machine, refrigerator, and vacuum cleaner, to the hair dryer, fan, garage door opener, and disk drive in a computer. In an automobile, motors adjust seats, raise an antenna, operate windshield wipers, adjust mirrors, run fans, start the engine, and someday may replace the internal combustion engine under the hood. There is probably no device more useful for doing work than electric motors, and their pervasiveness will surely grow.

Fundamentally, an electric motor converts electric energy to rotational energy. Rotation results from magnetic forces between a rotating part (the rotor) and a stationary part (the stator). There are many designs. In the simplest, the rotor is an electromagnet that rotates between the poles of a permanent magnet (Figure 1). The N-poles and S-poles of the electromagnet are determined by the direction of current flow. In the illustration, attraction and repulsion between poles on the rotor and stator cause the electromagnet to rotate clockwise. When the unlike poles approach each other, the direction of current flow is reversed, causing the poles on the electromagnet to change. The alternating magnetic attraction and repulsion between poles keeps the electromagnet rotating in the same direction.

Electric Generators

An electric generator for operating lights is a common sight on many bicycles. The generator has a coil of wire that rotates between the poles of a magnet and looks very much like a motor. Whereas a motor converts electric energy into rotational energy, a generator converts mechanical energy to electric energy. On a bicycle, the tire rubs against a wheel attached to the rotating coil. Some agent, in this case the cyclist pedaling, does the work to turn the coil with the reward being electric energy. In a large electric power plant an electric generator working on the same physical principle is driven by a large steam turbine.

Electric generators are based on the principle that an electric charge experiences a force when it moves in a magnetic field. Electrons in the metallic wires of

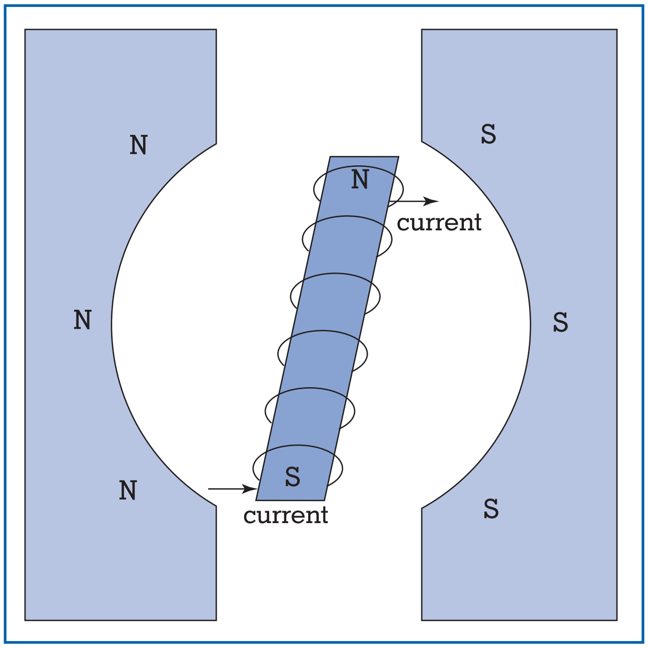


Figure 1.
An elementary motor consists of a rotating electromagnet and a fixed permanent magnet.

the rotor of a generator experience a force when the coil is rotated in a magnetic field. An electric current is produced in a light bulb, for example, when it is connected to the open ends of the wires making up the coils of the rotor.

Magnetic Levitation of Vehicles

Whereas electric motors utilize both attraction and repulsion of magnetic fields, there are energy applications that rely only on attraction or repulsion. Magnetic levitation of vehicles is a good example. A vehicle riding on a track or roadway experiences frictional forces that oppose the movement. Any scheme that can reduce the frictional forces offers improved energy economy. Magnetic levitation involves magnetic forces that hold a vehicle above a roadbed so that the vehicle appears to float on a cushion of air. It does not. It floats on a "magnet" cushion. One method of suspension capitalizes on the idea that like poles repel. The poles are produced by electromagnets rather than using permanent magnets. Another scheme is based on the principle that a metal experiences a force when in the magnetic field of an electromagnet that is energized by an electric current that changes rapidly with time.

Once a vehicle is levitated, it is not in material contact with a track, so cannot be propelled by wheels. The propulsion system, like the levitating system, is based on magnetic principles that are identical to the principles in an ordinary electric motor. The road bed is appropriately configured so that the magnetic field exerts a force on a current-carrying coil secured to the vehicle. In a real sense, the vehicle and magnetic road bed constitute a motor, albeit a linear motor (as opposed to the rotational motor shown in Figure 1). This is called a linear indication motor (LIM). The National Aviation and Space Administration (NASA) believes that this scheme could also be used to launch spacecraft into orbit. A magnetically levitated space vehicle accelerated to a speed of about 600 miles/hour would be catapulted from the ground. Once aloft, a rocket engine would take over and propel the spacecraft into orbit.

Magnetic controls

A magnetic field due to an electric current can be turned on and off simply by turning the current on and off. A piece of iron attached to the end of a spring having the other end fixed can be moved with a magnetic field and returned to its initial position by the spring. The iron piece can then be used to actuate a switch or move a lever on a valve. Applications of this principle include electrically controlled valves in a washing machine and an electrically controlled switch for the starter in an automobile.

Superconducting Magnets

The magnetic field produced by an electric current is proportional to the current; doubling the current doubles the magnetic field. It would seem that an experimenter could achieve any desired magnetic field by creating the necessary electric current in a coil of wire. But wires like those in an electric toaster offer resistance to electric currents, resulting in the production of heat. If the heat is not removed, the wires will melt. Solving this problem is the driving force behind the effort to develop superconducting materials that offer zero electrical resistance to electric current.

Until 1987, achieving zero resistance required cooling electrical conductors to around the temperature of liquid helium (4.2 K or -268.8°C). Nevertheless, practical electromagnets using superconducting wires cooled to around 4 K have been used in the laboratory for several decades. A new class of superconducting materials requiring cooling to around 77 K (-196ºC) was discovered in 1987. These materials are difficult to fabricate, but are very attractive because of their substantially higher superconducting temperature. Superconducting materials could be used for wires for transmission lines bringing electricity from an electric power plant to a city. Electromagnets made from superconducting wires can carry much more electric current, and so generate much stronger magnetic fields.

Magnetic Materials

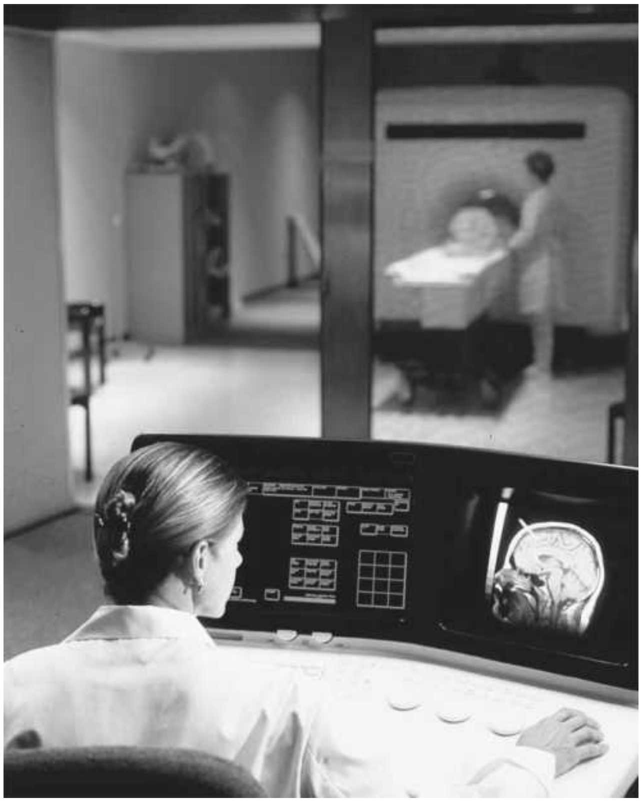
Because magnetic poles occur in N and S pairs, a magnet is referred to as a magnetic dipole. The net magnetic field produced by two magnetic dipoles depends on their orientations. If the N-poles of each point in opposite direction, their contributions to the net magnetic field tend to cancel. If the N-poles of each point in the same direction, their contributions to the net magnetic field tend to add, and the net magnetic field is larger than that of one alone. The more magnets there are with their N-poles pointing in the same direction, the greater is the net magnetic field. In a real sense, iron atoms are like tiny magnetic dipoles. In an unmagnetized piece of iron, the tiny atomic magnetic dipoles are randomly oriented and there is no net magnetic field produced by the iron. But if the dipoles are given a preferred orientation, a net magnetic field around the iron results. The preferred orientation can be achieved by putting the iron in a magnetic field caused by a magnet or an electric current. This is a way of producing a permanent magnet.

Magnetic Recording

The binary system of numbers requires only two numbers, usually designated 0 and 1. Any device having two distinct states can be used to represent the two numbers. A finger pointed up could represent 1; pointing down it could represent 0. Zero and one can be represented with the N-poles of a magnetic dipole pointing in opposite directions. Information is recorded on a floppy disk or magnetic tape in binary fashion. The disk or tape is coated with a magnetizable material. Tiny local areas are magnetized with either the N-pole up or down to represent 0 or 1 . The device that reads the information detects the orientations and translates the information.

Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is a revolutionary diagnostic tool for producing images of the interior of a human body. It works because a proton, the sole constituent of a hydrogen atom, behaves like a tiny magnetic dipole. There is a copious source of hydrogen atoms in the body because many components contain water, and every water molecule has two hydrogen atoms. Normally, the atomic dipoles are oriented randomly, but they align when in a strong magnetic field. When stimulated by a short burst of radio frequency electromagnetic radiation, the atomic dipoles are deflected away from the direction of the magnetic field. Following the short burst of radiation, the dipoles rotate (precess) in ever-decreasing circles around the direction of the magnetic field, and in doing so emit a detectable signal before returning to a state of equilibrium. The detected signal can be converted into an image using computer technology. The technique has revolutionized the diagnosis of problems associated with muscles and joints in the human body.


Technicians administer an MRI. (Photo Researchers, Inc.)

Joseph Priest

See also: Oersted, Hans Christian.

BIBLIOGRAPHY

Elster, A. D. (1986). *Magnetic Resonance Imaging: A Reference Guide and Atlas*. Philadelphia: Lippincott.

Feynman, R. P. (1985). *QED: The Strange Theory of Light and Matter*. Princeton, NJ: Princeton University Press.

Livington, J. D. (1996). *Driving Force: The Natural Magic of Magnets*. Cambridge, MA: Harvard University Press.

Macaulay, D. (1988). *The Way Things Work*. Boston: Houghton Mifflin.

Vranich, J. (1991). *Supertrains: Solutions to America's Transportation Gridlock*. New York: St. Martin's Press.

MAGNETOHYDRODYNAMICS

Magnetohydrodynamics (MHD) is a promising technology for electric bulk power generation. MHD is accomplished by forcing an electrically conducting fluid or a plasma through a channel with a magnetic field applied across it and electrodes placed at right angles to flow and field (Figure 1). An MHD plant can be directly fired with coal and there are no moving parts. To achieve extra high efficiencies, MHD is combined in a binary thermodynamic cycle with a conventional steam plant to add an extra 40 percent to the total power output and to boost the overall combined efficiency into the 60 percent range. The high temperature MHD process extracts part of the heat energy in the plasma at the high temperature end. The gas leaving the MHD generator, still at relatively high temperatures, is then used in a conventional bottoming steam plant.

A schematic coalfired MHD generator in Figure 2 is shown using a combustion plasma at up to 3,000 K, seeded with alkali salts for high conductivity, fed directly into the generator with a superconducting magnet providing up to 7 tesla. The combustion products, still at about 2,000 K, then pass through an air preheater, where the hot gas exhausted from the MHD generator preheats the air for the combustor leading to the high temperatures required to create the plasma.

To use lower gas temperatures (under 1,500 K), noble gases like argon and helium offer very high electrical conductivities. Recycling the noble gas leads to the "closed cycle" MHD process.

Nuclear reactors have been used for small generators providing the high temperature and pressure plasma to drive an MHD process, and a nuclear plant could offer a considerable reduction in pollution because of high overall efficiency without CO_2 emission.

EXPERIMENTAL DEVELOPMENTS

In the early part of the nineteenth century Michael Faraday (1832) conducted MHD experiments using

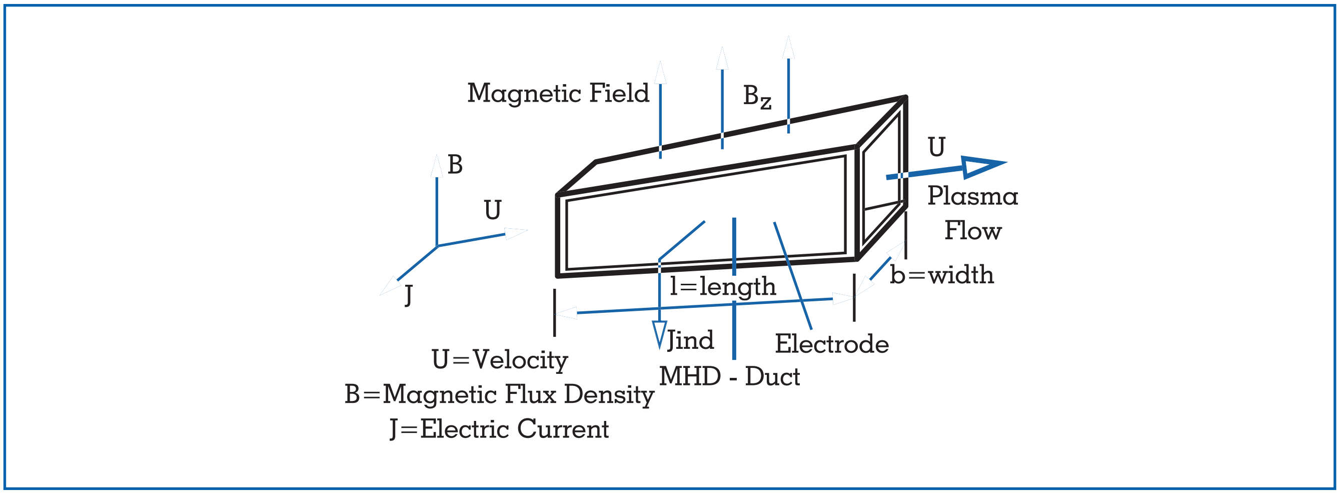


Figure 1.
Linear MHD Generator Channel.

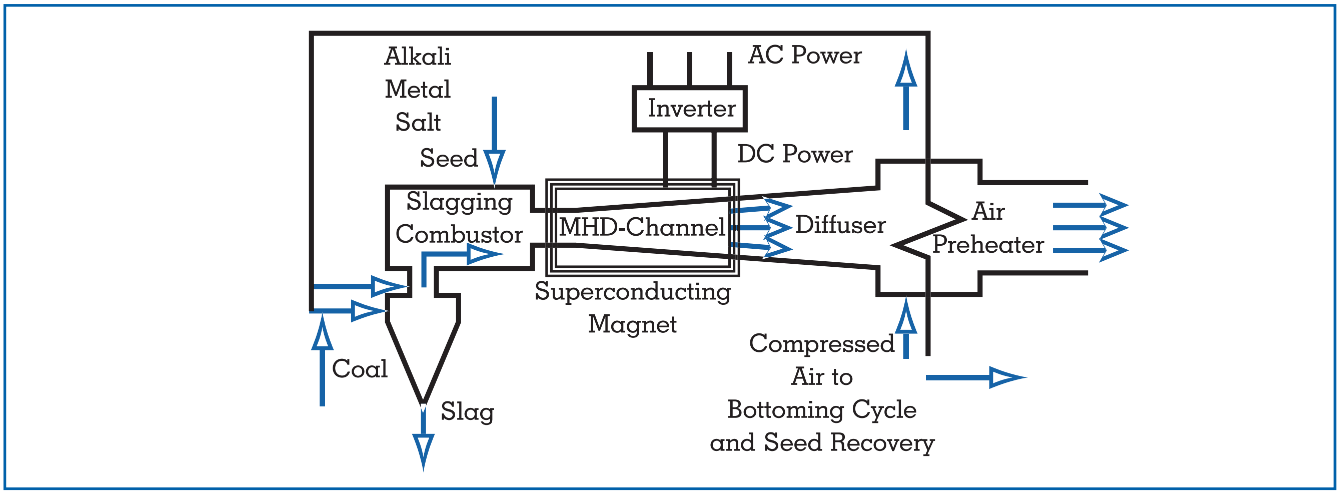


Figure 2.
MHD Linear Generator Flow Chart.

the brackish water of the river Thames flowing through the Earth's magnetic field. The first successful power generation experiment, developed by Richard Rosa in 1959, generated 10 kW with a timber walled channel on the AVCO "Mark 1" facility in Boston, Massachusetts. This success and the possibility of cheap MHD power led in the 1960s to national programs in Britain, the Soviet Union, The Netherlands, France, Germany, Poland, Italy, India, Australia and Israel. In 1965 the AVCO "Mark 5" generator successfully generated 32 MW over a one minute run using alcohol at 45 kg/sec fired with oxygen. AVCO later developed a sophisticated coal fired MHD channel for a 2,000 hour test program and demonstrated technical feasibility under the most stringent conditions.

In 1972 in Moscow, a large experimental facility, the "U-25," used a 250 MW natural gas combustor and generated 20 MW. The Soviets have been using very successfully mobile, pulsed MHD generators throughout the Soviet Union, for seismic studies.

MHD programs in the United States are concentrated in two major facilities. A "Component Development and Integration Facility" is located in Butte, Montana, and a "Coal Fired Flow Facility" at the University of Tennessee to studies coal fired MHD, slag processing, seed handling and downstream systems.

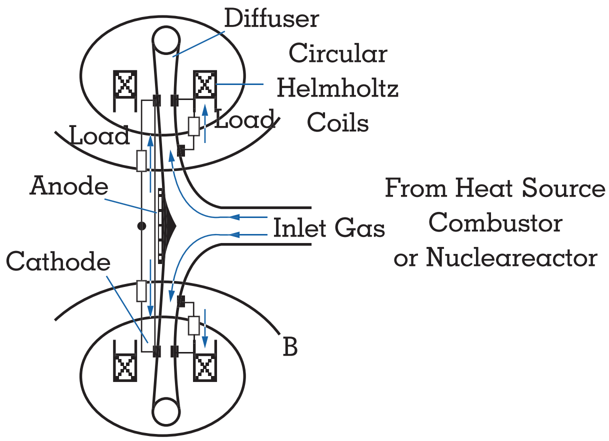


Figure 3.
MHD Disc Generator.

A circular disk MHD generator geometry requires a lower cost circular ring magnet structure as shown in Figure 3. Work in the United States using shock tunnels with large discs proved the feasibility of very high enthalpy extraction. Continuous operation on a fossil fuel fired disk facility was demonstrated at the University of Sydney, Australia. The potential advantages offered by closed cycle disc systems was made evident in the 1980s and 1990s by the Japanese MHD research effort and by the Dutch team at the University of Eindhoven.

A great deal of work has been done in the United States, Russia, Israel, and France showing that low temperature operation is possible using liquid metal as the MHD driver. A liquid metal MHD-generator can be very much smaller because of the much higher conductivity of liquid metal.

POWER SYSTEM REQUIREMENTS

Studies carried out in the United States, Russia, and Japan indicate that a combined cycle MHD-steam plant should be able to achieve an overall power station efficiency of at least 60 percent which is about 20 percent more than offered by a conventional steam plant. This should be possible at capital costs comparable with existing steam plants.

As shown in Figure 4, the operating temperature range for MHD is beyond that of any other generating technology. MHD could still add up to 15 percent at the top end of other combined cycles. MHD is potentially a natural choice for conversion of high temperature energy output from future nuclear power plants whether fission or fusion driven. A combined Nuclear-MHD system design with high efficiency does offer two advantages: (1) a reduction in thermal pollution and (2) no CO_2 emission as with fossil fuel-driven plants.

MHD efforts in many countries, including the United States, have declined substantially, in Russia because of lack of funds and in general because the high costs envisioned in setting up a full-scale power station. If funds become available to set up full-scale power stations and with the advent of high temperature nuclear heat sources, the many advantages provided by MHD may be realised.

Hugh Karl Messerle

See also: Combustion; Electric Power, Generation of Faraday, Michael; Hydroelectric Energy; Magnetism and Magnets; Nuclear Energy; Thermodynamics.

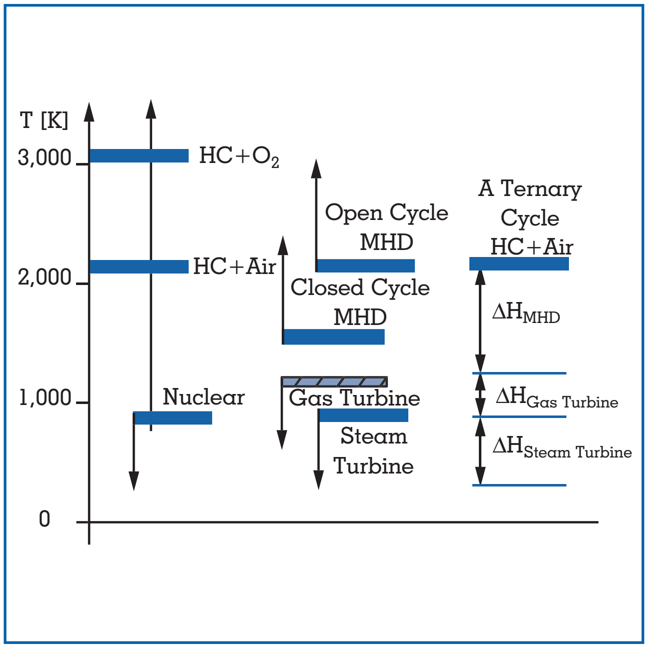


Figure 4.
Temperature Ranges for MHD, Gas Turbine, and Steam Turbine Plants.

BIBLIOGRAPHY

Faraday, M. (1832). “Experimental Researches in Electricity.” *First Series, Philosophical Transactions of the Royal Society*, pp. 125-162.

Messerle, H. K. (1995). *Magneto-Hydro-Dynamic Electrical Power Generation*. Chichester, England: John Wiley & Sons Ltd.

Petrick, M., and Shumyatsky, B. Y. (1978). *Open-Cycle Magnetohydrodynamic Electrical Power Generation.* Argonne, IL: Argonne National Laboratory.

Rosa, R. J. (1960). “Experimental Magnetohydrodynamic Power Generation.” *Applied Physics* 31:735-36.

Rosa, R. J. (1987). *Magnetohydrodynamic Energy Conversion.* Washington, DC: Hemisphere Publishing Corp.

Shioda, S. (1991). “Results of Feasibility Studies on Closed Cycle MHD Power Plants.” *Proceedings Plasma Technology Conference*, Sydney, Australia, pp. 189-200.

Simpson, S. W.; Marty, S. M.; and Messerle, H. K. (1989). “Open-Cycle Disk Generators: Laboratory Experiments and Predictions for Base-Load Operation.” *MHD An International Journal* 2(1):57-63.

Veilkov, E. P.; Zhukov, B. P.; Scheindlin, A. E.; Vengerskii, V. V.; Shelkov, E. M.; Babakov, Y. P.; Volkov, Y. M.; Zeigarnik, V. A.; Mtveenko, O. G.; and Kolyadin, N. M.

(1983). "Status and Prospects of Geophysical MHD Power." *Proceedings of the 8th International Conference on MHD Electric Power Generation, Moscow.*

Most of the work carried out in MHD power generation is published in the Proceedings of the eleven "International Conferences on MHD Electrical Power Generation" organized by the International Liaison Group on MHD Electrical Power Generation (ILG-MHD) sponsored by UNESCO and originally by IAEA and ENEA, in the annual series of the "Symposia on Engineering Aspects of Magnetohydrodynamics" (SEAM) in the United States and in the Japanese series of CIS MHD Symposia.

MANUFACTURING

Industrial production is the backbone for economic output in almost all countries. Over the past decades, manufacturing industrial production has been growing in most economies. The industrial sector is dominated by the production of a few major energy-intensive commodities, such as steel, paper, cement, and chemicals. In any given country or region, production of these basic commodities follows the general development of the overall economy. Rapidly industrializing countries will have higher demands for infrastructure materials, and more mature markets will have declining or stable consumption levels. The regional differences in consumption patterns (expressed as consumption per capita) will fuel a further growth of consumption in developing countries. In addition to labor costs and costs for raw materials in these "heavy" industries, energy is a very important production cost factor in the effort to achieve higher energy efficiency (see Table 1). Markets in the industrialized countries show a shift toward more service-oriented activities and hence non energy-intensive industries. Because of the great difference in energy intensity between energy intensive industries and all others, changes in output shares of these industries can have a major impact on total industrial energy use. Many commodities (e.g., food and steel) are traded globally, and regional differences in supply and demand will influence total industrial energy use. Production trends also depend on regional availability of resources (e.g., scrap) and capital. Manufacturing energy use also will depend on the energy efficiency with which the economic activities are done. In this article we will assess trends in industrial energy use and energy intensities, followed by a discussion of energy services, uses, and industrial technologies.

GLOBAL MANUFACTURING ENERGY USE

In 1990, manufacturing industry accounted for 42 percent (129 exajoules, EJ) of global energy use. Between 1971 and 1990, industrial energy use grew at a rate of 2.1 percent per year, slightly less than the world total energy demand growth of 2.5 percent per year. This growth rate has slowed in recent years and was virtually flat between 1990 and 1995, primarily because of declines in industrial output in the transitional economies of Eastern Europe and the former Soviet Union. Energy use in the industrial sector is dominated by the industrialized countries, which accounted for 42 percent of world industrial energy use in 1990. Industrial energy consumption in these countries increased at an average rate of 0.6 per year between 1971 and 1990, from 49 EJ to 54 EJ. The share of industrial sector energy consumption within the industrialized countries declined from 40 percent in 1971 to 33 percent in 1995. This decline partly reflects the transition toward a less energy-intensive manufacturing base, as well as continued growth in transportation demand, resulting in large part from the rising importance of personal mobility in passenger transport use.

The industrial sector dominates in the economies in transition, accounting for more than 50 percent of total primary energy demand, the result of the emphasis on materials production, a long-term policy promoted under years of central planning. Average annual growth in industrial energy use in this region was 2.0 percent between 1971 and 1990 (from 26 EJ to 38 EJ), but dropped by an average of -7.3 percent per year between 1990 and 1995.

In the Asian developing countries, industrial energy use grew rapidly between 1971 and 1995, with an annual average growth rate of 5.9, jumping from 9 EJ. It also accounted for the greatest share of primary energy consumption, between 57 percent and 60 percent. The fastest growth in this sector was in China and in other rapidly developing Asian countries. Growth in other developing countries was slightly lower.

The nature and evolution of the industrial sector vary considerably among developing countries. Some economies that are experiencing continued expan-

Sector	1973 Energy Intensity (primary energy)	1973 Energy Costs (share of production costs)	1985 Energy Intensity (primary energy)	1985 Energy Costs (share of production costs)	1994 Energy Intensity (primary energy)	1994 Energy Costs (share of production costs)
Iron & Steel	30.5 GJ/t	7%	27.8 GJ/t	11%	25.4 GJ/t	8%
Pulp & Paper	43.1 GJ/t	6%	42.7 GJ/t	6%	32.8 GJ/t	6%
Cement	7.3 GJ/t	40%	5.2 GJ/t	36%	5.4 GJ/t	33%
Primary Aluminum	N/A	14%	17.6 MWh/t	19%	16.2 MWh/t	13%
Petroleum Refining	6.2 GJ/t	4%	4.3 GJ/t	3%	4.5 GJ/t	3%

Table 1.
Energy Intensities and Energy Costs in Selected U.S. Industries.
NOTE: Energy intensity is expressed in primary energy, where the efficiency of electricity generation is assumed to be 33 percent. Energy intensity of primary aluminum production is given in MWh (1000 kWh).
SOURCE: Lawrence Berkeley National Laboratory, Berkeley, California.

sion in energy-intensive industry, such as China and India, show relatively unchanging shares of industrial energy use. In other countries, such as Thailand and Mexico, the share and/or growth of the transportation sector dominate. Many smaller countries have remained primarily agrarian societies with modest manufacturing infrastructure.

ENERGY INTENSITY TRENDS

In aggregate terms, studies have shown that technical efficiency improvements of 1 to 2 percent per year has been observed in the industrial sector in the past, Between 1975 and 1983 during and after the years of major oil price increases, U.S. industrial energy intensity declined by 3.5 percent per year. Between 1984 and 1994 industrial energy intensity declined by less than 1 percent on average (Brown et al., 1998). Figure 1 gives an overview of energy of economic intensity trends in the industrial sector in industrialized countries.

The trends demonstrate the capability of industry to improve energy efficiency when it has the incentive to do so. Energy requirements can be cut by new process development. In addition, the amount of raw materials demanded by a society tends to decline as countries reach certain stages of industrial development, which leads to a decrease in industrial energy use. The accounting of trends in structural shift, material intensity, and technical energy efficiency and their interactions can be extremely difficult. To understand trends in energy intensity it is important to analyze the structure of the industrial sector. Industrial energy use can be broken down into that of the energy-intensive industries (e.g., primary metals, pulp and paper, primary chemicals, oil refining, building materials) and the nonenergy intensive-industries (e.g., electronics and food). Reduction of energy intensity is closely linked to the definition of structure, structural change, and efficiency improvement. Decomposition analysis is used to distinguish the effects of structural change and efficiency improvement. Structural change can be broken down into intra-sectoral (e.g., a shift toward more recycled steel) and intersectoral (e.g., a shift from steel to aluminum within the basic metals industry). A wide body of literature describes decomposition analyses and explains the trends in energy intensities and efficiency improvement. Decomposition analyses of the aggregate manufacturing sector exist mainly for industrialized Western countries, but also for China; Taiwan; and selected other countries, including those in Eastern Europe. The results show that different patterns exist for various countries which may be due to specific conditions as well as differences in driving forces such as energy prices and other policies in these countries. More detailed analyses on the subsector level are needed to understand these trends better. Changes in energy intensities also can be disaggregated into structural changes and efficiency

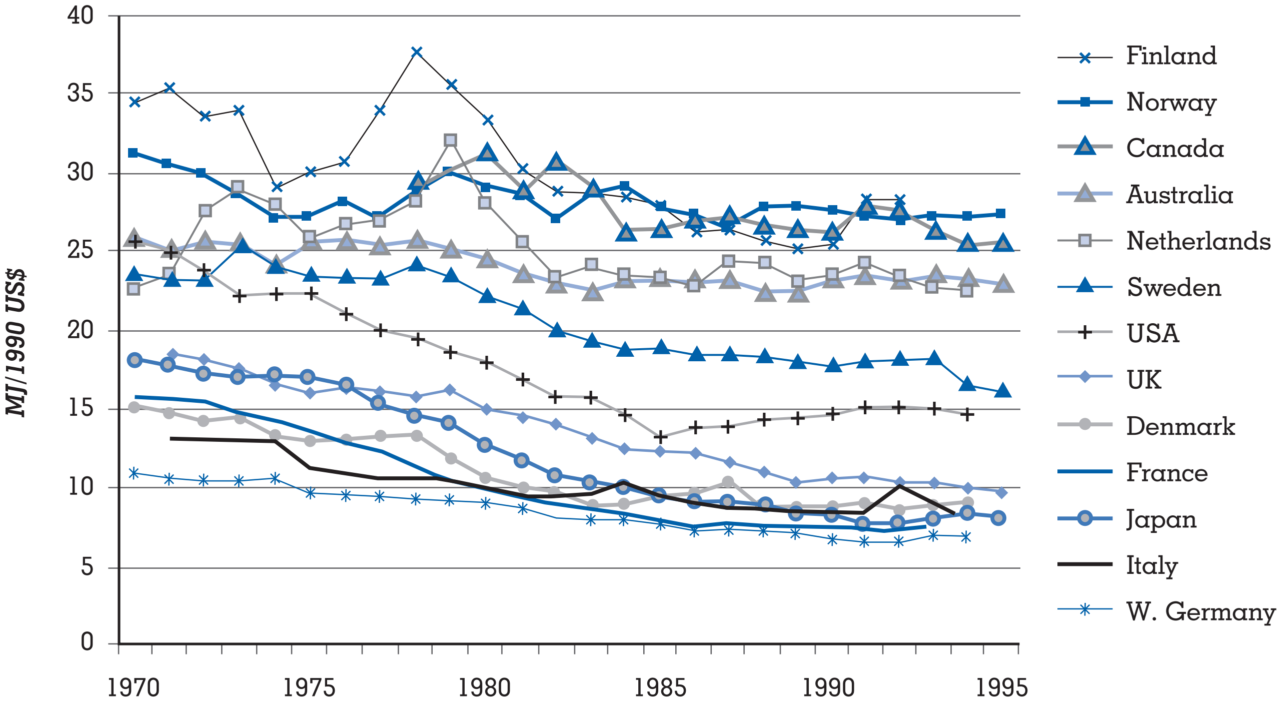


Figure 1.
Industrial sector economic energy intensity trends in selected industrialized countries, 1970–1995.
SOURCE: Lawrence Berkeley National Laboratory, Berkeley, California.

improvements at the subsector level. In the iron and steel industry, energy intensity is influenced by the raw materials used (i.e., iron ore, scrap) and the products produced (e.g., slabs, or thin rolled sheets). A recent study on the iron and steel industry used physical indicators for production to study trends in seven countries which together produced almost half of the world's steel. Figure 2 shows the trends in physical energy intensity in these countries, expressed as primary energy used per metric ton of crude steel. The large differences in intensity among the countries are shown, as well as the trends toward reduced intensity in most countries. Actual rates of energy efficiency improvement varied between 0.0 percent and 1.8 percent per year, while in the case of the restructuring economy of Poland, the energy intensity increased.

ENERGY SERVICES AND ENERGY EFFICIENCY

Energy is used to provide a service (e.g., a ton of steel) or to light a specified area. These services are called energy services. Energy efficiency improvement entails the provision of these services using less energy. About half of industrial energy use is for specific processes in the energy-intensive industries. On the other hand, various general energy conversion technologies and end uses can also be distinguished, such as steam production, motive power, and lighting. Hence, energy use in manufacturing can be broken down to various uses to provide a variety of services. A common breakdown distinguishes energy use for buildings, processes and utilities and boilers. The boilers provide steam and hot water to the processes, and the buildings. Due to the wide variety of indus-

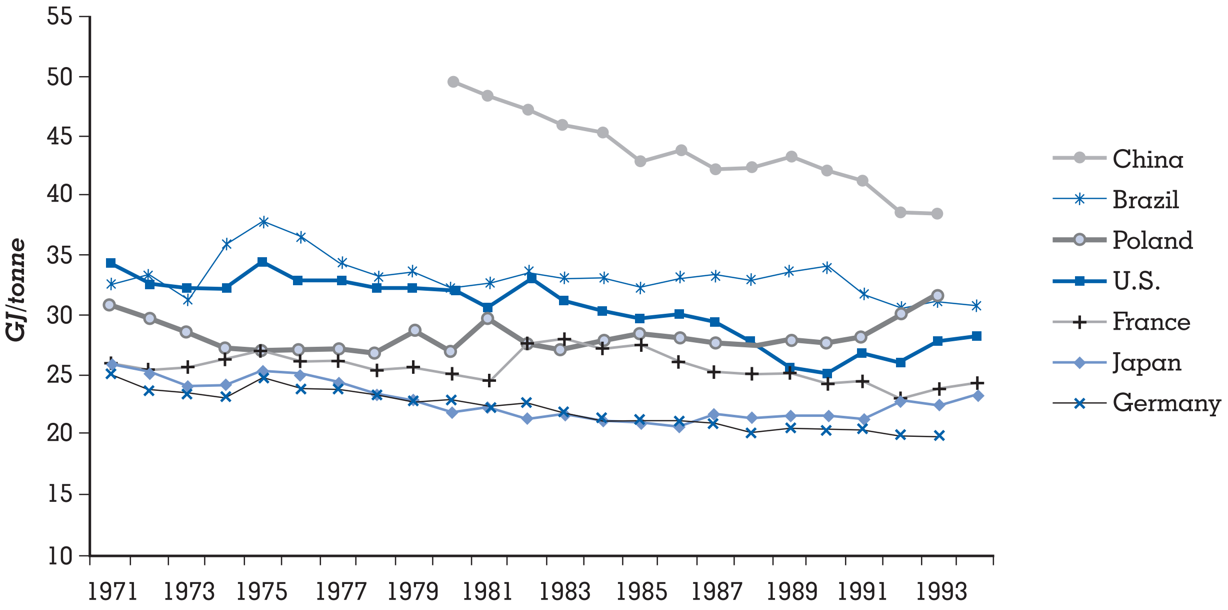


Figure 2.
Trends in physical energy intensity in seven countries between 1971 and 1994.
SOURCE: Lawrence Berkeley National Laboratory, Berkeley, California.

trial processes, we will limit our discussion to two energy intensive sectors—iron and steel, and the pulp and paper industries, as well as boilers, to illustrate important cross-cutting energy-consuming processes in industry.

Iron and Steel Industry

The first record of the use of iron goes back to 2500-2000 B.C.E., and the first deliberate production of iron began in about 1300 B.C.E. Small furnaces using charcoal were used. Evidence of such furnaces has been found in Africa, Asia, and Central Europe. The relatively low temperatures in the furnace lead to low-quality iron, and the slag had to be removed by hammering the iron. High-temperature processes started to be introduced in Germany in about 1300 C.E. The design of these furnaces is essentially the same of that of modern blast furnaces. The furnaces still used charcoal, and in 1718 the first use of coke is reported in the United Kingdom. The higher strength of coke allowed larger furnaces to be built, increasing energy efficiency. By 1790 coke iron making contributed to 90 percent of British iron production. Between 1760 and 1800, energy use declined by about 2 percent per year, mainly through the use of steam engines permitting higher blast pressures. During the nineteenth century, coke demand was further reduced by 1 percent per year. The development of the modern blast furnace after World War II resulted in an annual reduction of energy intensity of 3 to 4 percent a year, due to the use of improved raw materials, ore agglomeration, larger blast furnaces, and higher air temperatures. Today the blast furnace is the main process to make iron, and provides the largest raw material stream in steelmaking.

Steel is made by reducing the carbon content in the iron to levels below 2 percent. This reduces the brittleness of the material and makes it easier to shape. The first steelmaking process was invented in 1855 by Bessemer, and was in commercial operation by 1860. In the Bessemer converter air was blown through the hot iron, which oxidizes the carbon. This principle is still followed in modern steelmaking processes. In the United States the last Bessemer converter was retired in the 1960s. In the late nineteenth century the open-hearth furnace (OHF) or Siemens-Martin furnace was invented, which uses preheated air and fuels to oxidize the carbon and melt the steel. This process is currently found only in developing countries and in Eastern Europe. The United States was one of the industrialized countries that phased out OHF at a very late stage. In the 1980s the dominant process became the basic oxygen furnace (BOF), using pure oxygen instead of air. The BOF process was developed in Austria in the 1950s. The productivity of this process is much higher than that of OHF, as is the energy efficiency. An alternative process is the electric arc furnace (EAF). The EAF process is mainly used to melt scrap. Performance of EAFs has improved tremendously; fuel and oxygen are starting to be used besides electricity. In the future it is expected that the BOF and EAF processes will follow similar developmental paths.

Liquid steel is cast into ingots or slabs and shaped in rolling mills to the final product. Although most energy use is concentrated in ironmaking and steelmaking, reduced material losses and productivity gains in casting and shaping (e.g., continuous casting and, currently thin slab casting) have contributed to dramatic increases in the energy efficiency of steelmaking.

Today, the U.S. iron and steel industry is made up of integrated steel mills which produce pig iron from raw materials (iron ore, coke) using a blast furnace and steel using a BOF, and secondary steel mills (minumills), which produce steel from scrap steel, pig iron, or direct reduced iron (DRI) using an EAF. The majority of steel produced in the United States is from integrated steel mills, although the share of minimills is increasing, growing from 15 percent of production in 1970 to 40 percent in 1995. There were 142 operating steel plants in the United States in 1997, of which 20 were integrated steelmills and 122 were minimills. The integrated mills are most often located near or with easy access to the primary resources; for example, in the United States these are concentrated in the Great Lakes region, near supplies of coal and iron ore and near key customers such as automobile manufacturers.

The worldwide average energy intensity of steelmaking is estimated at gigajoules (GJ) per metric tone, although large variations occur among countries and plants (see Figure 2). Today the most energy-efficient process would use 19 GJ/per metric ton for integrated steelmaking, and 7 GJ/per metric ton for making steel out of scrap. Analyses have shown that many technologies exist that could improve energy efficiency further. For example, in the United States the potential for energy efficiency improvement is estimated at 18 percent, using proven and cost-effective practices and technologies. Under development are new technolgies that could considerably lower the energy intensity of steelmaking. Smelt reduction in ironmaking would integrate the production of coke and agglomerated ore with that of ironmaking, leading to reductions in production costs and energy use. The development of direct casting techniques that abandon rolling would increase productivity while reducing energy use further. Combined, these technologies could reduce the energy intensity of primary steelmaking to 12.5 GJ/per metric ton of steel, and for secondary steelmaking to 3.5 GJ/ per metric ton reductions of 34 percent and 50 percent, respectively. In the highly competitive and globalizing steel industry, manufacturers must continuously look for ways to lower their energy intensity and costs.

Pulp and Paper Industry

Paper consists of aligned cellulosic fibers. The fibers may be from wood or other crops, or from recycling waste paper. Starting with wood fibers, the fibers need to be separated from the wood, which is done by pulping the wood. The separation can be done by chemicals, by heat, or by mechanical means. In the chemical pulping process chemicals and hot water are used to separate the cellulosis from the ligno-cellulosis. The amount of pulp produced is about half the amount of wood used. Chemical pulping results in high-quality paper. In the mechanical process the wood is ground under pressure, separating the fibers from each other. In mechanical pulping the ligno-cellulosis is generally not removed, resulting in a lower quality of paper (e.g., paper used for newsprint) but a higher recovery (about 90%) of the used wood. In chemical pulping a lot of steam is used to heat the water and concentrate the chemical by-products. However, recovery of the by-products to be recycled in the process can actually produce sufficient steam for the whole paper mill. The mechanical process uses large quantities of electricity, while some processes can recover steam from the grinding process. Chemical pulp can be bleached to produce white paper. Waste paper is being pulped by mixing with water, after which ink is removed and the pulp refined. Paper recycling reduces the energy needs of the pulping process. Waste paper use in the production of paper varies widely, due to different structures of the industry in different countries.

While energy efficiency improvement options do exist in the pulping step, greater opportunities exist in the chemical recovery step. The most common pulping process in the United States is the Kraft pulping process. Black liquor is produced as a by-product. The chemicals are currently recovered in a recovery boiler, combusting the ligno-cellulosis. Because of the high water content, the efficiency of the recovery boiler is not very high, and the steam is used to generate electricity in a steam turbine and steam for the processes. Gasification of black liquor would allow use of the generated gas at a higher efficiency. This would make a Kraft pulp mill an electricity exporter (Nilsson et al., 1995).

In papermaking the pulp is diluted with water at a ratio of about 1:100. This pulp is screened and refined. The solution with the refined fibers (or stock) is fed into the paper machine, where the water is removed. In the paper machine, the paper is formed into a sheet, and water is removed by dispersing over a wire screen. At the end of the forming section, 80 percent of the water has been removed. The rest of the water is removed in the pressing and drying sections. While only a small amount of water is removed in the drying section, most energy is used there. Hence, those using such energy efficiency opportunities try to reduce the water content by increasing the water removal by pressing. In a long nip press, the pressing area is enlarged. The larger pressing area results in extra water removal. New technologies are under development aiming to increase the drying efficiency considerably. One technology—impulse drying—uses a heated roll, pressing out most of the water in the sheet; this may reduce the steam consumption of the paper machine by 60 to 80 percent.

Energy Efficiency Measure	*Typical Energy Savings (%)*	*Payback Estimate (years)*
Distribution System		
Reducing Steam Leaks	3-5%	
Insulating Steam Pipes	5-10%	0.3-1.7 year
Condensate Return	10-20%	<1 year
Process Integration & Heat Recovery	5-40%	2-
System Operation and Maintenance		
Water Treatment	10-12%	
Load Control	3-5%	
Decentralization Steam Supply	<40%	<4 years
Hot Standby		
Boilers		
Boiler Tune-Up	1-2%	
Combustion Air Preheating	<12%	<5 years
Boiler Feed Preheating	2-10%	<4 years
New Low-NOx Boiler Type	>5%	n.a.
Monitoring and Control	1-4%	<3 years

Table 2.
Energy Efficiency Measures and Estimated Improvement Potentials for Steam Boilers.
SOURCE: Prindle et al., 1995; Jones, 1997; CADDET, 1999.

The pulp and paper industry uses approximately 6 to 8 EJ globally. Because energy consumption and intensity depend on the amount of wood pulped, the type of pulp produced, and the paper grades produced, there is a great range in energy intensities among industrialized countries of the world. In Europe energy use for papermaking varied between 16 GJ per metric ton and to 30 GJ per metric ton of paper in 1989. The Netherlands used the least energy per metric ton of paper, largely because most of the pulp was imported. Countries such as Sweden and the United States have much higher energy intensities due to the larger amount of pulp produced. Sweden and other net exporters of pulp also tend to show a higher energy intensity. Energy intensity is also influenced by the efficiency of the processes used. Many studies have shown considerable potentials for energy efficiency improvement with current technologies (Worrell et al., 1994; Nilsson et al., 1995; Farla et al., 1997), such as heat recovery and improved pressing technologies.

Cross-Cutting: Steam Production and Use

Besides the energy-intensive industries, many smaller and less energy-intensive, or light, industries exist. Light industries can include food processing, metal engineering, or electronics industries. In light industries energy is generally a small portion of the total production costs. There is a wide variety of processes used within these industries. Generally a large fraction of energy is used in space heating and cooling, in motors (e.g., fans and compressed air), and in boilers. Industrial boilers are used to produce steam or to heat water for space and process heating and for the generation of mechanical power and electricity. In some cases these boilers will have a dual function, such as the cogeneration of steam and electricity. The largest uses of industrial boilers by capacity are in paper, chemical, food production, and petroleum industry processes. Steam generated in the boiler may be used throughout a plant or a site. Total installed boiler capacity (not for cogeneration) in the United States is estimated at nearly 880 million megawatts. Total energy consumption for boilers in the United States is estimated at 9.9 EJ.

A systems approach may substantially reduce the steam needs, reduce emissions of air pollutants and greenhouse gases, and reduce operating costs of the facility. A systems approach assessing options throughout the steam system that incorporates a variety of measures and technologies is needed (Zeitz, 1997), and can help to find low-cost options. Improved efficiency of steam use reduces steam needs and may reduce the capital layout for expansion, reducing emissions and permitting procedures at the same time. Table 2 summarizes various options to reduce losses in the steam distribution and to improve system operation and the boiler itself. In specific cases, the steam boiler can be replaced almost totally by a heat pump (or mechanical vapor recompression) to generate low-pressure steam. This replaces the fuel use for steam generation by electricity. Emission reductions will depend on the type and efficiency of power generation.

Another option to reduce energy use for the steam system is cogeneration of heat and power (CHP) based on gas-turbine technology as a way to substantially reduce the primary energy needs for steam making. Low- and medium-pressure steam can be generated in a waste heat boiler using the flue gases of a gas turbine. Classic cogeneration systems are based on the use of a steam boiler and a back-pressure turbine. These systems have relatively low efficiency compared to a gas turbine system. Steam-turbine systems generally have a power-to-heat ratio between 0.15 (40 kWh/GJ) and 0.23 (60 kWh/GJ) (Nilsson et al., 1995). The power-to-heat ratio depends on the specific energy balance of plant as well as energy costs. A cogeneration plant is most often optimized to the steam load of the plant, exporting excess electricity to the grid. The costs of installing a steam turbine system strongly depend on the capacity of the installation. Gas-turbine-based cogeneration plants are relatively cheap. In many places (e.g., The Netherlands and Scandinavia) gas turbine cogeneration systems are standard in paper mills. The power-to-heat ratio is generally higher than for steam turbine systems. Aero-derivative gas turbines may have a power-to-heat ratio of 70 to 80 kWh/GJ . Aeroderivative turbines are available at low capacities, but specific costs of gas turbines sharply decrease with larger capacities.

POTENTIAL FOR ENERGY EFFICIENCY IMPROVEMENT

Much of the potential for improvement in technical energy efficiencies in industrial processes depends on how closely such processes have approached their thermodynamic limit. There are two types of energy efficiency measures: (1) more efficient use in existing equipment through improved operation, maintenance or retrofit of equipment and (2) use of more efficient new equipment by introducing more efficient processes and systems at the point of capital turnover or expansion of production. More efficient practices and new technologies exist for all industrial sectors. Table 2 outlines some examples of energy efficiency improvement techniques and practices.

A large number of energy-efficient technologies are available (see Table 3) in the steel industry, including continuous casting, energy recovery, and increased recycling. Large technical potentials ranging from 25 to 50 percent exist in most countries. New technologies are under development (e.g., smelt reduction and near net shape casting) that will reduce energy consumption as well as environmental pollution and capital costs. A few bulk chemicals such as ammonia and ethylene represent the bulk of energy use in this subsector. Potentials for energy savings in ammoniamaking are estimated to be up to 35 percent in Europe and between 20 percent and 30 percent in Southeast Asia.

Iron and Steel
Heat recovery for steam generation, pre-heating combustion air, and high efficiency burners
Adjustable speed drives, heat recovery coke oven gases, and dry coke quenching
Efficient hot blast stove operation, waste heat recovery for hot blast stove, top gas power recovery turbines, direct coal injection
Recovery BOF-gas, heat recovery of sensible heat BOF-gas, closed BOF-gas-system, optimized oxygen production, increase scrap use, efficient tundish preheating
UHP-process, Oxy-fuel injection for EAF plants, and scrap preheating
Heat recovery (steam generation), recovery of inert gases, efficient ladle preheating
Use of continuous casting, 'Hot connection' or direct rolling, recuperative burners
Heat recovery, efficient burners annealing and pickling line, continuous annealing operation
Chemicals
Process management and thermal integration (e.g. optimization of steam networks, heat cascading, low and high temperature heat recovery, heat transformers), mechanical vapor recompression
New compressor types
New catalysts
Adjustable speed drives
Selective steam cracking, membranes
High temperature cogeneration and heat pumps
Autothermal reforming
Petroleum Refining
Reflux overhead vapor recompression, staged crude pre-heat, mechanical vacuum pumps
Fluid coking to gasification, turbine power recovery train at the FCC, hydraulic turbine power recovery, membrane hydrogen purification, unit to hydrocracker recycle loop
Improved catalysts (reforming), and hydraulic turbine power recovery
Process management and integration
Pulp and Paper
Continuous digester, displacement heating/batch digesters, chemi-mechanical pulping
Black liquor gasification/gasturbine cogeneration
Oxygen predelignification, oxygen bleaching, displacement bleaching
Tampella recovery system, falling film black liquid evaporation, lime kiln modifications
Long nip press, impulse drying, and other advanced paper machines
Improved boiler design/operation (cogeneration), and distributed control systems
Cement
Improved grinding media and linings, roller mills, high-efficiency classifiers, wet process slurry
Dewatering with filter presses
Multi-stage preheating, pre-calciners, kiln combustion system improvements, enhancement of internal heat transfer in kiln, kiln shell loss reduction, optimize heat transfer in clinker cooler, use of waste fuels
Blended cements, cogeneration
Modified ball mill configuration, particle size distribution control, improved grinding media and linings, high-pressure roller press for clinker pre-grinding, high-efficiency classifiers, roller mills

Table 3.
Efficiency Improvement Measures in Energy Intensive Industry
SOURCE: Worrell, Levine, et al., 1997.

Energy savings in petroleum refining are possible through improved process integration, cogeneration, energy recovery, and improved catalysts. Compared to state-of-the-art technology, the savings in industrialized countries are estimated to be 15 to 20 percnet, and higher for developing countries. Large potentials for energy savings exist in nearly all process stages of pulp and paper production (e.g., improved dewatering technologies, energy and waste heat recovery, and new pulping technologies). Technical potentials are estimated at up to 40 percent, with higher long-term potentials (see above). Energy savings in cement production are possible through increased use of additives (replacing the energy-intensive clinker), use of dry process, and use of a large number of energy efficiency measures (such as reducing heat losses and use of waste as fuel). Energy savings potentials of up to 50 percent do exist in the cement industry in many countries through efficiency improvement and the use of wastes such as blast furnace slags and fly ash in cementmaking.

In the United States various studies have assessed the potential for energy efficiency improvement in industry. One study has assessed the technologies for various sectors and found potential economic energy savings of 7 to 13 percent over the business-as-usual trends (Brown et al., 1998) between 1990 and 2010. Technologies like the ones described above (see Table 2) are important in achieving these potentials.

However, barriers may partially block the uptake of those technologies. Barriers to efficiency improvement can include unwillingness to invest, lack of available and accessible information, economic disincentives, and organizational barriers. The degree to which a barrier limits efficiency improvement is strongly dependent on the situation of the actor (e.g., small companies, large industries). A range of policy instruments is available, and innovative approaches or combinations have been tried in some countries. Successful policy can contain regulations (e.g., product standards) and guidelines, economic instruments and incentives, voluntary agreements and actions, information, education and training, and research, development and demonstration policies. Successful polices with proven track records in several sectors include technology development, and utility/government programs and partnerships. Improved international cooperation to develop policy instruments and technologies to meet developing country needs will be necessary, especially in light of the large anticipated growth of the manufacturing industry in this region.

SUMMARY

Manufacturing industry is a large energy user in almost all countries. About half of industrial energy use is in specific processes in the energy-intensive industries. On the other hand, various general energy conversion technologies and end uses can also be distinguished, such as steam production, motive power, and lighting. Opportunities and potentials exist for energy savings through energy efficiency improvement in all sectors and countries. Technology development, and policies aimed at dissemination and implementation of these technologies, can help to realize the potential benefits. Technologies do not now, nor will they in the foreseeable future, provide a limitation on continuing energy efficiency improvements.

Ernest Warrell

See also: Economically Efficient Energy Choices; Industry and Business Energy as a Factor of Production in; Industry and Business, Productivity and Energy Efficiency in.

BIBLIOGRAPHY

Ang, B. W. (1995) "Decomposition Methodology in Industrial Energy Demand Analysis." *Energy* 20(11):1081-1096.

Ang, B. W., and Pandiyan, G. (1997) "Decomposition of Energy-Induced CO_2 Emissions in Manufacturing." *Energy Economics* 19:363-374.

Brown, M. A.; Levine, M. D.; Romm, J. P.; Rosenfeld, A. H.; and Koomey, J. G. (1998). "Engineering-Economic Studies of Energy Technologies to Reduce Greenhouse Gas Emissions: Opportunities and Challenges." *Annual Review of Energy and the Environment* 23:287-385.

Center for the Analysis and Dissemination of Demonstrated Energy Technologies (CADDET). (1999). *CADDET Register on demonstration projects* Sittard, Neth.: Author/IEA.

De Beer, J.; Worrell, E.; and Blok, K. (1998). "Future Technologies for Energy Efficient Iron and Steel Making." *Annual Review of Energy and the Environment* 23:123-205.

De Beer, J.; Worrell, E.; and Blok, K. (1998). "Long-Term Energy-Efficiency Improvements in the Paper and Board Industries." *Energy* 23:21-42.

Farla, J.; Blok, J.; and Schipper, L. (1997). "Energy Efficiency Developments in the Pulp and Paper Industry." *Energy Policy* 25:745-758.

Howarth, R. B.; Schipper, L.; Duerr, P. A.; and Strom, S. (1991). “Manufacturing Energy Use in Eight OECD Countries, Decomposing the Impacts of Changes in Output, Industry Structure, and Energy Intensity.” *Energy Economics* 13:135-142.

IEA. (1997). *Indicators of Energy Use and Efficiency: Understanding the Link Between Energy and Human Activity.* Paris: Author/OECD.

Jones, T. (1997). “Steam Partnership: Improving Steam System Efficiency Through Marketplace Partnerships.” *Proceedings 1997 ACEEE Summer Study on Energy Efficiency in Industry*, Washington, DC: ACEE.

LBNL. (1998). *OECD Database.* Berkeley, CA, Lawrence Berkeley National Laboratory.

Li, J.-W.; Shrestha, R. M.; and Foel, W. K. (1990). “Structural Change and Energy Use: The Case of the Manufacturing Industry in Taiwan.” *Energy Economics* 12:109-115.

Nilsson, L. J.; Larson, E. D.; Gilbreath, K. R.; and Gupta, A. (1995). “Energy Efficiency and the Pulp and Paper Industry.” Washington DC: American Council for an Energy Efficient Economy.

Park, S.-H.; Dissmann, B.; and Nam, K.-Y. (1993). “A Cross-Country Decomposition Analysis of Manufacturing Energy Consumption.” *Energy* 18:843-858.

Price, L.; Michaelis, L.; Worrell, E.; and Khrushch, M. (1998). “Sectoral Trends and Driving Forces of Global Energy Use and Greenhouse Gas Emissions.” *Mitigation and Adaptation Strategies for Global Change* 3:263-319.

Prindle, W.; Farfomak, P.; and Jones, T. (1995). “Potential Energy Conservation from Insulation Improvements in U.S. Industrial Facilities.” *Proceedings 1995 ACEEE Summer Study on Energy Efficiency in Industry.* Washington, DC: ACEEE.

Ross, M. H., and Steinmeyer, D. (1990). “Energy for Industry.” *Scientific American* 263:89-98.

Schipper, L., and Meyers, S. (1992). *Energy Efficiency and Human Activity: Past Trends, Future Prospects.* New York: Cambridge University Press.

Sinton, J. E., and Levine, M. D. (1994). “Changing Energy Intensity in Chinese Industry.” *Energy Policy* 21:239-255.

WEC. (1995). *Energy Efficiency Utilizing High Technology: An Assessment of Energy Use in Industry and Buildings*, prepared by M. D. Levine, E. Worrell, N. Martin, and L. Price. London: Author.

Worrell, E.; Cuelenaere, R. F. A.; Blok, K.; and Turkenburg, W. C. (1994). “Energy Consumption by Industrial Processes in the European Union.” *Energy* 19:1113-1129.

Worrell, E.; Levine, M. D.; Price, L. K.; Martin, N. C.; van den Broek, R.; and Blok, K. (1997). *Potential and Policy Implications of Energy and Material Efficiency Improvement.* New York: UN Commission for Sustainable Development.

Worrell, E.; Price, L.; Martin, N.; Farla, J.; and Schaeffer, R. (1997). “Energy Intensity in the Iron and Steel Industry: A Comparison of Physical and Economic Indicators.” *Energy Policy* 25(7-8):727-744.

Worrell, E.; Martin, N.; and Price, L. (1999). “Energy Efficiency and Carbon Emission Reduction Opportunities in the U.S. Iron and Steel Industry.” Berkeley, CA: Lawrence Berkeley National Laboratory.

Zeitz, R. A., ed. (1997). *CIBO Energy Efficiency Handbook.* Burke, VA: Council of Industrial Boiler Owners.

MARKET IMPERFECTIONS

Modern economic theory provides a succinct description of the conditions under which the price system produces optimal outcomes in an idealized “laissez-faire” economy of perfectly competitive markets. This is the “First Welfare Theorem”; any competitive equilibrium is “Pareto optimal” (i.e., no agent in the economy can have his or her well-being increased except at the expense of another agent). Against this benchmark, the theory describes conditions under which policy interventions can, in principle, improve upon the performance of the unregulated market system. The possibility of such improvement arises from the existence of market “failures” or “imperfections,” which are factors in or features of particular markets that cause private decision-making to produce less-than-optimal economic outcomes.

Historically, the key imperfection in energy markets was thought to be “economies of scale,” or “declining average costs,” in electric power generation. This means simply that this was the kind of industry where a single firm’s costs of producing power would fall as its output was increased. Under this condition, unregulated market equilibrium with more than one competing firm would not result in economic efficiency, specifically the provision of power at minimum cost. Instead, power generation was a “natural monopoly”—a single firm could produce at lowest cost (in a given geographic area). This feature of power generation was the motivation for the U.S. system of privately owned, publicly regulated power companies, granted an exclusive license within a given service area with rates set by regulators. In the past several decades, however, technological change in electric power generation has resulted in a loss of economies of scale sufficient to motivate the deregulation or "restructuring" of this industry into a competitive form.

Currently, a different feature of energy markets having to do with the use of fossil fuels in energy production is recognized as entailing market failure. Among the assumptions required for the efficiency of competitive markets are that all commodities be both "rival" and "excludable." A rival commodity is one whose use by one precludes its use by another, while an excludable commodity is one whose production and consumption imposes no "side effects" on anyone not party to direct transactions involving the commodity. The market failures now most commonly ascribed to energy markets arise from the nonrivalry and nonexcludability of certain byproducts of energy consumption and production: so-called "environmental externalities," or side effects that have deleterious effects on health or welfare. Key examples are carbon monoxide emissions from vehicles, sulfur dioxide emissions from electric power generation, and emissions from power generation and from the use of fossil fuels. The creation of these by-products creates costs that are not reflected in market prices for energy products, so that in the absence of regulation or some other policy intervention there are no incentives to control them.

Sulfur dioxide emissions resulting from fossil fuel can have negative effects on urban air quality and create acid rain that harms aquatic life. These emissions are nonexcludable in that there is no private action that a particular individual can take to avoid this impact, and they are nonrival in that their effect on any one individual does not preclude or offset their effect on any other.

Another, more controversial, idea is that there may be market imperfections underlying the "energy-efficiency gap," the long-recognized apparent under-investment by consumers and firms in energy-efficient technology. Beginning in the 1970s, energy analysts used the term "market barrier" to refer to any of the various possible reasons for such underinvestment. Since then, there has been some effort to distinguish those "barriers" that constitute market imperfections. Although no consensus has emerged, there is general agreement that the energy-efficiency gap may arise in part from informational problems involved in private investment decisions. It is now widely recognized that the nonrivalry of information can result in market imperfections. To the extent that informational problems impeding optimal energy-efficiency investments are pervasive, the efficiency gap may be seen as another important example of an energy-related market imperfections.

POLICY RESPONSES

U.S. environmental regulations have for the most part tried to mitigate health and welfare effects from pollution through technological control strategies such as requiring installation of pollution reduction equipment. There has been a gradual increase of interest, however, in analyzing and attempting to correct environmental externalities within the paradigm of market failures or imperfections. In this paradigm, in the presence of market failures the government can under idealized assumptions reallocate resources to make some consumers better off while making none worse off—a "Pareto improvement." However, the recognition that in practice there will always be both winners and losers from a given policy led to the development of a less ambitious notion of the goals of policy. Thus, the focus of cost-benefit analysis is to determine how market failures justify policies in which some are better off, and these winners could in principle compensate the losers and still come out ahead. This is the "compensation principle."

We often think that solving the problem of environmental externalities means eliminating them altogether, but the cost-benefit approach applies a different criterion: emissions should be held to an optimal level, which is less than the unregulated level but in most cases not zero. The threshold is that the marginal damage from emissions should be equated to the marginal cost of abatement. This, in turn, naturally suggests the economic means of controlling emissions: the government should impose taxes or charges on them so that this marginal condition holds. This policy mechanism has been studied extensively as a way of reducing carbon emissions in an effort to mitigate global climate change. An alternative approach is for the government to issue permits to emitters that restrict the overall quantity of emissions; these permits would then be traded among emitters so that, overall, the cost of abatement would be minimized.

Alan H. Sanstad

See also: Government and the Energy Marketplace; Market Transformation.

MARKET TRANSFORMATION

The term "market transformation" first appeared in the literature in the early 1990s. The term emerged more as an abstraction than a concrete program strategy or model. Market transformation provides a vision of the ultimate objective of strategic interventions—markets that yield energy-efficient outcomes automatically as the result of normal market forces. Market transformation can be viewed as a catalyst for change—a means of intervening in imperfect markets to effect long-term changes to improve market performance with respect to energy efficiency.

While there is not a single, precise definition of market transformation in professional practice, the following definition captures the essential elements:

> Market transformation is a strategic intervention to achieve a lasting, significant share of energy-efficient products and services in targeted markets. Market transformation is essentially synonymous with marketing strategy as used in the private business world. A key distinction is that market transformation is motivated by the social objectives of improving the performance of markets to yield greater energy efficiency. And like marketing strategies in the private sector, it often requires ongoing measures to achieve and sustain desired market outcomes. For example, market transformation will occur in the U.S. clothes washer market when energy-efficient washers (often horizontal axis machines that use about one-third the energy and water of old vertical-axis, agitator machines) become the norm as they already are in Europe. After introduction of new machines in the market, ongoing marketing and related support may be required to sustain a significant market share.

The overall goal of market transformation is to increase the share of energy-efficient products and services through fundamental, enduring changes in targeted markets. This goal also serves to improve the economy and reduce negative social and environmental effects that result from energy use. Intervention is needed because of market imperfections that do not allow the market on its own to provide an optimum level of energy-efficiency goods and services. In most cases intervention requires an evolving mix of strategies and implementation measures over an extended period.

Jay Joseph and Don Gardner collect an old refrigerator after delivering a new one. In 1993, the Sacramento Municipal Utility District paid customers $100 for their old refrigerator plus an additional $100 rebate to buy an energy efficient model. (Corbis-Bettmann)

Market transformation as a strategy to improve market performance has become more important as most energy utility markets deregulate and restructure. Restructuring is resulting in more competition and reliance on market mechanisms in energy markets that have been highly regulated. Utility energy efficiency programs of the 1980s and 1990s, called demand-side management (DSM) programs, are being abandoned and market transformation is being introduced to fill the void. However, the roots of market transformation lie within the regulated energy industries.

The goal of most DSM programs has been relatively narrow: to reduce energy and power demand to avoid investments in new power plants or transmission and distribution systems. DSM was used within the context of integrated resource planning to yield the lowest cost of energy services by avoiding more costly construction and operation of supply-side power plants. DSM was considered a resource comparable and substitutable for supply-side resources (hence the name—integrated resource planning). Individual utilities have typically implemented DSM for their own customers, as ordered by public utility commissions or other regulatory bodies. All utility customers generally have shared the costs for DSM programs because regulators mandated such programs and consequently provided cost-recovery mechanisms for utilities.

Integrated resource planning and demand-side management have declined in the wake of the movement to restructure and deregulate energy markets. DSM has evolved to be more market-based, as program designers sought lasting change, and utilities reduced program costs and shifted some of the remaining costs to the direct beneficiaries of DSM programs.

As DSM evolved, program managers realized that their efforts could have much greater impact if they went beyond the service territories of single utilities to encompass regional and national markets. While not termed "market transformation," there were several early initiatives that took this approach, including the Manufactured Housing Acquisition Program in the Pacific Northwest and the Power Smart Program, which originated in British Columbia and was later adopted elsewhere. These and other state, regional, national and even international collaborations, with multiple parties contributing funds and expertise, have tried to change building practices, introduce new products and change market shares. Collaborations, such as the Consortium for Energy Efficiency (CEE) in the United States, have been making larger changes in the markets for target technologies. Examples include promoting super-efficient refrigerators, clothes washers, and motors. More broadly, the Energy Star Program of the United States Environmental Protection Agency (EPA) and Department of Energy (DOE) is an example of a market transformation effort that spans numerous household and commercial appliances and applications—from home computers and air-conditioners to energy ratings of commercial buildings. Energy Star is a labeling program that identifies the most energy-efficient technologies within a given appliance or application category.

Market transformation initiatives typically require collaboration among a diverse set of stakeholders, including manufacturers, retailers, utilities, research and development (R&D) organizations, government, and public-interest efficiency advocates. In recent years the need for coordination and collaboration among such a diverse set of actors to achieve consensus has led to the development of specific U.S. regional market transformation organizations, including the Northwest Energy Efficiency Alliance (NEEA), the Northeast Energy Efficiency Partnerships, Inc. (NEEP), and the Midwest Energy Efficiency Alliance (MEEA).

These organizations vary significantly in their structure, funding and operation. However, their overall approach to market transformation is similar. Market transformation typically includes the following steps (not necessarily in sequence):

- Identify needs through market analysis and research—the markets where opportunities exist to increase market share of products and services that respond to customer needs and deliver superior energy-efficient performance.
- Identify market participants (manufacturers, retailers, consumers) and stakeholders (such as consumer advocacy groups, trade organizations, and government).
- Form collaboratives and define roles among key market participants and stakeholders to lead and manage the market transformation initiative.
- Establish funding to cover costs of the initiative (program costs).
- Define program goals for target products or services within the chosen market.
- Establish market baselines against which intervention(s) will be evaluated.
- Design strategies and measures for the initiative, including a transition strategy that may be an exit or continued intervention such as advertising and education.
- Implement measures.
- Track market performance and evaluate results of the initiative.
- Continue, modify or end initiative as indicated by monitoring and evaluation results.

Implementation of market transformation programs requires adoption of coordinated measures targeted to various market participants over a fairly long period. The duration of market transformation programs depends on numerous factors, including the complexity of the market; customer response; support of manufacturers, distributors and retailers; and time required for manufacturers to change manufacturing operations. Experience with past DSM programs and early market transformation programs suggests that periods of five to ten years or more are needed to transform energy efficiency markets.

Typical measures used with market transformation programs may include the following:

- marketing, such as media advertisements, point-of-purchase displays, utility bill stuffers and other promotions

- labeling (a key example is the Energy Star program established and operated by the U.S. EPA and DOE)
- consumer education
- professional training (e.g., sales associates, skilled tradespeople, contractors, manufacturers)
- research and development in support of program needs
- codes and standards (to codify energy-efficient technologies by establishing minimum performance standards)
- consumer rebates or other incentives to increase consumer acceptance
- manufacturer and retailer incentives
- technology procurement (specifying required performance of technologies and aggregating customers to create sufficient demand for suppliers to respond to performance requirements)
- other types of bulk purchasing or buyer aggregation to create market pull
- design competitions based on desired performance.

Market transformation collaborations involve multiple parties—each with different motivations and objectives. The target markets are typically broad in geographic scope (regional and national markets). These two factors alone pose major challenges for market transformation programs. The complexity and dynamic nature of markets pose a different set of challenges. While the challenges of market transformation are many and complex, the advantages of market transformation versus traditional DSM intervention are substantial. Market transformation focuses on systemic, complementary measures for market improvement, whereas DSM programs typically were isolated efforts that addressed much narrower symptoms of market imperfections. For this reason, market transformation is growing rapidly as a dominant model for publicly and privately supported energy-efficiency programs.

Dan W. York
Mark E. Hanson

See also: Demand-Side Management; Efficiency of Energy Use, Labeling of.

BIBLIOGRAPHY

American Council for an Energy-Efficient Economy. (1998). *Proceedings from the 1998 American Council for an Energy-Efficient Economy Summer Study on Energy Efficiency in Buildings, Volume 7, Market Transformation.* Washington, DC: Author.

Eto, J.; Prahl, R.; and Schlegel, J. (1996). “A Scoping Study on Energy-Efficiency Market Transformation by California Utility DSM Programs,” LBNL-39058. Berkeley, CA: Lawrence Berkeley National Laboratory.

Geller, H., and Nadel, S. (1994). “Market Transformation Strategies to Promote End-Use Efficiency.” *Annual Review of Energy and the Environment* 19:301–346.

Nadel, S., and Latham, L. (1998). *The Role of Market Transformation Strategies in Achieving a More Sustainable Energy Future.* Washington, DC: American Council for an Energy-Efficient Economy.

Synergic Resources Corporation. (1996). *Market Transformation in a Changing Utility Environment.* Washington, DC: National Association of Regulatory Utility Com-missioners.

York, D., ed. (1999). “A Discussion and Critique of Market Transformation: Challenges and Perspectives,” Report 186. Madison, WI: Energy Center of Wisconsin.

MASS TRANSIT

Mass transit is a transportation service available to the public for trips generally within metropolitan areas. As with all transportation services, an energy source is a critical input in the production of mass transit trips. Almost all of this energy consumption is derived from burning fossil fuels, a process that emits pollutants affecting human health, visibility, vegetation, and climate change. Transit service is increasingly being scrutinized for its pollutant emissions. Mass transit facilitates travel within densely developed, large urban areas. Because of the existence of mass transit, more intense development within an urban area can occur. Higher-density land use may enable reduced energy consumption when considering the settlement area as a whole.

FORMS OF MASS TRANSIT

The term “mass transit” generally refers to passenger vehicles that are common carriers in urban areas, as distinct from intercity travel. The terms “public transit” or simply “transit” also are frequently used. The major types of public transit are bus (rubber-tired vehicles), rail (running on tracks), and ferryboat. Within each type there are several subcategories.

Motor Bus

A rubber-tired, self-propelled transit vehicle using an internal combustion engine for power. Most use direct-ignition (diesel) engines, but gaso-

A seabus crosses a harbor in Vancouver, British Columbia. (Corbis-Bettmann)

Mode	*Vehicle Distance Operated*	*Passenger Boardings*
Bus	56%	58%
Rail rapid transit	19%	30%
Suburban rail	8%	5%
Streetcar/LRT	1%	3%
Demand response	12%	1%
Other*	3%	3%
Total	**100%**	**100%**

Table 1.
Service Supplied and Consumed, by Mode, in the United States, 1997.
NOTE: "Other" includes ferryboat, inclined plane, automated guideway, monorail, aerial tramway, and cable car.
SOURCE: U.S. Department of Transportation, 1999.

line, propane, natural gas, and other fuels are used as well. Buses are available in varying sizes and capacities, varying from 10 passengers to 150 or more passengers, including articulated and biarticulated vehicles. They can be designed for city service (fewer seats, more doors and standing room) or long-distance express service (often with no standees permitted). Buses can operate on most streets open to traffic, including expressways, and also can be operated on exclusive rights-of-way such as high-occupancy-vehicle (HOV) expressway lanes, bus lanes on city streets, or busways with on-line stations.

Electric Bus

When powered by an electric motor, generally taking power from overhead wires ("catenaries") such a vehicle is called an electric trolleybus or a trackless trolley. An electric bus taking power from current in the ground has been developed by Ansaldo-Breda. Battery-powered buses also have been developed, although all of these have been smaller than standard-size buses. In the late 1990s, "hybrid" electric-internal-combustion buses were placed in service. These vehicles have a smaller than normal internal-combustion engine, which continually recharges a battery pack. The batteries power an electric motor, which provides supplementary power when needed. Fuel cells, which can be thought of as chemical batteries, also are used on an experimental basis to power transit buses.

Streetcar or Tram

The streetcar or tram, powered by electricity from overhead catenaries and running on rails on city streets, was the backbone of mass transit service from 1890 until the widespread deployment of the diesel bus in the early post-World War II period. In general, the only streetcar lines in North America that survived the transition to the bus were those in high-demand corridors, or those operating in their own right-of-way, such as a roadway median, tunnel, or subway. Since the 1970s, there has been increasing deployment of new streetcar service, now referred to as light rail transit (LRT). These services tend to use larger, articulated vehicles and frequently operate partly on exclusive rights-of-way such as disused rail corridors or expressway medians.

Rail Rapid Transit

Also known as metro, subway, elevated, underground, or heavy rail, this higher-capacity rail service is distinguished by its use of trained vehicles (several

	Public Transit Bus			*Passenger Car*		
	BTU per vehicle mile	***BTU per passenger mile***	***passenger mile per vehicle mile***	***BTU per vehicle mile***	***BTU per passenger mile***	***passenger mile per vehicle mile***
1970	31,796	2,472	12.9	9,301	4,896	1.9
1980	36,553	2,813	13.0	7,915	4,166	1.9
1990	36,647	3,735	9.8	6,183	3,864	1.6
1997	38,101	4,318	8.8	5,822	3,639	1.6
% change 1970-97	20%	75%	-31%	-37%	-26%	-16%

Table 2.
Average Energy Intensity and Occupancy Rates for U.S. Public Transit Buses and Passenger Automobiles, 1970–1997.
SOURCE: Davis, 1999.

cars attached together) and third-rail electric power (a live connection at grade). Because of its unprotected power source, rail rapid transit cannot be operated on city streets, and runs in elevated or underground rights-of-way. Unlike suburban rail, rail rapid transit service generally operates within cities and on a frequent schedule. Some of the new rail rapid transit systems, such as BART in San Francisco, have a metropolitan scope and distant stop spacing, and thus are similar to suburban rail.

Suburban Rail

Also known as commuter rail, this type of service generally provides long-distance—50 km (31 mi.) or more—routes to the far reaches of a metropolitan area using railroad rights-of-way. These operations use either diesel or electric power. Service is primarily oriented toward commuters from suburbs to central city locations and is concentrated in the peak commuting hours. Express bus operations also provide similar services, without the need for a separate right-of-way.

Ferryboat

In metropolitan areas adjacent to bodies of water, ferries can be used as mass transit—that is, providing frequent service useful for local travel. Geography limits ferryboat use to a select group of cities. Twenty-one cities in the United States and its territories have transit ferryboat service. Another example is Vancouver's SeaBus, which provides very frequent shuttle service between the center city and North Vancouver.

Table 1 shows the frequency distribution of U.S. mass transit service supplied (vehicle distance operated) and service consumed (passenger boardings) by transit mode in 1997. The official U.S. transit statistics also include the "demand response" mode. This type of transit consists of minibuses or vans operating by request, rather than on fixed routes, and typically available only to a portion of the public, such as elderly or disabled people. As shown in Table 1, demand response accounts for 12 percent of service operated but only 1 percent of boardings. Buses accounts for the majority of both service supplied and service consumed. Removing the few largest rail systems from the totals would reveal even greater significance of the bus mode.

IMPACT OF MASS TRANSIT ON ENERGY CONSUMPTION

The potential of mass transit to provide transportation services with low energy consumption relies on the high capacity of transit vehicles, since these vehicles have higher energy consumption per vehicle distance traveled compared to private motorized vehicles or nonmotorized modes. Therefore the occupancy rate of transit service is a key factor in determining its energy efficiency. This rate can be measured by the ratio of person distance traveled to vehicle distance traveled.

The energy efficiency of transit per passenger distance traveled depends both on its usage rate (person travel per vehicle travel) and its fuel efficiency (fuel consumption per vehicle travel). The transit vehicles with the greatest potential for energy efficiency gains

Year	Transit Bus	Transit Rail	Passenger Car	Intercity Rail	Intercity Bus
1970	2,472	2,453	4,896	3,677	1,051
1980	2,813	3,008	4,166	3,176	1,069
1990	3,735	3,453	3,864	2,609	944
1997	4,318	3,253	3,639	2,458	872
1970-97	75%	33%	-26%	-33%	-17%

Table 3.
Average Energy Intensity per Passenger Mile for Various U.S. Passenger Modes, 1970–1997.
SOURCE: Davis, 1999.

may not be those that can carry the most people per vehicle. Rather, the greatest potential gains can be had by matching vehicle capacity to travel demand. In fact, a high-capacity mode that is little used can increase energy consumption relative to the passenger automobile. On the other hand, a minibus with ten passengers may represent a significant reduction in energy use compared to the case of those ten passengers traveling by car.

The data in Table 2 dramatically illustrate the difference between vehicle fuel intensity and passenger fuel intensity from 1970 to 1977. (Fuel intensity, fuel use per distance traveled, is the inverse of fuel efficiency.) In 1970, the transit bus mode in the United States used 50 percent less energy per passenger mile than the automobile mode (passenger cars, not including light trucks). By 1997, transit buses used 19 percent *more* energy on average per passenger mile. How did this happen? The fuel efficiency of transit buses declined by 20 percent, while fuel efficiency improved by 37 percent for passenger cars. Occupancy declined in both cases, but twice as much for transit buses. The net result was a significant reduction in fuel use per passenger mile for cars despite declining occupancy, and an increase in fuel intensity for transit buses.

The evolution of transit bus compared to transit rail energy intensity in the United States from 1970 to 1997 is shown in Table 3. Transit bus and rail had similar energy intensities in 1970, but by 1997, energy use per passenger mile had increased by 33 percent for transit rail and by 75 percent for transit buses. This increase is yet more significant given that energy intensity for passenger cars declined 26 percent over the same period and given that intercity rail and intercity bus operations also showed energy intensity reductions of 33 percent and 17 percent, respectively.

In 1997, the intercity bus was the least energy-intense of the modes shown in Table 3. Its energy intensity per passenger mile was one-fifth of transit bus energy intensity. The vehicles used for the two operations are similar, although operating conditions are different. Intercity buses operate mostly on expressways, and transit buses mostly on local streets. Still, the fact that transit bus energy intensity increased 75 percent while intercity bus energy intensity decreased 17 percent suggests that changes in operating patterns, rather than changes in technology, accounted for much of the difference in energy use.

In the United States in 1997, transit accounted for a small fraction of transportation energy consumption, largely because it played such a small role in the total travel market, representing only 1.8 percent of trips. Transit accounted for only 0.7 percent of transportation energy consumption (bus 0.4%, rail 0.2%, and suburban rail 0.1%).

Transit represents an increase in energy use compared to walking or bicycling. For countries with a significant amount of nonmotorized transport, increasing transit use may mean increasing energy use. However, the increased energy consumption may be associated with significant improvements in urban passenger transport, which can produce large economic benefits. It is generally easier for transit modes to compete successfully with nonmotorized modes than with the private automobile for new passengers. But as demand for higher-speed travel increases, those customers may switch to private motorized vehicles.

COMPARISON OF MASS TRANSIT MODES

In the largest metropolitan regions, such as in Tokyo, New York, London, and Moscow, the huge demand for transit makes rail transit both indispensable and energy-efficient. But because transit energy efficiency is so dependent on matching supply with demand, the greater flexibility of the bus mode may have an advantage in this aspect. Bus service characteristics can be easily altered as to time of service, location of service, and size of vehicle used. For rail transit, not much change is possible in the short run.

The rail modes require a significant investment of both fiscal resources and energy in infrastructure,

Year	Diesel	Gasoline	Alternative Fuels						Total
			Compressed Natural Gas	Liquefied Natural Gas	Methanol	Propane (Liquid Petroleum Gas)	Other	Total	
1992	684,944	32,906	1,009	191	1,583	2,487	12	5,282	**728,414**
1993	678,511	37,928	1,579	474	4,975	2,098	197	9,323	**735,085**
1994	678,226	43,921	4,835	1,450	12,269	1,871	492	20,917	**763,981**
1995	678,286	42,769	10,740	2,236	11,174	3,686	865	28,701	**778,457**
1996	692,714	41,495	15,092	2,862	7,268	5,235	4,353	34,810	**803,829**
1997	716,952	41,547	23,906	4,030	965	5,150	7,771	41,822	**842,143**
1998	700,081	38,399	30,915	3,246	783	5,112	2,865	42,921	**824,322**

Table 4.
Mass Transit Fossil Fuel Consumption by Fuel (thousands of gallons).
SOURCE: American Public Transit Association, 2000.

including track, power supply, right-of-way, stations, and structures. This investment can produce a higher quality of service, but it commits the urban area to use of the rail mode and its relatively large vehicles.

The bus mode can use a broader range of vehicle sizes and does not require any separate infrastructure. Where right-of-way is shared with other vehicles, much of the energy used in producing the facilities may have been expended without consideration of use by mass transit. Where there is a need for exclusive bus infrastructure to bypass congested areas, busways can be constructed to improve service.

It is common outside of Europe and North America for bus service to be provided by private operators. The higher service efficiencies often achieved by private transit providers can translate into greater energy efficiency. While contracting out of transit service is possible for rail, it has been more common for bus operations.

Why has the load factor for transit decreased? Transit demand has declined due to higher incomes, higher automobile ownership, and a decrease in the share of jobs and population living in areas that are strong transit markets. At the same time, transit supply has increased, spurred by the growth of government subsidies for transit operation and capital investment. The long-standing problem of intense peaking of transit demand means that large vehicles are needed for only a few hours during the day, only to run nearly empty during off-peak hours. Transit agencies in the United States have been generally reluctant to use smaller vehicles. Privatized transit operations outside of London, England, were quick to adopt minibuses. Transit operation in the United States has largely become regionalized, with providers' operating areas spreading out over vast territories. Because all taxpayers in the region are typically required to contribute to transit via designated sales, property, or motor fuel taxes, transit agencies feel a political obligation to provide at least a minimum amount of service in all parts of the region, no matter how weak the transit demand.

Most urban rail service is electric-powered and most urban bus service is diesel-powered, although diesel rail and electric bus operations do exist, as noted above. The efficiency and environmental impacts of electricity depend greatly on the source of electric power. Although electric vehicles produce no tailpipe emissions, generation of electricity can produce significant emissions that can travel long distances. For example, coal-powered electricity plants produce particulate emissions that travel halfway across North America. Urban buses also can be powered by a variety of alternative fuels.

NEW TECHNOLOGIES

Applications of engineering and computer technology have the potential for reducing fuel use, greenhouse gas emissions, and toxic emissions from mass transit. They also may help in increasing transit load factors. Technology that enables greater operating efficiency of transit also has the potential to reduce fuel consumption per passenger distance traveled. The introduction of electronic fare cards in New York City prompted the development of new fare policies, including free transfer from bus to rail. The result was a 36 percent increase in bus use between 1997 and 1999. With only a 9 percent increase in bus service during the same period, the fuel efficiency of transit increased dramatically.

Traffic signal preemption and other technologies that seek to speed transit can have a similar effect. Faster bus travel means less fuel consumption per vehicle distance traveled. However, faster transit service increases demand for transit, potentially increasing the transit load factor.

The “conventional” fuels used for transit applications include gasoline, diesel fuel, and electricity. Alternatives to these fuels have been sought to reduce energy consumption, pollutant emissions, greenhouse gas emissions, and use of imported fuels. The conventional fuels for internal-combustion engines are the most energy-dense fuels: petroleum and diesel fuel.

Alternative fuels often are handicapped by the need to develop an alternative fueling infrastructure. Most transit fleets have their own fueling stations, so they are prime candidates for early introduction of new fuels. And transit agencies have experimented with many different fuels. Alcohol-fueled engines proved unreliable for heavy-vehicle applications. Los Angeles found that methanol and ethanol corroded bus engines, scrapped some of the vehicles, and converted the remainder to diesel in 1998-1999. The most popular alternative fuel has been compressed natural gas (CNG). This fuel offers lower pollutant emissions than diesel, but at a higher total cost than diesel. In addition to requiring new vehicles with natural-gas-powered engines and storage areas for fuel cylinders, using CNG requires an investment in new fueling and compression facilities and modifications to depots and maintenance facilities. Considering fueling, inspection, and maintenance, operating costs typically are higher for CNG than for comparable diesel vehicles.

Between 1992 and 1998, alternative fuels increased from less than 1 percent to more than 5 percent of total mass transit fossil fuel consumption in the United States (see Table 4). The share of alternative fuel consumption that was CNG increased from 19 percent to 72 percent over the same period.

One promising new technology is the hybrid electric vehicle, which combines an internal-combustion engine and an electric motor powered by batteries. The engine recharges the batteries, and the electric motor provides additional power during acceleration. These vehicles do not have the severe range and capacity limitations of battery-powered electric vehicles and do not have long refueling (recharging) times. Hybrids equipped with regenerative braking are well suited for city bus routes requiring extensive stop-and-go driving.

Ambient particulate matter (soot) has been linked to increased mortality. Diesel engines are high emitters of particulates. Particulate is also formed in the atmosphere from other pollutants, including those emitted by gasoline engines. As the public has become more aware of this problem, diesel engines have become much cleaner. By adding emissions control devices and electronic controls and by modifying engine designs, a large reduction in the emissions of particulate matter and nitrous oxides (precursors to both soot and smog) has already been achieved. In 2000, the U.S. Environmental Protection Agency proposed new regulations requiring very-low-sulfur diesel fuel and further large reductions in particulate and nitrous oxide emissions. Low-sulfur fuel reduces sulfate particulate emissions and also enables the introduction of emissions control devices that would be damaged by high-sulfur fuel. These regulations require diesel vehicles that have an emissions profile similar to today’s CNG buses.

In the longer term, more exotic technologies, such as fuel cells powered by hydrogen, may be feasible. These technologies are far from being economically feasible, but rapid progress is being made. However, as conventional vehicles become cleaner, the relative emissions-reduction benefits from alternative fuels declines.

LAND USE AND SOCIAL AND POLITICAL FACTORS

Matching the supply of transit service to its demand is the primary determinant of the energy efficiency of transit. The demand for transit depends on income levels and land use. As average income rises, so does the value of travel time, and therefore the cost of time spent traveling. Higher incomes also make automobile use more obtainable.

Land use is the other key determinant of transit use. Public transit requires a concentration of trips in the same time and place. A concentration of residential and commercial land use, such as that typically

A conductor on the number five train at Grand Central Terminal in New York on the eve of a possible strike by the subway and bus workers, December 1999. (Corbis Corporation)

found in cities, is necessary to generate significant transit demand. Concentrated land use has another, perhaps more important effect. To function efficiently, the private automobile must be provided with a significant amount of space for both storage and movement. In many cities there is insufficient space for the automobile, making automobile travel expensive. The largest costs of urban automobile travel are storage (parking) costs and travel time costs, due to traffic congestion. Under these circumstances, mass transit modes can compete effectively. They provide travelers with similar or lower total travel time (including time spent walking to and waiting for transit vehicles), and the avoided cost of parking may be several times greater than the cost of the transit fare.

However, most North American urban development since 1945 has been designed around the space needs of the automobile. A generous quantity of off-street parking is required for all new residential and commercial development. Streets are designed to be wide enough to accommodate free-flowing traffic, even in peak periods. Expressways have been constructed to connect all parts of the metropolitan region. The result has been the development of urban areas that are convenient for motoring. The same set of changes also has made them inhospitable to any mode of daily transport other than driving. The low population density of residential development makes postwar neighborhoods difficult to serve by public transit. More important, the prevalence in North America of widely scattered office parks with ample free parking has ensured that transit cannot compete for trips to those destinations, since driving is inexpensive and transit is either not available or takes up to twice as long.

Cities that were large (1 million or more) before 1920 have cores with a large concentration of jobs, expensive parking, and a large transit share of trips. The New York metropolitan area is the premier example, accounting for nearly 40 percent of all U.S. transit trips. Although newer areas have been able to increase transit ridership, none has transit shares approaching what is still common in the older large cities.

Per capita transportation energy consumption in a city such as New York is much lower than the U.S. average as a result of much lower car use. Although high-quality public transit service is one explanation for this result, parking costs, bridge and tunnel tolls, and the convenience of walking are equally important. Public transit is vital for the transportation needs of New York City residents. But many of the trips that residents of the typical U.S. metropolitan area take by auto are taken by New Yorkers on foot, or not at all. It is the lower number of auto trips, rather than a one-for-one substitution of transit trips for auto trips, that has a large impact on reducing energy use. Transit service in New York City is well used, making it energy-efficient despite the slow speed of surface transit travel.

In the United States, outside of the core areas of older major cities, transit has become the transport of last resort. There is a substantial social stigma attached to using transit, due to the low income levels of transit patrons in most U.S. cities. Transit customers also sometimes fear their vulnerability to crime, especially while waiting at bus stops.

INTERMODAL TRAVEL

Many transit trips involve transfers between transit vehicles. All else being equal, passengers would prefer not to transfer. However, the mode of access to transit is of interest. The vast majority of transit customers walk to transit and then walk again to their final destination. It is difficult to develop high transit demand in an urban environment that is not conducive to walking.

Park and Ride

A recent trend is the development of park and ride areas for transit passengers to drive to transit. These facilities help make transit accessible to a wide geographic area. Because emissions control devices are ineffective when cold, the auto trip to the transit station may produce nearly as many pollutants as a direct trip to the final destination. Further, park and ride lots occupy a lot of space, making transit stations unfriendly to pedestrians and reducing the opportunities for station-area real estate development. Finally, those driving to transit must still pay the full costs of automobile ownership. In fact, the only reason they are likely to take transit is if the cost of station-area parking and transit fare is less than the parking cost at the final destination. This explains why park and ride areas are most successful on transit lines serving urban cores with high parking costs. Most workers in North America, however, pay nothing out of pocket for parking at work.

Bicycling and Transit

Bicycling to transit is another solution for providing access. With no fossil fuel consumption and no pollution, bicycling to transit is attractive from an environmental point of view. Personal fitness is a major motivation for many bicyclists. Public health authorities concerned about physical inactivity have started to promote bicycling. Bicycle access to train stations is very popular in Japan and the Netherlands and some places in Germany. Since bicyclists can travel faster than local transit on some routes, the bike and ride option is most attractive for accessing express services such as suburban rail or express bus. Bike stations that provide bike storage, rental, and repair services and changing facilities adjacent to transit stations have been developed in three California cities.

Permitting bicyclists to take bicycles aboard transit vehicles allows them to use the bicycle for the trip to their final destination as well. Although many transit agencies are reluctant to make room for bicycles during crowded peak hours, railcars designed with bicycle storage have been deployed in Europe and North America. In the 1990s, hundreds of buses in North America were equipped with bicycle-carrying racks.

Despite the energy, emissions, and public health benefits of combining bicycle use with transit, few people in North America take advantage of this combination. Many urban roads are designed for high-speed operation and have narrow lanes. These designs often scare bicyclists from using the roads, a fear sometimes reinforced by motorists and police who do not believe that bicyclists have a right to use roads. Secure, sheltered bicycle storage at transit stations is rare, and therefore the threat of bicycle vandalism and theft deters others from using bicycling as an access mode to mass transit.

Paul M. Schimek

See also: Batteries; Behavior; Bicycling; Diesel Cycle Engines; Diesel Fuel; Electric Vehicles; Emission Control, Vehicle; Engines; Fuel Cells; Fuel Cell Vehicles; Gasoline and Additives; Gasoline Engines; Hybrid Vehicles; Locomotive Technology; Petroleum Consumption; Railway Passenger Service; Traffic Flow Management; Transportation, Evolution of Energy Use and.

BIBLIOGRAPHY

American Public Transportation Association. (2000). *Public Transportation Fact Book 2000.* Washington, DC: Author. <http://www.apta.com/stats/>.

Davis, S. C. (1999). *Transportation Energy Data Book: Edition 19.* ORNL-6958. Oak Ridge, TN: Oak Ridge National Laboratory. <http://www-cta.ornl.gov/Publications/Tedb.html>.

Kenworthy, J. R.; Laube, F. B.; Barter, P.; Newman, P.; Raad, T.; Poboon, C.; and Guia, B. (2000). *An International Sourcebook of Automobile Dependence in Cities, 1960-1990.* Boulder, CO: University Press of Colorado.

Parcells, H., and Replongle, M. (1992). *Linking Bicycle/Pedestrian Facilities With Transit (National Bicycling and Walking Study. Case Study No. 9).* Washington, DC: U.S. Department of Transportation, Federal Highway Adminstration.

Pickrell, D. (1999). "Transportation and Land Use." In *Essays in Transportation Economics and Policy: A Handbook in Honor of John R. Meyer*, ed. J. A. Gomez-Ibanez, W. B. Tye, and C. Winston. Washington, DC: Brookings Institution Press.

Shoup, D. C. (1995). "An Opportunity to Reduce Minimum Parking Requirements." *Journal of the American Planning Association* 61:14-28.

U.S. Department of Transportation, Federal Transit Administration (1999). *National Transit Database 1997.* Washington, DC: Author. <http://www.fta.dot.gov/ntl/database.html>.

MATERIALS

Energy and materials were developed together in early civilizations, beyond the use of fire for cooling and heating. In the courtyard of the stepped pyramid at Sakahra, built near Cairo, Egypt, in about 2600 B.C.E., there are a series of carved relief panels in stone, showing that the ancient Egyptians had mastered the smelting and working of metals with heat energy (fire), as well as many other technical skills. Glass melting was discovered at least 10,000 years ago, and fired ceramics (pottery) even much earlier.

Solid materials are essential to the production and transmission of energy; in the next section gives examples of materials used in these processes. The article then focuses on the large energy requirements needed to produce the metals, ceramics, glasses, and electronic materials (silicon and germanium) that our technological civilization demands, and finishes with an overview of the environmental problems encountered in trying to satisfy this vast demand for materials.

MATERIALS IN AUTOMOBILES AND TRUCKS

Advances in materials science have been a leading factor contributing to improvements in vehicle fuel economy and performance since the 1973 oil embargo. Better tire materials have reduced rolling resistance, better materials for moving parts have reduced friction, and most importantly lighter body parts have reduced the overall weight of vehicles.

The average 1993 model American car weighed about 3,200 pounds (1,450 kg) and delivered 28 miles per gallon of fuel (171 km/liter), representing a 19 percent reduction in weight and a 44 percent reduction in fuel consumption as compared to the average 1975 model. This reduction was achieved by substituting polymers for interior metallic trim, aluminum engines for cast iron, and careful design attention to weight factors. Weight reduction and improved engine design contributed greatly to better fuel economy. There is further potential for weight reduction through increased substitution of polymer matrix composites, aluminum, and ceramics for metals in vehicles.

Metal	Form or Process	Energy 10^{12} J/kg
Iron	Steel	28
Bismuth	Bulk	30
Lead	Ingot	31
Thallium	Sponge	36
Zinc	Electrolytic	70
Sodium	Bulk	107
Copper	Refined Bar	128
Cobalt	Electrolytic	144
Nickel	Electrolytic	167
Cadmium	Electrolytic	178
Tin	Ingot	221
Aluminum	Electrolytic, ingot	284
Tantalum	Powder	289
Tungsten	Powder	402
Magnesium	Electrolytic	416
Mercury	Liquid	459
Titanium	Sponge	474
Indium	Bulk	590
Cesium	Bulk	612
Hafnium	Sponge	768
Zirconium	Sponge	1390
Silver	Bars	1710
Rhenium	Powder	3600
Beryllium	Cast	6100
Gallium	Electrolytic	13,500
Gold	Bars	68,400

Table 1.
Energy Required to Produce Purified Metals

MATERIALS IN ENERGY PRODUCTION, TRANSMISSION, AND STORAGE

Metals are of overwhelming importance in these applications. Steel, which is iron containing carbon and many different metallic additions, is still the most used metal. Aluminum is increasingly used in many applications because of its light weight and resistance to chemical attack. Copper is required for transmission lines, wiring, and generators because of its high electrical conductivity. There are many specialized uses of other metals: manganese, vanadium, and molybdenum as alloying elements to improve strength and chemical durability of steel; uranium and plutonium in nuclear reactors; silver in electrical contacts; and tungsten as filaments in lamps.

Ceramics, including concrete, are useful especially in structures, reactors, as refractories in combustion of fuels, and as nuclear fuel. Porcelain insulators on transmission lines are an example of a specialized application of ceramics.

Electronic materials are needed for computers and control devices; purified silicon is the basic material for these applications. In addition silica glass (SiO_2) is an insulator, aluminum an electrical conductor, and polymers are reactive materials for patterning in these devices. Control of every step of energy production and transmission is now completely dependent on electronics.

ENERGY USE: REFERENCES AND METHODS

The energy use data in the tables come from three Battelle-Columbus reports and an article by H. H. Kellogg on "Energy Considerations in Metals Production in the Encyclopedia of Materials Science and Engineering." The Battelle-Columbus reports have detailed descriptions of the processing of all the materials listed in the tables, with methods from mining through separation to purification, and cost estimates at each processing step. There is a remarkable amount of valuable information on each material in these reports. The section on metals also relied heavily on the article by Kellogg.

The estimates for energy use can be separated into the following components:

$$\text{energy use} = F + E + S - B$$

where F is the heating values of fuels used, E is the fuel equivalent of electrical energy (from the U.S. average fuel equivalent of $11.1(10)^6$ joules per kilowatt hour), S is the fuel equivalent of supplies and chemical reagents consumed in the processing, and B is the fuel equivalent of by products and surplus steam. The units of energy use are given in joules per kilogram of final material produced in the tables; the conversion to English units is: divide J/kg by 1.16 to get Btu/ton (British thermal units per 2,000 pounds).

The energy equivalent of one barrel (159 liters) of crude petroleum is about $6.6(10)^9$ joules. Table 1 shows that the production of a metric ton (1,000 kg or 2,200 pounds) of steel requires about four barrels of oil; a ton of aluminum requires about forty barrels

Metal	*Energy 10^{12} J/kg*
Manganese	58
Chromium	71
Silicon	89
Niobium	220
Vanadium	570

Table 2.
Energy to Produce Metal for Alloying with Iron (Ferrometals)

of oil, and a ton of gold more than ten thousand barrels of oil. Table 2 shows the energy needed to produce metals for alloying with iron to produce steel.

ENERGY FOR MINING

The amount of energy required for mining the ores and minerals needed to make materials depends on their depth in the ground, and processing and separation methods. Ore and minerals lying near the surface need only be excavated by shovel or dredge, and thus require low energy expenditure (10^{11} J/kg). Fine grinding (0.1 mm) requires up to $3(10)^{11}$ J/kg. Loading, elevation out of a mine, and transporting the ore can require up to $5(10)^{11}$ J/kg.

If the ore consists of separate grains containing the desired material, it can be separated from undesired minerals by physical methods such as flotation, sedimentation, or magnetic separation. For metals this step can lead to 80 to 95 percent concentration of the value of the ore. Ceramic raw materials such as sand and clay can often be found pure enough in nature so that no concentration is needed.

If the desired material is not in separate grains, chemical treatment of the ore is required for metals, and for purification of ceramics.

ENERGY USE FOR PROCESSING OF METALS

Table 1 shows estimated values of energy to produce a kilogram of reasonably pure metals, or for metal useful for practical applications. The range of a factor of more than 2,000 between steel and gold depends on the concentration in ore, the chemical processing needed, and amount of technology development for the particular metal. Steel production has been developed during about three millennia, and iron ore is highly concentrated and can be reacted directly to steel alloys. Gold is widely dispersed in low concentrations, and requires intensive chemical treatment of ores.

Metal	*Consumption in millions of kilograms*	*Fraction as scrap*
Iron and steel	97	0.35
Aluminum	4.0	0.34
Copper	2.9	0.45
Lead	1.1	0.55
Zinc	0.95	0.29
Nickel	0.13	0.27
Platinum group	0.10	0.12
Gold	0.087	0.57
Magnesium	0.08	0.09
Titanium	0.036	0.34
Silver	0.033	0.72

Table 3.
Total Consumption of Certain Metals in the United States in 1981

Table 3 shows the total consumption of some metals in the United States.

If the metallic compound in the ore can be selectively leached by acid or base without dissolving much of the remaining ore, then the energy requirement is only about 10^{12} J/kg. Examples are leaching of oxide ores of copper, zinc, or uranium with sulfuric acid.

Much more energy is needed if the entire rock matrix must be chemically dissolved to free the metallic compounds. Examples are the separation of aluminum oxide (Al_2O_3) from bauxite ore by dissolution by strong base at elevated temperature and pressure (Bayer process), with an energy requirement of about $8(10)^{12}$ J/kg, and smelting of nickel ores by heating to produce a molten nickel-iron alloy and oxide slag, with energy up to $(10)^{13}$ J/kg needed. These high energy requirements result from the fuels needed to heat furnaces or reactors to the high temperatures of these chemical reactions, and for the energy equivalent to make the chemical reagents employed, such as acids, bases and iron alloys.

The grade of metallic ore is the percentage of metal (native and chemically combined) in the ore. For example, high grade iron ore can contain up to 65 percent iron, whereas usual gold ores contain less than 0.001 percent gold. The weight of ore that must be processed is inversely proportional to the grade of the ore. For example, about $7(10)^{13}$ J/kg are required for mining and concentration of hard rock ore containing 0.6 percent to copper, and $4(10)^{16}$ J/kg are needed to mine and concentrate a gold ore containing 0.001 percent gold. Thus, the grade of ore is a major factor in energy consumption.

High grade ores are used first, and are already substantially depleted for most metals. With time one might expect the energy use in Table 1 to increase. Improved technology can offset this increase somewhat; belt conveyors in mines and computer control of grinding are examples. Ordinary rocks contain small quantities (parts per million by weight) of many different metals, and have been suggested as a source for rare metals and those with depleted ores. For example, many rocks contain about 0.01 percent copper; it would require about one thousand times the energy to recover copper from these ores as from presently-available ores (Table 1). The large energy demand precludes the use of these rocks for producing copper, because of the low price of copper. Gold, however, is more valuable, and would justify a larger energy input.

In the concentrated ores most metals are in chemical compounds, as oxides or sulfides. Reducing these compounds to the metallic state in the final stage in producing metal can be accomplished by chemical processes or electrolysis. Two examples of chemical reduction are

Copper sulfide

$$CuS + O_2 = Cu + SO_2 \tag{1}$$

Steel blast furnace

$$2Fe_2O_3 + 4C + O_2 = 4Fe + 4CO_2 + \text{heat} \tag{2}$$

where $2Fe_2O_3$ is hematite coke ore. Because the oxides and sulfides of many metals are stable, their chemical reduction is difficult, and they are reduced to metal by electrolysis (electrowinning); examples are zinc, aluminum, and magnesium.

The electrolytic processing of concentrated ore to form the metal depends on the specific chemical properties of the metallic compound. To produce aluminum about 2 to 6 percent of purified aluminum oxide is dissolved in cryolite (sodium alumino-fluoride, Na_3AlF_6) at about 960°C. The reduction of the alumina occurs at a carbon (graphite) anode:

$$2Al_2O_3 + 3C = 4Al + 3\,CO_2 \tag{3}$$

Magnesium is reduced from a mixture of magnesium, calcium, and sodium chlorides. Electrolysis from aqueous solution is also possible: zinc, copper, and manganese dissolved as sulfates in water can be reduced electrolytically from aqueous solution.

Process energies are found by subtracting energies for mining and concentrating from the values in Table 3. The free energies of formation of the metal oxides are a measure of the total (theoretical) energy required to reduce the metal from the oxide. The ratio of the actual process energy to the free energy of formation is a rough measure of the efficiency of the reduction process. The free energies of formation are a measure of the chemical stabilities of the oxides; stable oxides such as aluminum, magnesium, and titanium intrinsically require more energy for reduction than from less stable oxides (or sulfides) of copper, lead, and nickel.

The most efficient processes in Table 1 are for steel and aluminum, mainly because these metals are produced in large amounts, and much technological development has been lavished on them. Magnesium and titanium require chloride intermediates, decreasing their efficiencies of production; lead, copper, and nickel require extra processing to remove unwanted impurities. Sulfide ores produce sulfur dioxide (SO_2), a pollutant, which must be removed from smokestack gases. For example, in copper production the removal of SO_2 and its conversion to sulfuric acid adds up to $8(10)^{12}$ J/kg of additional process energy consumption. In aluminum production disposal of waste cryolite must be controlled because of possible fluoride contamination.

As global warming develops, the formation of large quantities of carbon dioxide (see Equations 1 and 3) may become a problem in metals production. There is no simple or inexpensive way to reduce these emissions. Switching to electrolysis processing of steel, or a different electrode reaction for aluminum, would involve unacceptably large energy use and cost. Chemical absorption of carbon dioxide may be necessary in a wide variety of chemical and energy-producing processes, at enormous cost.

The scale of production also influences efficiency. Small-scale batch processing for metals such as titanium, tungsten, and zirconium leads to higher energy use and costs.

RECYCLING OF METALS

Reuse of waste metals generated from metal fabrication and from discarded products (scrap) can save large amounts of energy, particularly for metals that have high energy use in production, such as aluminum. The low fractions of energy used to produce metals from scrap for aluminum, certain sources of copper, and nickel show the value of recycling these metals.

The purity of the scrap mainly determines the fraction of energy needed to produce metal from it, and the value of recycling. Clean copper scrap need only be remelted and cast to form recycled copper; if the copper is contaminated with organic materials and other metals, more complex separation processes are needed that are similar to production from ores. It is easier to remelt the steel of a car driven in Arizona compared to one rusted by the road salt in snowy areas. Scrap that is produced as a by-product of metal processing can be easily recycled, and it can be collected from relatively few locations. There has been a strong effort to educate both householders and industrial users to separate scrap and return it to waste collectors, leading to a supply of reasonably separated scrap.

Despite the efforts of many communities to encourage recycling, there is still a large amount of metal that is not recycled. Only an estimated 30 percent of aluminum is recycled, as compared with up to 50 percent for precious metals. Landfills contain large amounts of metals, especially large use metals (Table 3) such as iron, aluminum, and copper, and more metals continue to accumulate in landfills. As the cost of disposal increases and ores of metals such as copper become of lower grade, it may be economically feasible to “mine” landfills for metals. Development of new technologies for treatment and separation of waste materials is needed to make this mining economical.

ENERGY USE IN PRODUCING CERAMICS

Traditionally ceramic raw materials have been dug out of the ground and used with little or no treatment or purification. Sand, fireclay, talc, and gypsum are examples. The energy expenditure for producing these materials is therefore small. Some of these materials can be found naturally in high purity. Silica sands (SiO_2) with less than 100 ppm (parts per million by weight) of impurities are known, and some clay deposits are nearly pure kaolin. Minerals such as feldspar, kyanite, and kaolinite (clay) can be purified by washing or solution treatments at near ambient temperatures, with low energy expenditure.

Many ceramic products require firing at high temperatures, and the fuels required to reach and sustain these temperatures are major factors in the energy consumed to make these products. Portland cement is made by firing a mixture of compounds, mainly carbonates, sulfates, and silicates to form the desired calcium silicate products. The firing is done in a rotary furnace or kiln, so that a fraction of the raw materials become liquid. As the resultant calcium silicates cool, they go through a large volume change that causes the cement particles to break into smaller sizes.

Concrete is made from a mixture of about equal parts of sand, gravel, and cement, plus some added water to give a mixture that flows. The low energy expenditures to make these raw materials mean that concrete is a material that requires very low energy; the only additional energy is a small amount for transport and for mixing the constituents. Concrete requires about one-third the energy expenditure for steel, and one thirtieth that for aluminum (Table 1). In Western Europe concrete has replaced metals and wood in many applications, because forests are depleted and energy costs are higher than in the United States. Examples are in building; American homes still use a wood frame, but in Western Europe almost all homes are made from concrete or stone. Electrical transmission poles in the United States are made of wood or aluminum, and in most of the rest of the world these poles are made of concrete because of their lower energy requirements. Considerable energy savings could occur by substituting concrete for metals in a variety of applications. Concrete has excellent compressive strength but is weak in tension or bending. By reinforcing concrete with steel bars the concrete building or structure has good strength in bending and tension as well as compression.

Other applications of ceramics require clay, either raw or purified, sand, and feldspar. Brick, porcelain, and white wares are made from these raw materials; the main expenditure in making these products is in firing the mixtures of powders to a dense solid. Ordinary brick made from fire-clay requires a small amount of energy; even refractory brick for high temperatures and chemical durability, made partly from purified oxides such as alumina or chrome ore, requires only about the same energy to make as an equivalent weight of steel.

Glass for containers is made continuously in a large tank or furnace. Raw materials (sand, soda-ash, limestone) are fed in at one end of a gas-fired furnace. The molten glass slowly passes through the furnace at high temperature (1,200°C-1,300°C) to homogenize it and remove bubbles. At the exit end the molten glass is fed into special molds in two stages to make containers of desired sizes and shapes. Flat glass (windows) is also made in a continuous furnace; a glass layer leaving the furnace is spread onto a bath of molten tin (float glass) to provide smooth surfaces. Lamp bulbs are made from a continuous furnace; a ribbon of glass is fed to blowers that blow the bulbs into a mold (ribbon machine). All these processes are continuous with large furnaces for melting, and so are energy efficient, using about $20(10)^{12}$ J/kg, a factor more the ten lower than the energy required to make an equivalent weight of aluminum. Nevertheless, aluminum containers have replaced glass for many purposes, because aluminum is easier to handle, and harder to break. Polymer (plastic) containers are also popular because of low cost, chemical durability, and ease of handling. As energy costs increase, aluminum containers will become less attractive than glass; the raw material (petroleum) for polymers may also become more expensive, leading to a return of glass as the primary container material.

For many specialized uses glass is made in small batches, so the energy costs are much higher than for the continuous furnaces. Special processes, such as for drawing fibers, casting optical components, and making laser glass, require highly purified or controlled raw materials, leading to much higher energy requirements than for continuously made glass.

Many ceramic applications are high value and small volume, so energy expenditure is high. Ferroelectric magnets, electronic substrates, electro-optics, abrasives such as silicon carbide and diamond, are examples. Diamond is found naturally, and made synthetically by the General Electric Company at high pressure and temperature. Synthetic diamonds for abrasives require less energy to make than the value in Table 4; nevertheless, the market is carefully divided between natural and synthetic diamonds.

Large quantities of uranium oxide are required for nuclear reactor fuel. The uranium ore must be carefully purified and processed to desired shapes, causing high energy expenditure.

Single crystals of synthetic quartz are made by crystallization from aqueous solution at temperatures and pressures well above ambient. The crystallization is slow and carefully controlled, so energy costs are high.

The energies for producing some gases are listed in Table 5 for comparison with those for other materials.

RECYCLING OF CERAMICS

Bulk ceramics such as building materials, porcelain, and concrete are not recycled, because of the low energy required to make them and the difficulty of collecting, transporting, and reforming them into useful shapes. Some glass is recycled in the form of "cullet," which is waste glass. The amount of cullet in a glass furnace is rigidly controlled, because the final product of the furnace must have just the right viscosity for the automatic machinery (container mold, tin bath, or ribbon machine) that forms the glass. Glass manufacturers are unwilling to build tanks to accept waste glass, because its variable composition leads to uncontrollable variations in the viscosity of the glass. Viscosities of silicate glasses are highly sensitive to impurities, especially water and alkali (sodium and potassium) compounds.

ELECTRONIC MATERIALS

Silicon wafers are the basis for electronic circuits. The silicon must be highly purified, then grown as a single crystal containing a small amount (a few parts per million) of additions to give either negative carriers (electrons from phosphorous or arsenic) or positive carriers (holes, from boron or aluminum). These processes require temperatures above the melting point of silicon (1,414°C) and careful control of several processing steps. The energy expenditure for making silicon for wafers (chips) is about the same as that for germanium, given in Table 4. The energy of about $2{,}500(10)^{12}$ J/kg is greater than that required to make such a valuable metal as silver, or zirconium, which is strongly bonded in compounds, because of the highly complex processing and high purity required for the semiconductors. Subsequent processing of silicon wafers to form devices on the wafers for practical use is highly specialized, carefully controlled, and expensive in cost and energy.

Material	*Form or process*	*Energy 10^{12} J/kg*
Sulfur	Frasch	2.2
Antimony	Bulk	7
Graphite (carbon)	Refined, bulk	40
Tellurium	Bulk	96
Phosphorous	Bulk elemental	200
Selenium	Bulk solid	340
Iodine	Solid	620
Germanium	Semiconductor grade	2500
Diamond	Natural	830,000

Table 4.
Energy to Produce Semi-Metals and Semiconductors

Gas	*Energy 10^{12} J/kg*
Oxygen	4.2
Argon (liquid)	4.9
Bromine	17
Chlorine	21
Ammonia	45

Table 5.
Energy to Produce Commercial Gases

ENVIRONMENTAL CONCERNS

There is great interest in the more energy efficient production of materials to reduce costs and environmental damage. The products of materials production receiving the greatest attention are sulfur dioxide, fluorides, and carbon dioxide. The production of energy from fossil fuels, especially coal and oil, leads to production of sulfur dioxide, which causes much damage locally and at long distances. It leads to respiratory problems and damage to plants, especially trees, and can acidify soils and lakes, damaging them for growing plants and animals. Sulfur dioxide can be scrubbed from flue glass at considerable expense, but much of it is still discharged into the atmosphere. Sulfur dioxide is a by-product in much of the materials production discussed here; sulfide ores (copper) when oxidized produce sulfur dioxide (Equation 2), and some raw materials for cement contain sulfates.

Fluorides are used in many materials processes, and can poison the environment when they are discarded. Examples are cryolite (sodium aluminofluoride, Na_3AlF_6) used to dissolve aluminum oxides for electrolysis, and hydrofluoric acid (HF) used in etching lamp bulbs and semi-conducting circuits. Today lamp bulbs are etched much less than they used to be to reduce fluoride disposal; not much has been done to reduce the amount of cryolite for aluminum production.

Some heavy metals and semi-metals are quite toxic (chromium, lead, and antimony) and expensive care is needed to prevent them from being dispersed in the environment. Lead in gasoline and paint has been almost completely eliminated; its use in storage batteries has resisted efforts to find a suitable substitute.

The discharge of carbon dioxide from combustion of fuels from vehicles, and from processes such as steelmaking, cement production, and much other materials production has increased the concentration of carbon dioxide in the atmosphere. Some computer models demonstrate that this increase is responsible for an increase in the mean temperature of the surface of the Earth, and there are numerous predictions of further temperature increases as more carbon dioxide is discharged into the atmosphere. There are other claims that this result is not proven. Reduction of carbon dioxide emissions is highly difficult and expensive. If the connection between carbon dioxide emissions and global warming is proven more conclusively and a carbon dioxide reduction plan is instituted, materials industries will feel a great impact because they consume about 20 percent of all industrial energy.

Reduction of overall energy use is one solution to the above problems. It requires money, technical advances, political power, and courage; some reduction has been achieved, but much more is needed to reduce emissions of gases. One solution being advanced is use of processes to produce energy that do not emit gases. Hydropower has been exploited about as fully as possible, and supplies only a small fraction of total energy needs. Other sources such as wind and solar power are still much too expensive.

One energy source that first appeared to be highly attractive was nuclear power. The problem with nuclear power is that some costs were hidden in its initial development. Especially pernicious is the disposal of uranium oxide fuel after it has become depleted. It can be reprocessed, but at considerable expense, and the product plutonium can be used for weapons. In the United States the plan is to bury depleted uranium from reactors, but many persons are not convinced that burial is safe. Much work has been done on encapsulation of radioactive waste in glass; the problem of reactor waste remains.

SUMMARY

The energy required to produce materials varies widely, gold requires more than two thousand times the energy to produce the same weight of steel, and diamonds two hundred thousand times the energy required to make ordinary brick. Factors in energy use are the quality (concentration) of ore, the complexity of processing, and the technological development of processing.

Some recycling of metals occurs; much more is possible, and substitution of materials requiring less energy for those requiring more has much potential.

Reduction of environmental pollution requires lower energy use and new technology to decrease emission of gases such as sulfur dioxide and carbon dioxide, and to prevent toxic fluoride, heavy metal, and radioactive wastes from discharging into the environment.

Robert H. Doremus

See also: Building Design; Climatic Effects; Drilling for Gas and Oil.

BIBLIOGRAPHY

Battelle-Columbus Laboratories, Energy Use Patterns in Metallurgical and Nonmetallic Mineral Processing. (1975). *High Priority Commodities*, PB 245 759; (1975). *Intermediate Priority Commodities* PB 246 357; (1976). *Low-Priority Commodities*, PB 261 150.

Doremus, R. H. (1994). *Glass Science*. New York: Wiley.

Fine, H. A., and Geiger, G. H. (1993). *Handbook on Material and Energy Balance Calculations in Metallurgical Processes*. Warrendale, PA: TMS.

Kaplan, R. S., and Ness, H. (1986). “Recycling of Metals: Technology.” In *Encyclopedia of Materials Science and Engineering*, ed. M. B. Bever. Cambridge, MA: MIT Press.

Kellogg, H. H. (1986). “Energy Considerations in Metals Production.” In *Encyclopedia of Materials Science and Engineering*, ed. M. B. Bever. Cambridge, MA: MIT Press.

Kingery, W. D.; Bowen, H. K.; and Uhlmann, D. R. (1976). *Introduction to Ceramics*. New York: Wiley.

Mayer, J. W., and Lau, S. S. (1990). *Electronic Materials Science*. New York: Macmillan.

Shackelford, J. F. (1997). *Introduction to Materials Science for Engineers*. New York: Macmillan.

MATTER AND ENERGY

The entire observable universe, of which the Earth is a very tiny part, contains matter in the form of stars, planets, and other objects scattered in space, such as particles of dust, molecules, protons, and electrons. In addition to containing matter, space also is filled with energy, part of it in the form of microwave radiation.

INERTIA, MASS, AND ACCELERATION

Matter itself has energy, called “rest energy.” What distinguishes matter-energy from other forms of energy is that all matter has inertia and is subject to the force of gravity when at rest as well as when in motion. Inertia measures the resistance of an object to being accelerated by a force, and the inertia of an object at rest is proportional to its mass.

According to a law of physics first formulated by Isaac Newton and later modified by Einstein in his general theory or relativity, any object with mass can be accelerated by applying a force to it. Physicists use the term “acceleration” not only to describe the speeding up or slowing down of an object but also for changing its direction. A car going around a curve at constant speed is accelerating because its direction changes.

If you flick a small plastic ball on a table top with your finger, thereby exerting a small force on the ball, you will see it move rapidly from its resting position. But if you do the same with a steel ball of the same size, the same flick of the finger (the same force) will produce noticeably less motion. The steel ball has greater mass and therefore greater resistance to being accelerated. The ratio of the accelerations of two objects experiencing the same force is equal to the ratio of their masses.

The basic unit of mass is the kilogram, which is the mass of a standard platinum cylinder located in the city of Paris. A kilogram has a weight of 2.2 pounds. The basic unit of energy is the joule, which is equal to the kinetic energy that a one-kilogram object has when it is moving at a speed of 1.41 meters per second or the amount of potential energy the object has when lifted to a height of 0.102 meters.

PROPERTIES OF MATTER

Matter on earth commonly takes the form of solids, liquids, or gases, but may also be in the form of plasmas, which are "ionized" gases, that is, gases in which some of the atoms of the gas have lost one or more of their electrons. These electrons move within the plasma.

Gases, liquids, and solids have different physical properties. A gas fills its container, so that if a certain amount of gas is transferred from a small container into a large one, the gas will expand to fill the new container. If there is a hole in the top of a container filled with gas, the gas will escape. A liquid keeps the same volume when transferred from one container to another, but takes the shape of the new container. On Earth, a liquid has a flat, horizontal surface, If there is a hole in its container below that surface, the liquid will spill out. A solid keeps both its shape and its volume when transferred from one container to another.

Solids, liquids, and gases all change their volumes when the temperature is changed. All gases and nearly all solids and liquids tend to expand when their temperature is raised. When heat is applied to a solid, its temperature normally goes up, but at a certain temperature it can change its state (phase) to a liquid while the temperature remains constant. Similarly, as heat is applied to a liquid, its temperature normally rises until a certain temperature, when it changes its state to a gas. On still further heating, the gas will expand if the container has a movable piston to let it expand; otherwise, the pressure of the gas will increase. Eventually, if enough heat is applied, the gas can become partially ionized, that is, it can turn into a plasma.

The sun is a partially ionized gas or plasma. It has no container to hold it together. Instead, the enormous gravitational forces in the sun do the job. Even on Earth, the atmosphere (a gas) does not escape to outer space because of gravity. On the surface of the moon, gravity is only one-sixth as strong as on the surface of the earth. This is not strong enough to hold gases on the moon, so the moon has no atmosphere.

According to Albert Einstein's theory of relativity, no object with mass can travel as fast as the speed of light in empty space (in vacuum). So another definition of matter is anything that is subject to gravity and is either at rest or traveling slower than the speed of light in vacuum. The speed of light in vacuum, denoted by the special symbol c, is a constant speed (186,000 miles per second, or 300,000 kilometers per second) and is the speed in vacuum of any quantum (packet) of energy that has no mass. (At this speed, light can travel seven times around the world in a little under a second.) Light, and anything else that travels at the speed of light in a vacuum, is not considered to be matter. However, all things that travel at the speed of light, including light itself, possess kinetic energy.

The theory of quantum mechanics says light has some properties of a wave (for example, a wavelength), but its energy is concentrated in little packets called photons. These quanta, or particles of light, do not have mass and always travel at the speed of light in vacuum. Light travels slower in a material medium (such as glass) than in a vacuum because the photons get absorbed and reemitted by the atoms of the medium, thereby slowing down the progress of the light wave.

Is the universe matter dominated or energy dominated? The very early universe was "energy dominated" in the sense that only a small fraction of the energy of the universe at that time was in the form of the rest energy of matter and antimatter. Much of the energy of the universe was in the form of electromagnetic radiation. It is not known whether the universe originally had an excess of matter over antimatter or whether the excess of matter developed as time went on. Most cosmologists favor the idea that reactions in the early universe led to the excess of matter, but the mechanism is not known. As the universe expanded it cooled, until, about 300,000 years after the Big Bang, the quanta of radiation (photons) no longer had enough energy to create electrons and positrons. As the universe continued to expand, the radiation cooled still further, until today all that is left of it is the microwave background radiation at a temperature of 2.7 K. The amount of energy in the background radiation plus all the light energy from stars is only a tiny fraction of the rest energy of matter. Therefore, today, the visible universe appears to be matter dominated. However, it is an open question whether other, as yet unseen, forms of energy exist that would make the universe energy dominated.

Forces, such as the gravitational force, are transmitted by means of quantities that scientists call "fields." All matter is influenced by forces carried by gravitational fields. Electrically charged particles are influenced by forces carried by electromagnetic fields. The gravitational and electromagnetic fields are not matter, but they have energy.

As far as is known, ordinary matter is made of tiny building blocks called elementary particles. For example, an atom is made up of a nucleus surrounded by one or more electrons. As far as scientists have been able to determine, the electrons are elementary particles, not made of anything simpler. However, an atomic nucleus is not elementary, but is a composite particle made up of simpler particles called protons and neutrons. (The lightest nucleus is the nucleus of ordinary hydrogen, which consists of only a single proton.) Today, physicists believe that even protons and neutrons are not elementary but are composite particles made up of still simpler building blocks called quarks.

At the present time, quarks are believed to be elementary particles. All the particles in an atom, whether elementary or not, are particles of matter and possess mass. Electrons, protons, and neutrons can also exist outside of atoms.

In addition to ordinary matter, scientists have evidence for the existence in the universe of "dark matter." Some of the dark matter is ordinary matter, such as dust in outer space and planets going around other stars. Astronomers cannot see ordinary dark matter because any light coming from such matter is too faint to be observed in telescopes. However, most of the dark matter in the universe is believed not to be ordinary matter. At the present time it is not known what this mysterious dark matter is, or what it is made of. Scientists know that this dark matter exists because it exerts a gravitational force on stars (which are made of ordinary matter), causing the stars to move faster than they otherwise would. According to present estimates, there is perhaps five times as much dark matter in the universe as ordinary matter.

GRAVITY

Gravity is a force that acts not only on matter but on anything, such as light, that possesses energy. A gravitational force cannot speed up light or slow it down, but it can accelerate light by changing its direction. Light from stars directly behind the sun can be seen on earth during an eclipse of the sun because the sun's gravity bends some starlight around it. An observation of this effect was first made in 1919 during a solar eclipse, and the amount of bending observed was in agreement with the predictions of general relativity.

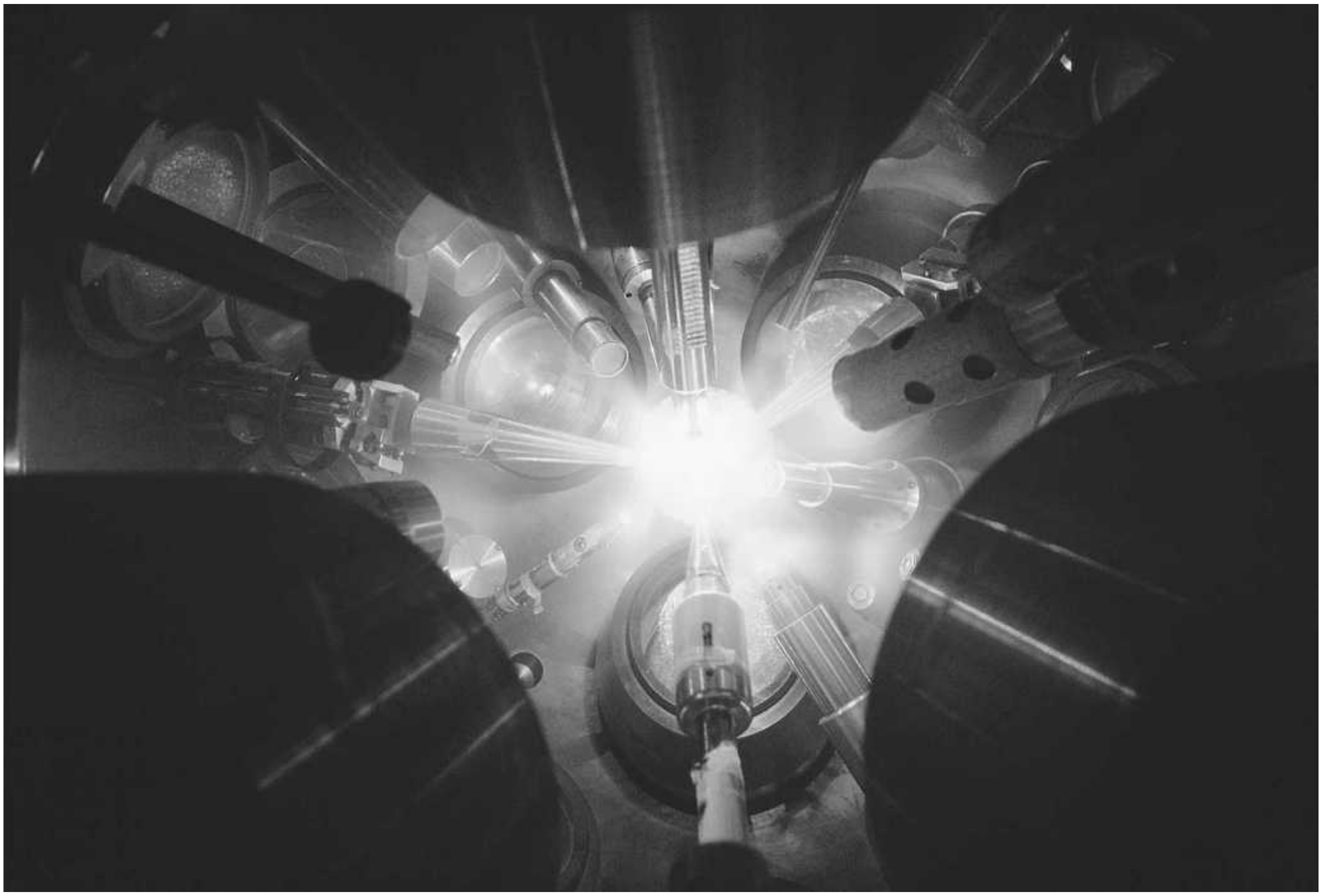
Gravity can also give energy to light or take energy away from light. If light from a laser is directed down from the top of a building to the ground, the light will gain a small amount of energy by "falling" in the gravitational field. Scientists can measure this slight energy gain as an increase in the frequency (decrease in the wavelength) of the light. Conversely, if the laser light shines upward, the light loses energy, and its frequency slightly decreases.

According to the theory of relativity, all matter has a kind of energy, called rest energy, denoted by the symbol E. If an object at rest has a mass denoted by m, its rest energy E is given by Einstein's famous formula $E = mc^2$. Because the speed of light c is such a large number, a small amount of matter contains a large amount of rest energy.

It has been noted that a one-kilogram object moving at 1.41 meters per second has a kinetic energy of 1 joule. The rest energy of the object is 90 million billion times as great. In fact, the rest energy of ordinary objects is so large that some people dream of unlocking that energy and converting it into more useful forms of energy, such as kinetic energy and heat. The laws of physics do allow matter to be converted into energy and energy into matter. However, at present, no way is known to convert the rest energy of matter entirely into energy except by "annihilation" in a collision with a form of matter known as antimatter.

PROPERTIES OF ANTIMATTER

Earth and the sun, and, as far as is known, the stars and planets in the rest of the visible universe, are made of ordinary matter. However, according to a theory first proposed by Paul Dirac in 1928, for every kind of particle of ordinary matter that exists in nature, there can exist an antiparticle made of antimatter. Some antiparticles have been discovered: for example, the antiparticle of the electron, called the positron, was discovered in 1932 in cosmic rays falling on earth and have also been created in experiments performed in the laboratory. Antimatter is very simi-



Lasers are focused on a small pellet of fuel. This is an attempt to create a nuclear fusion reaction for the purpose of producing energy. (Corbis-Bettmann)

lar in some of its properties to ordinary matter, while other properties are quite different. For example, an electron is a particle of ordinary matter with negative electric charge. However, although the positron has a mass equal to that of an electron, it has positive electric charge. The mysterious dark matter of the universe is not the antimatter of ordinary matter.

According to general relativity, both matter and antimatter are attracted by gravitational forces. However, as yet no experiment has succeeded in showing that antimatter falls under Earth's gravity. The reason is that only small particles of antimatter, such as antiprotons, antineutrons, and positrons, have been created in the laboratory. The electric and magnetic forces acting on these particles are much stronger than the gravitational forces and mask the effects of gravity.

The most spectacular difference between a particle and an antiparticle is that, as the result of a collision, the particle and antiparticle can both annihilate into pure energy. For example, if an electron and a positron collide, they may destroy each other (annihilate) into radiation. This is an example in which the rest energy of matter and antimatter is converted entirely into another form of energy. Conversely, under some conditions the kinetic energy of rapidly moving particles can be converted into new particles of matter, usually together with particles of antimatter. Because antimatter is rare in the universe, nobody has to worry about our earth colliding with enough antimatter to annihilate the earth, although some particles on earth are annihilated by antiparticles from outer space.

It is because of the rarity of antimatter that we cannot use annihilation of matter as a source of kinetic energy, heat, light, and other forms of energy. Of course, scientists can create antimatter, but they have to supply the energy to create it. When the created antimatter annihilates, the scientists get back only the energy that they put in. It is actually much worse than that, because creation of antimatter is a very inefficient process, and most of the input energy is wasted. Furthermore, it is very difficult to store antimatter. It cannot be stored in any container made of matter, as it will annihilate with the walls of the container. Antimatter has to be contained by electromagnetic forces in a vacuum.

CONVERTING MATTER TO ENERGY

Although it is impractical to convert the rest energy of matter entirely into other forms of energy, nevertheless, a small fraction of rest energy is converted in chemical and nuclear reactions. For example, if hydrogen is burned in oxygen (a chemical reaction), the product is water plus heat and light. A scientist can describe this process by saying that burning converts chemical energy into heat and light. However, the process can be looked at in another way. If careful measurements are made, it is found that the mass of the water is slightly less than the sum of the masses of the original hydrogen and oxygen. So it can also be said that "the burning process" converts a small amount of rest energy of the hydrogen and oxygen into heat and light.

Normally, electrons in an atom are "bound" in the atom by the attractive electrical forces between the electrons and the atomic nucleus. A certain amount of energy must be applied to an atom to release an electron from the atom, thereby ionizing it. This amount of energy is called the "binding energy" of the electron in the atom. The binding energy of the electrons in hydrogen and oxygen is slightly different from the binding energy of the same electrons when the oxygen and hydrogen are combined in water. A change in binding energy causes a change in rest energy of the same amount. This difference in binding energy (or rest energy) is the source of the heat and light when hydrogen is burned.

Just as electrons are bound in atoms, so are protons and neutrons bound in atomic nuclei, but the binding energy of the protons and neutrons is far greater. Consequently, changes in the binding energy in nuclei are a much greater source of heat and light than changes in the binding energy of electrons. More than a million times as much matter can be converted into energy in a nuclear reaction as in a chemical reaction, and even in such a nuclear reaction only about 0.1 percent of the matter is converted into energy.

Most of the energy of the sun comes from changes in binding energy when hydrogen is converted into helium in nuclear reactions. When very light nuclei, such as hydrogen nuclei, are combined to produce nuclei having less total mass than the very light nuclei, energy is released. The process is called "nuclear fusion." The energy released in the sun in nuclear fusion is what causes the sun to shine.

When very heavy nuclei, such as those of uranium and plutonium, are split into lighter nuclei having less total mass than the very heavy nuclei, energy is released. The process is called "nuclear fission." In either nuclear fission or nuclear fusion, much of the converted rest energy emerges as kinetic energy, heat, and light.

The explosion of an atom bomb (a uranium or a plutonium bomb) and the operation of a nuclear reactor are cases of energy released in nuclear fission, the first in a very fast process and the second in a slower, controlled way. In a nuclear reactor, a small fraction of the rest energy of the uranium or plutonium is converted into heat. The heat is then used to turn water into steam, which drives a turbine attached to an electric generator in order to generate electricity (electrical energy). In a hydrogen bomb, most of the energy released comes from nuclear fusion. Scientists have tried for almost fifty years to build a fusion reactor. Although scientists have been able to generate a small amount of heat by controlled fusion, they have not succeeded in generating large amounts of heat from controlled fusion in a profitable way.

Don Lichtenberg

See also: Einstein, Albert; Hydrogen; Molecular Energy; Nuclear Energy; Nuclear Fission; Nuclear Fusion; Thermodynamics; Units of Energy.

BIBLIOGRAPHY

Clark, J. O. W. (1994). *Matter and Energy: Physics in Action.* New York: Oxford University Press.

Einstein, A. (1961). *Relativity*, tr. R. Lawson. New York: Three Rivers Press.

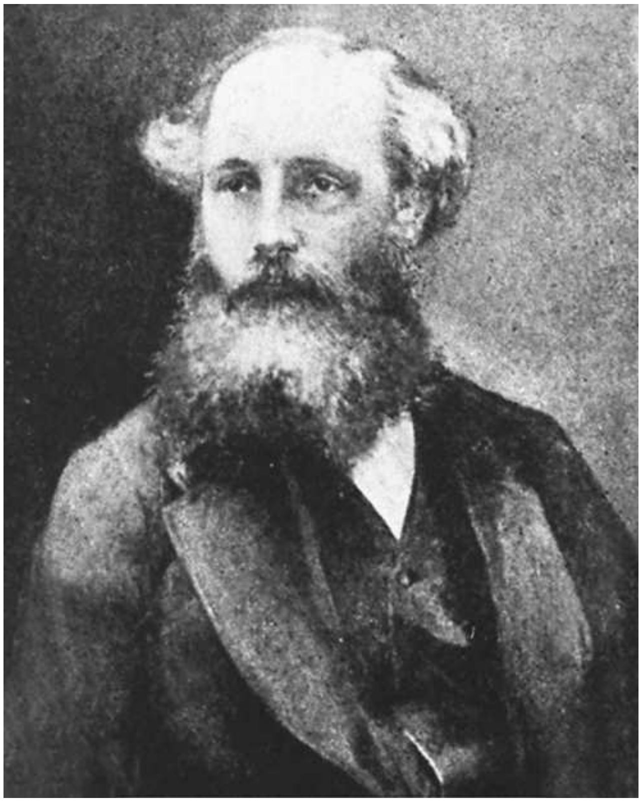
Schrödinger, E. (1953). "What Is Matter?" *Scientific American* 189:52-57.

MAXWELL, JAMES CLERK (1831-1879)

James Clerk Maxwell is the one theoretical physicist between Isaac Newton and Albert Einstein of a stature comparable to theirs. Maxwell's contributions to science ranged over many areas, of which the two greatest were his creation of the electromagnetic theory of light, and his work on molecular physics, gas theory, and statistical mechanics. He entered the scientific scene in the early 1850s, immediately after the principle of conservation of energy had been established. Its impact is seen everywhere in his work.

A descendant of a distinguished Scottish family, the Clerks of Penicuik, and, by an illegitimate line, of the ninth Lord Maxwell, he was born in Edinburgh but lived much of his life at his estate in Galloway in southwest Scotland, where he inherited 2,000 acres of rich farmland. From the ages of ten to nineteen he was educated in Edinburgh, entering the University of Edinburgh in 1847. At nineteen he went on to Cambridge University to take the rigorously severe mathematical *tripos*, from which he graduated in 1854 second in order of merit. He became Fellow of Trinity College in the following year and then, in 1856, at twenty-five, was appointed professor of natural philosophy at Marischal College, Aberdeen. In 1858 he married Katherine Mary Dewar, daughter of the principal. Though lacking any prior scientific training, she became an enthusiast in experimental research and worked closely with Maxwell on several experiments, first in color vision and then in physics. They had no children.

In 1860 Maxwell became a professor at King's College, London, where he served for five years. He retired from professorial life in 1866, at the age of thirty-five, to spend six years writing his famous *Treatise on Electricity and Magnetism* (1873). During the same time he also produced his small but important *Theory of Heat* (1871). In 1871 he was appointed Cavendish Professor of Experimental Physics at Cambridge and was responsible for designing and setting up the Cavendish Laboratory. Maxwell died of abdominal cancer in 1879 at age forty-eight.



James Clerk Maxwell. (Library of Congress)

ELECTROMAGNETIC THEORY

When Maxwell began studying electricity and magnetism in 1854, the field was in a state of confusion. The laws of electric and magnetic force had been established by Charles Augustin de Coulomb in the 1780s, and impressive mathematical structures had been built on them. However, the triumph was unsettled by Hans Christian Oersted's discovery in 1820 of electromagnetism—a peculiar twisting action exerted by an electric current on a magnet. This departure from Newtonian attractions and repulsions met two contrasting reactions. André Marie Ampère sought to reinterpret Oersted's force as a disguised form of attraction. Michael Faraday treated it as primary and related it geometrically to properties of lines of magnetic and electric force.

It is wrong to see Maxwell's achievement as one of merely translating Faraday's ideas into precise mathematical language. Though he once described Faraday as "the nucleus of everything electric since 1830," two other men, William Thomson (Lord Kelvin) and Wilhelm Weber, were equally influential.

From Faraday Maxwell gained a way of thinking; from Thomson, the first mathematizations of Faraday's ideas and several groundbreaking connections to the concept of energy; from Weber, the remarkable insight that the ratio of the two kinds of force, electrostatic and electromagnetic, somehow involves a velocity.

Between 1855 and 1868 Maxwell devoted great effort (five substantial papers) to clearing up the confusions in electromagnetism. The outcome was the dramatic discovery that light is an electromagnetic phenomenon, and the prediction—twenty-seven years before they were detected by Heinrich Hertz—of radio waves. Crucial was Maxwell's devising in 1861 of a speculative "ether" transmitting Faraday's lines of magnetic force. To his astonishment he found that this ether would transmit waves. Using some measurements by Weber and Friedrich Kohlrausch, Maxwell then calculated their velocity and found, to his even greater astonishment, that it was just equal to the velocity of light. Thus the great discovery was made and thus began the great intellectual metamorphosis, shaped by Maxwell and Einstein, in which the velocity of light was transformed from an isolated quantity into a universal fundamental constant influencing every part of physics.

The essence of Maxwell's later development of his theory was in the electromagnetic equations and the idea that electric and magnetic energies, instead of being located on charged bodies, are disseminated through space. That he could so quickly discard his ether model was closely related to the new doctrines of energy. Rather than attempt to explain light or electromagnetism in terms of a mechanism, Maxwell demonstrated that one set of unexplained equations describes both. Philosophically, the theory became a theory of relations. In this line of thought, Maxwell was strongly influenced by his mentor at Edinburgh, Sir William Hamilton, who held that all human knowledge is of relations rather than absolutes.

Maxwell's *Treatise on Electricity and Magnetism* (1873) covered every branch of the science and was a source of ideas and discoveries for fifty years to come.

GASES, MOLECULES, AND STATISTICS

In 1859 Maxwell, who had just completed a famous essay on the structure of the rings of Saturn, chanced to read a paper by Rudolph Clausius on gas theory. Maxwell had proved that the rings had to be composed of large numbers of independent bodies constantly colliding with each other. Clausius, expanding on earlier work by James Prescott Joule and August Karl Krönig, proposed that in a gas the rapidly moving molecules are constantly colliding. His interest at once aroused, Maxwell in a few months had written the first of several papers that created the modern kinetic theory of gases.

Maxwell's and Clausius's innovations were of two kinds, mathematical and physical. Mathematically, the key to dealing with large numbers of molecules was statistics, used not as a means of processing scientific data but as a fundamental explanatory idea. Clausius recognized that molecules must travel a certain average distance between collisions—the mean free path—but restrictively assumed that they all have the same speed. Maxwell transformed the discussion by introducing his velocity distribution function, giving the proportion of molecules traveling with a particular speed. Armed with this mathematical weapon, he could attack many previously intractable physical phenomena. He obtained theoretical formulas for viscosity, diffusion, and heat conduction in gases that then could be compared with experimental data. One startling consequence was that the viscosity of a gas should be independent of its pressure. When this was confirmed in independent experiments by Oskar Emil Meyer and by Maxwell and his wife, it added tremendous credibility to the theory.

The work on gas theory had many extensions. In 1865 Johann Josef Loschmidt used estimates of the mean free path to make the first generally accepted estimate of atomic diameters. In later papers Maxwell, Ludwig Boltzmann, and Josiah Willard Gibbs extended the mathematics beyond gas theory to a new generalized science of statistical mechanics. When joined to quantum mechanics, this became the foundation of much of modern theoretical condensed matter physics.

Through his famous "demon" Maxwell addressed one mystery of energy physics: the relation between the first law of thermodynamics, which states that energy as a whole is conserved, and the second law, which states that mechanical energy will be gradually dissipated. Maxwell was the first person to realize and forcefully argue that the second law is a statistical rather than a dynamical truth. Following this clue, Boltzmann in 1872 found the exact formal expression relating entropy to probability. Their work, together with earlier reflections by Kelvin, framed a discussion of irreversibility in physics, embracing even the nature of time, that has continued to this day.

C. W. F. Everitt

See also: Ampère, André-Marie; Clausius, Rudolf Julius Emmanuel; Electricity; Electricity, History of; Faraday, Michael; Gibbs, Josiah Willard; Magnetism and Magnets; Molecular Energy; Oersted, Hans Christian; Thomson, William.

BIBLIOGRAPHY

Brush, S. G. (1976). *The Kind of Motion We Call Heat*, Vols. 1-2. Amsterdam: North-Holland.

Buchwald, J. T. (1985). *From Maxwell to Microphysics.* Chicago: University of Chicago Press.

Campbell, L., and Garnett, W. (1882). *The Life of James Clerk Maxwell.* London: Macmillan.

Everitt, C. W. F. (1975). *James Clerk Maxwell, Physicist and Natural Philosopher.* New York: Scribner.

Harman, P. (1998). *The Natural Philosophy of James Clerk Maxwell.* Cambridge, Eng.: Cambridge University Press.

Siegel, D. M. (1991). *Innovation in Maxwell's Electromagnetic Theory.* Cambridge, Eng.: Cambridge University Press.

Whittaker, E. T. (1954). *History of the Theories of Aether and Electricity*, Vols. 1-2. New York: Philosophical Library.

Julius Robert Mayer. (Library of Congress)

MAYER, JULIUS ROBERT (1814–1878)

In the early nineteenth century many scientists had glimmerings of the conservation-of-energy principle. The three most important among these were the Frenchman Marc Séguin, the American-born, well-travelled soldier of fortune Benjamin Thompson, and the chief engineer of the city of Copenhagen, Ludwig Colding.

The three men whose work later in the nineteenth century was crucial in bringing clarity to this principle were two Germans, the physician Julius Robert Mayer and the great polymath Hermann von Helmholtz, and the British amateur scientist James Joule. In a lecture delivered by Helmholtz on February 7, 1854, in Königsberg on "The Interaction of Natural Forces," he referred to Mayer as "the founder" in 1842 of the principle of conservation of energy and acknowledged Mayer's priority in this discovery over Colding (1843), Joule (1843), and Helmholtz himself (1847). Rudolf Clausius agreed with Helmholtz and put Mayer in touch with the British physicist John Tyndall, who quickly became Mayer's English champion in his long-drawn-out priority dispute with Joule and his British supporters, William Thomson (Lord Kelvin) and George G. Stokes.

MAYER'S LIFE AND CONTRIBUTIONS TO SCIENCE

Julius Robert Von Mayer was born in 1814 in Heilbronn, a small town on the Neckar river, halfway between Heidelberg and Stuttgart. Interested in science as a youth, he decided on a career in medicine, and in 1832 began his medical studies at the University of Tübingen. After completing his studies there in 1838, he received his M.D. degree.

In February 1840, Mayer embarked for a year as physician on a freighter carrying cargo to Jakarta in the East Indies, just south of the equator. In Jakarta he noticed that sailors' venous blood was a much brighter red than he had observed in patients back in Germany. He surmised that this change was due to the hot climate and the reduced oxidation needed to preserve normal body temperature. This stimulated him to thinking more generally about how heat affects human metabolism. This interest in heat, work, and what is now called energy became the passion of Mayer's life.

Mayer returned to Heilbronn in 1841, began his medical practice, and eventually became chief surgeon of the town. In his free time he did some experiments and struggled with difficult, abstract concepts in an attempt to understand the nature of energy. He knew so little physics from his one semester of the subject at Tübingen that many of the papers he submitted for publication were rejected as incompetent by the important scientific journals of the day. He was forced to publish most of his writings at his own expense, and so their circulation was confined primarily to Heilbronn residents.

The first published results of Mayer's work appeared in the March 1842 issue of *Liebig's Annalen der Chemie und Pharmacie* in an article entitled "Remarks on the Forces of Inanimate Nature." It contained a great deal of philosophy and some inaccurate science, and yet buried among these distracting elements was the essence of the conservation-of-energy principle. Perhaps more important to scientists, the article contained the first quantitative attempt to determine the relationship of the "calorie" (a unit of heat) to what is now called the "joule" (a unit of energy).

Mayer's calculations were based on his conviction that there is a definite quantitative relationship between the height from which a mass falls to the ground and the heat generated when it strikes the ground. Here he was stating his conviction that energy is conserved in this process if heat is considered a form of energy. He calculated the ratio of the joule to the calorie, and found that it was equal to 3.59 J/cal, which differs considerably from today's accepted value of 4.18 J/cal. His whole approach to the problem was correct, however, and his numerical result was certainly of the right order-of-magnitude, but he had used inaccurate values for some constants needed in his calculation.

Few details of his research were given in Mayer's 1842 paper, but in 1845 he published (at his own expense) his most original and comprehensive paper, "Organic Motion and Its Relation to Metabolism," in which he gave full details of his earlier work. In 1848 he also published, again privately, "Contributions to Celestial Dynamics," in which he made the interesting conjecture that the source of the energy radiated by the sun was its constant bombardment by high-energy meteors.

Mayer had married in 1842, and for the first few years his marriage was a very happy one, but then his life began to fall apart. Between May 1845 and August 1848 three of his children died. A nasty priority controversy with Joule became public when their claims were read and discussed by the Academy of Science in Paris. Mayer was upset that his writings on energy conservation were not more widely read, and that they often were not appreciated by the few scientists who did read them. And, as the final blow, he was accused by local scientists of being more a mad philosopher than a competent scientist.

Finally on May 28, 1850, in a fit of despair, Mayer threw himself out of his bedroom window to the street thirty feet below, but escaped without serious injury. He spent three years in mental hospitals and did little scientific work of value after his release in 1853, although he was able to return to limited service as a physician in Heilbronn.

Physicists around the world gradually came to appreciate Mayer's scientific work, but by this time they were unsure whether he was still alive and, if so, what his mental condition was. In his later years he finally reaped some fruit from his scientific labors. In 1859 he received an honorary doctorate from the University of Tübingen. This was followed in 1871 by his reception of the Copley Medal from the Royal Society of London, and then the Prix Poncelet from the Paris Academy of Sciences. It is unknown how appreciative Mayer was of this belated notoriety when he died of tuberculosis in 1878.

Joseph F. Mulligan

BIBLIOGRAPHY

Caneva, K. L. (1993). *Robert Mayer and the Conservation of Energy*. Princeton, NJ: Princeton University Press.

Friedländer, S. (1905). *Julius Robert Mayer*. Leipzig: T. Thomas.

Kuhn, T. S. (1977). "Energy Conservation as an Example of Simultaneous Discovery." In *T. S. Kuhn: The Essential Tension*. Chicago: University of Chicago Press.

Lindsay, R. B. (1973). *Robert Mayer, Prophet of Energy.* Elmsford, NY: Pergamon Press.

Schmolz, H., and Weckbach, H. (1964). *Robert Mayer, Sein Leben und Werk in Documenten.* Heilbronn: Anton H. Konrad Verlag.

Turner, R. S. (1974). “Mayer, Julius Robert.” In *Dictionary of Scientific Biography, Vol. 9*, ed. C. C. Gillispie. New York: Scribner.

MECHANICAL TRANSMISSION OF ENERGY

Mechanical devices are used to magnify the applied force (mechanical advantage), to magnify the distance moved, or to change the direction of the applied force. They of course cannot decrease the amount of work (force × distance) necessary to do a job; they only make it more convenient to do it. In many cases, without a machine, the job would be impossible.

There is generally considered to be five distinct simple machines: lever, wedge, wheel and axle, pulley, and screw. The transmission of energy by these simple machines is so basic that people use them with little understanding of the physical principles involved. Most learn their use intuitively, through experience, and consider their application just plain common sense.

THE BASICS

The history of the origin of simple machines is largely conjectural, but there also exists documentation of the ancient Egyptians using simple machines to build pyramids nearly 5,000 years ago. An inscription in a 4,000-year-old tomb tells of 2,000 men pulling a statue estimated at 132 tons into place. The mass of the 2,000 men would be about the same as the mass of the statue, and it would probably take that many because they moved it on sledges without wheels.

The use of simple machines has sometimes been taken as a definition of what separates humans from animals; however, some primates have been observed fashioning probes out of sticks to pry out or to reach food. One of the most powerful images depicting the use of tools as defining humanity is the opening scene in the movie *2001*. An ape has discovered the club and is bashing some bones. One of the bones flies upward and in slow motion morphs into a spaceship. The club or the hammer is such a basic tool that it does not even make it into the classical listing of the five simple machines. However, it is also a mechanical device that multiplies force and transmits energy.

In the transmission of energy by these simple machines, the conservation law always applies: The work input equals the work output. When work is done *by* a system, energy is transferred out of it; and when work is done *on* a system, energy is transferred into it. When two objects interact by way of a machine (e.g. a lever), the work out of one object equals the work into the other. The work done by a person forcing one end of a lever downward equals the work done lifting a load at the other end as the lever moves upward. In any practical situation, the frictional forces resisting motion will always increase the amount of force (and work) required to do a job.

The amount of work done on an object is determined by the force exerted on it multiplied by the distance it moves in the direction of the force. Therefore the key to figuring out how much the force is magnified by a simple machine is to compare distances moved. For example, if the end of a lever under a stone weighing 2,000 newtons moves upward 1 meter, the amount of work done lifting the stone is 1 meter × 2,000 newtons = 2,000 joules. An equal amount of work must be done by a person on the other end of the lever. If that end of the lever is pushed downward 2 meters, then the person needs to apply a force of only 1,000 newtons to do the same amount of work (2,000 joules) and lift the 2,000 newton stone.

In brief, the mechanical advantage of a lever or any machine equals the ratio of the distance the applied force moves to the distance the load moves.

THE LEVER

The lever is such a part of everyday activity that its application usually requires no conscious thought: the pop top on a soda can, a doorknob, a wrench, pliers, a fishing pole, a faucet with a handle that lifts, a wheelbarrow, fingernail clippers, and so on. The crank and winch used to pull a heavy boat out of the water up onto a trailer can be thought of as a lever in circular form.

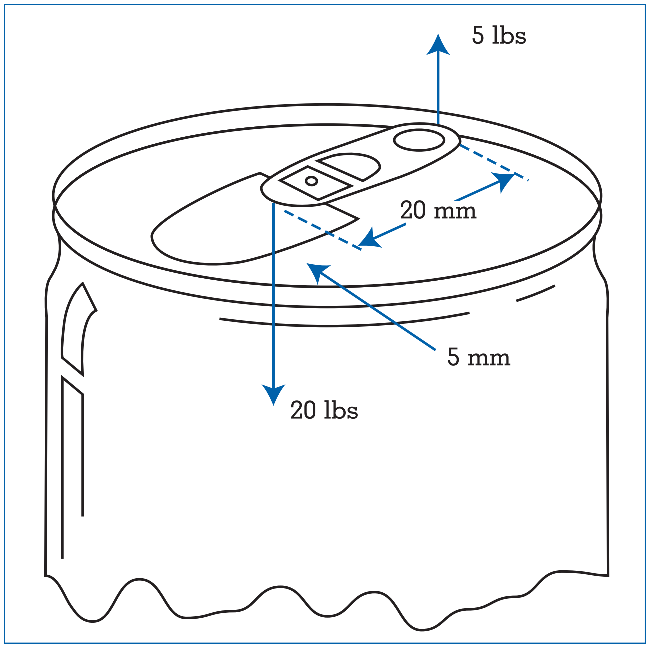


Figure 1.
Standard design of a can with a pop top.

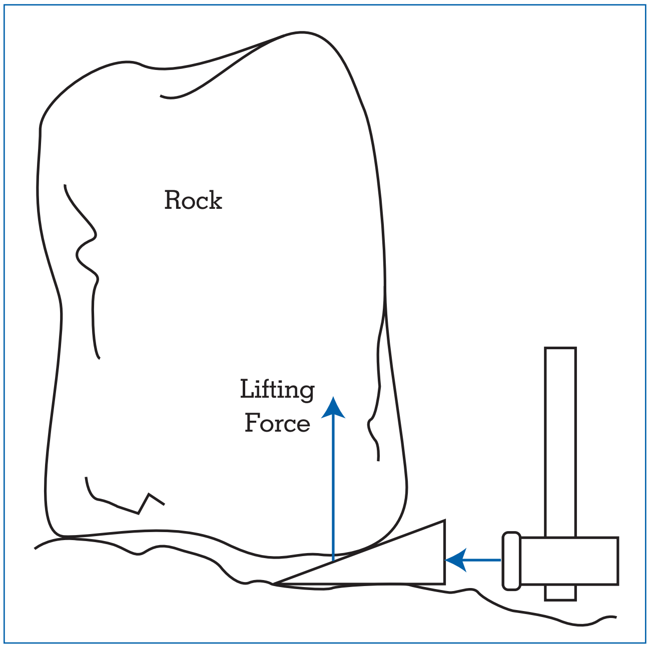


Figure 2.
Forces involved in the use of a wedge and hammer.

Basically, a lever is a solid object with an axis about which it rotates (fulcrum). As the lever rotates about its fulcrum, a point on the lever farther from the fulcrum moves a greater distance. The conservation of energy applied to the lever results in the fact that the output force times its distance from the fulcrum equals the input force times its distance from the fulcrum. A little experience lifting heavy loads with a lever soon teaches one that to maximize the output force, the load should be placed as close to the fulcrum as possible and the input force as far from the fulcrum as possible. To dramatize the nearly infinite possibility of the lever to magnify force, Archimedes said that if he had a lever long enough and somewhere to stand, he could move Earth.

As an illustration of the lever, consider the current design of the pop top on a popular brand of cola (Figure 1). The built-in opener is a piece of aluminum 25 mm long, made rigid by crimping its edges. The fulcrum is 5 mm from the end which presses into the top of the can to "pop" it open. This leaves 20 mm from the fulcrum to the end, which the user lifts up on with a force of about 5 pounds to open the can. (Since ratios of quantities are involved, there is no problem with mixing English and metric units.) The ratio of distances is 20 mm/5 mm = 4, which means that the opener presses into the top of the can with a force of 4 × 5 pounds = 20 pounds. If the built-in opener is missing, it is necessary to open the can by pressing a small object such as a key into the top with a force of 20 pounds. It is not impossible but it is awkward, and a slip could be messy.

THE WEDGE, THE INCLINED PLANE, AND THE SCREW

These three simple machines change the direction of the applied force as well as magnify it. Each one's operation can be understood by nearly the same physical principles.

A wedge is fairly easy to understand. One side of a heavy rock can be lifted a small amount by pounding a wedge under it, as illustrated in Figure 2. If friction is neglected, the force pushing the wedge under the rock is magnified by the ratio of the distance the wedge moves to the amount the rock is lifted. This follows from requiring the work done by the driving force to equal the work done in lifting the rock. This magnification is the ratio of the length of the wedge to its width, and is obviously greater the smaller the angle of the wedge. The friction force of the wedge against the rock decreases the available lifting force, but it is to

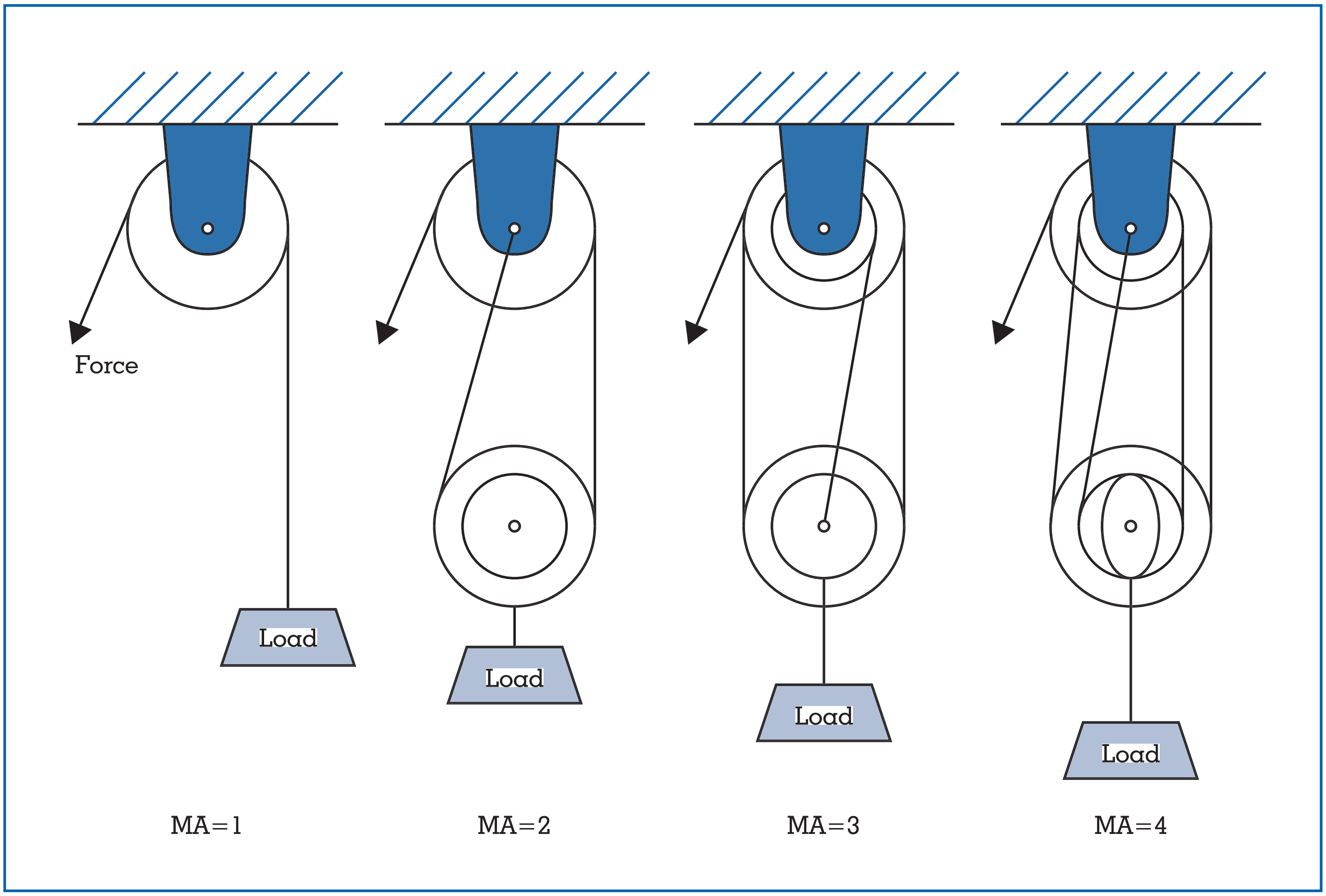


Figure 3.
Pulley systems with varying mechanical advantages.

some extent an advantage. It holds the wedge in place even when no driving force is applied. A wedge is usually used in combination with a hammer.

The hammer itself is an important simple machine used to magnify force. The ape in *2001* was shown inventing the hammer or the club. The user applies a force on the hammer through a relatively large distance, giving it kinetic energy. The hammer comes into contact with an object, compressing it or moving it some short distance. The average delivered force depends on how much compression or motion of the object there is and on whether the hammer rebounds off the object. The mechanical advantage is approximately the ratio of the distance the hammer moves to the distance the wedge, for instance, moves. A person might apply a relatively small force to a hammer for a 1-meter swing. If it drives the wedge 1 millimeter, then the mechanical advantage is 1,000. No wonder that striking a wedge with a hammer can lift an object weighing several tons.

The inclined plane is stationary, and the load to be lifted is moved; whereas it is the other way around for the wedge. The force is magnified by the ratio of the distance the load is pushed to the height through which it is lifted. Again friction will increase the amount of force that has to be applied. If the coefficient of friction is large enough (at least as great as the tangent of the angle of incline), then the load will not slip back down the incline when the pushing force is released. For most materials (coefficient of friction greater than, 0.6 say), an angle less than 30 degrees will keep the load from slipping back on its own. This allows the pushers to take a break to regain their strength before finishing the job. Or they can push the load up the incline a little at a time by lunging at it. In this way they play the role of the hammer relative to the wedge as they collide with the heavy load.

The screw is an inclined plane that is conveniently wrapped around a circular cylinder. The incline of the screw is in the form of a helix similar to a spiral staircase. The mechanical advantage is the ratio of the distance the driving force moves in a circle as it rotates the screw to the distance the load is lifted. A screwdriver with a large handle will provide a larger mechanical advantage, as will a screw with threads closer together. By using a screw jack with a long enough handle, a person can easily lift an automobile or a house. Again, friction conveniently keeps the screw from backing up on its own so that the user can just leave the jack supporting the load, assured that the jack will stay in place.

THE PULLEY

A fixed pulley is a device for changing the direction of an applied force. A common form is a mounted wheel with a rim around which a rope passes. In a very primitive form it could be a vine looped over a tree branch. A pull downward on the rope (vine) results in lifting a load on the other end. Neglecting friction, the mechanical advantage of the single fixed pulley is 1; the load moves the same distance as the applied force.

A combination of fixed and movable pulleys can achieve larger mechanical advantages. The diagrams in Figure 3 show systems of pulleys with mechanical advantages of up to 4. The mechanical advantage is always the ratio of the distance the applied force moves to the distance the load moves. Another, and perhaps easier, way to determine the mechanical advantage of a system of pulleys is to count the number of rope segments that support the load. That number is the mechanical advantage.

Friction is a very important factor in the actual mechanical advantage of a system of pulleys because the frictional losses are compounded each time the rope passes over a pulley. For a typical coefficient of friction of 0.03 (greased shaft with no ball bearings), the system with a theoretical mechanical advantage of 4 would be decreased by friction to an actual mechanical advantage of about 3.5. The largest mechanical advantage that could be obtained with pulleys having a coefficient of friction of 0.03 is about 16 no matter how many pulleys are used. A coefficient of friction of more than 0.333 (such as a vine over a tree branch) results in an actual mechanical advantage of less than 1 for any system of pulleys.

The friction coefficient of 0.03 used above is so low due to the advantage of a wheel on an axle. The friction force at the axle of the pulley wheel is overcome by a smaller force applied at the rim in accord with the principle of the lever.

The friction forces in a pulley system will never hold the load by themselves, but the effort required can be quite small. In the case of a block and tackle used to lift an engine out of an automobile, the weight of the chain hanging from the last pulley may be enough to keep the engine in place. By wrapping a rope once around a post, a cowboy can hold a raging bull in check.

GEARS

Gears are used almost entirely in rotary motion applications, and as such it is easier to discuss the mechanical advantage as a multiplication of torque rather than as a multiplication of force. The work involved in rotary motion is torque times angle; whereas for the linear motion discussed above, it is force times distance.

Torque arises when a force is applied so that it tends to rotate an object about an axis. The force must have a component at right angles to the axis and at some distance from the axis. The torque produced is the product of the force component and its perpendicular distance from the axis. The units of torque turn out to be the same as the units of energy: force × distance. However, torque is not energy. An angle is the ratio of two lengths (arc length/radius for the angle in radians) and has no units; thus the work in rotary motion (the product of torque and angle) has the appropriate units for work.

Upon comparing the work input with the work output for a gear system similar to the one shown in Figure 4, the mechanical advantage is found from the ratio of the angles turned by the respective shafts as the gears engage. This ratio in turn is equal to the ratio of the numbers of teeth on each gear. For example, if a gear (pinion) with 10 teeth drives a gear with 40 teeth, the mechanical advantage is 4—that is, the torque imparted to the large gear is 4 times the torque input by the small gear. There is a commensurate speed reduction; the large gear will rotate once for each four revolutions of the small gear.

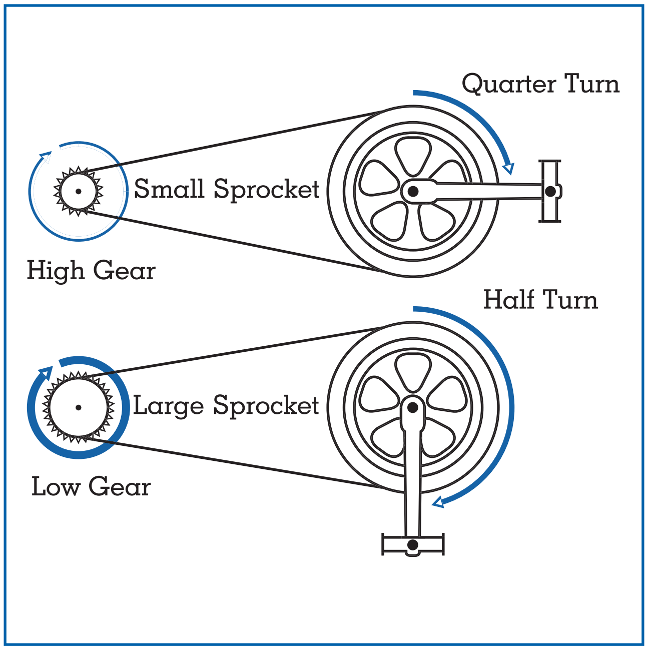


Figure 4.
Gear set with a 12/24 ratio.

Gears may be used to increase the available torque, as in most applications of electric or internal- combustion engines, or to increase the amount of motion, as in a bicycle.

In an automobile, except for the highest gear, the transmission reduces the rotation rate of the drive shaft relative to the engine speed. The differential gears further reduce the rate of rotation. For a typical automobile cruising at 60 mph, the engine runs at about 2,000 rpm while the wheels turn at about 800 rpm giving a mechanical advantage of 2,000/800 = 2.5.

For a typical 21-speed bicycle, on the other hand, the lowest gear ratio is 1.0 and the highest is nearly 3.5. In the highest gear the wheels rotate 3.5 times for each rotation of the pedals. The torque with which the rear wheel propels the bicycle is less than the torque exerted on the pedals by a factor of 3.5. The propulsion force is further reduced by the fact that the radius of the pedals is less than the radius of the rear wheel. The advantage of a bicycle is basically that the wheels move faster than the pedals, coupled with the fact that it takes very little force to overcome the rolling friction of the wheels on a hard, smooth surface. This allows the rider to move faster than a pedestrian using the same energy. However, there is no advantage in force even in the lowest gear. To climb a really steep incline with a bicycle, it is better to get off and walk. From a standing start the greatest acceleration is achieved by pushing the bicycle.

BELTS AND CHAINS

Another function of the transmission of energy by mechanical means is to transfer the motion of a rotating shaft to another shaft at a distant location. A historical application was to get the rotation of a waterwheel coupled to a mill located safely away from the stream in order to grind grain. Another was to drive several rotating machines in a factory from a single large steam engine. Before the invention of the electric generator and the electric motor this type of problem was solved by the use of belts. Today it would be ridiculous to use belts in this way. Think how complicated it would be to connect the power company's spinning steam turbine to a refrigerator in a home by means of belts. Electricity makes it easy. The power is distributed electrically through relatively small wires to drive individual small electric motors.

Today belts are used in automobiles to drive auxiliary devices such as air conditioning, power brakes, power steering, the alternator, and the cooling fluid pump. Belts also can be found in household appliances such as vacuum sweepers, on lathes in machine shops, or inside copying machines.

A belt drive is an inexpensive solution for transmitting mechanical energy over short distances. Compared with machining precision gears, fashioning two pulleys to be linked by a belt is not technologically demanding. The mechanical advantage is the ratio of the diameter of the pulley on the load to the pulley on the driver. The length of the belt makes no difference. By adjusting the sizes of the two pulleys, the mechanical advantage can be conveniently changed. Also, one twist in the belt will reverse the direction of rotation. One disadvantage of a belt is that the amount of torque it can deliver is limited to the product of three factors: the coefficient of friction between the belt and the pulley, the force producing tension in the belt, and the radius of the load pulley.

A V-belt greatly increases the deliverable torque, since the wedging of the belt in the sheave groove increases the force of contact between the surfaces (N) far above the tension force (P). The driving action takes place through the sides of the belt rather than the bottom, which normally is not in contact with the sheave at all. This is yet another example of the application of the wedge as a force multiplier.

For a situation where large torques are involved, such as a bicycle drive, a chain linkage is superior to a belt. A person putting all his or her weight on a pedal probably would make most belt systems slip. Another advantage of a chain over a belt is that a chain is more efficient, mainly because it does not require any ambient tension. The return side of a chain drive has only enough tension to support itself. Furthermore, the chain links are equipped with rollers, which can rotate as they contact the teeth, reducing the frictional forces and wear.

The mechanical advantage of a chain linkage can be calculated by counting the teeth on the load sprocket and the drive sprocket. The output torque is found by multiplying the input torque by the ratio of the number of load teeth to the number of drive teeth. A chain drive is also compact compared to a belt. Imagine trying to arrange 21 speeds on a bicycle derailleur using belts.

A serious disavantage of belts and chains for transmitting energy is that they can be quite dangerous. Whereas the low torque of a bicycle chain rarely results in bad injuries when trouser cuffs or shoestrings get caught in the chain, high-torque industrial and agricultural machinery (such as mechannical reapers with numerous belts and chain drives) is another matter. They have caused many grave inuries and loss of limbs because of the tremendous torque that engines and motors transmit to belts and chains. Therefore, as a safety measure, almost all new equipment comes equipped with guards to prevent accidental injuries either from broken belts and chains whipping out with tremendous force, or the careless actions of workers.

Don C. Hopkins

See also: Bicycling; Drivetrains; Electric Power, Generation of; Electric Power Transmission and Distributive Systems; Engines; Flywheels; Kinetic Energy; Propellers; Steam Engines.

BIBLIOGRAPHY

Adkins, J. (1980). *Moving Heavy Things.* Boston: Houghton Mifflin.

Barnes, M.; Brightwell, R.; Von Hagen, A. L.; and Page, C. (1996). *Secrets of Lost Empires.* New York: Sterling.

National Geographic Society. (1986). *Builders of the Ancient World.* Washington, DC: Author.

Macaulay, D. (1988). *The Way Things Work.* Boston: Houghton Mifflin.

"Machines and Machine Components." (1973). *Encyclopaedia Britannica, Macropeadia*, Vol. 11, pp. 230-259. Chicago: University of Chicago.

MEITNER, LISE
(1878–1968)

In a 1959 lecture at Bryn Mawr College in Pennsylvania, Lise Meitner reflected that "Life need not be easy, provided that it is not empty." Life was not easy for any Jewish woman scientist in Germany in the first half of the twentieth century, and Meitner certainly had her own experience in mind when she made this statement.

Lise Meitner grew up in the Vienna of Emperor Franz-Josef and horsedrawn trolley cars. She was born there in 1878 into a well-to-do Jewish family and decided at an early age that she wanted to be a scientist like Madame Curie. (Later Albert Einstein would call her "the German Madame Curie.") In 1901, she entered the University of Vienna. There, where serious women students were considered odd, she was treated rudely by many of her fellow students. In 1905 she was only the second woman in the university's history to receive a Ph.D. in science.

In 1907, she went to Berlin to study under Max Planck, promising her devoted parents that she would return to Vienna in six months at the most. She stayed in Berlin for thirty-one years. In Berlin, Meitner met Otto Hahn, a professor of chemistry, and took an unpaid position assisting Hahn with his research on the chemistry of radioactive substances. At that time women were not allowed to work in the Chemical Institute, and she had to set up her laboratory in a carpenter's workshop outside the Institute.

While continuing work with Hahn at the new Kaiser Wilhelm Institute for Chemistry in Berlin-Dahlem, beginning in 1912 Meitner served as assistant to Max Planck at the Institute for Theoretical Physics at the University of Berlin, and in 1918 was appointed head of the physics department at the Kaiser-Wilhelm Institute.

Lise Meitner. (Library of Congress)

MEITNER'S CONTRIBUTIONS TO NUCLEAR ENERGY

In their years together Hahn and Meitner did significant research on beta- and gamma-ray spectra. They discovered the new element protoactinium-91 and, at Meitner's suggestion, took up, and made great progress with, work on neutron bombardment of nuclei that Enrico Fermi had commenced in Rome. In 1938, this research was suspended when Adolph Hitler annexed Austria and Meitner had to flee Germany.

Based on the strong recommendations of her German physics colleagues, Meitner received a research position in the Stockholm laboratory of Manne Siegbahn, the Swedish physicist who had received the 1924 Nobel Prize in Physics for his precision measurements on X-ray spectra. Siegbahn provided laboratory space for Meitner, but no suitable equipment for her to continue the research she had started in Berlin, and little encouragement for her work.

Meitner was left very much to herself in the Stockholm Physics Institute, which had little research in progress when she arrived in 1938. Meitner's stipend from the Nobel Foundation was a pittance, and she was forced to do without many of the little comforts she had grown used to in Berlin. It was a difficult time for her, relieved only by regular letters from Hahn containing news of what had once been their joint research effort.

At the end of 1938, Hahn sent her a description of his experiments on the interaction of neutrons with uranium. He and a young chemist, Fritz Strassman, had determined that one of the reaction products was clearly barium. Meitner was so excited about this that she showed Hahn's letter to her nephew, physicist Otto Frisch. Their discussions on the topic gave birth to the idea of nuclear fission.

Frisch then demonstrated in his laboratory the tremendous release of energy accompanying fission, and a short paper by Meitner and Frisch in the British journal *Nature* in 1939 revealed the momentous concept of nuclear fission to the scientific world. It provided a new source of energy for the Earth, while at the same time introducing the possibility of a new weapon capable of unbelievable destructive power.

The step from nuclear fission to a nuclear chain reaction and the atomic bomb was, in principle, quite straightforward. In practice, however, it consumed more time and money than was ever foreseen. Although it was her basic insight that eventually led to the fission bomb dropped on Hiroshima, Meitner refused to work on the bomb and, for humanitarian reasons, hoped that it would not work.

When the question of the award of a Nobel Prize in Physics for the discovery of nuclear fission arose at the end of World War II, it was complicated by the fact that both Hahn and Strassmann were chemists. Another complication was that the Nobel Prize Committee had always considered radioactivity and radioactive atoms the responsibility of their chemistry committee—despite the fact that the discovery of fission had been interdisciplinary from beginning to end. The Swedish Academy of Science was divided on whether the Chemistry Prize should be given jointly to Hahn and Meitner, or to Hahn alone. Finally they decided by a close vote to give the 1945 chemistry prize solely to Otto Hahn.

The physics prize was still in question, and many nominators were strong in their support of Meitner as the recipient. She had continued to correspond with Hahn and advise him from afar on experiments to be performed in Berlin, but when Hahn and Strassmann published their paper on barium as a reaction product of neutron bombardment of uranium, Meitner's name was not included as a coauthor. The Nobel Committee finally decided to award no physics prize for the discovery of nuclear fission in 1945, and gave the prize for that year to Wolfgang Pauli for his theoretical discovery of the Exclusion Principle.

After the war, although now famous, Meitner continued her research in Stockholm, interrupted only by trips to receive honorary degrees and other scientific accolades. She shared in the prestigious Enrico Fermi Prize awarded by the U.S. Atomic Energy Committee in 1966. She retired to Cambridge, England, in 1960, to be near her nephew, Otto Frisch, and died there in 1968 at the age of ninety. Like so many people all over the world during the Hitler period, Meitner's life had been far from easy, but no reasonable person would ever be tempted to call her life empty.

Joseph F. Mulligan

BIBLIOGRAPHY

Crawford, D. (1969). *Lise Meitner, Atomic Pioneer.* New York: Crown Publishers.

Crawford, E.; Sime, R. L.; and Walker, M. (1996). "A Nobel Tale of Wartime Injustice." *Nature* 382:393-95.

Frisch, O. R. (1970). "Lise Meitner." *Biographical Memoirs of Fellows of the Royal Society* 16:405-420.

Frisch, O. R. (1974). "Lise Meitner." In *Dictionary of Scientific Biography*, Vol. 9, ed. C. C. Gillispie. New York: Scribner.

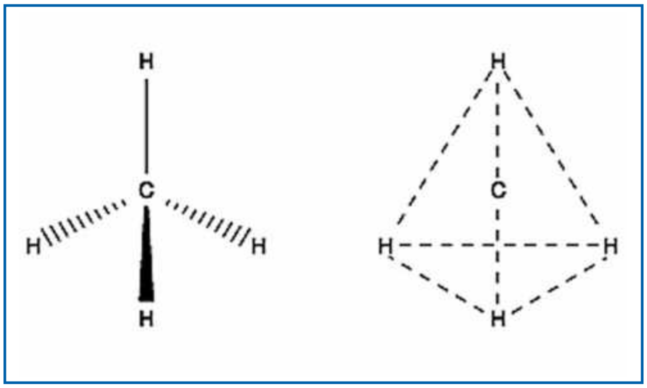
Meitner, L. (1964). "Looking Back." *Bulletin of the Atomic Scientists* 20:2-7.

Meitner, L., and Frisch, O. R. (1939). "Disintegration of Uranium by Neutrons: A New Type of Nuclear Reaction." *Nature* 143:239.

Sime, R. L. (1996). *Lise Meitner: A Life in Physics.* Los Angeles: University of California Press.

METHANE

The hydrocarbon methane (CH_4) is the major component of natural gas (around 90 percent) that is found in oil and gas wells throughout the world. Since the beginning of time, methane has also been produced by a number of biological sources—both natural and human—by the decomposition of organic material. From 1800 to 2000, atmospheric concentrations of methane, which are approximately 0.00017 percent, have grown around 150 percent. However, the patterns of methane emission is highly irregular and, for reasons yet unclear, the rate of increase slowed considerably from 1980 to 2000. The major natural releases of methane are from wetlands (marsh gas) and termites; the major human releases are from energy use, rice paddies, gaseous emissions from animals, human/animal wastes, landfills and biomass burning. Methane research is proceeding in two major directions: the energy course, looking for ways to make bioconversion of wastes to methane more economically attractive as an alternative fuel; and the environmental course, looking for ways to limit its release into the atmosphere since it is a much more potent greenhouse gas than carbon dioxide—a thermogenic effect four to six times that of carbon dioxide. The shared goal is finding ways to "harvest" for energy production much of the methane now being released into the atmosphere.



Molecular structure of methane. (Gale Group)

In the energy sector, many coal mines are looking at ways to put the methane produced as a result of the mining process to work instead of venting it into the atmosphere. The U.S. Environmental Protection Agency estimates that up to 40 percent of the methane that migrates to the atmosphere can be used for power generation (electricity and heat), injection into pipeline systems, methanol production, or onsite applications like coal drying. Besides methane sales revenue and greenhouse gas reductions, the removal of methane from coal seams could serve the vital function of decreasing the risk to workers of firedamp—methane-air mixtures igniting inadvertently.

A researcher checks a sealed poy of methane in Kajima Corp.'s newly developed system designed to break down organic waste. (Corbis-Bettmann)

Other energy sector concerns are methane emissions from unburned fuel, and from natural gas leaks at various stages of natural gas production, transmission and distribution. The curtailment of venting and flaring stranded gas (remotely located natural gas sources that are not economical to produce liquefied natural gas or methanol), and more efficient use of natural gas have significantly reduced atmospheric release. But growth in natural gas production and consumption may reverse this trend. Methane has the highest ratio of hydrogen to carbon of any fossil fuel (4:1), so switching to natural gas is increasingly seen as an attractive option for cleaner air and carbon dioxide reduction.

Unlike natural gas production, the biological production of methane is attractive from a sustainability perspective. Whereas the methane that sits in underground natural gas reservoirs is finite, the methane production from biological sources is potentially very large as well as renewable. Anaerobic bacteria digestion of organic materials, in the absence of air, can produce a biogas that is 60 to 70 percent methane (state-of-the-art systems have reported producing 95 percent pure methane), the rest being carbon dioxide and other trace gases. Burning this gas can provide energy for cooking and space heating, or electricity generation.

Bioconversion of manure-to-methane is accomplished with biogas devices called digesters: organic material fed into the digester tank is heated to increase the natural decomposition rate by microorganisms, and then a pipe carries the biogas to where it will be used. There is an outlet for digested residues that usually are used as fertilizer. Several demonstration projects are taking place that are showcasing the technology. The Mason Dixon Dairy located in Gettysburg, Pennsylvania, produces enough methane from cow manure to meet its power needs, with excess power sold to the local utility. However, wide-scale development is unlikely because the low margins characteristic of farming do not support the additional capital investment to build such an operation.

Landfill gas-to-energy is another promising way to reduce atmospheric release and provide inexpensive energy to local industry and communities. Boilers for steam heat, hot water and the generation of electricity can be designed to burn a blended fuel or fueled exclusively on landfill gas. Because large-scale landfill methane recovery projects are most economical when an industrial facility is located nearby, landfill gas-to-energy projects are few and can only provide a limited amount of energy.

Controlling methane release from wetland, rice paddies and gaseous emissions from animals is more problematic. The release from rice paddies and wet lands is slow, intermittent and takes place over a wide geographic area, and thus very difficult to control. Gaseous emissions from agricultural animals contribute to atmospheric accumulation of methane due to fermentative digestion that produces methane in the rumen (stomach). Although the rate of release is highly variable, affected by factors such as quantity and quality of feed, body weight, age and exercise, beef cattle and draught animals are suspected of contributing around 50 percent, dairy cows around 20 percent, and sheep around 10 percent. Higher quality feed standards, which increase the efficiency of nutrient use, are being recommended as a way to curtail these emissions.

Methane from renewable biological sources will never be a major energy resource, yet it can be a valuable addition to the energy supply mix. Nevertheless, whether methane comes from fossil fuel reservoirs or from bioconversions, it is certain to provide useful energy for many years to come.

John Zumerchik

See also: Biofuels; Capital Investment Decisions; Climatic Effects; Natural Gas, Consumption of; Natural Gas, Processing and Conversion of; Natural Gas, Transportation, Distribution, and Storage of.

BIBLIOGRAPHY

Buell, P., and Girard, J. (1994). *Chemistry: An Environmental Perspective*. Englewood Cliffs, NJ: Prentice-Hall.

Leng, R. A. (1993). "Quantitative Ruminant Nutrition—A Green Science." *Australian Journal of Agricultural Research* 44:363-80.

Olah, G. A., and Molnar, A. (1995) *Hydrocarbon Chemistry*. New York: John Wiley.

METHANOL

Robert Boyle, an Irish chemist noted for his pioneering experiments on the properties of gases, discovered methanol (CH_3OH) in 1661. For many years methanol, known as wood alcohol, was produced by heating hardwoods such as maple, birch, and hickory to high temperatures in the absence of air. The most popular modern method of producing methanol, which is also the least costly, is from natural gas (methane) by the direct combination of carbon monoxide gas and hydrogen in the presence of a catalyst. Methanol also can be produced more expensively from oil, coal, and biomass.

Methanol accounted for less than one hundredth of one percent of total transportation fuel consumption in 1999. However, it is a high-octane, clean-burning liquid fuel that has been an important component of rocket fuel, and has been the fuel of choice for Indianapolis-500 race cars since 1965. Internal combustion engine vehicles can operate on methanol alone, but M85 (85% methanol and 15% unleaded gasoline), with an octane rating of 102, is a more common automotive fuel since it requires less dramatic vehicle modifications.

Automobiles that are designed to run on methanol need a few modifications to become flexible fuel vehicles (vehicles that run on either gasoline or methanol). First, for the fuel tank, fuel lines and fuel-injection equipment, the vehicle needs noncorrosive materials such as stainless steel and high-fluorine elastomers. Second, since methanol is a lower energy density fuel, fuel injectors must be larger to provide greater volumes of fuel, and vehicles must be equipped with larger fuel tanks to achieve a range comparable to a gasoline vehicle. Third, a fuel sensor that detects fuel composition is needed to relay information to the on-board computer. And finally, the lower volatility and higher heat vaporization of methanol requires a special starting system for convenient cold weather start-ups.

By 1996, California had about 13,000 cars and 500 buses and trucks running on methanol, which was more methanol vehicles than the rest of the United States combined. The driving force for methanol vehicle growth was stringent California emission regulations designed to improve air quality. Since methanol vehicles cost only marginally more than gasoline-only vehicles, and the combustion of M85 produces less smog-forming and toxic air pollutants than comparable gasoline-powered vehicle—around 30 percent of the non-methane hydrocarbons (HC), 20 percent of carbon monoxide (CO), and 10 percent of the nitrous oxide (NO_x) of conventional gasoline engines—fuel-flexible methanol vehicles were seen as a very cost-effective means of improving air quality. However, the near-term future for methanol vehicles looks bleak. Los Angeles had to scrap its methanol transit fleet in the late 1990s because of the corrosive effect of the fuel on the heavy-duty diesel engines. The oil companies also responded to the methanol threat by developing much cleaner-burning reformulated gasoline. Combined with the tremendous advances in vehicle technology, such as improved fuel injection, advanced computer controls, variable valve timing, and electrically heated catalysts, some standard model year 2000

THE UNTAPPED METHANE RESERVES: HYDRATES

Since methane is almost always a byproduct of organic decay, it is not surprising that vast potential reserves of methane have been found trapped in ocean floor sediments. Methane forms continually by tiny bacteria breaking down the remains of sea life. In the early 1970s it was discovered that this methane can dissolve under the enormous pressure and cold temperatures found at the ocean bottom. It becomes locked in a cage of water molecules to form a methane hydrate (methane weakly combined chemically with water). This "stored" methane is a resource often extending hundreds of meters down from the sea floor.

Some petroleum geologists believe that there may be more methane trapped in hydrates than what is associated with natural gas reserves. However, as an energy source, there is considerable uncertainty whether this methane can ever be recovered safely, economically, and with minimal environmental impact. The Russians have experimented with the use of antifreeze to break down hydrates at some onshore locations in Siberia. But perhaps a more promising approach would be to pipe warm surface water to the bottom to melt the hydrates, with a collector positioned to convey the gas to the surface. Another approach might be to free methane by somehow reducing the pressure on the methane hydrates.

Regardless of the "harvesting" method, before these vast methane hydrate reserves can become a viable energy source, ways must be found to minimize the impact to the ocean floor and ocean-bottom ecosystems, and to limit the amount of methane escaping into the atmosphere.

gasoline-powered automobiles not only can meet the California Air Resource Board's Ultra Low Emission Vehicle standards (HC: 0.040, CO: 1.7, NO_x: 0.040 grams/mile), but also the proposed Super Ultra Low Emission Vehicle standards (HC: 0.02, CO: 1.0, NO_x: 0.010 grams/mile).

In the rest of the United States, the primary use of methanol is as a chemical feedstock and in the synthesis of methyl-t-butyl ether (MTBE), the most widely used gasoline additive that boosts octane and reduces the level of emissions. To reduce carbon monoxide emissions, the Clean Air Act of 1992 designated thirty-nine areas in the United States where gasoline sold from November through February must contain 2.7 percent oxygen. MTBE, with an octane rating of 116, is the primary oxygenate additive. Oil companies favored MTBE over ethanol, which has an octane rating of 108, since gasoline formulated with MTBE runs more smoothly with less knock because of its higher boiling point, and MTBE does not vaporize out of the gas tank as quickly as ethanol. The downside for MTBE, however, is that it adds five to ten cents a gallon to the cost of gasoline and reduces gas mileage by around 10 percent. There have also been concerns about the toxicity of any spilled MTBE that may get into the water supply.

As a fuel, none of the four ways of producing methanol can compete with petroleum at its 1999–2000 price of $15 to $30 a barrel. The per-barrel price of oil would have to surpass $40 for methanol made from natural gas to become a competitive alternative, over $50 if the methanol was produced from coal, and close to $70 if it was made from wood. Although methanol is a cleaner burning fuel than gasoline and diesel fuel, when produced from natural gas, it generates more carbon dioxide emissions (which is suspected of contributing to global warming) than diesel fuel.

Although methanol's future is bleak as a fuel for internal combustion engines, its future is much brighter for fuel cell vehicles. Fuel cell vehicles run on hydrogen, yet due to hydrogen's low energy density, it is expensive to transport and store. Thus, auto makers are looking at ways to extract hydrogen from methanol through a device called a steam reformer. Steam reformers combine methanol with steam and heat to produce hydrogen, carbon dioxide and trace amounts of carbon monoxide (that must be removed by an oxidation reactor downstream of the reformer):

$$CH_3OH + H_2O \rightarrow XO_2 + 3H_2$$

By 1999, General Motors, Daimler-Chrysler, Toyota, and Nissan all had demonstration fuel cell vehicles operating on methanol, with plans to start introducing vehicles into the market by 2005. Auto makers have shown a preference for methanol over gasoline primarily because of the likelihood of the sulfur content in gasoline poisoning some of the catalysts used in the fuel cell.

Nevertheless, methanol still faces a major technological hurdle: the endothermic (requires heat) nature of the methanol steam reformer. Heat must come from an additional reactor to burn some of the fuel or exhaust gases from the fuel cell stack. It also takes time to reach operating temperature, which may be an unacceptable compromise for potential auto buyers accustomed to start-and-go vehicles. Additionally, if methanol for fuel cell vehicles began to capture a significant share of the transportation fuel market, it would require major investments in infrastructure. Gasoline and methanol cannot share the same distribution system since methanol can be highly corrosive and presents a greater fire hazard. And since methanol is much more toxic than gasoline, colorless and nearly odorless, greater precautions are needed to lower the risk of groundwater contamination from leaking storage tanks.

John Zumerchik

See also: Alternative Fuels and Vehicles; Biofuels; Climatic Effects; Emission Control, Vehicle; Methane; Synthetic Fuel.

BIBLIOGRAPHY

Greene, D. L. (1996). *Transportation and Energy*. Landsdowne, VA: Eno Transportation Foundation, Inc.

Olah, G. A., and Molnar, A. (1995). *Hydrocarbon Chemistry*. New York: John Wiley.

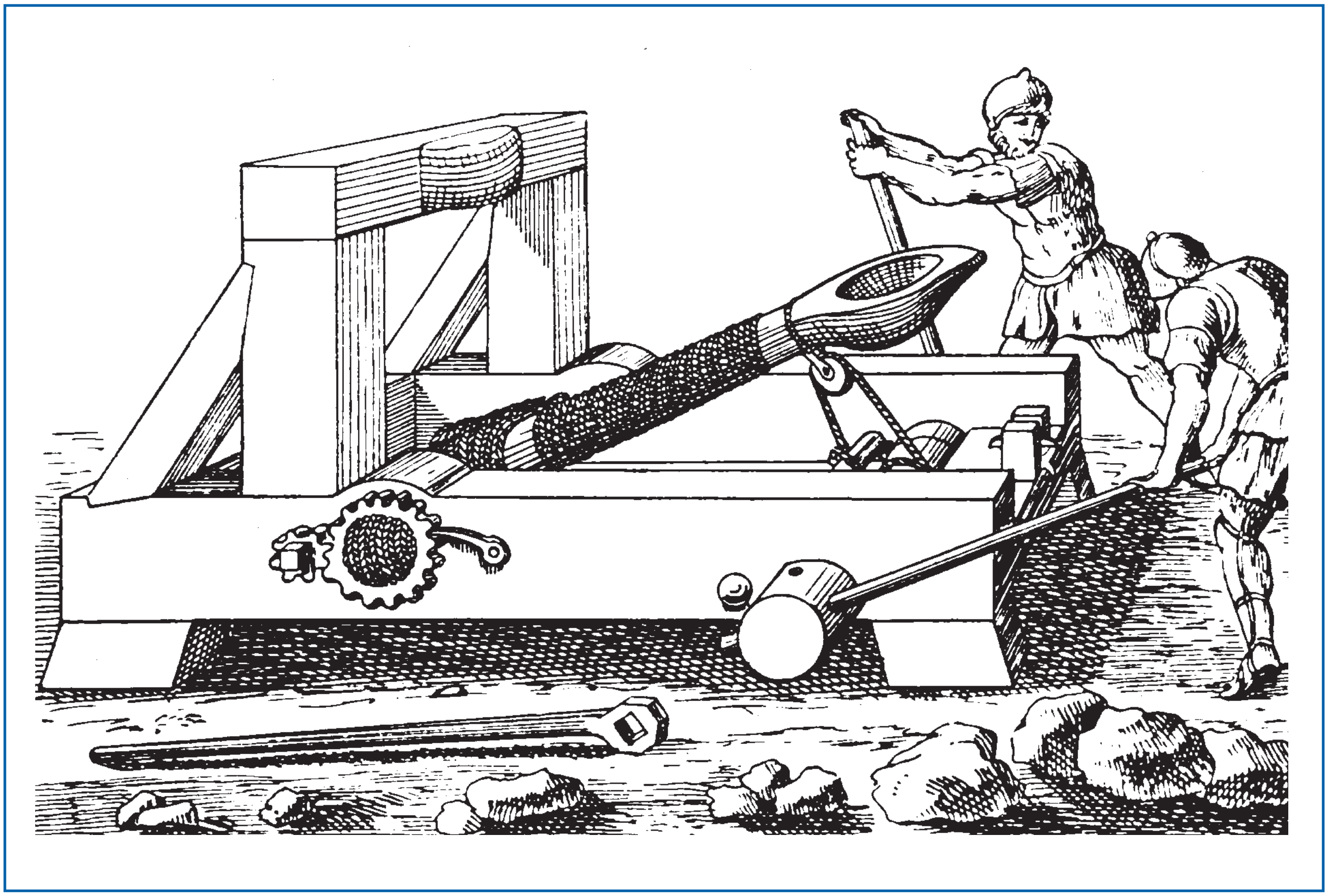
MILITARY ENERGY USE, HISTORICAL ASPECTS

According to historian Vaclav Smil, the destructive energy of military weapons has increased by sixteen orders of magnitude over the past five thousand years. The exploitation of inanimate energy sources has also resulted in the increased technical specialization of men-at-arms, with corresponding changes in the relationships between industry, the state, and military forces. Conquests were once limited by the availability of food for the pillaging troops; naval exploits were at the mercy of the winds, tides, and aggregate human energy of rowing crews. The discovery and subsequent harnessing of ever more efficient sources of energy unshackled the militaries of the Western world, but forced soldiers, sailors, and civilians to reexamine and redefine the place of the military in larger society. By the late nineteenth century, Western militaries were highly industrialized and bureaucratized institutions with intricate divisions of labor. Scientific enterprises, the purview of lone investigators and small collectives of enthusiasts before the nineteenth century, increasingly came under the direct control of state authorities.

ANCIENT HISTORY

Organized military aggression was limited to the use of human muscle power for thousands of years. Warfare was a matter of close-quarters combat, much of it hand-to-hand, and the energy expended in this activity was limited to the physical endurance of the participants. Personal kinetic energy weapons—slings, bows, and crossbows—increased slightly the fighting range of the combatants, but again the energy expended to kill others was limited to the strength of the warrior. The use of the horse as a combat mount increased the destructive energy available to the military, as cavalry were used to bear down on unprotected footsoldiers, although special weapons such as the pike, or architectural protections such as obstructions or fort walls, could limit the impact of the mass of the charging horse and rider.

Siege ballistic weapons, the largest of the early kinetic energy weapons, were developed and deployed to counter the protective walls and other stubborn defenses of cities and fortifications. The most dramatic early example of this kind of device was the catapult, invented in 399 B.C.E. in the siege workshops organized by Dionysus I of Syracuse. The early catapult used mechanical elastic energy to hurl projectiles and smash high fortification walls or disrupt formations of men. Although the catapult did not significantly extend the destructive range of the attacker, this device did allow for the use of projec-



Early example of a catapult. A simple lever is turned into a potent long-distance weapon. (Corbis Corporation)

tiles of ever greater mass, with a corresponding increase in the destructive kinetic energy applied to an opponent. Other offensive devices employed by Dionysus included the battering ram and siege tower, elements of the so-called "offensive-defensive inventive cycle," a kind of arms race that pitted fortress and city defenders throughout Europe and East Asia against marauding attackers. Although the application of kinetic energy to fortress walls could influence the course of a siege, the availability of biomass energy (food) for the consumption of attackers or defenders often determined the ultimate outcome of these stalemates that were so common in European and Near Eastern warfare before the fifteenth century.

The immediate availability of biomass energy, in the form of forage and food, placed serious constraints on military operations before 1850. Feudal political arrangements in Europe and Asia facilitated the growth of large armies capable of protracted warfare, but only at great human and monetary expense. The long-standing constraints on marching range and military strength were gradually addressed as emerging national governments created more effective systems of taxation and resource allocation to support military forces. Logistics, a system of energy distribution for the benefit of military forces, developed into a kind of science as emerging states recruited large standing armies and engaged in open warfare against other organized political groups. Although troops relied heavily on local food supplies (often with disastrous consequences for the local civilian population, as was the case during the Thirty Years' War in Central Europe from 1618 to 1648), the growth of logistics support institutions in the centuries that followed relieved some of the energy-demand burden from non combatants. The presence of logisticians, either civilian contractors or soldiers consigned to a support role, precipitated wider military reforms and the changing role of the army and navy in society at large.

Harnessing the energies of explosive materials, first in the form of gunpowder, produced a revolution in warfare, albeit a very slow one. Explosive powders were commonplace in Asia before the fifteenth century, although dynastic China did not use this material in an effective military capacity. Early hand-cannons, sometimes made of leather-wrapped iron cylinders, were little more than launch tubes for unrefined (and inaccurate) projectiles. The military firearm evolved slowly, beginning with the invention in the fifteenth century of the matchlock (or arquebus), followed by the invention of the German wheel-lock in 1515. The flintlock, subject to various mechanical improvements until the weapon was eclipsed in the nineteenth century by the breech-loading rifle, was a standard infantry arm by the close of the eighteenth century.

With subsequent improvements of these weapons, the explosive energy released by the combustion of gunpowder was harnessed to yield longer projectile ranges with more accuracy. State interest in firearms technology led to government sponsorship of scientific research into aspects of chemistry and ballistics in an effort to better understand the mechanics of explosive energy. The alignment of guilds and other productive tradesmen with military institutions in the sixteenth century foreshadowed a restructuring of the relationship between armies and industry in the years to come.

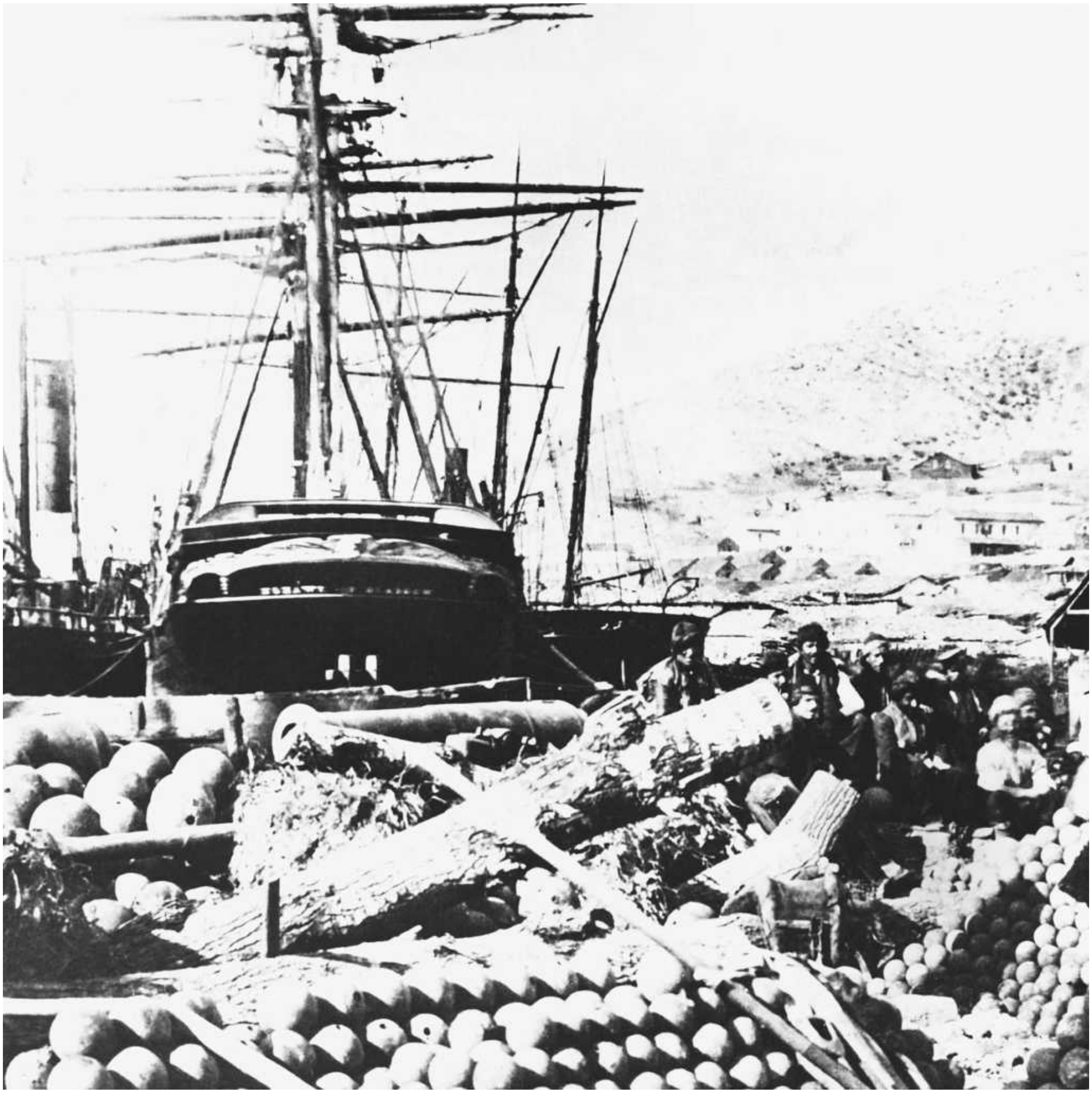
Gunpowder technologies heightened the competition between defense and offense, and raised the stakes for sovereign polities. Vauban, a French military engineer in the employ of Louis XIV, devised an innovative fortification system to respond to the threat of new siege techniques developed since 1400. Among these new threats was the increasing power of portable artillery and, more ominously, an improved form of "sapping," a process of undermining the walls of a fortification with explosives deposited by tunneling enemies. In the 1490s, the first portable siege guns were deployed as part of a larger military expedition organized by Charles VIII; within a century, small cannons could be found in the arsenals of every European power, as well as various kingdoms in India and Asia. Vauban's geometric fortifications were designed to maximize the firepower of the defenders with ramparts that would effect devastating cross-fires on any enemy bold enough to approach the high walls. The importance of specialists, such as miners, to the military endeavor stratified armies and increased the costs of warfare, but the effectiveness of new destructive energies made this kind of investment worthwhile.

THE AGE OF STEAM

One of the hallmarks of the industrial revolution, the steam engine, had important military uses. After nearly two centuries of labor-augmenting industrial use, steam power matured into a viable energy source for military applications in the nineteenth century. The same rails that were used to transport goods between the rural border regions and the urban centers of the Western world were also used for the rapid transportation of troops. In India, the "famine and security" lines built with the assistance of the British government in the late nineteenth century linked provincial territories with the administrative centers where military forces were housed. These rail lines were used to move food and other vital supplies to impoverished areas for relief of the population, or to move troops to quell civil unrest. Railroads proved to be important military assets during the American Civil War (1861–1865) and the Wars of German Unification (1864–1871); the rails served as arteries of support for the combatants. The rapid mobilization of troops in 1914, which eventually led to the horrors of stalemate on the Western Front, was facilitated by the efficiency of Western European railroads.

Guncotton (nitrocellulose) and nitroglycerine, substances that exponentially increased the explosive energy available to mankind, were developed in 1840. In 1867 Alfred Nobel found that nitroglycerine liquid could safely be absorbed in a clay-like substance called kieselguhr. This solid and relatively safe form of explosive became known as dynamite. In 1875, Nobel mixed nitroglycerine liquid with collodion cotton to make an explosive gelatin with both mining and military applications. These early forms of high explosive were refined into more powerful and destructive substances, including French Poudre B (1884) and ballistite (1888). By the 1880s, nitroglycerine (in various forms) and other nitrated organic compounds were important components of munitions manufacturing in Europe and the United States.

The development of high-energy explosives in the nineteenth century corresponded to a period of great



The war of 1854–1856 was fought mainly in the Crimean peninsula between Russia on the one hand and Great Britain, France and Turkey. Shown are cannonballs used during the war. (Corbis-Bettmann)

innovation in artillery design and manufacturing, as more powerful and longer-range guns were introduced into the arsenals of the West. The ranges of these weapons increased from about two kilometers in the 1850s to more than thirty kilometers in the 1890s. The production of the steel and powder necessary for the deployment of these weapons required a tremendous amount of energy, most of it derived from fossil fuels.

The expansion of the colonial empires between 1870–1900 involved the deployment of troops and military hardware to the far reaches of the world, and placed enormous energy demands on the economies of the imperial powers. The so-called tools of empire—steamships, the railroads, and canals—were energy-intensive projects with important military uses. The first oceangoing steam-powered naval vessels were commissioned in the 1850s, and were quickly demonstrated as important military tools. During the Crimean War, the British government nationalized a fleet of private steamships to transport men and supplies to the Ottoman Empire, an arrangement made possible by a special "militarization" clause inserted in the charter agreements between the British government and commercial shipping companies. Private and military fleets grew substantially in the decades that followed; during the period from 1870 to 1910, steam vessels became increasingly more energy-efficient and powerful. Around 1880, the triple-expansion steam engine was invented, followed by the introduction of the quadruple-expansion steam engine in the 1890s. These innovations made warships more fuel efficient with longer cruising ranges. This also meant that remote colonial possessions became more strategically important as coal refueling stations for modern naval forces.

The harnessing of electrical energy had a profound effect on the conduct of military operations. Perhaps most importantly electricity revolutionized communications, with consequences for the conduct of military affairs. The British telegraph network of the late nineteenth century connected the vast reaches of the empire to London, and other imperial powers were often beholden to the British system for news and diplomatic correspondence. In 1901, the first trans-Atlantic "wireless" communications system was demonstrated by Guglielmo Marconi, who promptly approached representatives of the American, British, and Italian navies with the hope of selling his invention. Wire-based telegraph and telephone systems, as well as the "wireless" radio, were adopted by the world's armed forces between 1900 and 1914. The defeat of a Russian expeditionary force in Eastern Asia at the hands of the Japanese army (1904–1905) was widely attributed to the latter's decisive use of the telephone and telegraph to coordinate the movement of troops and the distribution of supplies. The availability and proper application of inanimate energy, it seemed, was vital for the correct distribution of biomass during protracted campaigns, a necessary condition for victory.

Despite the obvious implications that new and more efficient inanimate energy resources held for the world's armies and navies, it was some time before military planners began to consider the importance of energy resources and infrastructure. In 1830 revolutionaries in Paris attempted to paralyze the state government by plunging the city into darkness by attacking the gasworks, the principle distribution center for the gas supply that fed the city's street lamps. The growing popularity of central electric plants in the United States and Europe around 1900 raised questions about the vulnerability and strategic importance of these facilities. By 1914, electric power plants, which supplied crucial power to the war industries, were being protected from sabotage by militiamen and professional soldiers. When called upon, military forces were deployed to disrupt the activities of striking coal miners and others who threatened economic stability or national security. In other instances, military forces were sent abroad to protect energy resources from the predations of other states.

OIL AND THE INTERNAL COMBUSTION ENGINE

Large-scale crude oil exploitation began in the late nineteenth century. Internal combustion engines, which make use of the heat and kinetic energy of controlled explosions in a combustion chamber, were developed at approximately the same time. The pioneers in this field were Nikolaus Otto and Gottleib Daimler. These devices were rapidly adapted to military purposes. Small internal-combustion motors were used to drive dynamos to provide electric power to fortifications in Europe and the United States before the outbreak of World War I. Several armies experimented with automobile transportation before 1914. The growing demand for fossil fuels in the early decades of the twentieth century was exacerbated by the modernizing armies that slowly introduced mechanization into their orders of battle. The traditional companions of the soldier, the horse and mule, were slowly replaced by the armored car and the truck in the early twentieth century.

Internal combustion and electricity wrought a number of changes in naval architecture. Electric power was introduced on warships in the 1880s, allowing for the construction of larger ships, with better light and ventilation, deeper decks, improved artillery fire control, and improved efficiency in han-

British pilots on the Western Front swoop down on a German formation during actual combat in World War I. (Corbis Corporation)

dling ammunition and steering the ever-larger gun turrets. Highly energy-efficient oil-burning turbine engines were introduced on naval vessels about 1900. In 1906, the British navy commissioned the Dreadnaught, widely hailed as a revolutionary warship design, powered by turbine engines, equipped with heavy steel plate armor, and fitted with powerful long-range cannons. Twentieth-century battleships dwarfed the steam-powered vessels of the previous century in terms of displacement and destructive power.

Various schemes of powered flight were finally realized at Kitty Hawk, North Carolina, in 1903 when the Wright brothers successfully demonstrated a self-propelling flying machine. The Wright brother's plane was a breakthrough because of its lightweight aluminum internal combustion engine. Within a few years, strategic planners were speculating about the military importance of the airplane, although there was some debate as to whether the machines would ever pose a threat to static fortifications or field armies. Aircraft were used by the belligerents during World War I, ordered to perform reconnaissance and harassing work. Improvements in engine design, metallurgical science, and aerodynamic theory resulted in net increases in the energy efficiency of aircraft. By the mid-1930s, both airships and airplanes could perform a range of military functions, serving as interceptors, bombers, transports, and reconnaissance craft.

MILITARY ENERGY USE IN THE TWENTIETH CENTURY

The great demand for fossil fuels, a consequence of the rapid industrialization of emerging world powers in the early twentieth century, created political frictions between states vying for energy supplies. The United States, Britain, France, and Russia benefited from direct control of oil- and coal-producing territories, while Germany, Japan, and other modernizing nations were forced to import great quantities of precious energy. Historians have suggested that Japanese expansionism in the 1930s was an expression of energy insecurity; regardless of the causes, Japanese aggression in the Pacific, culminating with the attack on Pearl Harbor, Hawaii, on December 7, 1941, resulted in a protracted war, during which an unprecedented amount of energy was consumed by the combatants. The importance of energy resources, especially fossil fuels, affected the military strategies of all involved. German campaigns in North Africa were planned with the intention of liberating the Suez Canal from British control and permitting the Nazis access to Middle Eastern oil supplies. Allied bombing raids on Ploetsi, Romania, struck the oilfields of that region, a move designed to deny the Germans and Italians access to that energy. The German army's drive to seize the Caspian oilfields in 1941, an operation that compromised the effectiveness of the siege of Stalingrad, was an attempt to secure energy resources for the war effort. The Japanese invasion of Dutch Indonesia was undertaken to obtain the precious oil reserves found among the islands of the archipelago.

World War II was ultimately a contest between economies, and victories were a direct result of effective resource mobilization. The atomic bombs dropped on Hiroshima and Nagasaki in August 1945 released a tremendous amount of energy in the form of heat and radiation; the development of that weapon required a substantial economic and energy investment. After 1945, states went about building up strategic reserves of important natural resources; renewed national environmental and conservation efforts began with concerns about security. In some instances, the interests of the state, the military, and industry aligned, resulting in the execution of energy policies designed to protect the stability and security of the state. In the 1960s, for example, France undertook an aggressive nuclear energy development program in response to the agitations of the domestic coal-mining unions; it was feared that the miners had developed close ties with the French Communist Party. Civil nuclear power was the crossover manifestation of a technology originally developed for military purposes, but adapted for civilian use. Fuel cells and other high-yield, portable power-generation devices, developed in the mid-twentieth century and designed for use in space or other hostile and isolated environments, have both civilian and military applications.

The availability of ever-more-efficient kinds of inanimate energy has made the coordination of men and machines easier. With more efficient and powerful forms of energy generation, the coordination of the movements of man and machines became more effective. Two hundred years ago, soldiers were dependent upon verbal and visual signals for direction; the effective use of electrical communications technologies has resulted in the increased scope and scale of military operations. Digital computer technology, developed in the latter half of the twentieth century, found a ready audience among military officers seeking to maximize their control over subordinates and improve the collection and distribution of intelligence information.

The improvement of human control over inanimate forms of energy, put to use to military ends, has improved the logistics and coordination aspects of armies and navies, and increased the overall destructive capacity of humanity. Energy-efficient propulsion systems have reduced the costs and increased the ranges of various forms of transportation, both military and civilian. For the military, energy is both a blessing and a vulnerability, requiring ever-more-specialized soldiers and more expensive equipment to remain effective in the face of competition from other modern military forces.

Shannon A. Brown

See also: Communications and Energy.

BIBLIOGRAPHY

Brodie, B., and Fawn, M. (1962). *From Crossbow to H-Bomb*. Bloomington, IN: Indiana University Press.

Cipolla, C. M. (1966). *Guns, Sails and Empires*. New York: Pantheon Books.

Ferrill, A. (1985). *The Origins of War: From the Stone Age to Alexander the Great*. London: Thames and Hudson.

Gibbs-Smith, C. H. (1970). *Aviation: An Historical Survey*. London: HMSO.

Hall, B. S. (1997). *Weapons and Warfare in Renaissance Europe*. Baltimore: Johns Hopkins University Press.

Headrick, D. R. (1981). *The Tools of Empire*. New York: Oxford University Press.

Headrick, D. R. (1988). *The Tentacles of Progress*. New York: Oxford University Press.

McNeill, W. H. (1982). *The Pursuit of Power*. Chicago: University of Chicago Press.

McNeill, W. H. (1989). *The Age of Gunpowder Empires, 1450–1800*. Washington, DC: American Historical Association.

Mumford, L. (1934). *Technics and Civilization*. San Diego: Harcourt, Brace and Company.

Nye, D. E. (1998). *Consuming Power: A Social History of American Energies*. Cambridge, MA: MIT Press.

Pacey, A. (1974). *The Maze of Ingenuity*. Cambridge, MA: MIT Press.

Pacey, A. (1990). *Technology and World Civilization*. Cambridge, MA: MIT Press.

Parker, G. (1988). *The Military Revolution*. Cambridge: Cambridge University Press.

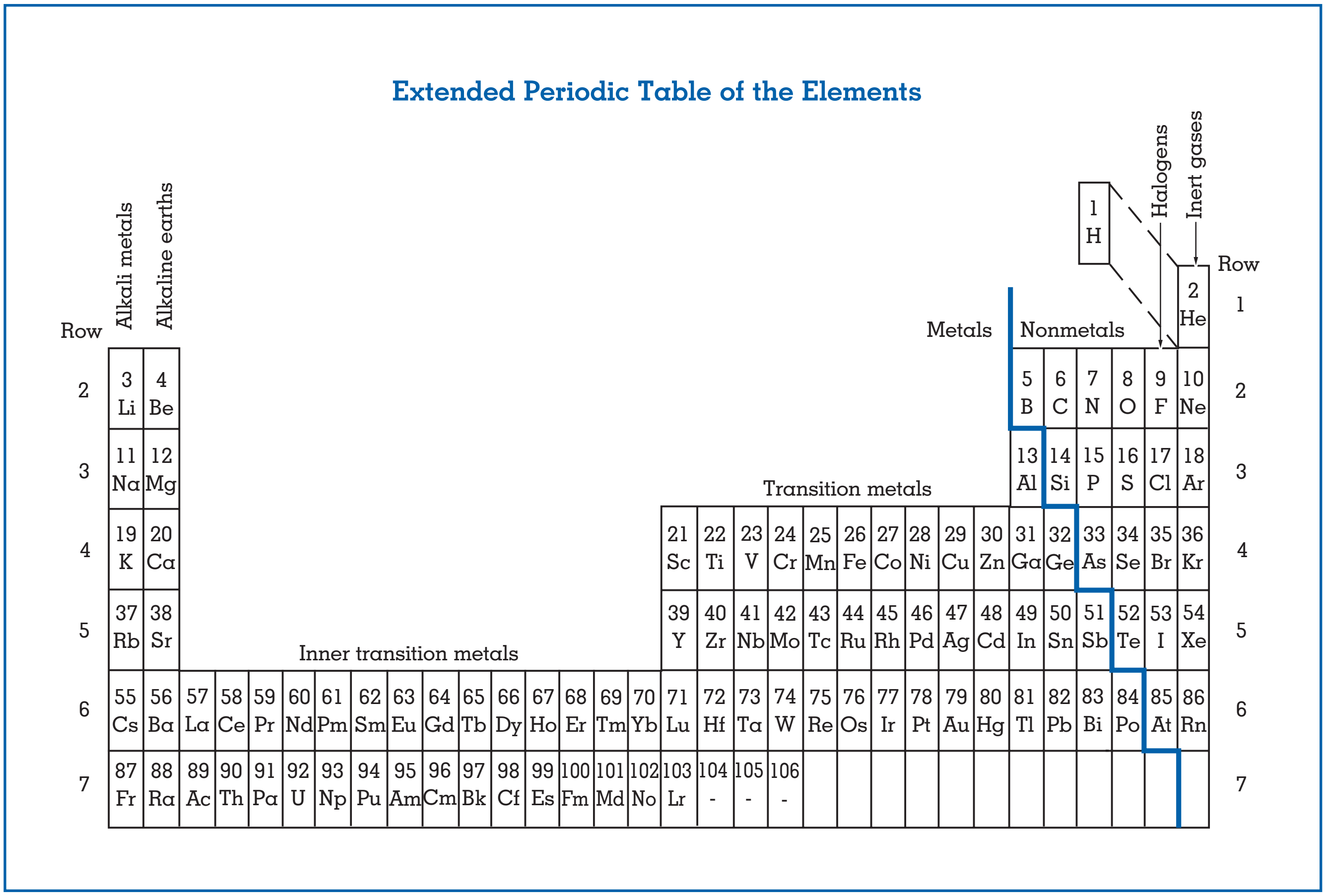
Pick, D.. (1993). *War Machine: The Rationalisation of Slaughter in the Modern Age*. New Haven, CT: Yale University Press.

Smil, V. (1994). *Energy in World History*. Boulder, CO: Westview Press.

White, L. (1978). *Medieval Religion and Technology*. Berkeley: University of California Press.

MOLECULAR ENERGY

The fusion reactions that power the stars are nuclear reactions, in which one type of atom is converted into another. The extremely high heat of our Sun provides energy so intense that when atoms collide, they may collide with enough force to allow their nuclei to fuse together, a process that in itself releases tremendous energy. At moderate temperatures, as on our temperate planet Earth, when two elements bump into each other, they do not ordinarily contain enough energy to fuse. However, under the right circumstances, they may bind together to form a new compound. What



Extended Periodic Table of the Elements

Row	Alkali metals	Alkaline earths																													Halogens	Inert gases	Row
1																													1 H			2 He	1
2	3 Li	4 Be																									5 B	6 C	7 N	8 O	9 F	10 Ne	2
3	11 Na	12 Mg																									13 Al	14 Si	15 P	16 S	17 Cl	18 Ar	3
4	19 K	20 Ca															21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr	4
5	37 Rb	38 Sr															39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe	5
6	55 Cs	56 Ba	57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn	6
7	87 Fr	88 Ra	89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr	104 -	105 -	106 -													7

Figure 1.
When the elements are arranged in order of atomic number, a remarkable repetition of chemical properties occurs. The periodic table arranges the elements so that those in any one vertical column possess similar properties.

holds the two elements together is called a chemical bond. The energy involved in forming a typical chemical bond is ten million times smaller than that involved in nuclear reactions. Yet it is chemical bonding that allows for the existence of the highly organized and complex collection of chemical entities found here on Earth, even for life itself.

ELECTRONIC STRUCTURE OF THE ATOM

An atom is composed of a nucleus, which contains two types of relatively massive particles: the positively charged proton and the neutral neutron. The nucleus is surrounded by very light, negatively charged electrons equal in number to the number of protons, so that the overall charge on the atom is neutral. The number of protons in an atom, its atomic number, defines it as a chemical element. There are 105 elements, each with a different number of protons in its nucleus, and hence a unique atomic number. For example, all atoms with six protons are by definition carbon; those with seventy-nine protons, gold. The electrons surround the nucleus in distinct spatial patterns and possess specific energies. These energies and spatial patterns define the element's chemical behavior. They are in large part responsible for determining which elements will react with each other to form molecules. When atoms come close enough for their outer electrons to interact, attractions among the atoms may occur that are strong enough to hold them together to form a chemical bond.

In the mid-seventeenth century, when accurate atomic weights for a large number of elements were

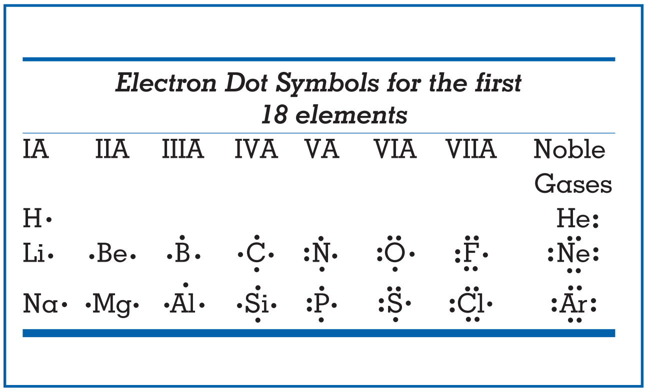


Figure 2.
Electron-dot symbols for the first eighteen elements. This scheme, invented in the early twentieth century by G. N. Lewis, provides a rough but useful tool for predicting the availability of an atom's valence electrons for chemical bonding.

becoming increasingly available, it was noted that there is a periodic recurrence of similar properties among the elements, both in terms of their characteristics as elements (are they, for example, metallic or nonmetallic?) and their behavior in chemical reactions (reactive or stable?). (See Figure 1.) Through the latter half of the nineteenth century natural philosophers amassed an ever-increasing collection of empirical rules, called "valence rules," to describe the formation of chemical bonds and to systematize the characterization, synthesis, and use of new compounds. The periodic table, developed in the mid-nineteenth century, provided a basis for predicting the nature of chemical bonds and the likely composition of a growing list of compounds. However, the approach was largely intuitive and descriptive. No underlying theoretical principle existed to explain chemical bonding.

In the early part of the twentieth century, G. N. Lewis observed that chemical bonding seemed to favor a state in which the atoms in stable compounds, by "sharing" electrons, achieved the stable electron distribution exhibited by the nonreactive noble gases, so-called because they are almost always found as pure elements in the gas state. He proposed that the electrons shared between two elements act as an electromagnetic "glue" to hold the two atoms together. The positive nuclei are attracted to the negative electrons; the electrons spend most of their time between the two nuclei, where the negative charge on the electrons screens the repulsive forces between the two nuclei. This attraction between positive and negative charges and the shielding of the positive nuclear charges by the electrons create a lower-energy state for the two atoms. The energy required to pull these two atoms apart again is called the bond energy. To pull apart a mole of H_2 molecules into two moles of H atoms requires 432 kJ of energy. Lewis's schemes for predicting bond formation among elements, the octet rule and electron-dot diagram, although incomplete, provided a strong predictive tool for describing covalent bonding and are still useful. (See Figure 2.)

The advent of quantum mechanics—developed over the first half of the twentieth century—provided the theoretical framework for our modem understanding of the chemical bond. This body of work predicates that light, ordinarily described in terms of electromagnetic waves, also can be understood as a collection of quanta. And, likewise, moving particles can possess wave-like properties. (This interchangeability is known as "wave-particle duality.") Infrared light quanta are similar in scale (and energy) to molecular bonds; the wavelengths of electrons in atoms and molecules are similar in scale (and energy) to the wavelengths of ultraviolet and visible light.

At the turn of the twentieth century, the planetary model for the atom, with electrons orbiting the nucleus, held sway, despite its fatal flaws: The laws of classical physics stipulate that the orbiting, charged electron should radiate energy and gradually spiral and crash into the nucleus. In 1913 Niels Bohr proposed discarding this classical notion and limiting the electron orbits to those having angular momenta that are integer multiples of Planck's constant h divided by 2π. Although not completely correct, Bohr's theory of the hydrogen atom paved the way for quantum mechanics and won him a revered place in science history.

In 1926, Austrian physicist Erwin Schrödinger made the next leap. He expanded the suggestion by Louis de Broglie (1924) that likened the wave behavior of the hydrogen electron, with its fixed energy values, to a guitar string. The string vibrates at a distinct frequency and has harmonics at wavelengths related to string length by integer numbers, indicating that the guitar string's vibrations are "quantized": The string can vibrate only at those frequencies characterized by integer numbers. To set up an analogous mathematical description of systems such as the electron in the hydrogen atom, Schrödinger described a nonclassical wave function that contains information on the whereabouts of an electron in an atom (or molecule). The wave function (actually, its square) tells us the *probability* of finding the electron at a given point in space. The probability distribution for an electron in a given energy level is known as an orbital.

We now know that electrons in atoms can hold only particular energies and that their probable whereabouts are described by Schrödinger's wave function. The energies and probable locations depend on integer numbers, or "quantum numbers." Quantum numbers describe the energy and geometry of the possible electronic states of an atom. These states, in turn, determine the chemical behavior of the elements—that is, how chemical bonds can form.

The periodicity of the elements derives from the rules dictating the manner in which electrons in atoms fill the energy levels. These levels can be roughly collected as a series of "shells," which one might think of as floors in a multistory building. Each floor, or shell, lies at a discrete energy; higher floors lie at higher energies. Quantum numbers dictate the number of orbitals in each shell. Each orbital can hold no more than two electrons. Electrons begin occupying the lowest-energy shell first; once a shell is filled, the additional electrons must go into the next, higher-energy shell. The most stable (i.e., the lowest-energy) configuration for an atom is to have completely filled electron shells. An atom with one or two vacancies in its highest shell is stabilized by the addition of one or two electrons. An atom with only one electron in its highest shell is stabilized by the release of that lone electron. Atoms interact to form chemical bonds by capturing, losing, or sharing electrons. The shared electrons hold groups of atoms together to form molecules of distinct geometry and character.

The ionization energy, electron affinity, and orbital occupancy determine the chemical behavior, or reactivity, of the elements. The uppermost (highest-energy) occupied orbitals are called the valence orbitals; the electrons occupying them are the valence electrons. An element's ionization energy, the energy required to remove an electron from a neutral atom, is related to its reactivity: A low ionization energy means that the valence electron is readily removed, and the element is likely to become involved in bonding with another atom by losing an electron to it. Electron affinity, the energy released upon capture of an electron by a neutral atom, is an important characteristic with respect to ionic bonding: The energy of an atom with high electron affinity is lowered as another electron approaches. If that approaching electron is attached to another atom, the two atoms can form a bond.

Electron Configurations for the First 20 Elements

Name	Atomic Number	Electron Structure
Hydrogen	1	$1s^1$
Helium	2	$1s^2$
Lithium	3	$1s^22s^1$
Beryllium	4	$1s^22s^2$
Boron	5	$1s^22s^22p^1$
Carbon	6	$1s^22s^22p^2$
Nitrogen	7	$1s^22s^22p^3$
Oxygen	8	$1s^22s^22p^4$
Fluorine	9	$1s^22s^22p^5$
Neon	10	$1s^22s^22p^6$
Sodium	11	$1s^22s^22p^63s^1$
Magnesium	12	$1s^22s^22p^63s^2$
Aluminum	13	$1s^22s^22p^63s^23p^1$
Silicon	14	$1s^22s^22p^63s^23p^2$
Phosphorus	15	$1s^22s^22p^63s^23p^3$
Sulfur	16	$1s^22s^22p^63s^23p^4$
Chlorine	17	$1s^22s^22p^63s^23p^5$
Argon	18	$1s^22s^22p^63s^23p^6$
Potassium	19	$1s^22s^22p^63s^23p^64s^1$
Calcium	20	$1s^22s^22p^63s^23p^64s^2$

Table 1.
Bohr diagrams showing the electronic configurations for the first ten elements, along with representative bond formation.

Atoms are held together to form molecules either by the attraction between oppositely charged atoms (ionic bond) or by the sharing of electrons among positively charged atomic nuclei (covalent bond). In the language of quantum mechanics, a "bonding molecular orbital" is formed between the atoms in a molecule by the combination of atomic orbitals in a way that increases the likelihood of finding electrons between the nuclei. Both the attraction between opposite charges and the screening of repulsive forces between the positive nuclei by the negative electrons lower the overall energy of the system. The bonding

1. Electrons in an atom differ in energy depending on the orbital they occupy—that is, their average location in space relative to the nucleus.
2. Electrons closest to the nucleus have the lowest energy.
3. Electrons can obtain only certain discrete energy values in a stable atomic state.

molecular orbitals are lower in energy than the isolated atomic orbitals from which they arise.

BOND FORMATION BY THE FIRST TEN ATOMS

The "first floor" of the atom can hold only two electrons. When two hydrogen atoms, each with one electron, come together, each nucleus is in effect "sharing" two electrons, and as far as each nucleus is concerned, its shell is filled. The helium atom already contains two electrons; its first floor is filled. That is why helium is unreactive. When two helium atoms collide, rather than stick together, they simply bounce apart. (See Table 1.)

The second shell can hold up to eight electrons. Lithium (Li), the first element in the second row of the periodic table, has atomic number 3. With three electrons, lithium's first shell is filled and its second shell has only one electron. Lithium is highly reactive and readily gives up its lone second-shell electron. In fact, all elements in the same column as lithium, called the alkali metals, have a single valence electron and share these properties. Because an alkali metal gives up its valence electron readily, the bonds formed between alkali metals and other elements are ionic: The electrons are not shared equally. In sodium chloride, NaCl, sodium carries a positive charge and chlorine a negative charge.

Electrons are not only charged, they also have a characteristic physicists call "spin." Pairing two electrons by spin, which has two possible values, "up" or "down," confers additional stability. Beryllium (Be, atomic number 4) has two spin-paired electrons in its second shell that are easily given up in chemical reactions. Beryllium shares this characteristic with other elements in column two, the alkaline earth metals. These atoms also generally form ionic bonds. Boron (B, atomic number 5) will readily give up three electrons.

Carbon (C, atomic number 6) is the most versatile of all elements and is found abundantly in compounds produced by living organisms. With four unpaired electrons, it can make as many as four covalent bonds to fill its outer shell to eight, producing extremely stable molecules. Carbon can form branched and cross-linked chains, and it can form especially stable double bonds and "aromatic" bonds (highly stable, delocalized bonds found in planar ring compounds, like benzene) that can absorb energy in the visible and ultraviolet range of the electromagnetic spectrum.

Nitrogen (N, atomic number 7) has five outer electrons, two of which are paired. It therefore has three electrons to share in chemical bonds. Oxygen (O, atomic number 8), with six outer electrons, has two paired and two single electrons. Oxygen combines readily with other elements to gain two electrons to fill its outer shell, in a process called oxidation. Reactions between oxygen and carbon are especially favorable. (Carbon-oxygen compounds, particularly carbon dioxide, CO_2, are stabilized to a very low energy.)

Fluorine (F, atomic number 9) has seven outer electrons, one unpaired. Because it needs to obtain only one electron to fill its outer shell and gain stability, it is highly reactive. Neon (Ne, atomic number 10), on the other hand, has a filled outer shell. Like helium and its other column mates, the noble gases, neon does not readily react with any element.

ENERGY AND CHEMICAL CHANGE

In chemical reactions, when the atomic configurations of molecules are changed, matter is neither created nor destroyed (Law of Conservation of Matter). The identity and number of atoms remain unchanged. When methane gas (CH_4) is burned, its atoms don't disappear; they combine with oxygen (O_2) in the air and are transformed into carbon dioxide (CO_2) and water vapor (H_2O):

$$CH_4 + 2\,O_2 \rightarrow CO_2 + 2\,H_2O + \text{heat}$$

Likewise, in chemical change, energy is neither created nor destroyed (Law of Conservation of Energy). When charcoal is burned the potential energy stored in the carbon-carbon bonds is released as heat. Although in this reaction the *forms* of matter and energy are changed, both are ultimately conserved; the number of atoms and the amount of energy in the universe remain unchanged.

Energy is the capacity to do work. Potential energy is the energy possessed by an object as a result of its position. Heat, another form of energy, ran be thought of in molecular terms as frictional losses of the uncoordinated motion of molecules.

In a heat-producing (exothermic) reaction the molecular energy of the products is lower than that of the reactants. In the combustion of methane, the energy stored in the bonds of the molecules CO_2 plus two molecules of H_2O is less than that in CH_4 plus two molecules of O_2. If the molecular energy of the products is greater than that of the products (an endothermic reaction), energy (often in the form of heat) must be added for the reaction to occur.

Oxidation reactions release energy because oxidized compounds have lower energy than the reduced compounds from which they were formed. This energy difference derives from the strong "greediness" of oxygen for electrons. Oxygen has a high electronegativity, the attraction of an atom for an electron in chemical bonding. The electronegativities of carbon and hydrogen are roughly the same. Similarly, in the oxygen molecule, the two oxygen atoms share their bonding electrons equally. In carbon dioxide, however, oxygen, which is more highly electronegative than carbon, holds its electrons much more tightly than carbon. The same is true for the other product molecule: water. It is a more stable (lower-energy) situation for the electrons to be held closer to the nucleus of the highly electronegative atom, oxygen, than it is for the electrons to be shared between atoms (carbon and hydrogen) that are less electronegative. Oxygen "wants" the electrons, while carbon and hydrogen are somewhat more "ambivalent" about holding on to them.

SPONTANEITY AND RANDOMNESS

Like a ball at the top of an incline, chemical reactions seem to have a tendency to naturally "roll downhill" from a state of higher to a state of lower potential energy, with the potential energy converted to kinetic energy along the way (motion for the ball, heat for the chemical reaction). However, while most spontaneous chemical reactions are indeed exothermic, many spontaneously occurring chemical changes actually absorb heat, notably ordinary table salt dissolving in water.

In determining if chemical change will occur spontaneously, one must consider not only whether heat is released (the reaction rolls downhill energetically), but also whether it creates order or disorder. Both the lowering of energy and an increase in disorder are changes that tend to occur spontaneously. The measure of the randomness of a system, its "entropy," increases, for example, when a solid or liquid is converted to a gas, when a solid or a liquid is dissolved in solution, with increasing weakness between bonds, or with an increasing number of particles. Sodium chloride (table salt) dissolves spontaneously in water. Energy must be added to break the salt's crystalline bonds, but the increase in entropy (disorder) produced as hydrated sodium and chloride ions are freed from the ordered NaCl crystal into solution drives the reaction forward. The importance of entropy in determining whether a chemical reaction will proceed spontaneously increases as the temperature is increased. For example, ice is the lowest-potential energy state for water. However, at temperatures above 0°C entropy becomes a driving force for melting, and water goes from a low-entropy crystalline phase to a more disordered liquid phase in which the water molecules can move more freely.

OXIDATION AND REDUCTION

Reactions that involve the loss and gain of electrons are termed reduction-oxidation, or Aredox," reactions. The two processes of reduction (loss of an electron) and oxidation (gain of an electron) are complementary and always occur together: One substance's loss is another's gain. Reduced forms of matter—often fuels such as sugar, fat, gasoline, and charcoal—are high in potential energy. Oxidized forms, such as carbon dioxide and water, are low in potential energy.

The term "oxidation" derives from its original use to describe reactions involving combination with oxygen. The broader definition includes reactions in which an element or compound gains oxygen (in the combustion of methane, the carbon gains oxygen atoms and is thus "oxidized"), loses hydrogen, or loses electrons. The working definition for reduction includes reactions in which an element or a compound loses oxygen atoms, gains hydrogen atoms, or gains electrons. The oxidation of fuels—by engine, stove, or living organism—releases energy, often in the form of heat, that can be used to perform work. Green plants convert the light energy of the sun, by reducing carbon dioxide during photosynthesis, into potential energy stored in hydrocarbon bonds:

$$6\ C0_2 + 6\ H_2O + \text{solar energy} \rightarrow C_6H_{12}O_6(\text{glucose}) + 6\ O_2$$

Redox reactions occur in the reduction of ores (metal oxides) into pure metals and the corrosion (oxidation) of pure metals in the presence of oxygen and water. Rusting iron, $4Fe + 3O_2 + 6H_2O \rightarrow 4Fe(OH)_3$, is a good example of metal oxidation. Strong oxidizing agents can be used as antiseptics (hydrogen peroxide, H_2O_2) or bleaches (sodium hypochlorite, NaOCl).

A voltaic cell produces electrical energy through spontaneous redox chemical reactions. When zinc metal is placed in a solution of copper sulfate, an electron transfer takes place between the zinc metal and copper ions. The driving force for the reaction is the greater attraction of the copper ions for electrons:

$$Zn + Cu^{2+} \rightarrow Zn^{2+} + Cu + \text{energy}$$

To exploit the energy produced in this reaction, the "half reactions" are separated. The oxidation reaction is carried out at a zinc electrode ($Zn \rightarrow Zn^{2+} + 2$ electrons) and the reduction reaction is carried out at a copper electrode ($Cu^{2+} + 2$ electrons $\rightarrow$ Cu metal). Electrons flow through a metal wire from the oxidizing electrode (anode) to the reducing electrode (cathode), creating electric current that can be harnessed, for example, to light a tungsten bulb.

BIOENERGETICS

Biological systems ultimately rely on energy from the Sun to build complex, high-potential energy molecules used to store fuel (glucose), provide structure (collagen), transmit signals (hormones, neurotransmitters), or carry information (DNA). Plants harvest the Sun's energy directly through photosynthesis to produce glucose from carbon dioxide and water. Whether by means of a vegetarian, carnivorous, or omnivorous diet, animals, too, benefit from the Sun's energy, albeit less directly. Living organisms couple energy-releasing reactions to energetically unfavorable reactions in synthesizing complex compounds that are high in potential energy. The primary energy-producing reaction in living organisms (those that use oxygen) is the combustion of sugar:

$$C_6H_{12}O_6\ (\text{glucose}) + 6\ O_2\ (\text{oxygen gas}) \rightarrow 6\ CO_2\ (\text{carbon dioxide}) + 6\ H_2O + \text{energy}$$

Biological glucose combustion is coupled through a complex set of reactions to the synthesis of the energy-storing molecule adenosine triphosphate, ATP. This elaborate metabolic pathway ensures that much of the energy released, 686 kilocalories per mole of glucose, is stored in small parcels of useful energy in the phosphate bonds of ATP, while allowing as little energy as possible to be wasted as heat. Glucose metabolism can be broken down into three parts: glycolysis, the citric acid cycle, and the respiratory chain. In glycolysis, also known as anaerobic fermentation, glucose is broken down into pyruvic acid in about ten steps, none of which requires oxygen. Over the course of these steps a significant portion of the energy released is taken up by the formation of ATP or by the reduction of special molecular intermediaries, nicotinamide adenine dinucleotide, NAD^+, and flavin adenine dinucleotide, FAD:

$$NAD^+ + H_2O + 52.7\ \text{kcal/mole} \rightarrow NADH + H^+ + \tfrac{1}{2}\,O_2$$

$$FAD + H_2O + 36.2\ \text{kcal/mole} \rightarrow FADH_2 + \tfrac{1}{2}\,O_2$$

The citric acid cycle, a nine-step process, also diverts chemical energy to the production of ATP and the reduction of NAD^+ and FAD. In each step of the citric acid cycle (also known as the Krebs cycle) a glucose metabolite is oxidized while one of the carrier molecules, NAD^+ or FAD, is reduced. Enzymes, nature's chemical catalysts, do a remarkable job of coupling the oxidation and reduction reactions so that energy is transferred with great efficiency.

The reduced carrier molecules NADH and $FADH_2$ are recycled in the respiratory chain. The reoxidation of these intermediaries releases additional energy for use in the production of more ATP:

$$NADH + H^+ + \tfrac{1}{2}\,O_2 \rightarrow NAD^+ + H_2O + 52.7\ \text{kcal/mole}$$

$$FADH_2 + \tfrac{1}{2}\,O_2 \rightarrow FAD + H_2O + 36.2\ \text{kcal/mole}$$

Photosynthesis uses many of the same enzyme-driven steps found in glucose metabolism, only in reverse. The stages of photosynthesis termed the "dark reactions" are fueled not by the Sun but by ATP and reduced nicotinamide adenine dinucleotide phosphate, NADPH. This series of reactions, which takes place in the chloroplasts of the plant cell, synthesizes glucose from carbon dioxide. In the "light reactions," solar energy trapped by chlorophyll molecules is used to produce the continuous supply of ATP and NADPH (NADH in bacteria) necessary to power the dark reactions. The light-harvesting apparatus is composed of chlorophylls and beta-carotene, aromatic molecules with molecular orbitals that are highly Adelocalized," meaning that they are spread over a major portion of these large molecules. These delocalized molecular orbitals have energy spacings that are similar in energy to light in the visible range: Chlorophyll-*a* absorbs violet light; chlorophyll-*b* absorbs blue and, to a lesser extent, red; and the carotenes are long, antenna-like molecules that harvest light at wavelengths not picked up by the chlorophylls, in the orange and yellow portion of the visible spectrum. Green light is not absorbed, but is reflected, giving green plants their characteristic color.

Ellen J. Zeman

See also: Animal and Human Energy; Batteries; Conservation of Energy; Nuclear Energy; Solar Energy; Thermodynamics; Units of Energy.

BIBLIOGRAPHY

Asimov, I. (1965). *A Short History of Chemistry*. Garden City, NY: Doubleday.

Companion, A. (1979). *Chemical Bonding*, 2nd ed. New York: McGraw-Hill.

Dickerson, R., and Geis, I. (1976). *Chemistry, Matter, and the Universe*. Reading, MA: W. A. Benjamin.

Dickerson, R. E.; Gray, H. B.; and Haight, G. P., Jr. (1979). *Chemical Principles*, 3rd ed. Reading, MA: Benjamin/Cummings.

Hill, J. W. (1992). *Chemistry for Changing Times*, 6th ed. New York: Macmillan.

Hill, J. W., and Feigl, D. M. (1987). *The Chemistry of Life*, 3rd ed. New York: Macmillan.

Pauling, L., and Hayward, R. (1964). *The Architecture of Molecules*. San Francisco: W. H. Freeman.

Snyder, C. H. (1998). *The Extraordinary Chemistry of Ordinary Things*, 3rd ed. New York: John Wiley & Sons.

MONOPOLY, NATURAL

See: Regulation and Rates for Energy

MOTOR FREIGHT

See: Freight Movement

N

NANOTECHNOLOGIES

Nanotechnology is the technology of building things at a molecular scale, that is, where objects are measured in nanometers. Nanotechnology has the potential to improve both energy production technology, and the efficiency of end–use technology. Before this is possible, scientists must develop techniques to manipulate atoms and molecules well enough to build machines and structures. This means that the design of an object must specify each atom in the object, and all the chemical bonds between them. Such a design or object is called "atomically precise." In plants and animals, molecular biology has this capability, with certain substantial limitations. In the laboratory, atoms and molecules can be moved with the tip of a scanning probe, such as an STM (scanning tunneling microscope) or an AFM (atomic force microscope), but with other limitations. Other approaches include manipulation with electron beams, light waves ("optical tweezers"), and deposition in thin films (sputtering, molecular beam epitaxy).

HISTORY

In the 1950s, biologists (notably Francis Crick and James Watson) discovered the molecular basis for information coding in DNA and established that the workings of cells were molecular machines with understandable structure and function. Mathematician John von Neuman developed a mathematical theory of self-reproducing machines based on the biological theories.

In the 1960s, physicists such as Richard Feynman and Carver Mead showed that the speed and power efficiency of machines can improve with decreasing scale, particularly for solid-state electronics. A trend toward miniaturization in electronics was accelerated by the space program. Biologists began to decipher the genetic code.

In the 1970s, computers on a chip (microprocessors) appeared. Electron microscopes neared molecular resolution. By the end of the decade, transistors could be made the size of a human cell (10 microns).

In the 1980s, biologists began manipulating ("recombining") DNA–they could now read and write information at the molecular scale. Physicists (notably Gerd Binnig and Heinrich Rohrer) invented "scanning probe microscopes" able to image individual atoms. Technologists (notably William Trimmer) began to build "micromachines" using techniques from microelectronics. Futurists (notably Eric Drexler) began describing nanomachines and using the term "nanotechnology." At the end of the decade, transistors could be made the size of a bacterium (1 micron).

In the 1990s, nanotechnology emerged as a distinct field, with its own journals, conferences, and funding programs. MIT conferred the first Ph.D. in nanotechnology (to Drexler). Micromechanics became a burgeoning commercial field. Although most of the activity was still nanoscience as opposed to nanoengineering, the explosion of computational power made designing and simulating molecular machines feasible. Drugs and industrial catalysts were routinely designed and simulated in this fashion. Genetically modified organisms were commercial technology. Nanotechnologists (notably Nadrian Seeman) built controllable, albeit very rudimentary, molecular machines. By the end of the decade, commercial electronics had one-tenth micron (100 nanometer) transistors, but switching with single molecules had been done in laboratories.

ENERGY

As the field of nanotechnology matures, the ability to manipulate atoms and molecules will develop into the ability to build machines that do so. This will revolutionize the generation, storage, transmission, and use of energy. For example, a major source of energy is the oxidation of fossil fuels. If the fuel is burned, the energy is thermalized, and must be recovered by a heat engine. The laws of thermodynamics together with engineering constraints give this process a poor efficiency, typically 30 to 50 percent. Efficiency is even worse for small (e.g., handheld) engines. Nanotechnology gives us the prospect of a different approach. Molecules of oxygen and fuel could be positioned rigidly and moved mechanically through the oxidation process, so that it becomes reversible (both practically and in the technical thermodynamic sense). This would raise the efficiency to near 100 percent, even in small engines.

Reversing the process yields an efficient method of energy storage. The process inputs combustion products such as carbon dioxide and water, and energy in the form of electricity or shaft power, and outputs oxygen and fuel (typically hydrogen or hydrocarbons).

Reversible fuel oxidation is as yet not well understood, and can be considered a major goal of nanotechnology in the energy area. Other forms of reversible conversion, such as shaft power to electricity and vice versa (generators and motors), are better understood and there are existing designs awaiting only construction capability. When of reversible oxidation becomes a reality, nanotechnology will enable the construction of heat engines that are small, powerful, clean, and as efficient as thermodynamics permits. One early application that is the subject of research is storage of hydrogen as a fuel.

The understanding of the electronic properties of nanostructures is one of the most rapidly advancing areas in science. This has two major implications: first, it will lead to the construction of nanocircuitry and nanocomputers that will use considerably less power than current computers while being faster and smaller; and second, it will lead to increasing efficiency and decreasing cost of photovoltaic power conversion (“solar energy”).

ATOMIC IMAGING AND MANIPULATION

A scanning-probe microscope consists of a sharply pointed object, preferably so sharp that its tip is a single atom, mounted on a block of piezoelectric material such as a quartz crystal. A voltage is applied to the block, which warps in response. This warping can be controlled to atomic dimensions, allowing the tip to be steered across a molecular sample. The proximity of the tip to atoms of the sample can be sensed by various means, allowing a computer to build up a picture of the sample by scanning the tip across it. Kinds of scanning probe microscope include STM (scanning tunneling microscope) in which a current tunneling from the sample to the probe is measured, the AFM (atomic force microscope) in which the probe presses on the sample and the resulting force is measured, near-field optics in which the probe is an optical funnel focusing or detecting photons with much greater precision than their free-space behavior would allow, and many others.

Scanning probes can also be used to manipulate atoms and molecules individually, placing the tip in contact with the subject atom and pushing or pulling (atoms stick to the tip by virtue of the van der Waals force).

COMPUTATIONAL NANOTECHNOLOGY

Computer-aided design (CAD) and simulation of molecular structures is a rapidly advancing and widely applicable field. Molecules can be designed with molecular CAD programs; early programs allowed the user to specify the type and place of each atom or of substructures (“moieties”) from a library. Research in molecular CAD is now focusing on the automation of parts of the process, following in the footsteps of a similar development in design software for digital electronics (e.g., for microprocessors).

Simulation of molecules can be done at the quantum mechanical level, as is necessary to determine the electronic properties of molecules, to analyze covalent bonds or simulate bond formation and breaking. However, quantum mechanical simulation is extremely computationally intensive and is too time-consuming for all but the smallest molecular systems.

A more practical approach for larger systems is molecular dynamics. In this method, the properties of bonds are determined through a combination of quantum-mechanical simulation and physical experiments, and stored in a database called a (semi-empirical) force field. Then a classical (non-quantum) simulation is done where bonds are modeled as spring-like interactions. Molecular dynamics simulations are appropriate for studies involving the properties of molecules as physical structures and shapes (including "docking" and the catalytic properties of biomolecules in solution, and the structural properties and energy dissipation mechanisms of nanomachine parts in operation).

EARLY COMMERCIAL NANOTECHNOLOGY

Commercial nanotechnology in the 1990s was limited by the lack of a general synthetic capability. It characterized by a plethora of techniques for building nanostructures, including a number of methods based on gas-phase nucleation (e.g., laser pyrolysis, sputtering), methods from synthetic chemistry, including dendrimers and fullerenes, and methods from molecular biology, including DNA synthesis and the use of DNA as a structural material, and protein engineering. In general, the limitations on DNA/protein methods are in the ability to design and predict structures, where the limitations on the other methods are on their physical synthetic capabilities.

Methods capable of detailed atomic manipulation were confined in the 1990s to the laboratory, since they were generally incapable of producing commercially useful amounts of product. There are a few exceptions to this, in applications where a few carefully constructed molecules can be useful, such as chemical sensors, laboratory equipment, and so forth.

THE FUTURE OF NANOTECHNOLOGY

The field of nanotechnology is advancing rapidly, so it is not practical or useful to make short-term predictions about its specific form and capabilities. At most it can be noted that one of the most rapidly advancing subfields is nanoelectronics (sometimes referred to as "molectronics," for molecular electronics).

For longer-term projections, a common practice is to compare the field with others that are advancing as rapidly. Computers, both hardware and software, and biotechnology are cases in point. Both fields bear on nanotechnology and contribute to its advance. In computers, the watershed capability was the stored-program von Neumann architecture. In nanotechnology, it is expected to be another von Neumann design, the self-reproducing system. In practical terms, this simply means that nanomachines need to be capable of making parts for nanomachines, just as a conventional machine shop is capable of making parts for its own machines. Once this is achieved, it will be possible to build commercial quantities of product with atomic precison, and in particular, to have commercial quantities of product that consists of working nanomachines.

J. Storrs Hall

BIBLIOGRAPHY

Crandall, B. C., and Lewis, J., eds. (1992). *Lewis Nanotechnology: Research and Perspectives.* Cambridge, MA: MIT Press.

Drexler, K. E. (1992). *Nanosystems: Molecular Machinery, Manufacturing, and Computation.* New York: Wiley.

ten Wolde, A., ed. (1998). *Nanotechnology: Towards a Molecular Construction Kit.* The Hague: Netherlands Study Centre for Technology Trends.

NATIONAL ENERGY LABORATORIES

The Department of Energy operates seventeen major national laboratories and thirteen minor facilities in the United States that carry out energy research and development, basic science, and defense weapons work. The combined budgets for these laboratories exceed $6 billion annually, with a scientific and technical staff of more than thirty-thousand.

Each of these laboratories is a government-owned/contractor-operated facility selected from industry, academia, and university consortia. As of 1999, the most prominent civilian contractor was Lockheed Martin, the operator of Oak Ridge and Sandia, and the major academic institution was the University of California, administrator of Los Alamos, Lawrence Berkeley, and Lawrence Livermore.

ATOMIC ENERGY COMMISSION YEARS

The origins of the national laboratory network can be traced back to the late 1940s and the beginning of the atomic age. At the end of World War II, the scientific community, particularly the staffs from the Manhattan Project laboratories, lobbied Congress for civilian control of atomic power. Toward this end, the federal government transferred authority from the Army to the newly established Atomic Energy

Commission in 1946. Atomic research and development, both for defense and peaceful use, was to be solely in the hands of the Atomic Energy Commission. Its mission was to "assure the common defense and security," and to "improve the public welfare, increasing the standard of living, strengthening free competition, in private enterprise, and promoting world peace."

After the Russians achieved the first detonation of a nuclear device in 1949, there was a great fear of falling behind in the nuclear weapons race, setting in motion an all-out effort to develop a thermonuclear device. Prominent scientists like Edward Teller and Ernest O. Lawrence were instrumental in establishing the Lawrence Livermore Laboratory to join Los Alamos as a major weapon design laboratories. This strong nuclear science emphasis made the 1950s and 1960s an era of great advances in civilian nuclear energy as well as nuclear weapons and nuclear propulsion for the military. Nuclear weapons not only became more powerful, but more tactical and versatile, including a variety of designs for short, intermediate, and long-range missiles.

The laboratories at Argonne and Oak Ridge were busy designing nuclear reactors in the 1960s. Once the electric utility companies saw that nuclear power could compete with fossil fuel plants on purely economic grounds, more than one hundred reactors were being designed or under construction by the mid-1960s. But confronted by safety and environmental concerns, the nuclear juggernaut started to run into trouble shortly before the energy crisis of 1973—ironically at a time when the need for alternatives to oil and natural gas was greatest.

With so much of the work at the laboratories involving the design of nuclear reactors, the future of many of the laboratories, particularly Argonne and Oak Ridge, looked bleak. But the energy crisis brought plenty of new energy research and development funding to all the laboratories. It also gave many of the laboratories an opportunity to diversify beyond nuclear energy.

DEPARTMENT OF ENERGY YEARS

When the Department of Energy (DOE) was established in 1977, along with its many missions, it was granted authority over a coveted prize: oversight of all the laboratories and research facilities from the predecessor agencies. These included the Bureau of Mines research laboratories at Bartlesville, Morgantown, Pittsburgh, and Laramie from the Federal Energy Administration, and the National Laboratories that were managed by the Atomic Energy Commission (1947–1974) and the Energy Research and Development Administration (1974–1977).

By the late 1990s the number and the size of the laboratories had grown tremendously. From 1977–1997, the federal government's investment in the laboratories was in excess of $100 billion, in seventeen major laboratories and thirteen minor laboratories.

Since the end of the Cold War, the focus of all the laboratories has moved beyond weapons, accelerators, and energy-related research to encompass almost every imaginable field of basic and applied science. Most of the laboratories are increasingly being pushed by Congress to create partnerships with industrial firms to commercialize laboratory-developed technology in the hope that it will improve the overall competitiveness of the U.S. economy.

AMES LABORATORY

To accompany the physics program of the Manhattan Project, Frank H. Spedding, an expert in the chemistry of rare Earth metals, agreed to set up a chemical research and development program in 1942. The Ames Project went on to develop new and far less expensive methods for both melting and casting uranium metal. Based on this successful work, in 1947 the Atomic Energy Commission established Ames Laboratory to produce high-purity uranium metal in large quantities for atomic energy. Although still heavily involved in material research, the research scope expanded considerably through the years to include engineering, environmental, mathematical, and physical sciences.

ARGONNE NATIONAL LABORATORY

The Atomic Energy Commission designated Argonne National Laboratory as the first national laboratory on July 1, 1946. Argonne was the lead laboratory for nuclear reactors, instrumental in designing and building the first nuclear powered submarine, the *USS Nautilus* in 1954, and the first nuclear reactor that completely powered the town of Arco, Idaho, in 1955.

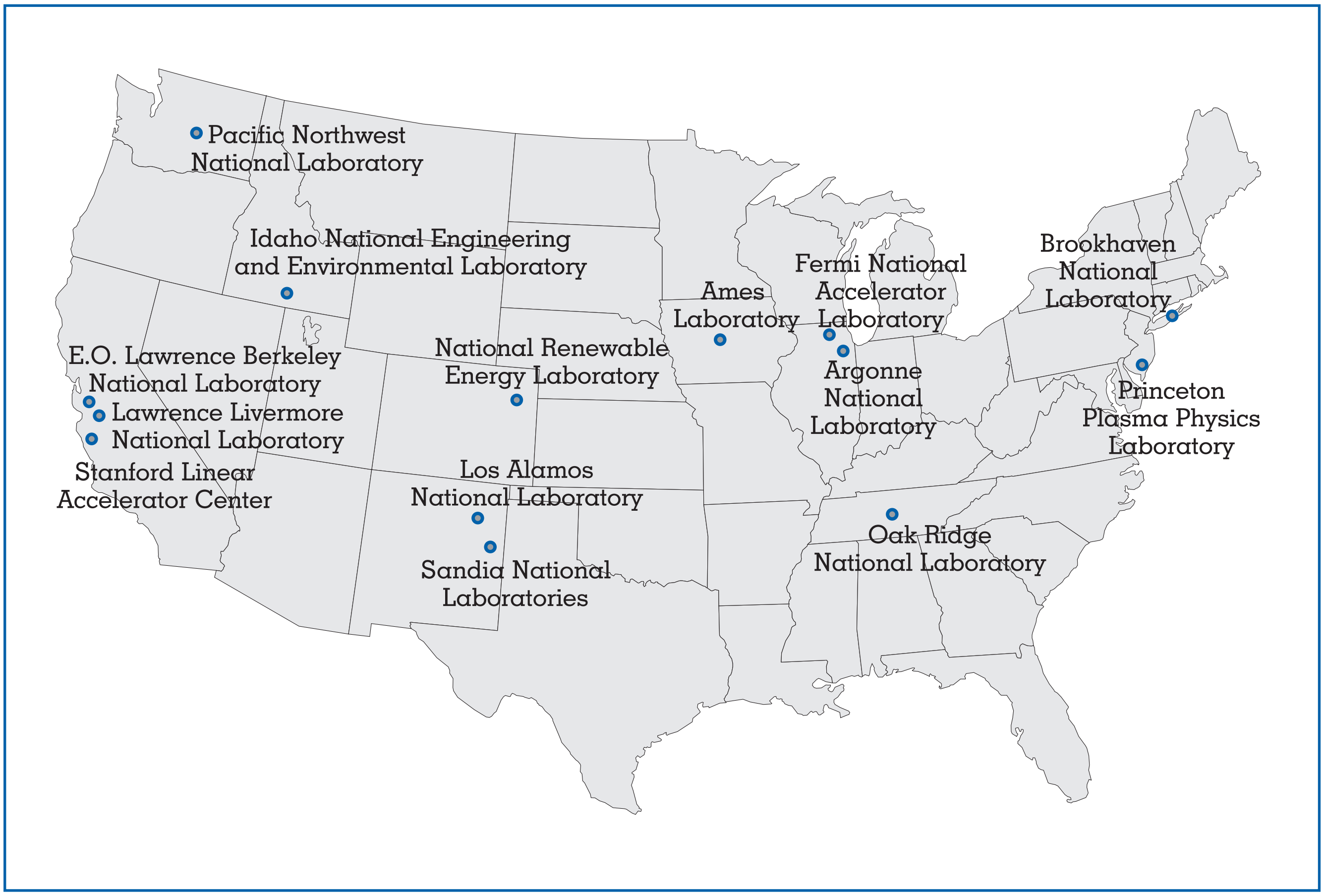


Figure 1.
Locations of the major U.S. National Energy Laboratories.
SOURCE: U.S. Department of Energy website <http://www.doe.gov/people/labsmap.htm>.

Through the 1960s and 1970s, Argonne scientists continued to design and build nuclear reactors, but as nuclear power became engulfed in controversy, federal support dwindled. In response, Argonne began to emphasize other areas such as batteries, magnetohydrodynamics, solar, and fusion in 1977. From the early 1980s through the late 1990s, Argonne continued its diversification, encompassing fields such as materials, medical research, and transportation fuels.

BROOKHAVEN NATIONAL LABORATORY (BNL)

BNL was also founded in 1947 by the Atomic Energy Agency. The management contract was awarded to the nonprofit educational consortium called Associated Universities Inc. with the goal being to create a regional laboratory with the advanced facilities too costly for Universities to build and maintain. Initially BNL concentrated on developing small nuclear reactors for peaceful scientific exploration, and also was one of the pioneers in nuclear medicine. The scope of the laboratory has evolved in the late 1990s beyond the basic sciences to include areas such as environmental research and energy technology.

FERMI NATIONAL ACCELERATOR LABORATORY (FERMILAB)

Fermilab, where groundbreaking for its first linear accelerator began in December 1968, is the premier high energy physics facility in the world. Its mission is to advance the understanding of the fundamental nature of matter and energy. Universities Research Association Inc., a consortium of eighty-six research

Aerial view of the world's largest particle accelerator, located at the Fermi National Accelerator Laboratory in Batavia, Illinois. (Corbis-Bettmann)

universities, operates the laboratory for the DOE. Fermilab is home to the Tevatron, the most advanced accelerator in the world, that can accelerate particles at 0.950 TeV. Huge amounts of electricity, typically running anywhere from $10 to $20 million a year, are needed to power the laboratory and conduct experiments. The future role of Fermilab was uncertain when the Superconducting Super Collider was approved by President Reagan in January 1987. But because federal budgetary problems led to the termination of SSC funding in 1993, it guaranteed the role of Fermilab as the foremost experimental particle research facility in the United States for many years to come.

IDAHO NATIONAL ENGINEERING LABORATORY (INEL)

INEL's origins were as one of the satellite facilities for the Manhattan Project in the 1940s. Since the early 1950s, INEL has developed and operated fifty-two nuclear reactors for the commercial power and national security sectors. INEL has diversified through the years, but the areas of concentration continue to be energy, environmental management (methods for characterizing, treating, and storing radioactive and hazardous waste), and national security. As with the other major laboratories, INEL is committed to the transfer of technology to the private sector.

LAWRENCE BERKELEY LABORATORY (LBNL)

The roots of the LBNL can be traced back to the 1920s, and the pursuit of the secrets of the nucleus. Ernest O. Lawrence, built the first large cyclotron (a particle accelerator) on the Berkeley campus of the University of California in 1931. Unlike most the other labs, LBNL's beginnings depended on the support of philanthropists who saw the promise in Lawrence's work. Seeking private sector support, an onerous chore in good times, was all the more difficult for Lawrence because of the Depression. Nevertheless, by the late 1930s Lawrence had raised enough money to plan his 184-inch diameter cyclotron. It was too big to build on campus, so he moved his Radiation Laboratory to the hill above the campus. After World War II Lawrence and Brobeck built the 7 billion electron volt proton accelerator (called the Bevatron), specifically designed to produce antiprotons, which it did in l955. But with no more flat land in the Berkeley hills, the lead for more powerful accelerators passed to SLAC (Stanford Linear Accelerator Center), the Fermilab, and to CERN.

As the lead in high-energy accelerators passed to Fermi Laboratory in the 1950s, LBNL moved beyond high energy physics to become a collection of interdisciplinary groups working in many diverse fields. By 1977, because of the energy crisis, the focus continued to change with the addition of two new divisions, one for Energy and Environment and another for Earth Sciences, which included geothermal energy and the disposal of nuclear wastes. By the late 1990s these two divisions together had grown to account for about one quarter of the laboratory's budget.

LAWRENCE LIVERMORE NATIONAL LABORATORY (LLNL)

In an attempt to develop the hydrogen bomb before the Russians, a second weapons laboratory, Lawrence Livermore, was established in July 1952 to handle the additional work that would be necessary to stay ahead of the Russian nuclear weapons program. The administrator chosen was the University of California. For the next forty-five years, this LLNL was a formidable competitor to Los Alamos in the development of nuclear weapons. But much like most of the other major national laboratories, its focus also shifted away from nuclear weapons to basic science to fields like magnetic and laser fusion energy, non-nuclear energy, biomedicine, and environmental science. By the late 1990s, half of the laboratory's budget was nondefense related as the shift away from nuclear weapons continued.

LOS ALAMOS NATIONAL LABORATORY (LANL)

In 1943, under the direction of J. Robert Oppenheimer, the Los Alamos National Laboratory, referred to as Project Y at the time, was the site chosen to develop the technology and build the bombs that were instrumental in ending World War II. Managed since its beginning by the University of California, the laboratory's original mission to design, develop, and test nuclear weapons has broadened and evolved through the years. During the Cold War, from the 1950s to the 1980s, the work primarily involved nuclear weapons and the Strategic Defense Initiative. But as the Cold War ended, programs in energy, nuclear safeguards, biomedical science, environmental protection and cleanup, material science, and other basic science programs were added and enhanced. National security remains at the core of LANLs mission, yet the laboratory continues to try to diversify toward more basic science research, and the development and commercialization of emerging technologies.

NATIONAL RENEWABLE ENERGY LABORATORY (NREL)

In the belief that solar energy was a key to energy independence, the Solar Energy Research Institute (SERI) was established in 1977. Meanwhile Rockwell International, the operator of the nearby Rocky Flats Plant, was awarded a federal grant to build and test many different small wind turbines in various configurations. SERI and Rocky Flats wind energy programs were merged in 1984 and transferred to SERI. Both programs prospered in the late 1970s because of very generous federal appropriations, but after federal funding for solar energy research declined in the 1980s, SERI, renamed the National Renewable Energy Laboratory in 1991, was forced to expand its mission well beyond photovoltaics, solar industrial technologies, solar thermal electric, and wind energy to a greater emphasis on biomass electricity, biofuels, and geothermal power. In a renewed effort to transfer more technology for commercialization, Battelle Memorial Institute and Bechtel Corp. began managing the laboratory in collaboration with Midwest Research Institute, the original administrator, in November 1998.

OAK RIDGE NATIONAL LABORATORY (ORNL)

Ground breaking for the "Clinton Laboratories," what Oak Ridge was originally called, occurred in February 1943. Whereas the Manhattan Project at Los Alamos served as the center of weapons design, ORNL devised the techniques to produce and purify the large quantities of fissionable uranium and plutonium. ORNL was a natural choice due to the abundant hydroelectric power that the Tennessee Valley Authority could provide. During the war, the lab was managed by DuPont and the University of Chicago, but in December 1947 the Atomic Energy Commission chose Union Carbide, whose tenure lasted until 1984. Argonne was designated the lead laboratory for reactor design, so ORNL concentrated on isotope production and radiochemical separations, biological research, and specialized reactors for the Navy and Air Force. The promise of electricity "too cheap to meter" in the 1950s and 1960s led to a renewed nuclear reactor emphasis. But as the energy crisis erupted and the promise of nuclear power faded in the 1970s, ORNL diversified into other areas such as energy conservation, synthetic fuels, and solar power. The diversification continued into the 1980s and 1990s with an expansion of environmental research, and an international outreach as ORNL scientists began advising developing countries on ways to secure the energy needed for economic growth.

PACIFIC NORTHWEST LABORATORY (PNL)

Battelle Memorial Institute took over management of what was then called Hanford Laboratories in 1965, and the research facility was separated from Hanford site operations (the Hanford site is where much of the weapons grade plutonium was manufactured from the 1940s to the late 1980s) and renamed the Pacific Northwest Laboratory. PNL is a unique multiprogram laboratory in that part of the facilities are Battelle Memorial Institute (established by the estate of Gordon Battelle, a wealthy industrialist, in 1929) facilities, and part are DOE facilities. When Battelle took over management, they invested $50 million in private research facilities adjacent to the government laboratory.

At first PNL focused on nuclear technology and the environmental and health effects of radiation, but through the years expanded its mission to cover nearly every field of basic science to solve problems in the areas of environment, energy, and national security. Environmental issues and cleanup still encompass two-thirds of PNL work in the 1990s, but PNL has strengthened its role in regional electric power issues as well.

PRINCETON PLASMA PHYSICS LABORATORY (PPPL)

In 1951 Professor Lyman Spitzer, Jr., conceived of an idea of plasma being confined in a figure-eight-shaped tube by an externally generated magnetic field. Called the "stellarator," the Atomic Energy Commission decided to fund it, which established Princton University's controlled fusion effort. Magnetic fusion research was declassified in 1958 so that all nations could collaborate. The PPPL has managed to secure the lion's share of federal plasma and fusion research funding from the 1960s through the 1990s, and with it developed the Tokamak Fusion Test Reactor, that operated at PPPL from 1982 to 1997. Early in the next millenium, an advanced fusion device—the National Spherical Torus Experiment—will begin operation. Although much of the promise of fusion has been unfulfilled to date, some of the knowledge gained in fusion research has found applications in other areas such as materials science, chemistry, and manufacturing.

SANDIA NATIONAL LABORATORIES (SNL)

Sandia National Laboratories began as an ordinance design, testing, and assembly facility in 1945 as part of what is now Los Alamos National Laboratory. AT&T began managing the laboratory on November 1, 1949, and continued to do so until 1993 when the Martin Marietta Corp., now Lockheed Martin, took over. The primary mission of the laboratory was to provide engineering design for all non-nuclear components of the nation's nuclear weapons, but by the late 1990s the laboratory's mission had expanded to include all defense systems, energy security, environmental challenges, and national technological challenges facing industry.

STANFORD LINEAR ACCELERATOR CENTER (SLAC)

The Stanford Linear Accelerator Center, administered by Stanford University, was founded in 1962 as a center for experimental particle physics, but it took until 1966 for its first linear accelerator to be completed. The Stanford Synchrotron Radiation Laboratory, built a decade later, became part of SLAC in 1992. Unlike many of other national laboratories that greatly expanded their mission through the years, SLAC always remained a national basic energy research laboratory.

THE FUTURE OF THE LABORATORIES

The most valuable attribute of the laboratories is the vast human and physical resources that can be called upon to solve national problems of great complexity and scope. However, without a national problem or crisis on the horizon, as was the case in the 1990s, the vast funding needed to maintain this national resource comes into question. And since the laboratories are all funded almost exclusively by federal dollars, the ultimate customer is the U.S. taxpayer, a taxpayer who asks: What have I received for my investment? What can I hope for in the future?

In considering the future course of the laboratories, and the involvement of the federal government, there are two important realities to consider. (1) The number of laboratories and their funding has always increased. (2) No national laboratory has ever been closed. There is a good reason for both realities. The laboratories are located in so many different congressional districts that, much like military bases, finding the political will to close any of the laboratories would prove extremely difficult. Moreover, all the major laboratories have also diversified into many different areas of scientific research to further ensure greater and more stable funding. When there was talk about closing one of the defense laboratories as a “peace dividend” following the breakup of the Soviet Union in the early 1990s, Congress approved more funding, not less.

Critics contend that the laboratories are bureaucratically bloated and inefficient, citing numerous independent reports and audits over the last few decades warning that DOE ownership and operation does not work well. Bureaucracy has made it difficult for the laboratories to generate and carry out a viable mission. And following the Chinese spy scandal of 1999, the DOE has also been widely faulted for a failure to ensure the proper security at its weapons labs.

Some critics recommend that the weapons laboratories (Los Alamos, Sandia, Lawrence Livermore) come under the control of the Department of Defense, and that ownership of the other laboratories be sold to the highest bidder, or turned over to the administrator now running the laboratory. The new owner can then contract with public and private entities in the free market, or shut down the laboratories. They contend that this is the best way for the laboratories to create a vision with value and effectively carry out a mission.

Of course, much of the vision and mission problems of the laboratories can be traced back to funding. Federal funding strongly influences activities of the laboratories. Although much of the $5.0 billion allocated each year for basic science and energy research and development is steered toward the laboratories, the projects supported are decided more for political reasons rather than sound economic and scientific ones. This results in the majority of research dollars flowing toward marginal ideas, which is not the fault of the laboratories, yet the laboratories are held accountable for the dismal record of picking winners and losers that were, in reality, ill-conceived political and bureaucratic decisions. Take electricity generation. The winner—the fuel of choice—turned out to be the fuel that the DOE did not feature in its research and development portfolio: natural gas. Of the $60 billion (in 1996 dollars) spent from 1978 through 1996, only 1 percent ($787 million) went to natural gas, while 99 percent was spent on conservation ($13.3 billion), civilian nuclear ($20.1 billion), coal ($13.3 billion), solar ($5.1 billion), geothermal ($1.8 billion), wind ($900 million), other renewables ($2.8 billion), oil ($1.4 billion), and hydroelectric ($193 million).

Proponents of the laboratories counter that, despite these shortcomings, the laboratories serve a vital mission of undertaking the high risk and expensive investments that the private sector would never agree to invest in. Although natural gas research and development was minimal, DOE support accelerated technological advances on natural gas-fired turbines. Much of the research and development at the laboratories has provided a net social benefit to the nation and economy, work such as safe nuclear reactors and the development of sophisticated defense weapons.

The laboratories undertook the high risk and expensive investments that helped improve the efficiency of current technologies on both the consumption and production end. To lessen dangerous emissions at U.S. coal power plants, the laboratories helped develop better burning systems (e.g., pressurized fluidized bed system), and better scrubber systems. They have also participated in partnerships that have been responsible for a number of fuel cell technology breakthroughs. On the consumption side, the laboratories have helped to develop energy use standards for appliances that have significantly helped lower the nation's overall energy consumption.

Another area of success has been in applied materials research. Because of the integral nature of materials to advances in energy production and consumption, the laboratories have developed a number of toughened ceramics. When used as a replacement for steel, they will improve the energy performance characteristics of high-temperature applications for components of combined-cycle power plants and vehicle engines.

Proponents also point to the fact that thirty-one scientists associated with the laboratories have won Nobel prizes, and that the laboratories have received more "R&D 100" awards (award given annually to technology innovations that hold a strong prospect for commercial success) than any other institution.

John Zumerchik

See also: Government Agencies.

BIBLIOGRAPHY

Brown, E. (1996). *Frontiers: Research Highlights, 1946–1996*. Argonne, IL: Argonne National Laboratory.

Crease, R. P. (1999). *Making Physics: A Biography of Brookhaven National Laboratory*. Chicago: University of Chicago Press.

Crow, M. M., and Bozeman, B. (1998). *Limited by Design: R&D Laboratories in the U.S. National Innovation System*. New York: Columbia University Press.

Hoddeson, L., et al. (1993). *Critical Assembly: A Technical History of Los Alamos During the Oppenheimer Years, 1943–1945*. New York: Cambridge University Press.

Holl, J. M., and Fehner, T. R. (1994). *The Department of Energy, 1977–1994. Energy History Series*. Washington, DC: U.S. Department of Energy.

Johnson, L.; Schaffer, D.; and Schaffer, D. (1994). *Oak Ridge National Laboratory: The First Fifty Years*. Knoxville, TN: University of Tennessee Press.

Plastino, B. J., and Graves, D. P. (1998). *Coming of Age: Idaho Falls and the Idaho National Engineering Laboratory, 1949–1990*. Chelsea, MI: BookCrafters.

Shroyer, J. A. (1997). *The Secret Mesa: Inside Los Alamos National Laboratory*. New York: John Wiley & Sons.

Stelzer, I. (1996). "The Department of Energy: An Agency That Cannot be Reinvented." Washington, DC: American Enterprise Institute.

U.S. House Committee on Science, Subcommittee on Basic Research. (1995). *Alternative Futures for the Department of Energy National Laboratories "The Galvin Report" and National laboratories Need Clearer Missions and Better Management, A GAO Report to the Secretary of Energy*. Washington, DC: U.S. Government Printing Office.

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

See: Government Agencies

NATIONAL RENEWABLE ENERGY LABORATORY

See: National Energy Laboratories

NATIONAL SCIENCE FOUNDATION

See: Government Agencies

NATURAL GAS, CONSUMPTION OF

Natural gas is a mixture of naturally-occurring methane (CH_4) with other hydrocarbons and inert gases. The 2.3 trillion cubic meters (Tcm) or 81 trillion cubic feet (Tcf) of gas marketed and consumed globally in 1997 accounted for about 24 percent of the world's primary energy, ranking third among fuels after petroleum liquids (40%) and coal (25%).

The modern natural-gas industry has its origins in the nineteenth century as urban "gas works" that distributed synthesis gas (a mixture of carbon monoxide, hydrogen and carbon dioxide made by the incomplete combustion of coal, oil, or organic wastes in the presence of steam). Gas works illuminated London streets even before 1800, and subsequently

Gas lit lamps in use around 1911. (Library of Congress)

provided lighting, cooking, water- and space-heating for homes, businesses and public buildings. By the late nineteenth century, gas light was common in the central districts of cities and larger towns throughout North America and Western Europe, and even in such places as Buenos Aires, Cairo, St. Petersburg, Shanghai, and Sydney.

Between the World Wars, consumption in North America switched rapidly from synthesis gas to natural gas which, lacking carbon monoxide, was nontoxic and contained three to four times as much energy as synthetic gas per unit of volume. This shift resulted from the advent of thin-walled, seamless welded steel pipe and leak-proof pipe couplings, which permitted highly compressed vapors to be transmitted safely and efficiently over long distances, together with the discovery and production of large volumes of methane as an initially unwelcome byproduct of crude oil. By 1940, almost every major American gas had become a local distribution company engaged in the resale to retail customers of natural gas purchased mostly from oil companies. In western and central Europe, a similar transition from synthesis gas to natural gas waited until the last third of the twentieth century, when pipelines were laid from natural-gas fields in the North Sea, North Africa, and Western Siberia.

In North America, the local distribution companies that distribute and sell gas to retail customers tend to be distinct from both gas producers and the operators of long-line gas-transmission pipelines, although common ownership of businesses in two or three sectors is not uncommon. Gas distributors are generally treated as public utilities, whose retail prices and other terms of service are regulated by state or provincial authorities. Prior to the 1980s, distribution utilities held a legal monopoly on both the physical delivery and the sales of gas within their local service areas. Recently, however, there has been a trend among state/provincial regulators to unbundle gas

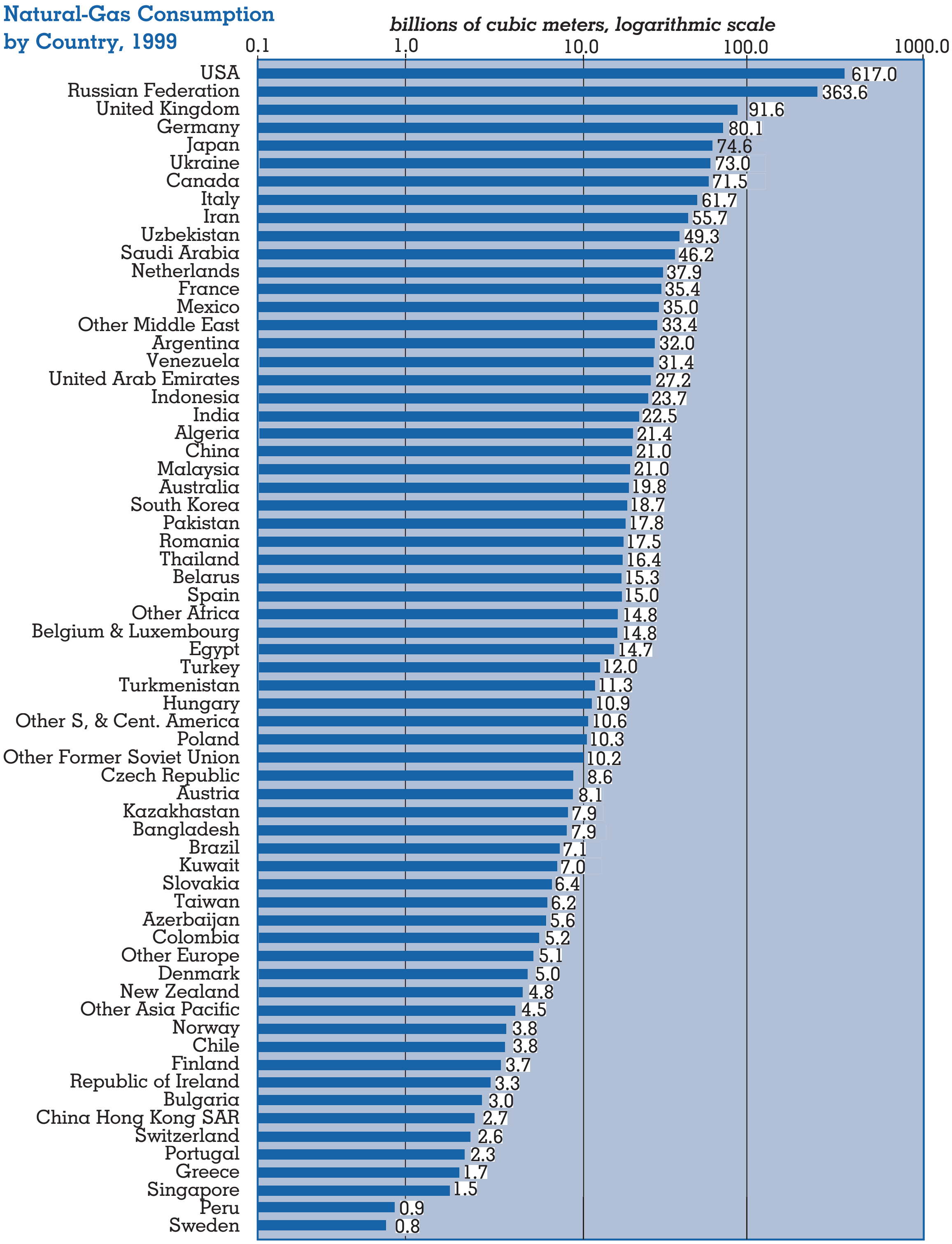

Natural-Gas Consumption by Country, 1999
billions of cubic meters, logarithmic scale
0.1
1.0
10.0
100.0
1000.0
USA 617.0
Russian Federation 363.6
United Kingdom 91.6
Germany 80.1
Japan 74.6
Ukraine 73.0
Canada 71.5
Italy 61.7
Iran 55.7
Uzbekistan 49.3
Saudi Arabia 46.2
Netherlands 37.9
France 35.4
Mexico 35.0
Other Middle East 33.4
Argentina 32.0
Venezuela 31.4
United Arab Emirates 27.2
Indonesia 23.7
India 22.5
Algeria 21.4
China 21.0
Malaysia 21.0
Australia 19.8
South Korea 18.7
Pakistan 17.8
Romania 17.5
Thailand 16.4
Belarus 15.3
Spain 15.0
Other Africa 14.8
Belgium & Luxembourg 14.8
Egypt 14.7
Turkey 12.0
Turkmenistan 11.3
Hungary 10.9
Other S, & Cent. America 10.6
Poland 10.3
Other Former Soviet Union 10.2
Czech Republic 8.6
Austria 8.1
Kazakhastan 7.9
Bangladesh 7.9
Brazil 7.1
Kuwait 7.0
Slovakia 6.4
Taiwan 6.2
Azerbaijan 5.6
Colombia 5.2
Other Europe 5.1
Denmark 5.0
New Zealand 4.8
Other Asia Pacific 4.5
Norway 3.8
Chile 3.8
Finland 3.7
Republic of Ireland 3.3
Bulgaria 3.0
China Hong Kong SAR 2.7
Switzerland 2.6
Portugal 2.3
Greece 1.7
Singapore 1.5
Peru 0.9
Sweden 0.8

Figure 1.

sales from supposedly naturally monopolistic transport functions of distributors, and to permit competition among gas marketers for retail sales, particularly to industrial and commercial customers.

GAS USE BY COUNTRY

Reliance on natural gas as a primary fuel varies widely among different national economies from the 1999 global average of 25 percent. In absolute quantity of gas consumed (see Figure 1), the United States is the largest single consumer, with natural gas accounting for 25 percent of its total energy consumption. The share of natural gas in total primary energy consumption in the European Union was similar to that of the United States, 23 percent, but varied from a high of 41 percent in the Netherlands, through 34 percent in Italy, 22 percent in Germany, 13 percent in France, to less than 2 percent in Sweden.

The number-two gas-consuming country, Russia, used about the same volume of gas per capita as the United States in 1999 but depended on gas for more than half (54%) of its total energy. It is worth noting that although the Russian Federation was the world's biggest gas producer, and international exporter, the entire eastern third of the country—Eastern Siberia and the Russian Far East, roughly from Krasnoyarsk to Vladivostok and Magadan—consumed less than five percent of its primary energy in the form of natural gas.

Figure 2 shows no systematic correlation between a country's dependence on natural gas and its degree of industrialization or per capita; GDP. Variation in relative dependence on gas was even wider among the less-developed and emerging economies outside of Europe and North America, from a high of 82 percent in Uzbekistan, 70 percent in Bangladesh and Algeria, and 50 percent in Argentina, to only 8 percent in India, 3 percent in China, and none at all in several African and Latin American countries.

NATURAL-GAS CONSUMPTION PATTERNS

In addition to widely differing degrees of dependence on natural gas, different regional economies exhibit dramatically contrasting consumption patterns for the gas that they do consume. Strong distinctions are evident, for example, between the diversified and finely-tuned natural-gas markets of high-income, high-latitude countries in Europe and America; the high-volume but relatively undifferentiated gas industry of Russia and other former Soviet republics; and the specialized liquified natural gas-based gas-consuming sectors of Japan, South Korea, and Taiwan.

North America and the European Union.

Figures 3 and 4 contrast the gas-consuming patterns of the world's two largest economies, the United States and Japan. The uses of gas tend to be the most diverse in high-latitude areas, where seasons are distinct and winters are cold, and in regions such as North America and the European Union, where economies are generally sophisticated. In these areas, a dense network of transmission and distribution pipelines makes gas directly available to a numerically large and finely differentiated population of potential residential, commercial, and industrial customers. There, the sales for residential, commercial, and institutional space heating peak in the winter months when the market value of gas is highest. "In contrast, peak demand for gas to generate electricity in combustion turbines and combined-cycle plants occurs in the summer, driven by power requirements for air-conditioning. Together, these climate-sensitive, seasonal components of gas demand constitute its most valuable portions, generating more than 60 percent of total sales revenue."

More than 60 percent of natural gas physically consumed in the course of a year is nevertheless attributable to purchases at lower, interruptible prices by industrial boiler-fuel users and electrical generators that are capable of substituting natural gas in off-peak months, when gas is available at prices competitive with those of "black fuels" (coal and heavy fuel oil). In addition to these relatively low-value, price-sensitive industrial gas uses is a wide range of intermediate-value demand categories for natural gas, such as in process and feedstock use.

In such diverse and sophisticated gas-using economies, security of supply and the efficient employment of producing assets depend upon an extensive network of pipelines that interconnect regions with diverse climates and diverse consumption patterns: winter-peaking and summer-peaking; demand that is climate-sensitive, business cycle-sensitive, and price-sensitive; customers who place a high premium on continuity of supply, and those who are relatively insensitive to risk of interruption. These parties depend to a different degree, and place

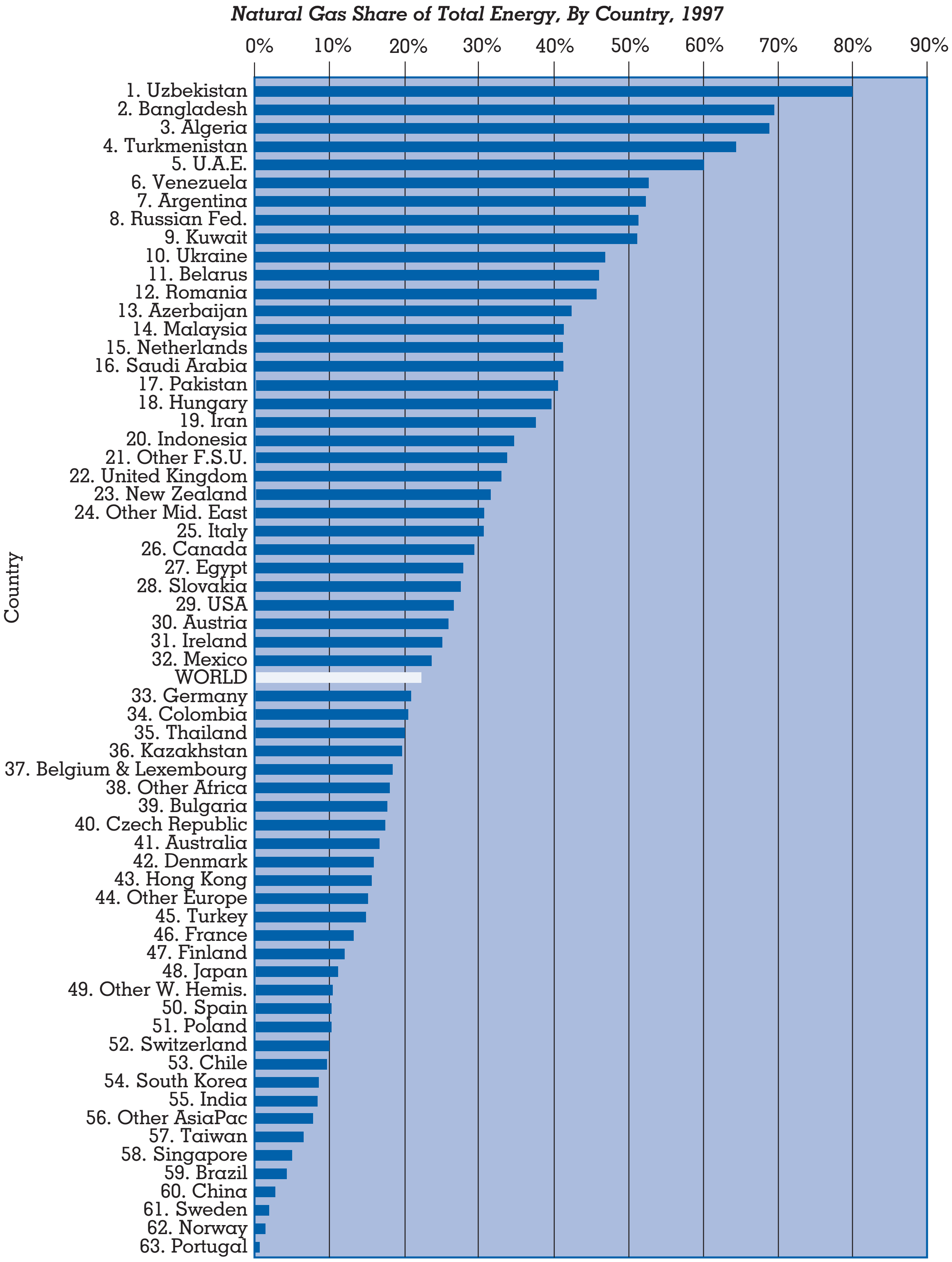

Natural Gas Share of Total Energy, By Country, 1997
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
Country
1. Uzbekistan
2. Bangladesh
3. Algeria
4. Turkmenistan
5. U.A.E.
6. Venezuela
7. Argentina
8. Russian Fed.
9. Kuwait
10. Ukraine
11. Belarus
12. Romania
13. Azerbaijan
14. Malaysia
15. Netherlands
16. Saudi Arabia
17. Pakistan
18. Hungary
19. Iran
20. Indonesia
21. Other F.S.U.
22. United Kingdom
23. New Zealand
24. Other Mid. East
25. Italy
26. Canada
27. Egypt
28. Slovakia
29. USA
30. Austria
31. Ireland
32. Mexico
WORLD
33. Germany
34. Colombia
35. Thailand
36. Kazakhstan
37. Belgium & Lexembourg
38. Other Africa
39. Bulgaria
40. Czech Republic
41. Australia
42. Denmark
43. Hong Kong
44. Other Europe
45. Turkey
46. France
47. Finland
48. Japan
49. Other W. Hemis.
50. Spain
51. Poland
52. Switzerland
53. Chile
54. South Korea
55. India
56. Other AsiaPac
57. Taiwan
58. Singapore
59. Brazil
60. China
61. Sweden
62. Norway
63. Portugal

Figure 2.

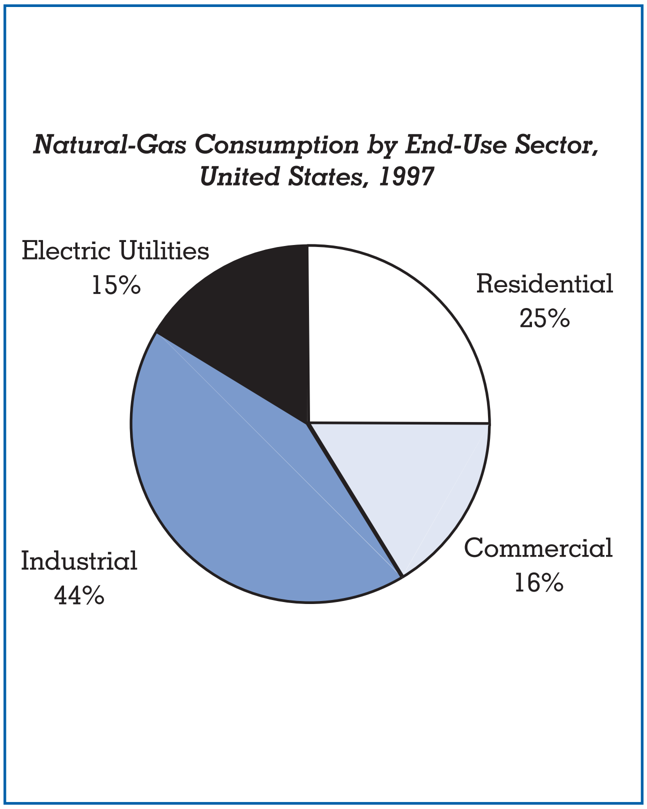


Figure 3.

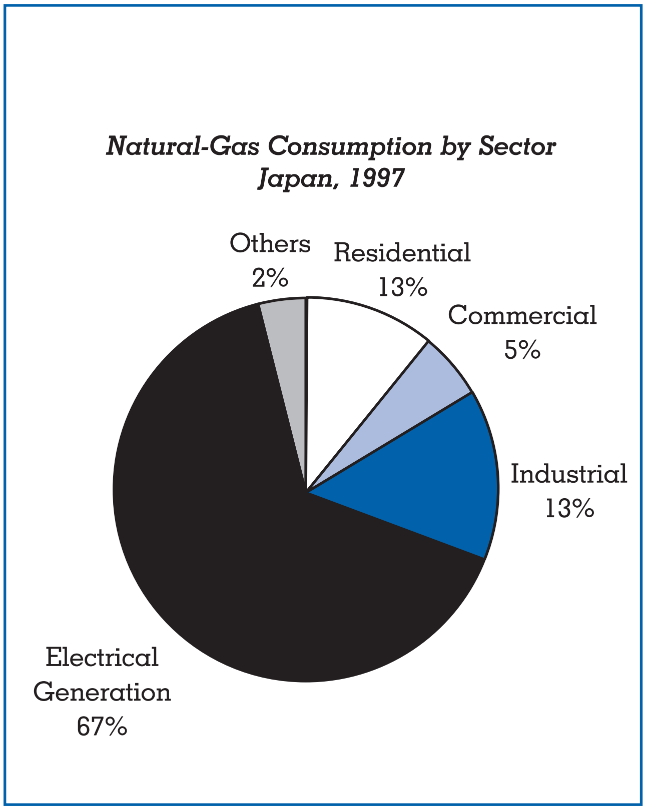


Figure 4.

different values on the access these lines give them to a large and widely distributed system of underground and other gas-storage facilities. The dispatch and allocation of producing, transmission, and storage capacity within the large and diverse population of producing assets and end-users is coordinated by a highly differentiated system of first-sale and wholesale, cash and forward prices for the gas commodity, and for auxiliary services.

Former Soviet Union

Russia, Ukraine, and other entities of the former Soviet Union [FSU] depend even more heavily on natural gas for their primary energy than do the high-income market economies of North America and the European Union. Moreover, their structure of gas demand differs considerably from that of Europe and America, along with the physical and institutional infrastructure of their energy sectors. Because of severe winters, space-heating and domestic water heating doubtless account for an even greater share of total gas consumption than in the West. Rigorous comparison is not possible, because space- and water-heating by means of gas is seldom served or metered separately for individual homes or flats, offices, or small enterprises. Most dwellings and workplaces do use gas or another primary fuel such as coal for space- and water-heating, but only indirectly through steam or hot water cogenerated at steam-electric stations or generated in stand-alone gas-fired district heat plants. District heat produced in this manner is typically distributed to apartment blocks and office towers through elevated and insulated ducts.

Gas for process and feedstock use, and industrial heat for metallurgy, smelting, and materials drying, and other manufacturing applications in the FSU is often delivered directly to the gas-using enterprise, but neither price nor any general system of end-use priorities is systematically used to dispatch or allocate transmission capacity seasonally or among competing domestic shippers, or gas deliverability among competing domestic users. Thus far, the prevailing strategy for accommodating supply and demand in these areas has been to strive to provide sufficient field deliverability and transmission capacity to accommodate the unconstrained aggregate demand of all connected customers, without causing pipeline pressure to collapse.

Japan, South Korea, and Taiwan

The natural-gas industries of Japan, Korea and Taiwan are based almost entirely on gas imported by tanker as liquified natural gas (LNG). These insular and peninsular economies produce almost no domestic natural gas, and until recent efforts to develop production on the Russian island of Sakhalin immediately north of Japan, import of natural gas by pipeline from nearby has not seemed a realistic alternative. In 1999, LNG imports provided about 12 percent of Japan's primary energy, 9 percent of South Korea's, and 6 percent of Taiwan's. In turn, these three countries together accounted for 79 percent of the world's international movements of LNG.

LNG imports to East Asia from Alaska, Southeast Asia, and the Middle East have generally been targeted to specific base-load electrical generating stations built adjacent to the receiving terminals. To minimize delivered costs per unit of fuel, the entire chain of physical facilities from the producing field, through the liquefaction plant at tidewater in the exporting country, the tankers and import terminals, to the receiving customers are tightly coordinated as to capacity and scheduling. The corresponding chain of commercial transactions is composed almost entirely of long-term "take-or-pay" contracts—which oblige the purchaser to pay for the full contracted volume, even if that volume is not or *can* not be taken—for fixed rates of delivery from specific physical sources to specific end-use facilities.

Japan is the world's second largest economy and the fifth greatest consumer of natural gas, behind the United Kingdom and Germany and ahead of Ukraine and Canada. It nevertheless lacks a network of transmission and distribution pipelines capable of delivering gas to diverse and dispersed customers, or reallocating it among them in response to shifting seasonal or business demand. As a result, delivered gas prices to households and industry (other than for base-load power generation) are nearly the highest in the world. Not surprisingly, consumption of natural gas in the residential, commercial, and industrial sectors, and for peak-load electric generation, is exceptionally small.

In the 1990s, Japan's smaller insular and peninsular neighbors, Taiwan and South Korea, began to create integrated national gas-distribution systems that joined previously separate LNG terminals and the potential markets between them. This strategy, particularly in Korea, is consciously directed at creating a more competitive bulk market for gas and expanding the use of gas for space-heating, electricity peaking supply, and industrial fuel.

COMMERCIALIZATION STRATEGIES FOR "STRANDED GAS"

Natural gas is frequently found in association with crude oil, in the search for crude oil, or fortuitously at great distances from developed gas markets or from existing gas-transmission infrastructure-providing access to such markets. In the first half of the twentieth century, carbon black—high-quality soot used as colorant in printing inks and as an additive to rubber in tires—was a leading "scavenger industry" for stranded gas in North America. Later in the century, manufacture of fertilizer, particularly ammonia and urea, created a major part of the early demand for gas along the U.S. Gulf Coast, in Alaska's Cook Inlet basin and in China. The oil industry has recently devoted great effort to promoting gas-to-liquids (GTM) conversion systems to make stranded gas into motor gasoline or diesel fuel. Several GTM technologies are firmly proved, but thus far appropriate market conditions for their commercial application have been hard to find. All of these initiatives are seeking opportunities to convert abundant, low-cost gas into a higher-valued commodity that is liquid or solid at ambient temperatures, and thus can be moved in "normal" tankers, barges or railcars, rather than requiring costly transcontinental pipelines or cryogenic (super-cooled) transport systems. Other notable applications for stranded gas are the local generation of electricity for local consumption and, particularly in the Middle East and North Africa, desalinization of seawater.

INTERFUEL SUBSTITUTION AND COMPETITION

The broad variance in the amount of energy consumed as natural gas, and the diverse mixes of consumption patterns in different countries, illustrate important characteristics of energy supply and demand generally:

- Neither gas nor any other primary fuel or energy source is technically or economically indispensable to

modern civilization. Society's energy "needs" are for heat, light, motive power, information media, and small hydrocarbon building blocks for construction of larger organic-chemical molecules. Primary energy in one form of another is the world's most abundant resource and, in the aggregate and at any human scale, inexhaustible. Economical means are already well established, rapidly proliferating, and improving to transform liquid, solid, or gaseous fuels (and "non-fossil" energy forms) one into another, or into electricity.
- Commercially exploitable natural gas is distributed unevenly in the earth's crust, but is everywhere in relentless competition with other fuels and energy forms over practically all of its actual and potential uses. A substantial share of existing fuel-consuming equipment, mostly in industry or used to generate electricity, has installed dual or multi-fuel capacity. Considering (1) those additional facilities that could economically be retrofitted to use another fuel or energy source, (2) facilities that are nearing the end of their economic lives and subject to replacement by alternatively powered installations, and (3) imminent investment decisions regarding new producer and consumer durables, at least half of the world's energy use is attended by active, near-term interfuel competition.

With appropriate changes in intake, burner, and exhaust hardware, natural gas is readily substitutable for liquid petroleum, coal, and other fuels in almost every stationary (i.e., non-transport) application. Common stationary uses include space heating and cooling, electrical generation, metallurgy, pulp and paper manufacture, petroleum refining, materials drying, and food processing. Natural gas in compressed form or as LNG does indeed serve as transport fuel in motor vehicles, ships, and railway locomotives, and can be adapted even for aircraft, but such mobil employments are less common. Methane and the havier hydrocarbon components of natural gas- ethane and propane-also compete with naphtha, gas oil, and synthesis gas from coal as a feedstock for making the fertilizers ammonia and urea, and "primary petrochemicals" such as ethylene, propylene, methanol, vinyl chloride and acetonitrile. These primary petrochemicals serve as building blocks for further processing into plastics, solvents, pharmaceuticals and other intermediate chemical products.

TRENDS IN GAS CONSUMPTION

Because of the greater difficulty and expense of storing or transporting fuel in gaseous form, markets have historically tended to treat natural gas as less valuable than liquid petroleum products per unit of heating value. However, at the turn of the twenty-first century, world consumption of natural gas is increasing at nearly twice the rate of increase for total primary energy for two reasons: emissions and efficiency.

Natural gas will continue to be substituted for oil and coal as primary energy source in order to reduce emissions of noxious combustion products: particulates (soot), unburned hydrocarbons, dioxins, sulfur and nitrogen oxides (sources of acid rain and snow), and toxic carbon monoxide, as well as carbon dioxide, which is believed to be the chief "greenhouse gas" responsible for global warming. Policy implemented to curtail carbon emissions based on the perceived threat could dramatically accelerate the switch to natural gas.

Natural gas also has an efficiency advantage in electricity generation. The economic and operational superiority of gas-fired combustion turbines and combined-cycle machines (and prospectively, the superiority of gas-powered fuel cells) relative to coal- and nuclear-powered steam turbines made the combination of natural gas and natural gas turbines the supply favorite of most electric utilities in the 1990s.

Arlon R. Tussing

See also: Natural Gas, Processing and Conversion of; Turbines, Gas.

BIBLIOGRAPHY

BP Amoco. (1998). *Statistical Review of World Energy.* London: Author.

Campbell, N., ed. (1995). *Fundamentals of the Natural Gas Industry.* London: The Petroleum Economist.

Energy Information Administration. (1979–2000). *Natural Gas Annual.* Washington, DC: U.S. Department of Energy.

Institute of Energy Economic. (2000). *Handbook of Energy and Economic Statistics in Japan.* Tokyo: The Energy Conservation Center.

Institute of Gas Technology. (1999). *Natural Gas in Nontechnical Language.* Tulsa, OK: PennWell Books.

Tussing, A. R., and Tippee, B (1995). *Natural Gas: Evolution, Structure, and Economics.* Tulsa, OK: PennWell Books.

Tussing, A. R., and Van Vactor, S. A. (1998). "South Korea's Thirst for Gas." *Finanical Times Energy Economist,* March.

U.S. Bureau of the Census. (1975). *Historical Statistics of the United States: Colonial Times to 1970.* Washington, DC: U.S. Government Printing Office.

NATURAL GAS, PROCESSING AND CONVERSION OF

Natural gas is an important energy source consisting mainly of methane gas. It is usually found commingled with deposits of crude oil and also in standalone deposits where the gas has migrated to, leaving the associated petroleum in some other location. Methane is also produced by decaying vegetation (swamp gas) in some coal mines and in land fills, but these sources generally are not suitable for commercial use.

When natural gas comes out of the ground (see Figure 1), it typically consists of 75 to 95 percent methane, with small quantities of ethane, propane, and butane. It may also contain water vapor, carbon dioxide, nitrogen, oxygen, and sulfurous gases such as hydrogen sulfide. Unlike petroleum, which needs to be separated and refined into a variety of fuels and petrochemical products, the nature and general purity of natural gas makes processing far less complex. Most natural gas from the wellhead needs little processing. Nonhydrocarbons are removed from the contaminated alkanes (hydrocarbons containing only single carbon-carbon bonds) by absorption. The heavier hydrocarbons tend to liquefy at the high operating pressures needed for natural gas pipelining. If the natural gas is to be liquefied (LNG), the hydrocarbons are separated by absorption or low temperature distillation.

Before entering the pipeline, the gas is also adjusted in the field to achieve a uniform heating value of 1,000 Btu (British thermal units) per cubic foot. And for safety, because natural gas is odorless and colorless, an odorant is added to provide a distinctive and disagreeable smell that is easy to recognize.

CONVERSIONS

The clean-burning nature of natural gas has for many years made it the fuel of choice for heating and cooking. If its energy content per cubic meter were comparable to liquid fuels like such as diesel and gasoline, it would be ideal as a transportation fuel as well. However, the void is wide. Whereas gasoline and diesel deliver 110,000 to 120,000 Btu per gallon, an equivalent volume of natural gas delivers only about 134 Btu per gallon. Thus, there is great interest in finding ways to efficiently compress or liquefy natural gas so that the same low-emission benefits found in the residential, industrial, and electricity generation markets serviced by pipelines can also be enjoyed by areas without pipeline service and by the transportation sector.

Once water vapor, sulfur, and heavy hydrocarbons are removed, natural gas can be compressed or liquefied. As a transportation fuel, the high methane content gives natural gas its high octane rating (120–130) and clean burning characteristics, resulting in the dual benefit of high engine performance and low pollution. There are no sulfur or particulate (smoke) emissions. Currently, there are two types of natural gas vehicles: exclusively natural gas and bifuel. The latter operate on natural gas or either diesel or gasoline, and the fuel can usually be changed with the flip of a switch.

Compression or conversion for greater use in the transportation market is promising for two reasons: First, natural gas is usually cheaper than liquid fuel and, second, there exist large quantities of stranded gas—remotely located natural gas sources that are not economical to use because tanker or pipeline transportation costs can be over four times as much as for crude oil. Often this gas is recompressed and injected back into the oil-producing zones to help maintain reservoir pressure and optimal crude oil flow to the wellhead. In some cases this gas is wasted by being flared, but this practice is increasingly frowned upon. The demand for cleaner-burning transportation fuels, and the advances in gas turbines that have dramatically improved the efficiency of natural gas powered electricity generation have renewed interest in developing ways to compress or liquefy this gas to lower its shipping cost. Liquefaction can mean cooling the natural gas until it condenses at –187°C (at atmospheric pressure) or converting it chemically to a suitable liquid fuel. Both of these schemes entail considerable energy costs.

Some environmentalists have also touted natural gas as a way station on the road to a hydrogen fuel (carbon dioxide-free) economy. As seen in Table 1, per unit of energy released, natural gas generates about 23 percent less carbon dioxide than gasoline and about 30 percent less than heavy fuel oil. This is helpful in reducing greenhouse emissions, but the other excellent properties of natural gas are even

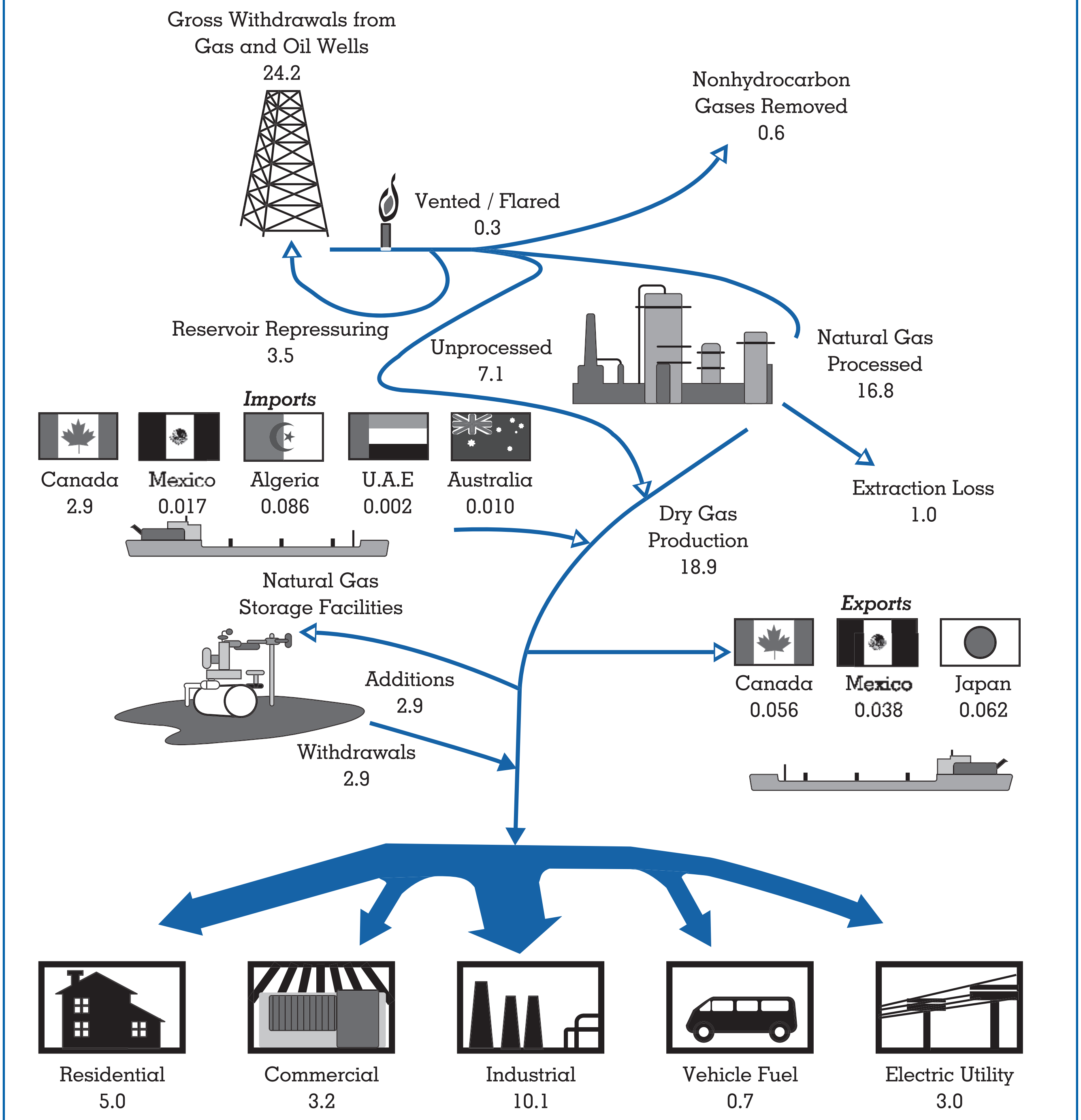


Figure 1.
Natural gas supply and disposition in the United States in trillions of cubic feet.

	Hydrogen	*Natural Gas (as Methane)*	*LPG (as Propane)*	*Gasoline (as Octane)*	*Methanol*
Molecular Weight	2	16	44	114	32
Heat of Combustion kilocalories / gram	34.16	13.30	12.06	11.46	5.42
Grams of carbon per gram of fuel	0.00	0.75	0.82	0.84	0.38
Grams CO_2 evolved per gram of fuel	0.00	2.75	3.01	3.09	1.38
Grams CO_2 evolved per kilocalorie	0.00	0.207	0.250	0.270	0.254

Table 1.
Fuel Combustion Properties

more important from the standpoint of economic efficiency. However, if the natural gas has to be converted chemically to methanol, the high-octane, clean-burning advantage is maintained, but the carbon dioxide advantage is lost.

As seen in Table 1, on a weight basis, hydrogen has the highest heat of combustion of all fuels. Hydrocarbons are less than half as energetic, the lighter, more hydrogen-rich molecules having somewhat higher heating values than the heavier, more carbonaceous fuels. However, on a volumetric basis, the heavier fuels win out, as their higher density (specific gravity) more than off-sets the difference in heat of combustion per unit weight.

COMPRESSED NATURAL GAS

Table 2 compares the heating value of 20-gal tanks of natural gas at different pressures. Pipelines deliver natural gas at a relatively low pressure of 60 pounds per square inch (psi). A 20-gallon vehicular tank filled at this pressure would provide only about 11,000 Btu versus about 2.4 million Btu for a diesel fuel tank of the same size. Thus, the gas must be compressed and stored in a welded bottle-like tank at 3,000 to 3,600 psi to provide any reasonable range. A 20-gallon 3,000 psi tank will provide about 20 percent of the diesel heating value, so to begin to approach the range of gasoline or diesel vehicles, some of the newer vehicles offer advanced tanks capable of holding gas compressed to 5,000 psi.

Although the heating value of a 20-gal 3,000 psi tank of natural gas is only 20 percent of the same volume of diesel fuel, the mileage range comparison will likely be better than the volumetric ratios because the natural gas engines can achieve higher performance.

There are two ways of refueling compressed natural gas (CNG): time-fill or fast-fill. For a time-fill compressor, it is necessary to develop a pressure only slightly greater than the vehicle storage pressure—the gas flows from greater pressure to less pressure. Time-fill compressor stations can require a couple of hours to refuel, which is a major inconvenience for most motorists. However, it is the best choice for many fleets that can be refueled overnight with one fill-post for each vehicle to be refueled.

Public stations are of the fast-fill type, typically to satisfy the desire of customers to refuel quickly. The biggest problem for fast-fill operations is the lack of space and high cost. Large capacity high pressure storage is needed to fast-fill vehicles because the larger and greater the pressure, the faster the fill-up.

All CNG ground storage, vehicle storage and refueling equipment must meet stringent industry and government safety standards for both normal operation and crashes. The controls include monitors of critical pressures and temperatures from the pipeline to the storage tank, and the flow of gas from ground storage to vehicle storage. Once the compressor reaches discharge pressure (fill-up complete), a control then automatically turns off the compressor. CNG is then delivered to the engines as low pressure vapor (ounces to 300 psi). Since natural gas cylinders are much thicker and stronger than gasoline or diesel tanks, the safety record of natural gas vehicles is equal or better than conventionally fueled vehicles.

Almost all the major car, bus, and truck manufacturers have developed compressed natural gas engines and vehicles. These manufacturers have been able to offer better performance (due to higher octane) and far lower emissions of nitrogen oxides, carbon monoxide, particulate matter, and carbon dioxide to the atmosphere. In 1998, Honda introduced the cleanest internal combustion engine vehicle ever commercially produced: the natural gas Civic GX with emissions at one-tenth the state of California's Ultra Low Emission Vehicle standard. Primarily due to the high octane of natural gas, Honda achieved these results without sacrificing performance.

Despite the environmental benefits of natural gas vehicles, large numbers of compressed natural gas stations need to be built or compressed natural gas will never be more than a niche fuel servicing large fleets of buses, cabs, and delivery trucks that can be fueled at a central location. Nonroad short-range vehicles such as forklifts, backhoes, street sweepers, and airport ground support equipment are also ideally suited for natural gas use.

The high mileage, local routes, and regular returns to a central refueling point make local transit buses an ideal application for CNG vehicles. Another option for local driving and commuting is home refueling. For around $2,500, a system can be plumbed directly into a home's natural gas supply for refueling in the garage or driveway. Though the benefit of home refueling is a tremendous benefit for the majority of drivers, the CNG disadvantages of shorter range, slower refueling, and few refueling stations far outweigh its advantages. American car buyers want the mobility to go anywhere at any time. Whereas most gasoline cars and trucks can go 250 miles or more on a tank of fuel, natural gas vehicles typically can go about half that distance. Consumers are reluctant to switch to a CNG vehicle that requires refueling twice as often with so few refueling options.

The industry has developed higher compression tanks to expand the range, and more fast-fill stations are becoming available, yet the prospects of the majority of service stations adding compressed natural gas refueling anytime in the near future are bleak. The oil companies, which control most of the service stations and over 60 percent of America's natural gas reserves, are not eager to make the massive infrastructure investment to cannibalize the billions of dollars they have tied up in refineries, pipelines, and service stations designed to deliver gasoline and diesel fuel. However, though not willing to lead, they are certain to follow. If consumers purchase the vehicles, the oil companies will naturally invest in the infrastructure. Supply will follow demand. Moreover, if it turns out that fuel cell vehicles that run on hydrogen are the future, compressed natural gas vehicles could be the logical bridge between petroleum and hydrogen because hydrogen is also a compressed gas. The same infrastructure can deliver both.

Pressure	*Million BTUs*
Ambient (14.7 psi)	0.00267
60 psi	0.01009
3000 psi	0.546
5000 psi	0.910
diesel oil	2.4

Table 2.
Energy in a 20-Gallon Tank of Natural Gas (in million BTUs)

More stringent clean-air regulations and enforcement was the primary reason for natural gas vehicle growth in the 1990s, and is highly likely to play a part in future growth. According to the Natural Gas Vehicle Coalition, more than 20 percent of all new orders for transit buses were for natural gas-fueled vehicles in 1998. These new vehicles require a significant capital investment, yet often justify this higher initial investment by reducing air pollutants, lower maintenance costs, and offering fuel savings of about 30 percent compared to gasoline and diesel.

LIQUEFIED NATURAL GAS

Natural gas is liquefied by cooling it to its liquid state (approximately –260°F), at either the wellhead, central facility, or on-site. Since liquefaction reduces its volume by a factor of about 600, it becomes economical to ship by tanker.

To remain a liquid at a reasonably low pressure, liquefied natural gas (LNG) must be maintained at below at least –117°F. Insulated storage tanks alone cannot maintain these very cold temperatures. LNG is stored at its boiling point to take advantage of "autorefrigeration." Just as the temperature of water does not rise above its boiling point (212°F) with increased heat (it is cooled by evaporation), LNG is kept near its boiling point if kept at a constant pressure. As long as LNG vapor boil off is allowed to leave the storage tank, the temperature will remain constant. The pressure and temperature in the tank will rise when the vapor is not drawn off.

During the liquefaction process, usually much of the oxygen, carbon dioxide, sulfur compounds and water are removed so that liquefied natural gas (LNG) is nearly 100 percent methane. LNG takes up one-six-hundredth the volume of natural gas, with a density less than half that of water.

Although LNG is as safe or safer than gasoline and diesel fuel, and emits less harmful emissions when burned, it has three major drawbacks: It is expensive to produce, requires a larger and heavier fuel tank (about 1.5 gallons of LNG per gallon of gasoline and 1.7 gallons per gallon of diesel to achieve the same range), and is not the best fuel for vehicles used rarely or intermittently because of vapor boil-off over time. The best applications for LNG are heavy-duty vehicles (trucks and buses) that are heavily used, and vehicles that can store larger fuel tanks, or are not inconvenienced by need for more frequent refueling.

LNG tanks use low pressure (less than 5 psi), yet need double-wall construction so that insulation between the walls keeps the LNG cool. For the large tanks, a cylindrical design with a domed roof is used, but for smaller quantities (70,000 gallons or less), storage is in horizontal or vertical vacuum-jacketed tanks at pressures any where from less than 5 psi to over 250 psi.

Because much of the world lacks the natural gas resources and transportation pipelines of the United States, remote natural gas must be liquefied and transported by ship. Gas-rich countries want to capture stranded gas by liquefying and shipping it to gas-poor regions as LNG. The gas-poor countries enter into contracts so that a long-term supply is available to warrant the investment in the electricity-generating infrastructure. The overall investment is enormous, not only in the liquefaction plant, but in the refrigerated tankers and the regasification plant at the delivery site.

Sometimes LNG is the only option in regions and countries where political issues constrain pipeline development.

Shipments of LNG began in the early 1960s and continued to expand so that by 1995 there were over 65 ships transporting almost 68 million tons of LNG, with each equipped with a specialized refrigeration system to keep LNG cool enough to stay in its liquefied state. Transportation was estimated to reach 107 million tons by 2,000, with the major exporters being Malaysia, Abu Dhabi, and Qatar, and the major importers being Japan, Korea, and Europe. Since OPEC production quotas limit petroleum production, which by extension limits revenue, LNG has also developed into an attractive export commodity for OPEC countries since current production agreements do not extend to natural gas. Several major projects to expand LNG trade went to contract in the late 1990s. New LNG processing facilities have been built or are under construction in Oman, Qatar, Nigeria, and Trinidad, with Japan, South Korea, Taiwan, and Thailand being the largest customers committing to purchase output from the new facilities.

There has never been an LNG tanker accident; yet, with growing shipments, there is growing concern about a tanker accident since an explosion and fire occurring in a crowded harbor could be disastrous. However, while an LNG tanker may contain the energy equivalent of several Hiroshima atomic bombs, the damage would hardly be comparable because the LNG energy cannot be released quickly. For a detonation, LNG must first be mixed with air in the correct flammability ratio, and near a tank rupture the mix would probably be too rich to explode. Further, the liberation rate of LNG as a gas would be determined by the heat transfer rate to the boiling liquid. Thus, any accident would likely be a large deflagration, not a horrific explosion.

CHEMICAL CONVERSIONS

As an alternate to LNG, natural gas can be chemically converted to methanol, chemical feedstocks (such as ethylene), gasoline, or diesel fuel. Most processes start with the conversion of methane to synthesis gas, a mixture of carbon monoxide and hydrogen. This can be done partial oxidation, an exothermic reaction:

$$CH_4 + 0.5\ O_2 \rightarrow CO + 2\ H_2$$

or by steam reforming, an endothermic reaction:

$$CH_4 + H_2O \rightarrow CO + 3\ H_2$$

Shortly after World War I, Badische Anilin patented the catalytic conversion of synthesis gas to methanol, and Fischer and Tropsch (F-T) announced a rival process in which an iron catalyst converted synthesis gas into a mixture of oxygenated hydrocarbons. Later,

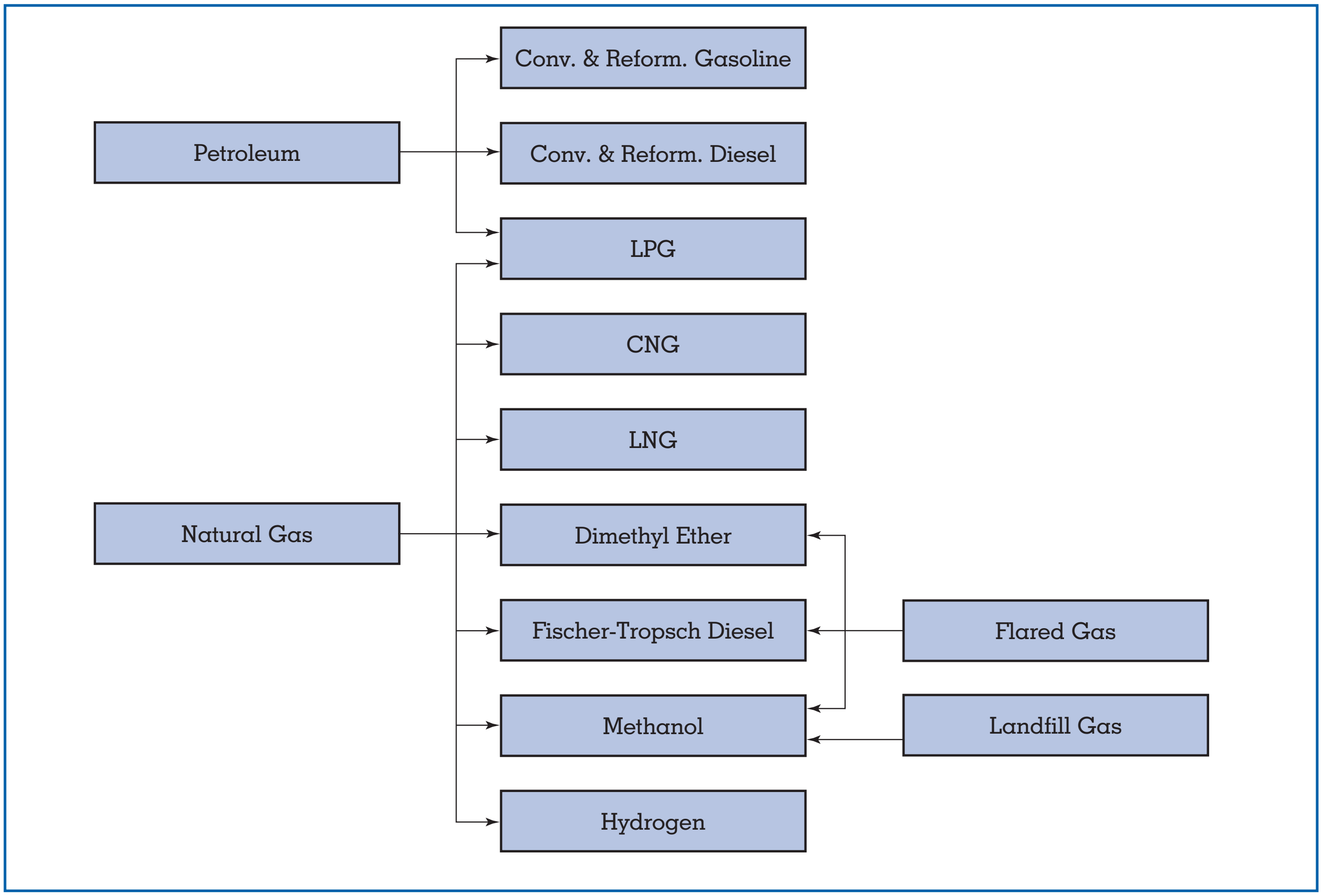


Figure 2.
Alternative fuels derived from natural gas.

improved F-T catalysts produced a liquid resembling a very paraffinic (waxy) crude oil.

Most of the early commercialization used coal as the synthesis gas feed-stock. Use of stranded natural gas feed to a F-T refinery was finally innovated by Shell-Mitsubishi in a small, 10,000 bbl per day refinery in Sarawak, Malaysia in 1993. F-T liquids are refined in the usual manner to produce gasoline and diesel fuel/kerosene of very high quality.

Similarly, a natural gas-to-methanol-to-gasoline process was finally developed by Mobil as a result of the 1973 oil crisis. Methanol is transformed into gasoline range hydrocarbons using proprietary Mobil synthetic zeolite catalysts. This process was commercialized at a small New Zealand refinery in the 1980s. Methanol is also the feedstock of choice in the production of the oxygenated additives needed to produce today's cleaner-burning gasoline blends.

While several other processes have been developed to convert natural gas to liquid fuels (GTL), these technologies are generally uneconomical composed to using the crude oil feedstocks. About one-third of the energy in natural gas is lost in converting it into liquid fuels, so highly distressed gas prices or government subsidies are needed for GTL to be competitive.

Fischer-Tropsch Diesel

One of the most promising GTL fuels is Fischer-Tropsch diesel. Fischer-Tropsch diesel offers lower emissions without compromising fuel efficiency, creating distribution problems (new infrastructure), or requiring a greater investment in equipment for fuel storage and refueling. Depending on the price premium for GTL, the Fischer-Tropsch diesel can be used as a fuel or blended with traditional diesel that is not compliant with Federal or California standards. Used alone, GTL diesel reduces hydrocarbons by over 20 percent, carbon monoxide over 35 percent, nitrous oxide about 5 percent, and particulates around 30 percent. Although GTL diesel is more expensive than traditional diesel fuel, it seems to be a promising short-term solution for the fuel industry to meet the California heavy-duty diesel engine standard that goes into effect in 2004. ARCO, Exxon, Chevron, and Texaco are all in the process of developing pilot plants.

The major advantage of Fischer Tropsch diesel, compared to natural gas, lies in its liquid nature. It does not need special infrastructure and compression like CNG does, and unlike LNG, once converted, it is a liquid fuel that can be treated like any other liquid fuel. However, because the GTL process is more complex than traditional refining, it requires low-cost natural gas priced at less than $1 per million BTUs to remain cost-competitive. Without stranded gas, sources sold at a large discount compared to crude oil, GTL diesel would be considerably more expensive than traditionally refined diesel fuel.

Hydrogen

Many transportation experts feel that hydrogen is the fuel of the future. It has a high energy content and many environmental advantages. However, before hydrogen becomes an economical alternative fuel, ways to produce hydrogen on a large scale will need to be developed. Conversion from natural gas is widely viewed as a promising option for two reasons: Hydrogen-rich natural gas can be converted more cleanly than coal, and natural gas requires less energy input than a conversion from water.

NATURAL GAS AS A FUEL ADDITIVE

Many GTL-derived fuels are being considered for blending with gasoline and diesel to achieve emission reductions of particulate matter (PM), carbon monoxide (CO), nitrogen compounds (NOx) and nonmethane hydrocarbons (NMHC). The most promising fuels converted from natural gas are methanol and ethers such as dimethyl ether (DME) and methyl-t-butyl ether (MTBE).

Like LNG, the natural gas-to-methanol fuel market relies on stranded gas as feedstock. The advantages of conversion to methanol is that it requires far less specialized infrastructure than LNG since the final product is a 110-octane liquid that ships in regular tanks, and does not need regasification. And because of a plentiful natural gas supply in the United States, methanol derived from natural gas as a fuel additive is a promising future market. Methanol has neither the environmental problems of methyl-t-butyl ether (MTBE), nor the evaporating qualities of ethanol.

John Zumerchik
Herman Bieber

BIBLIOGRAPHY

American Petroleum Institute. (1987). *Liquid Fuels from Natural Gas*. Petrol Information, API 34-5250. Washington, DC: Author.

Chen, N. Y.; Garwood, W. E.; and Dwyer, F. G. (1989). *Shape Selective Catalysis in Industrial Applications*. New York: Marcel Dekker.

Dry, M. E. (1990). Fischer-Tropsch Synthesis over Iron Catalysts, Spring 1990 A.I.Ch.E. Meeting, Orlando, Florida. March 18–22, 1990.

Sofranko, J. A. (1988). *Gas to Gasoline: The Arco GTG Process*. Bicentennial Catalyst Meeting, Sydney, Australia.

Tussing, A. R., and Tippee, B. (1995). *The Natural Gas Industry: Evolution, Structure, and Economics*, 2nd ed. Tulsa, OK: PennWell Publishing.

U.S. Energy Information Administration. (1998). *Annual Energy Outlook*. Washington DC: United States Department of Energy.

U.S. Energy Information Administration. (1998). *International Energy Outlook 1998, With Projections Through 2020*. Washington, DC: Author.

Yerchak S., and Wong, S. S. (1992). *Mobil Methanol Conversion Technology Process*. IGT Asian Natural Gas Seminar, Singapore, pp. 593–618.

NATURAL GAS, TRANSPORTATION, DISTRIBUTION, AND STORAGE OF

Transportation of natural gas across state lines from production to consuming areas is a function of interstate pipeline companies. The modern U.S. natural gas industry also includes natural gas exploration and production companies, intrastate pipelines, local distribution companies (LDCs), end-users and, the most recent addition to the industry, marketers.

HISTORICAL BACKGROUND

Transportation of natural gas through pipelines began in the United States in the early part of the nineteenth century. One of the first known uses occurred in 1821 with the building of a system of metallic lead pipes to transport natural gas from a nearby shallow well to commercial establishments in Fredonia, New York. Gas lights—burning gas made from coal—illuminated the streets of Baltimore beginning in 1816.

By 1900, natural gas had been discovered in seventeen states, mostly as a byproduct of oil exploration. Lacking a viable long-distance transportation system to move it to market, however, natural gas, until the 1920s, was used mainly for lighting city streets or was vented into the air when found with oil.

Today, the natural gas industry—responsible for locating, producing, transporting and distributing gas to end-users—is a major contributor to the U.S. economy, employing more than 170,000 workers nationwide. And the fuel, in addition to being used in a variety of commercial and industrial applications—including one of the fastest growing markets, the generation of electricity—is the primary source of energy for space heating in more than fifty million American homes and provides some 25 percent of all the energy consumed in the United States.

This is largely due to the discovery and development of major natural gas fields in the U.S. Southwest, mid-continent, on- and offshore areas of the Gulf of Mexico and Canada—and the development of safe and efficient interstate natural gas transmission pipelines to transport natural gas to markets across the country. Some 77 percent of the natural gas consumers use is produced domestically.

NATURAL GAS PRODUCTION

The first step in the movement of natural gas from production areas to consumers (see Figure 1) begins at the wellhead. Most natural gas wells require pumping to bring the natural gas to the surface. From there, small-diameter gathering lines, connected to clusters or a series of natural gas wells, carry the gas to pipelines or to facilities where it is processed to remove valuable hydrocarbon liquids such as ethane, propane, butane, iso-butane or natural gasoline. Prior to entering the pipeline transmission system, and during actual transmission, the natural gas stream also may undergo processing to remove water vapor, solids and other elements that may interfere with efficient transportation of the fuel.

NATURAL GAS TRANSPORTATION

Natural gas pipelines are generally defined as intrastate, those transporting gas to markets within state boundaries, or interstate, those crossing state lines. Natural gas from Canada, a growing source of natural gas consumed in the United States, also enters the U.S. through interconnections with interstate pipelines.

Today, a network of more than 300,000 miles of interstate natural gas pipelines serves markets across the U.S. Construction of this network began in the 1920s, but large-scale expansion was limited by the technology of the day, the Great Depression and, finally, World War II.

Following War II, advancing technology in steelmaking and the manufacture of pipe, along with increased energy demand, led to rapid growth of the nation's natural gas transportation network. Today, large-diameter, high-quality steel pipe, normally ranging from twenty to forty-two inches in diameter, is constructed at pipe mills, generally in twenty and forty-foot "joints" (sections). Transported by truck or train to construction sites, the joints are welded together and buried in trenches to form the long-distance pipelines necessary to transport natural gas from production basins to distribution and other customers across the nation.

Compressor stations, generally located fifty to sixty miles apart along interstate pipelines, house reciprocating or gas turbine engines that drive compressors to propel the gas through the pipeline. The compressor engines, usually powered by a small portion of the natural gas flowing through the pipelines, compress the gas to an average of 700 to 950 pounds per square inch. In the very early years of the natural gas transmission industry, sections of pipe made of iron or steel were bolted together, often resulting in the loss of major amounts of gas in transit. The advent of modern technology, welded pipe seams, and pipeline protection and monitoring programs have reduced the amount of lost and unaccounted for gas during interstate transportation to negligble quantities, generally a fraction of one percent of all the gas transported.

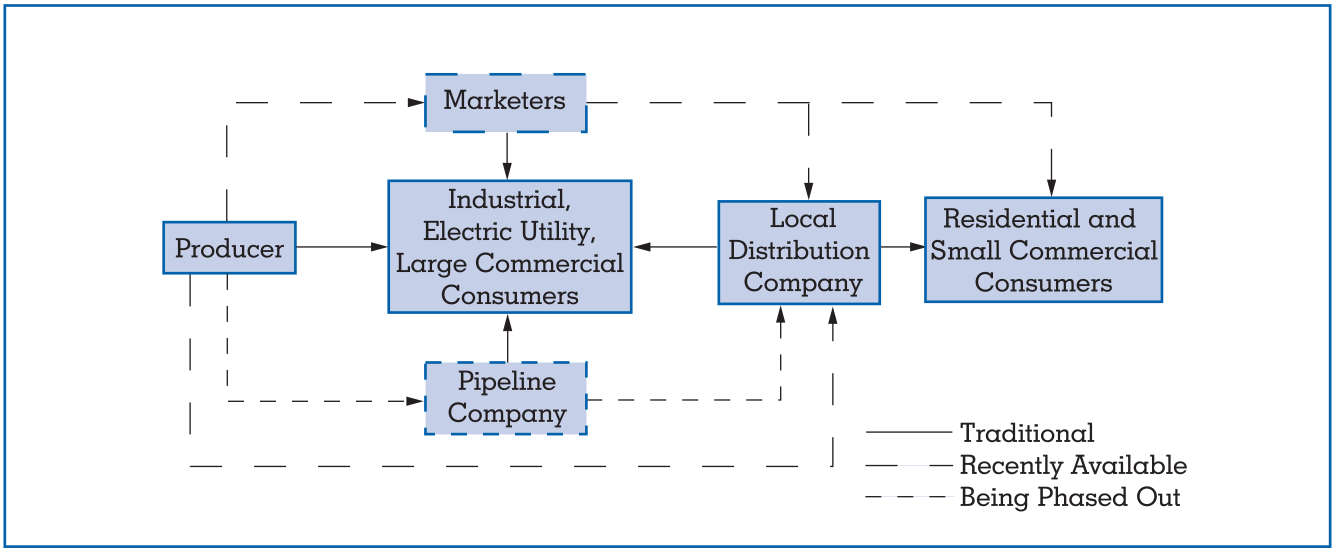


Figure 1.
Transaction paths for natural gas purchases.

BALANCING SUPPLY AND DEMAND

Because of their role as the link between producing and market areas, interstate natural gas pipelines play a crucial role in balancing supply and demand. To do so, the interstate pipelines continually monitor pipeline performance. These companies rely heavily on computers to gather, analyze and retain information on the performance of their pipelines, and to control the flow of gas in remote sections of the line.

Interstate pipelines also use computer simulation programs to calculate pipeline capacity, pressures, horsepower, fuel and other physical characteristics and properties of their systems. Using this information and incorporating variables such as ambient temperatures, facility outages, and changes in market patterns, transmission companies can run daily studies to determine how much natural gas their systems will deliver under expected operating conditions.

Another tool the interstates use to maintain their pipelines is a device known as an intelligent “pig.” Propelled through the pipeline with the gas stream, these devices, taking thousands of measurements with electronic sensors that can be analyzed later by computers, can inspect pipeline interior walls for corrosion or other defects and remove accumulated debris from a section of pipeline. Pipelines also use state-of-the-art coating and cathodic protection to battle corrosion.

Advancing technology has had an impact on the interstate pipeline industry in other ways. Interstate pipeline shippers now have access to electronic bulletin boards (EBBs) which are developed and maintained by the pipelines and can be accessed by users to purchase transportation services, check on billing, determine status of pipeline utilization, arrange for the storage of gas transported through a pipeline and gain other information that allows the pipeline’s capacity to be used efficiently.

NATURAL GAS STORAGE

Underground storage fields play a key role in smoothing the peaks and valleys of natural gas supply and demand. Natural gas transported by pipeline, for instance, can be injected into storage fields near market areas during periods of low demand in the summer and withdrawn during periods of high demand in the winter, thus reducing the need to build additional pipeline capacity to meet high demand periods and allowing pipelines designed to operate at high load factors to run at efficient levels all year. Storage can also be used to prevent flaring of gas and other waste when production exceeds market demand.

Recent uses for natural gas storage include: utilizing storage to provide transportation balancing services, to take advantage of variations in prices, to maintain wellhead production and to meet new sales commitments by marketers.

A crane moves soil next to sections of 42-inch pipes to be used for a gas pipeline. (Corbis-Bettmann)

The three basic types of underground storage facilities for natural gas are depleted oil and gas formations, salt caverns, and aquifer formations. Depleted oil and gas formations, the most commonly used storage sites, are generally classified as seasonal supply reservoirs, where gas is injected in summer and withdrawn in winter. Salt caverns, the large majority of which are located in the Gulf Coast region, are well-suited to storing gas and provide very high withdrawal and injection rates compared with their working gas capacity (the amount of natural gas inventory that can be withdrawn to meet customer needs). Aquifer formations, water-only reservoirs conditioned to hold natural gas, are generally classified as seasonal supply reservoirs and are generally more expensive to develop and maintain than depleted oil or gas reservoirs.

NATURAL GAS DISTRIBUTION

The final link in the physical transportation of natural gas is distribution, the delivery of natural gas by local distribution customers (LDCs) to local consumers. LDCs deliver gas to consumers from storage or from wellheads accessed by pipelines.

Metering gas as it enters their systems, LDCs lower it to pressures suitable for their customers and add the familiar odor that identifies natural gas, otherwise odorless, to the consumer.

For most of the twentieth century, LDCs, either investor owned or municipally owned, have had exclusive rights or franchises to distribute gas in specified geographic areas. Regardless of ownership, LDCs are regulated, either by state public utility/service commissions or local government agencies, to assure adequate gas supply, dependable service and reasonable prices for consumers.

Today, distribution of natural gas involves some 1,200 LDCs that deliver natural gas to more than 160 million American consumers in all fifty states.

RESTRUCTURING OF THE NATURAL GAS INDUSTRY

Government, recognizing the need to provide dependable service at reasonable prices, has been involved with the gas industry since the mid-1800s. Local regulation gave way to state regulation as the issues surrounding exploration, production, transportation, and distribution became increasingly complex. As improved technology allowed pipeline's to carry large volumes of natural gas over great distances, the federal government increasingly became involved in regulation of this segment of the industry.

The first major federal regulatory event affecting natural gas pipelines was the Natural Gas Act of 1938, which gave the Federal Power Commission jurisdiction to regulate three areas: (1) pipeline sales of gas purchased from producers and resold to local distribution companies in interstate commerce; (2) transportation in interstate commerce; and (3) the facilities used for such sales and such transportation.

The second major federal regulatory event was a Supreme Court decision in 1954 which resulted in the imposition of wellhead price regulation of gas sold in interstate commerce.

Until the 1970s, however, the structure of the industry remained relatively constant. Gas producers explored for and produced natural gas and sold it to pipeline companies. The pipeline companies transported the gas across the country and, for the most part, sold it to local distribution companies, who then sold the gas to end users. The price producers sought for their gas from interstate pipelines was regulated by the federal government, as was the price interstate pipelines imposed on gas sold to LDCs. State or local government agencies in turn regulated the price charged by LDCs to end users.

In 1978, the structure of the industry began a period of dramatic change. Following a severe winter and an energy crisis that closed schools, businesses, and industries—and threw more than a million people out of work, Congress passed the National Energy Act (NGPA) and the Public Utility Regulatory Act. This legislation, assuming the nation was running out of natural gas, said that gas could no longer be used for "low-priority" uses and that national policy should promote conservation and shift demand to alternative fuels. The NGPA, to provide producers with incentives for exploration and development of new supplies of gas, called for phased and gradual deregulation of the price of newly discovered gas.

In the early to mid-1980s, retail gas prices began to rise as the more expensive "new" gas constituted an increasing percentage of the pipelines' average cost of gas. This drove consumer prices above the level that would exist in a competitive market, and demand for natural gas was subsequently reduced as large industrial customers switched to other fuels. Also reducing demand were the Fuel Use Act, which prohibited the use of natural gas as a boiler fuel, increased conservation by residential and commercial customers, warmer-than-normal winters, and an economic recession.

In 1985, the Federal Energy Regulatory Commission (FERC) began a series of regulatory actions designed to improve the competitiveness of the natural gas market and give the customers of interstate pipeline companies more service options and thus allow ultimate consumers to benefit from deregulation of wellhead prices.

By 1993, the structure of the interstate pipeline industry had undergone dramatic changes. The interstates, which once had acted as both transporters and merchants of natural gas, "bundling" the sale and transmission of gas into one service, were required to separate these services and give pipeline customers the opportunity to contract for only those services they needed.

With "unbundling," interstates no longer own the gas transported on their pipeline systems. They simply transport it for third parties who range from LDCs to natural gas marketers. (Emerging in the 1980s in response to the deregulation of prices, marketers serve as intermediaries between gas buyers and all other segments of the industry.) And buyers of natural gas can now negotiate price and contract terms with different gas suppliers while contracting separately with the interstates for transportation, storage and other services.

Restructuring has also affected LDCs. Many LDCs, for example, given increasing flexibility by regulators, now allow large volume customers to select the most cost-effective and efficient mix of supply, transportation, storage, and other services.

PRICING OF NATURAL GAS

Several factors determine the ultimate price consumers pay for natural gas. They include

Pipeline Transportation Pricing: The Federal Energy Regulatory Commission (FERC) regulates the rates, terms and conditions of service provided by interstate natural gas pipelines. The FERC, with the issuance of its "restructuring" rule, Order 636, in April 1992, set a policy goal to separate the transportation of gas from the sale of gas and to put all natural gas suppliers and gas purchasers on equal footing. Before restructuring, as described earlier, the interstates bought gas from producers and sold it to LDCs and other end users. The rates they charged were "bundled" and included charges for a variety of services including supply, storage, and transportation. Now the interstates must provide all transportation services on an equal basis, whether the gas being transported is purchased from the pipeline, a pipeline marketer or another gas supplier.

Wellhead Pricing: With passage of the Natural Gas Wellhead Decontrol Act of 1989, all remaining wellhead price controls on natural gas were lifted. As a result, federally mandated natural gas wellhead prices no longer exist. With this, the competitive marketplace has become the most important factor affecting natural gas wellhead prices, along with weather, demand, pricing of competitive fuels and competition for supplies.

The Spot Market: This segment of the natural gas market consists of transactions of a short duration that can be rearranged quickly if gas prices change. The spot market developed in the early and mid-1980s, when there was a substantial supply of gas and long-term supply contracts fell into disfavor. Spot market prices fluctuate depending on availability of supply. Most LDCs do not rely exclusively on spot-market transactions, depending on time of year. In the event of unexpected demand surges, however, LDCs and other customers may be forced to purchase spot market gas.

Futures Prices: Natural gas futures contracts began trading on April 3, 1990, on the New York Mercantile Exchange (NYMEX). Sabine Pipe Line Company's Henry Hub near Erath, Louisiana, is the delivery site for these contracts. Henry Hub consists of an interconnection of seven interstate pipelines, two interstates and one gathering system. These interconnections allow natural gas to move from major production areas to major consumption areas. The transportation of natural gas to and from Henry Hub is contracted separately by the seller and buyer, respectively.

Price Indexing: Some gas prices, by contractual agreement, reflect changes in the spot, futures or other markets, such as heating oil. Price indexing is often used to reflect that the parties to the contract believe the price to be paid is related in some way to the particular index used.

Residential Pricing: Residential natural gas rates are for natural gas service and for the gas commodity. The price of the gas commodity comprises about one-third of the total price a residential customer pays, on average. The remainder of the bill includes amounts for transmission and distribution of gas, system maintenance, safety and inspection programs, customer service, metering, billing and other costs.

THE FUTURE OF THE NATURAL GAS INDUSTRY

Additional Regulatory Change

The natural gas industry, undergoing fundamental changes in recent years, continues to face additional change. Current issues include discussion about lifting the federally-set caps on pipeline pricing for certain services and letting the market determine pricing on these services, as well as a number of new pipelines proposed to import significant new volumes of Canadian gas into the United States.

On the state level, several pilot programs now allow customers to choose gas suppliers other than their LDCs.

Supply Outlook

Debate continues over just how much natural gas remains in North America. Part of this debate centers on the definition of gas reserves—the amount of gas in a given area that is recoverable and gas resource, the total amount of gas in the ground. Gas production to date, as might be expected, has most commonly been from easy-to-produce conventional sources, such as the mid-continent or the highly permeable sands of the Gulf Coast.

A considerable amount of gas is located in areas of lower permeability or where the geology is poorly understood, making it more difficult to earn a profit using current technology and production methods. New technology may open these sources to future development and production.

Additionally, new supplies of natural gas from Canada and Alaska could significantly increase available supply, along with imports of Mexican gas and the potential for liquefied natural gas from overseas.

Demand Outlook

Demand for natural gas, in all markets—residential, commercial, and industrial—is projected to grow into the foreseeable future, particularly in the electric power generation market and the industrial sector. Total natural gas use in the United States is projected to grow from 20.1 quadrillion British thermal units in 1992 to 26.1 by 2010, an average growth rate of 1.6 percent per year.

Robert V. McCormick

See also: Natural Gas, Consumption of; Natural Gas, Processing and Conversion of; Oil and Gas, Drilling for; Oil and Gas, Exploration for; Oil and Gas, Production of.

BIBLIOGRAPHY

Capozza, F. C. *Energy Futures Trading in the '90s,* 2nd ed. Red Bank, NJ: Waterman Associates, Inc.

Commoner, B. (1983). "A Nearly Perfect Fuel." *New Yorker Magazine,* May 2.

Energy Policy Act Transportation Study. (October 1995). Washington, DC: Energy Information Administration.

Natural Gas: The Fuel, The Future, The Regulation. Detroit, MI: ANR Pipeline Company.

Wright, J. C., ed. (1994). *The Universal Almanac, 1994.* Kansas City, MO: Andrews and McMeel.

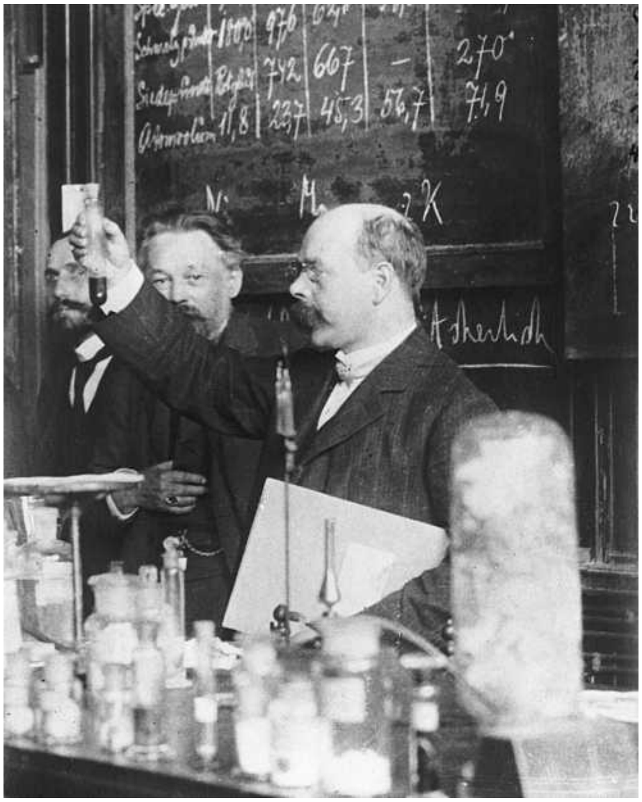
NATURAL GAS REGULATION AND RATES

See: Regulation and Rates for Electricity

NERNST, WALTHER HERMANN (1864–1941)

Walther Hermann Nernst was born in Briesen on June 25, 1864, into a prominent Prussian family. The line of his ancestors can be traced back to the era of the first Prussian king in the seventeenth century. His most famous ancestor is Hermann Nernst, who received an award for conveying the news to the king's family about the Germans victory against Napoleon in the Battle of Waterloo. Nernst lost his mother, Ottilie Nerger, very early. His father, Gustav Nernst, was a district judge. Nernst was a very impulsive person and known for his impatience. He died of a heart attack in his country house in Zibelle on November 18, 1941.

When Nernst attended secondary school, his chemistry teacher awakened in him a passion for chemistry. This left a deep impression on the young pupil, who started early to do his own chemistry experiments at home, in the cellar. Nernst often changed universities during his study of physics, as he wanted to take part in selected lectures by the most influential comtemporary scientists. First he visited the physics lectures of Heinrich Friedrich Weber in Zurich in 1883 and then he went to Berlin for a semester to take thermodynamics lessons from Hermann von Helmholtz, the most famous German physicist at that time. After another semester in Zurich, Nernst studied for the fourth semester in Graz, where Ludwig Boltzmann founded the statistical interpretation of thermodynamics. Nerst, finished his education there and started his Ph.D. program in collaboration with Albert von Ettinghausen, one of Boltzmann's students. Nerst, finished his Ph.D. requirements in Wurzburg in 1887 under the supervision of Friedrich Whilhelm Georg Kohlrausch, who entered a completely new physical field with his research on the ionic conductance of liquids. Nernst adopted Boltzmann's atomistic view (or mechanical picture) of natural phenomena and combined it with the energetic view that he became familiar with during his collaboration with Wilhelm Ostwald in



Walther Hermann Nernst in his laboratory. (Corbis-Bettmann)

Leipzig. It was due to Svante Arrhenius that Nernst became highly interested in the new topic of ionic theory that was established by Ostwald. Fascinated by this topic, Nernst habilitated in Leipzig in 1889 on the electromotive activity of ions. He incorporated into this work Arrhenius' theory of dissociation and Jacobus Henricus van't Hoff's osmotic theory of solutions. Two years later Nernst became an assistant professor in physical chemistry at the University of Gottingen. In 1905 he was offered an appointment as the successor of Hans Landolt and started his work at the University of Berlin. In the same year he discovered the third law of thermodynamics during the time he was lecturing on the thermodynamical treatment of chemical processes. Nernst worked in many fields of physical chemistry—for example the osmotic theory of galvanic elements, the law of distribution, equilibrium at high temperatures and high pressures, specific heats at high and low temperatures, calorimetry, IR radiation, chain reactions, electrochemistry, and photochemistry.

Of fundamental importance to Nernst's discovery of the Third Law of Thermodynamics is Nicholas Leonard Sadi Carnot's work in the middle of the nineteenth century establishing thermodynamics by combining the laws of motion with the concept of action. At the same time Helmholtz found that the principle of energy conservation is also valid for thermal energy and established together with Robert Julius Mayer and James Joule the first law of thermodynamics. A short time later Rudolf Clausius and Carnot introduced the concept of entropy and established the second law of thermodynamics. The chemical affinity—the extent to which a compound is reactive with a given reagent—was known of since attempts of alchemists to produce gold in the Middle Ages. The importance of the concept of chemical affinity was first recognized by van't Hoff. He showed that it was possible to measure the chemical affinity via free energy. Helmholtz and Josiah Gibbs independently established an exact mathematical relationship between the total energy (the energy content of a system) and the free energy (the capacity of a system to perform work). According to Marcellin Pierre Berthelot the total and the free energy should be the same during the electrochemical processes in the galvanic cells Nernst was investigating. Nernst's experiments showed that this statement was not exactly valid for moderate temperatures, and the deviation became larger as the temperature was increased. Therefore he assumed that both energies should be equal at zero value of temperature. Finally, he got the idea that the difference between these two energies asymptotically approaches zero as the absolute temperature approaches zero. This is the first formulation of the new law. Nernst was led to this law, which he and other scientists spent a long time investigating, while he was searching for mathematical criteria for the description of chemical equilibrium and the spontaneity of chemical reactions. That why some chemical reactions are spontaneous while others are not was already known a century before Nernst's own discovery. Since 1900 it was known that a thermodynamical calculation of the chemical equilibrium could not be performed using only the thermal data of the current thermodynamics because the integrated form of the Gibbs-Helmholtz Equation—which allowed calculation of the maximum yield of work during a thermodynamical process—contained an undetermined integration constant.

Nernst's contribution to the solution of this problem was to give this undetermined constant a new interpretation. This new interpretation, however, required an additional assumption for the description of the free energy exchange in the vicinity of the absolute zero value of temperature. He recognized that the work function for the state transition of a system could not be calculated by means of energy differences. Rather, derivatives of the energy and the free energy with respect to temperature were necessary. Furthermore, Nernst assumed that the entropy approaches a constant value provided the absolute temperature approaches zero.

After Nernst's publication of his *Heat Theorem* all resources of the Institute for Physical Chemistry at the University of Berlin were dedicated to its experiments. No experimental result was found to contradict Nernst's heat theorem. Its general validity, which was established by experiments in many subfields of physical chemistry, justified its inclusion among the laws of thermodynamics. This law was regarded by Arnold Sommerfeld as "the most ingenious extension the classical thermodynamics ever experienced in the twentieth century." Expressing Nernst's heat theorem in terms of entropy, this law means that the entropy difference between different states of a system tends to zero as the temperature reaches absolute zero. It can also be expressed as the law of unattainability of absolute zero temperature. Franz Eugen Simon provided Nernst's heat theorem a more elegant theoretical foundation and reformulated it by stating that the entropy differences disappear between all those states of a system that are in internal thermodynamic equilibrium. Very soon after he set forth the third law of thermodynamics, Nernst showed the importance of his discovery by calculating the chemical equilibrium using thermal data only. Nernst was awarded the Nobel Prize in chemistry in 1920 as recognition of his work in thermochemistry.

Tyno Abdul-Redah

See also: Carnot, Nicholas Leonard Sadi; Clausius, Rudolf Julius Emmanuel; Gibbs, Josiah Willard; Heat Transfer; Helmoltz, Herman von; Joule, James Prescott; Ostwald, Wilhelm; Thermodynamics.

BIBLIOGRAPHY

Gillispie, C. C., ed. (1978). *Dictionary of Scientific Biography,* Vol. 15, p. 432. New York: Scribner.

Einstein, A. (1942). "The Work and Personality of Walther Nernst." *Scientific Monthly,* February, p. 195.

Mendelsohn, K. (1973). *The World of Walther Nernst: The Rise and Fall of German Science, 1864–1941,* Pittsburgh: University of Pittsburgh Press.

Nernst, W. H. (1893). *Theoretische Chemie vom Standpunkte der Avogadroschen Regel und der Thermodynamik.* Stuttgart, Germany: F. Enke.

Nernst, W. H. (1906). "Ueber die Berechnung chemischer Gleichgewichte aus thermischen Messungen." *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen,* pp. 1–40.

Nernst, W. H. (1913). *Experimental and Theoretical Applications of Thermodynamics to Chemistry.* New Haven, CT: Yale University Press.

Partington, J. R. (1953). "The Nernst Memorial Lecture." *Journal of the Chemical Society,* p. 2853.

NEWCOMEN, THOMAS (1663–1729)

Fable or fame? Thomas Newcomen, like many inventors who preceded him in the steam revolution, has been clearly overshadowed in historical circles by the far more famous Scotsman, James Watt, who remains—incorrectly to some—known as the inventor of the steam engine. Watts engines arrived more than fifty years after Newcomen's successful mechanical works, and were considered improved versions of the Englishman's concepts. But this was precisely the basis of many inventors' successes, building upon their predecessors' efforts in the normal course of technological advancement. What is irrefutable is that both men, as well as others, can lay claim as pioneering "fathers" of the Industrial Revolution.

Newcomen came from the ranks of practical tradesmen, unlike many industrial inventors who tended to be noblemen, philosophers and royal protégés. The Newcomen family had had an impressive lineage and had held its manor from the twelfth century until misfortune dropped them into obscurity four centuries later. Yet, a work ethic was instilled by Newcomen's grandfather, who became a merchant venturer (owning several ships), a freeholder of Dartmouth, treasurer for his town, and a staunch Parliamentarian. Elias Newcomen, the father of Thomas, was also a freeholder and a merchant of Dartmouth, trading to distant areas with a ship that he had inherited.

It is reasonably certain that Thomas Newcomen was born in late January or early February 1663 in the family house in Dartmouth in Devon, England. He was schooled at home by the well-known nonconformist scholar, John Flavell, who played a key role in Newcomen's educational thinking. Although throughout his life Newcomen was proud of his common status as an ironmonger, some rivals attempted to credit his success to no more than good luck and chance. Many of his contemporaries doubted that he could be the sole author of so momentous an invention.

But going solo neither accurately summed up the style of Newcomen's inventiveness, nor his ability to learn from others and to work with them. In his teens, he served an apprenticeship before joining his partner, John Calley, in their shop at Dartmouth. By the time Newcomen began solving technical tin mining challenges at the beginning of the eighteenth century, he was already well established as an iron-monger, dealing in annual quantities as high as twenty-five tons per year. He apparently became acquainted with Thomas Savery, or at least his works, observing firsthand the pumping problems in the mines. He may also have read about earlier research on atmospheric pressure done by a Huguenot, Denis Papin, late in the seventeenth century.

Although Newcomen's predecessors may have had superior backgrounds and academic intelligence, his doubters possibly overlooked the practical advantages he and his partner brought to the research table. Through astute knowledge of every variety of metalwork in iron, brass, copper, tin, and lead, he and Calley possessed the right combination of commercial ability and highly practical and versatile craftsmanship. Historical data suggests that these categories may have been deficient among his forerunners in the matter of applying the principles of steam technology.

Yet, Newcomen's success came with great difficulty and frustration, taking ten years before achieving initial success at a South Staffordshire colliery (mine works) at Dudley Castle in 1712. He perfected a variation of Savery's work by creating a dramatic advance of an internal water injection system to cool the heated steam, creating a far more rapid and effective condensation and ensuing vacuum. There is some historical evidence that the arrival of this particular breakthrough came by fortunate accident, whereby an unintended leak may have led him to design his internal water-cooling system.

In another significant change, he connected one end of a large overhead rocking beam, or "great lever," to this piston. Realizing the risks of high pressure steam, he thus wisely relied on two basic forces to drive his engine—atmospheric pressure to plunge the piston down, and simple gravity to lift it back it up, through the weight of the counterbalancing rocker beam.

By modern standards, it was barely an engine at all, but it met the most important criteria of its time by dramatically meeting the challenge of pumping water from the mines, doing the equivalent work of twenty horses and forty men. The use of his excellent pumping engine spread throughout Europe, with over one hundred engines constructed within the life of the patent, which expired four years after his death on August 5, 1729, at the age of sixty-six. Well over fifteen hundred units were eventually built during the eighteenth century, despite their significant cost (over one thousand pounds), high consumption of fuel, and the requirement of an operating license.

The usefulness of Newcomen engines was both startling and unprecedented, with immediate impact benefiting the mining industry for the next century or more. The record for longevity for a Newcomen-type engine was probably set at the South Liberty colliery of the Ashton Vale Iron Company. It was built around 1750, and was still pumping from a depth of seven hundred feet it was until dismantled in 1900. Another remarkable example was at the Cannel Mine at Bardsley, where a Newcomen engine worked from 1760 to 1830, after which it rested in a derelict state for a full century. In 1930 it was acquired by Henry Ford and shipped to his museum in Dearborn, Michigan where it was restored and re-built.

While not nearly as sophisticated as later engines, the quantity of Newcomen engines built vastly exceeded the hundreds of Watt engines, with the more primitive types often used as substitutes for the newer, far more expensive versions.

Dennis R. Diehl

BIBLIOGRAPHY

Rolt, L. T. C., and Allen, J. S. (1997). *The Steam Engine of Thomas Newcomen.* Ashbourne, UK: Landmark Publishing.

NEWTON, ISAAC (1642–1727)

Isaac Newton was born at Woolsthorpe, near Grantham in Lincolnshire. He entered Cambridge University as a student in 1661. Although much is known of Newton's professional life, little is known about Newton's student life. He studied under Isaac Barrow, the Lucasian professor of mathematics. He was forced by the plague of 1665–1666 to return to Lincolnshire where, during the miraculous year of 1666, he forged the foundations for his considerable achievements in mathematics, optics, and dynamics.

After returning to Cambridge in 1667, Newton was elected Fellow of Trinity College. Two years later he succeeded Barrow as Lucasian Professor. In 1696 Newton moved to London. He served first as Warden and from 1699 to his death in 1727 as Master of the Royal Mint. He was elected a Fellow of the Royal Society of London in 1671, and the President of this society in 1703, a position he retained for the rest of his life. He also served two undistinguished terms as a Member of Parliament for the University of Cambridge (1689–1690 and 1701–1702). He was knighted in Cambridge in 1705.

In the period following the War of the Spanish Succession of 1714, Newton enjoyed a reputation as the most important natural philosopher of his day. His scientific output during the last twenty years of his life was restricted to the revision of his major works and in defending himself against his many critics. In his private life, he was modest, generous, and given to simple tastes; however was buried with great fanfare in Westminster Abbey.

In his scientific life, his behavior and demeanor were far different. Newton was hostile to any criticism and he was capable of ruthless behavior. The priority dispute with the German mathematician Gottfried Wilhelm Leibniz over the invention of the calculus is a case in point. In his capacity as President of the Royal Society, Newton appointed a committee of loyal Newtonians to investigate the matter. He then authored the committee's report, the infamous *Commercium Epistolicum* and submitted it as though it were an utterly impartial report in his own favor.

Newton has been regarded as the very exemplar of the modern scientist, employing a style of reasoning from "the phenomena of motions" that other scientific disciplines have attempted to emulate with limited success. However, his interests ranged far from mathematical physics. From 1669 to 1696 he pursued alchemy and chemistry with equal or greater passion. His knowledge of the Greek classics was profound. In his *The Chronology of Ancient Kingdoms Amended* (1728), he attempted to reconcile Jewish and pagan dates and to fix them absolutely from an astronomical argument about the earliest constellation figures devised by the Greeks. In *Observations upon the Prophecies of Daniel and the Apocalypse of St John* (1733), Newton also wrote on Judeo-Christian prophecy, sustained by the conviction that its decipherment was essential to the understanding of God.

Isaac Newton. (Archive Photos, Inc.)

As to his scientific output, Newton contributed to all extant branches of mathematics, but is renowned for his solutions to the problems in analytical geometry of drawing tangents to curves (differentiation) and defining areas bounded by curves (integration). Newton discovered that these problems were inverse to each other, and designed general methods of resolving problems of curvature—his "method of fluxions" (from the Latin meaning "flow") and "inverse method of fluxions." Fluxions were expressed algebraically, but

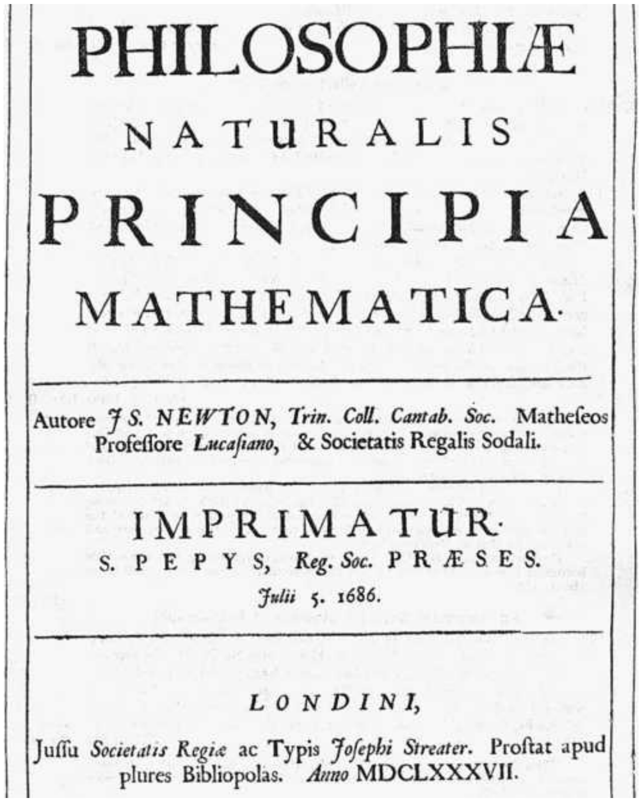
Newton made extensive use of analogous geometrical arguments. Late in life, Newton expressed a preference for the geometric style of the Classical Greeks, which he regarded as more rigorous.

During 1669, Newton worked out the details of his discovery of the decomposition of a ray of white light into rays of different colors by means of a prism. Newton's explanation of the theory of the rainbow followed from this discovery. These discoveries formed the subject matter of a series of lectures that he delivered as Lucasian Professor in the years 1669, 1670 and 1671. The results were communicated to the Royal Society in February 1672 and subsequently published in the *Philosophical Transactions.* Reactions to Newton's ideas were negative. The inability of the French physicist Edmé Mariotte to replicate Newton's prism experiments in 1681 entrenched the rejection of Newton's optical theory for a generation. Newton delayed the publication of his *Opticks* (published in 1704, revised in 1706), which was largely composed in 1692, until his critics were dead. The manuscript of his original lectures was printed in 1729 under the title *Lectiones Opticae.*

Through a curious set of circumstances, Newton failed to solve the problem of chromatic aberration, and so, he abandoned the attempt to construct a refracting telescope which should be achromatic, and instead designed a reflecting telescope, probably on the model of a small one that he had constructed in 1668. The form he used is known by his name today.

Newton's dynamics is presented in *Philosophiae Naturalis Principia Mathematica* (published in Latin in 1687, revised in 1713 and 1726, and translated into English in 1729). According to the famous story, on seeing an apple fall in his orchard sometime during 1665 or 1666, Newton conceived that the same force governed the motion of both the Moon and an apple. Newton calculated the force needed to hold the Moon in its orbit, as compared with the force pulling an object to the ground. He also calculated the centripetal force needed to hold a stone in a sling, and the relation between the length of a pendulum and the time of its swing.

These thoughts were put away until correspondence with Robert Hooke (1679–1680) redirected Newton to the problem of the path of a body subjected to a centrally directed force that varies as the inverse square of the distance. Newton calculated this path to be an ellipse, and so informed the astronomer Edmond Halley in August 1684. Halley's urgings prompted Newton to return to the problem raised by Hooke.



PHILOSOPHIÆ
NATURALIS
PRINCIPIA
MATHEMATICA.

Autore *JS.* NEWTON, *Trin. Coll. Cantab. Soc.* Matheſeos
Profeſſore *Lucaſiano*, & Societatis Regalis Sodali.

IMPRIMATUR.
S. PEPYS, *Reg. Soc.* PRÆSES.
Julii 5. 1686.

LONDINI,
Juſſu *Societatis Regiæ* ac Typis *Joſephi Streater*. Proſtat apud
plures Bibliopolas. *Anno* MDCLXXXVII.

Facsimile of Newton's *Philosophiae Naturalis Principia Mathematica.* (Corbis Corporation)

The result of Newton's labor was *Principia,* which introduced a radically new style of reasoning in science, one that treated scientific problems as though they were exercises in pure mathematics. Newton's work enabled scientists to study forces of different sorts without any inhibiting considerations as to whether such forces can actually (or do actually) exist in nature.

Book I of *Principia* is organized around fundamental laws and axioms, erecting on this foundation the mathematics of orbital motion around centers of force. Gravitation is identified as the fundamental force controlling the motions of the celestial bodies. Newton did not advance an explanation for gravity itself, much to the disappointment of his critics. Its manner of action and its cause were matters that Newton worked hard to distance from the domain of mathematical physics. Further to this, the first half of this book is not entirely original, since it rests upon the achievements of René Descartes, Galileo, and Christian Huygens. Newton's genius shines, however, in the second half of the book where the problem of two bodies is reduced to an equivalent problem of one body attracted to a fixed center. It is here that we find the masterpiece of this book—Newton's treatment of the three-body problem. It is true that his formulation of the general laws of mechanics ruled out a solution to this problem, but Newton's approach inspired future generations of physicists.

Book II investigates the dynamical conditions of fluid motion. Book III displays the law of gravitation at work in the solar system. It is demonstrated from the revolutions of the six known planets, including Earth, and their satellites, though Newton could never quite perfect the difficult theory of the Moon's motion. It is also demonstrated from the motions of comets. The gravitational forces of the heavenly bodies are used to calculate their relative masses. The tidal ebb and flow and the precession of the equinoxes is explained in terms of the forces exerted by the Sun and Moon. These demonstrations are carried out with precise calculations.

Newton's work in dynamics was accepted at once in Britain, though it was strongly resisted on the Continent until 1740 or so when the last of his critics conceded that, while gravitation itself was inconceivable, Newton's arguments were incontestable. During the eighteenth century, Newton's dynamics was extended and perfected by others, but its basic character was unchanged. It was only in the late nineteenth century that Newton's dynamics began to reveal its limitations. Where Newton regarded the law of gravitation as a significant dynamical result, Albert Einstein argued that this law is a geometric result, on a par with Galileo's law of inertia. The movement of a planet around the sun, for example, is to be seen, not in terms of the action of gravity, but in terms of the curvature of space.

Brian S. Baigrie

BIBLIOGRAPHY

Cohen, I. B. (1980). *The Newtonian Revolution.* Cambridge: Cambridge University Press.

Dobbs, B. J. T. (1975). *The Foundations of Newton's Alchemy or "The Hunting of the Greene Lyon."* Cambridge: Cambridge University Press.

Halley, E. (1687). "Review of Newton's Principia." *Philosophical Transactions* 186:291–297.

Herivel, J. W. (1965). *The Background to Newton's Principia. A Study of Newton's Dynamical Researches in the Years 1664–84.* Oxford: Clarendon Press.

Koyré, A. (1965). *Newtonian Studies.* Chicago: University of Chicago Press.

Maclaurin, C. (1748). *An Account of Sir Isaac Newton's Philosophical Discoveries, in Four Books.* London: A. Millar and J. Nourse.

Newton, I. (1934). *Sir Isaac Newton's Mathematical Principles of Natural Philosophy and His System of the World,* tr. A. Motte, rev. F. Cajori. Berkeley: University of California Press.

Newton, I. (1952). *Opticks, or A Treatise of the Reflections, Refractions, Inflections & Colours of Light.* New York: Dover Publications.

Newton, I. (1958). *Isaac Newton's Papers and Letters on Natural Philosophy and Related Documents,* eds. I. B. Cohen and R. E. Schofield. Cambridge: Harvard University Press.

Newton, I. (1959–1977). *The Correspondence of Isaac Newton*, eds. H. W. Turnbull, J. F. Scott, A. R. Hall. Cambridge: Cambridge University Press.

Newton, I. (1962). *The Unpublished Scientific Papers of Isaac Newton: A Selection from the Portsmouth Collection in the University Library, Cambridge*, ed. A. R. Hall and M. B. Hall. Cambridge: Cambridge University Press.

Newton, I. (1967). *The Mathematical Papers of Isaac Newton,* ed. D. T. Whiteside. Cambridge: Cambridge University Press.

Newton, I. (1975). *Isaac Newton's 'Theory of the Moon's Motion' (1702).* London: Dawson.

Pemberton, H. (1728). *A View of Sir Isaac Newton's Philosophy.* London: S. Palmer.

Stukeley, W. (1936). *Memoirs of Sir Isaac Newton's Life*, ed. A. H. White. London: Taylor and Francis.

Westfall, R. S. (1971). *Force in Newton's Physics: The Science of Dynamics in the Seventeenth Century.* London: Macdonald.

NITROGEN CYCLE

All life is dependent on nitrogen; nitrogen is a critical component of amino acids, protein, DNA, and RNA. While the atmosphere is 78 percent nitrogen, animals and plants cannot convert nitrogen directly from the atmosphere into a utilizable form. The nitrogen cycle is the process by which nitrogen from the atmosphere is converted to biological nitrogen compounds in plants and animals and then is returned to the atmosphere. The major steps in the nitrogen cycle are nitrogen fixation, nitrification, nitrogen assimilation, denitrification, and ammonification.

Nitrogen fixation is the conversion of nitrogen (N_2) in the atmosphere to nitrates (NO_3^-) and ammonia (NH_3). In nature, this is done primarily by microorganisms (bacteria and algae). The most important bacteria in this process are symbiotic bacteria that live on nodules on the roots of leguminous plants such as soybeans, peas, and alfalfa. The electricity in lightning also can cause a small amount of nitrogen to be fixed by directly combining nitrogen and oxygen. Nitrification is the process wherein ammonia is further converted to nitrates. Plants use nitrates to make proteins by nitrogen assimilation. Animals consume these plants for subsequent protein production. The cycle is completed by bacteria converting decaying plants, animals, and excretory products back into either atmospheric nitrogen (denitrification) or ammonia compounds (ammonification).

The natural supply of nitrogen available to plants from the nitrogen cycle is limited. To meet the growing demands for agriculture crops, nitrogen is added to soil in the form of fertilizers. An estimated one-third of the human population is fed as a result of the use of synthetic fertilizers.

To “fix” nitrogen in synthetic fertilizers requires hydrogen that is obtained by dissociation from water or from fossil-fuel hydrocarbons. Either way, this process is energy-intrusive. The situation is ominous to some scientists, since the supply of a fundamental world energy source—food—is dependent on nonrenewable energy sources that are being depleted. Although the addition of fertilizers is the largest human impact on the global nitrogen cycle, the burning of fossil fuels also accounts for an increase in the amount of fixed nitrogen in the atmosphere. Anthropomorphic activity is responsible for as much nitrogen fixation as that from natural sources.

There are environmental consequences to human influence on the nitrogen cycle. While the addition of nitrogen to the soil is required to feed the world population, it is estimated that 50 percent of the nitrogen added as fertilizers is washed off the soil and ends up in agricultural runoff. This can cause contamination of drinking water in agricultural areas. Excess fertilizers in waterways also can cause eutrophication, where excess nitrogen leads to events such as algae bloom, which robs other plants and animals of oxygen. Agricultural runoff has altered areas ranging from San Francisco Bay to the Baltic Sea to the Great Barrier Reef. In addition, excess nitrates in the soil can acidify the soil, causing loss of other soil nutrients.

Excess fertilizer and combustion processes also can increase nitrous oxide (N_2O) and nitrogen oxides (NO_X) in the atmosphere. Nitrous oxide is a powerful greenhouse gas, and nitrogen oxides lead to smog and acid rain. The production of fertilizers requires a great deal of energy. The use of fossil fuels to supply the thermal requirements for fertilizer production further increases emission of nitrogen compounds to the atmosphere.

Deborah L. Mowery

See also: Acid Rain; Agriculture; Biological Energy Use, Cellular Processes of; Biological Energy Use, Ecosystem Functioning of; Green Energy.

BIBLIOGRAPHY

Galloway, J. N.; Schlesinger, W. H.; Levy, Hiram, II; Michaels, A.; and Schnoor, J. L. (1995). “Nitrogen Fixation: Anthropogenic Enhancement-Environmental Response.” *Global Biogeochemical Cycles* 9:235–252.

Kinzig, A. P., and Socolow, R. H. (1994). “Human Impacts on the Nitrogen Cycle.” *Physics Today* 47(11):24–31.

Smil, V. (1997). “Global Population and the Nitrogen Cycle.” *Scientific American* 277(1):76–81.

Vitousek, P. M.; Aber, J.; Howarth, R. W.; Likens, G. E.; Matson, P. A.; Schindler, D. W.; Schlesinger, W. H.; and Tilman, G. D. (1997). “Human Alteration of the Global Nitrogen Cycle: Causes and Consequences.” *Ecological Applications* 7:737–750.

NORTH AMERICAN ELECTRIC RELIABILITY COUNCIL

See: Government Agencies

NUCLEAR ENERGY, BASIC PROCESSES OF

Nuclear energy, sometimes referred to as atomic energy, originates in the atomic nucleus, which is the extremely dense core at the heart of an atom. A large

Nuclear Power Plant reactor building and cooling tower (left). (Field Mark Publications)

amount of energy can be released by nuclei in two different ways: fission, in which a very large nucleus is induced to break apart into two smaller ones, and fusion, in which two very small nuclei combine.

A great deal of energy is released when a large nucleus undergoes fission, but for most nuclei the fission process is not easy to initiate. There are very few nuclei—uranium-235 and plutonium-239, in particular—that are relatively easy to fission. At present, commercial nuclear reactors use fission of uranium-235 as the energy source. A uranium-235 nucleus can be induced to undergo fission through interactions with a slowly-moving neutron. The uranium nucleus absorbs the neutron, thus becoming a uranium-236 nucleus, which then breaks apart into two smaller nuclei called fission fragments. In addition, neutrons are also released, and these neutrons can then induce fission of other uranium nuclei in a chain reaction. Neutron-absorbing materials are located in the reactor core to ensure that the chain reaction proceeds at the proper pace. If too few neutrons are available for further fissions, then the reaction slows down or stops. If too many neutrons are available, then the reactor core can overheat.

The energy from nuclear fission is released mainly as kinetic energy of the new, smaller nuclei and neutrons that are produced. This kinetic energy is essentially heat, which is used to boil water to generate steam that turns turbines to drive electrical generators. In a nuclear power plant, the electrical generation area is essentially the same as in a plant that burns fossil fuels to boil the water.

Nuclear fission is also involved in nuclear weapons. To create a bomb, the concentration of the isotope uranium-235 must be increased to at least 85 percent from its natural concentration of only 0.7 percent. This increase of concentration is difficult and expensive. In a typical nuclear reactor the uranium-235 concentration in the fuel is only 3 to 4 percent, and hence a nuclear reactor cannot explode like a bomb. In a nuclear bomb the chain reaction is uncontrolled, and a large number of uranium nuclei undergo fission in a very short period of time, producing a nuclear explosion.

Generation of electricity by nuclear fission power reactors has many advantages. A primary advantage is that very little uranium fuel is required—only about two-hundred tons annually for a typical reactor. To generate the same amount of electricity, a coal-fired plant needs 3 million tons or 15,000 times as much coal. Another advantage is that nuclear power produces no air pollution such as nitrogen oxides and sulphur dioxide, both of which contribute to acid rain, and no greenhouse gases such as carbon dioxide and methane that contribute to global warming. Nuclear power also produces no ozone and no particulate matter in the air, and hence, in terms of air pollutants, a nuclear plant is much "cleaner" than a fossil-fuel electrical plant.

Nuclear reactors, however, do generate highly radioactive waste. This waste, which consists primarily of the fission fragments and their radioactive-decay products, must be stored for many years before its radioactivity decays to a reasonable level, and the safe long-term storage of this waste is a matter of great concern and debate. Fortunately, the volume of waste that is created is only about 20 cubic meters annually from a reactor, compared with 200,000 cubic meters of waste ash from a coal-fired plant. When nuclear weapons were tested in the atmosphere, the radioactive products from the nuclear explosions were released into the air and fell to Earth as radioactive fallout.

Another concern about nuclear power plants is their decommissioning. Nuclear reactors have useful lifetimes of about thirty years, after which they need to be shut down permanently. The remaining fuel, which is very radioactive, will have to be removed and stored for many decades. The entire reactor building structure has also been made somewhat radioactive, as well as all the pipes, valves, etc., in the plant. There are a number of options for dealing with this radioactive material, ranging from immediate dismantling of the plant (with some of the work probably done by robots) to using the normal reactor containment structure as a long-term storage facility. In the United States, the Nuclear Regulatory Commission has required that plant owners set aside sufficient funds for dismantling.

The greatest concern that most members of the general public have about nuclear energy is the possibility of a catastrophic accident such as occurred in 1986 at Chernobyl in the Ukraine. If reactors have been properly designed and the staff trained with safety in mind, then the chance of a major accident is very slight. The Chernobyl disaster was caused by a combination of poor reactor design and insufficient training of the reactor operators, who violated many of the operating procedures related to safety.

Nuclear *fusion* is the energy-producing process that occurs in the sun and other stars. Small nuclei such as those in hydrogen and helium fuse together to produce larger nuclei, releasing energy. Nuclear fusion is not yet commercially viable as an energy source, since extremely high temperatures are required to initiate fusion, and containment of the fusing nuclei at these temperatures is difficult. However, if nuclear fusion ever becomes a usable energy source, a typical fusion power reactor would require less than a ton of fuel annually.

It is often stated that nuclear fusion will produce no radioactive hazard, but this is not correct. The most likely fuels for a fusion reactor would be deuterium and radioactive tritium, which are isotopes of hydrogen. Tritium is a gas, and in the event of a leak it could easily be released into the surrounding environment. The fusion of deuterium and tritium produces neutrons, which would also make the reactor building itself somewhat radioactive. However, the radioactivity produced in a fusion reactor would be much shorter-lived than that from a fission reactor. Although the thermonuclear weapons (that use nuclear fusion), first developed in the 1950s provided the impetus for tremendous worldwide research into nuclear fusion, the science and technology required to control a fusion reaction and develop a commercial fusion reactor are probably still decades away.

Ernie McFarland

BIBLIOGRAPHY

Ahearne, J. F. (1993). "The Future of Nuclear Power." *American Scientist* 81:24–35.

Atomic Energy of Canada Limited (AECL). (1999). *Nuclear Sector Focus: A Summary of Energy, Electricity and Nuclear Data.* Mississauga, Ontario, Canada: AECL

Atomic Energy of Canada Limited (AECL). (2000). <http://www.aecl.ca/english/energy/energy_f.html>.

Cordey, J. G.; Goldston, R. J.; and Parker, R. R. (1992). "Progress Toward a Tokamak Fusion Reactor." *Physics Today* 45(1):22–30.

McFarland, E. L.; Hunt, J. L.; and Campbell, J. L. (1997). *Energy, Physics, and the Environment*, 2nd ed. University of Guelph, Ontario: Department of Physics.

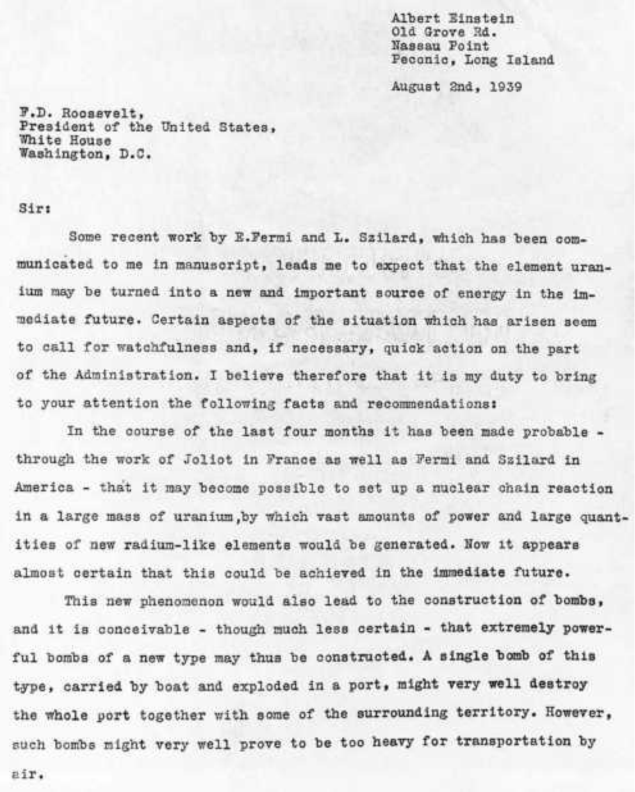
NUCLEAR ENERGY, HISTORICAL EVOLUTION OF THE USE OF

The history of nuclear energy is a story of technical prowess, global politics, unfulfilled visions, and cultural anxiety. The technology's evolution in the second half of the twentieth century progressed through several stages: theoretical development by physicists; military application as atomic weapons in World War II; commercialization by the electrical industry in several industrialized nations; proliferation (for military and non military uses) among less developed nations; crises spawned by power plant accidents, cost overruns, and public protests; and retrenchment and slowdown in the last few decades of the twentieth century. By far the most potent form of energy to be harnessed by humankind, nuclear power has not become the dominant form of energy because of the great economic costs and social risks associated with its use.

MILITARY ORIGINS

The concept of "atoms" dates back to the ancient Greeks, who speculated that the material world was comprised of tiny elemental particles, and for centuries thereafter alchemists attempted to unlock the secrets of the elements. But modern atomic science did not emerge until the turn of the twentieth century. In 1896 Henri Becquerel of France discovered radioactivity, and Albert Einstein calculated the mass-energy relationship ($E = mc^2$) in 1905. By the 1930s, scientists in several countries were making progress toward understanding nuclear reactions, including Ernest Rutherford and James Chadwick in Great Britain; Enrico Fermi in Italy; Niels Bohr in Denmark; and Ernest O. Lawrence in the United States. The key breakthrough came in December 1938, when German physicists Otto Hahn and Fritz Strassmann achieved the first controlled atomic fission, splitting atoms of uranium into lighter elements by bombarding them with neutrons, and releasing enormous amounts of energy in the process.

The news spread quickly, and became charged with implications as Hitler's Nazis began their march through Europe. Two days before the outbreak of World War II in Europe in 1939, Bohr and John Wheeler of Princeton published an academic paper on fission. Several leading physists fled Germany and Stalin's Soviet Union for the United States, including Hungarian refugee Leo Szilard. Fearful that the Nazis might build a powerful atomic bomb, Szilard and fellow Hungarian émigré Eugene Wigner convinced Einstein (then at Princeton University) to write President Roosevelt to warn of the possibilities of atomic weaponry and to suggest U.S. action.



Albert Einstein
Old Grove Rd.
Nassau Point
Peconic, Long Island

August 2nd, 1939

F.D. Roosevelt,
President of the United States,
White House
Washington, D.C.

Sir:

Some recent work by E.Fermi and L. Szilard, which has been communicated to me in manuscript, leads me to expect that the element uranium may be turned into a new and important source of energy in the immediate future. Certain aspects of the situation which has arisen seem to call for watchfulness and, if necessary, quick action on the part of the Administration. I believe therefore that it is my duty to bring to your attention the following facts and recommendations:

In the course of the last four months it has been made probable - through the work of Joliot in France as well as Fermi and Szilard in America - that it may become possible to set up a nuclear chain reaction in a large mass of uranium,by which vast amounts of power and large quantities of new radium-like elements would be generated. Now it appears almost certain that this could be achieved in the immediate future.

This new phenomenon would also lead to the construction of bombs, and it is conceivable - though much less certain - that extremely powerful bombs of a new type may thus be constructed. A single bomb of this type, carried by boat and exploded in a port, might very well destroy the whole port together with some of the surrounding territory. However, such bombs might very well prove to be too heavy for transportation by air.

The first page of a letter dated August 2, 1939, from Albert Einstein to President Franklin Delano Roosevelt. In the letter Einstein advises Roosevelt of the possibilities of nuclear research. (Corbis-Bettmann)

Einstein's letter (dated August 2, 1939) had little impact, however. What stirred the U.S. government into action were reports out of Great Britain in the early 1940s: one from German refugee physists Rudolf Peierls and Otto Frisch in 1940, which discussed the possibility of making a "super-bomb" from a uranium fission chain reaction; and a study from the top-secret British "MAUD committee" in 1941, which deemed a uranium bomb "practicable and likely to lead to decisive results in the war," and which urged the United States to make development of an atomic bomb its "highest priority."

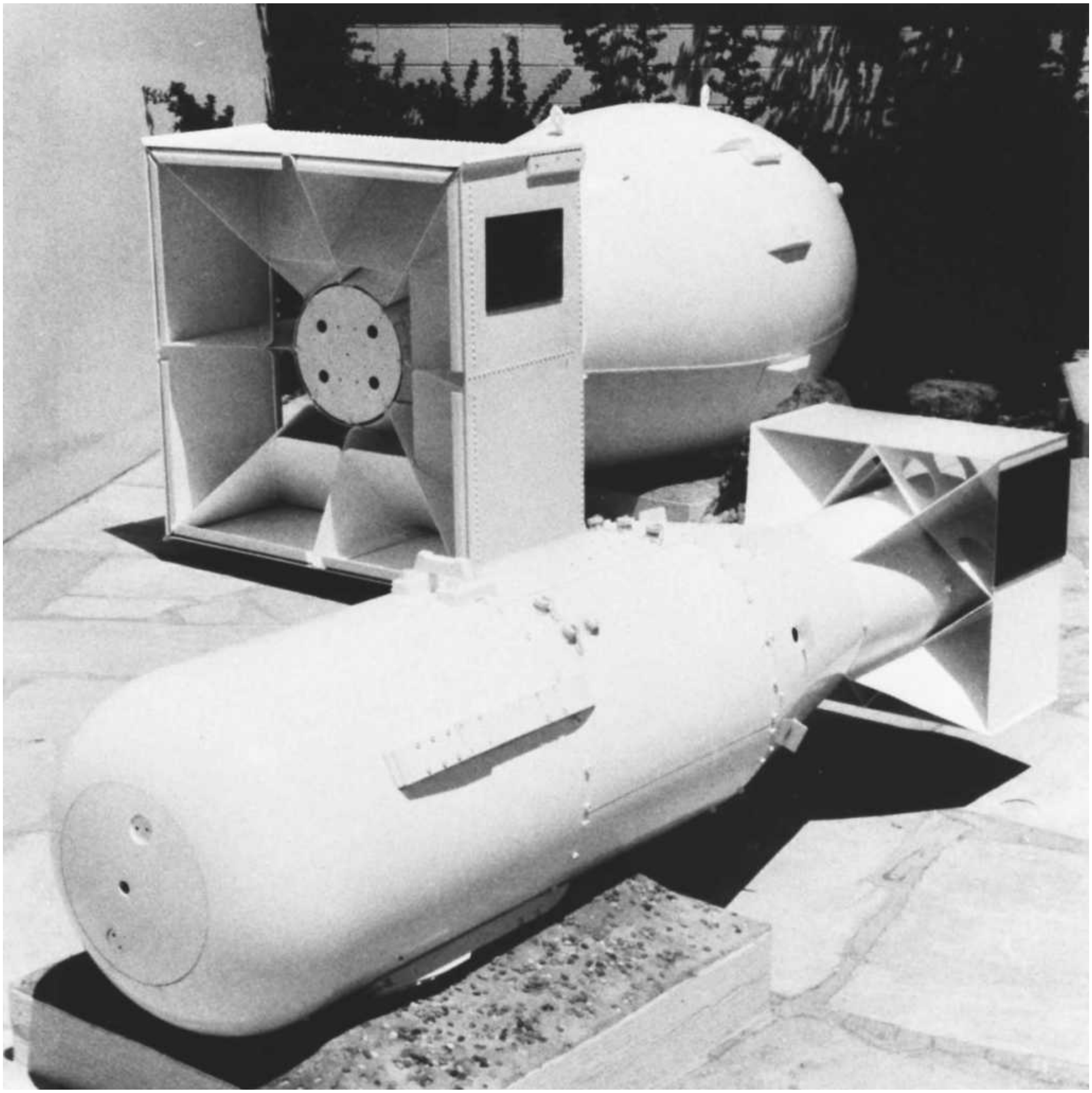
At the time, the scarcity of fissionable material—whether natural uranium or man-made plutonium—seemed the greatest barrier to atomic bomb production. Now determined to win the atomic bomb race, President Franklin Roosevelt approached Colonel (soon-to-be Brigidier General) Leslie Groves of the U.S. Army Corps of Engineers to head what was code-named the "Manhattan Engineer District" (later popularly known as the Manhattan Project), America's atomic bomb project. It was a massive, sprawling effort that ultimately encompassed several leading university laboratories, three giant manufacturing sites, tens of thousands of constructions workers, the world's best scientific talent, several major corporations, and some $2.2 billion of federal funds.

The project's two key scientific advisers were Vannevar Bush, an electrical engineer at head of the Office of Scientific Research and Development, and a former dean of the Massachusetts Institute of Technology; and James Conant, a chemist, chair of the National Defense Research Committee, and president of Harvard University. Groves also recruited Nobel laureate Arthur Compton to the project, who in turn recruited Wigner and Szilard as well as Nobel laureates Fermi and James Franck to his burgeoning research laboratory at the University of Chicago (code-named the "Metallurgical Laboratory"). Meanwhile, the army brought in Boston-based Stone & Webster as principal engineering contractor, and the giant Du Pont chemical firm, which had no experience in plutonium production but took on the work for costs plus $1. After Fermi—working in a racquetball court under the stands at the University of Chicago's Stagg Field—achieved the first self-sustaining nuclear reaction in December 1942, the Manhattan Project settled on water as a coolant and forged ahead with construction plans.

At Hanford, Washington, along the Columbia River, the project built three production piles (reactors) and four separation plants for separating plutonium from other elements. At Oak Ridge, Tennessee (near Knoxville), it built uranium-235 facilities utilizing three processes—thermal diffusion, gaseous diffusion, and electromagnetic separation; the latter, under the direction of Lawrence, proved to be the most successful. And at Los Alamos, New Mexico (near Santa Fe), the Manhattan Project in 1945 began building a facility for making both plutonium and uranium bombs under the direction of the brilliant Harvard- and Gottingen-educated physicist J. Robert Oppenheimer. Like the Chicago-based scientists before him, Oppenheimer and his researchers often clashed with Groves and the project engineers, who preferred to compartmentalize and control information about the project rather than exchange it freely among the scientists. At Los Alamos, Oppenheimer's approach prevailed.

Confident that the uranium bomb would detonate with a gun-type neutron device but less sure of the plutonium bomb's "implosion" detonation device (invented by physicist Seth Neddermeyer), the scientists tested a plutonium bomb—dubbed "Fat Man" in honor of Winston Churchill—at Alamogordo, New Mexico, on July 16, 1945. The awesome power of "Project Trinity" immediately inspired doubts and fears about a nuclear future, even among several Manhattan Project scientists. Szilard, Franck, and others urged the U.S. military to demonstrate the atomic bomb to the Japanese in an uninhabited area; but the momentum of the project, and political pressure for a rapid end to the war, were too great. On August 6, the U.S. B-29 bomber *Enola Gay* dropped a uranium atomic bomb ("Little Boy") that detonated over Hiroshima, Japan, at 8:15 A.M., killing an estimated 75,000 to 100,000 people instantly, and another 200,000 from radiation over the next five years. Three days later, the United States destroyed Nagasaki with a plutonium bomb, and the Japanese surrendered shortly thereafter. By building the largest government-business-university collaboration in its history, the United States had harnessed atomic energy and brought World War II to a rapid end.

Over the next several years, the atomic bomb helped usher in a new kind of geo political conflict: the Cold War. Rather than clashing with conventional weapons, the two postwar superpowers—the democratic United States and the Communist Soviet Union—relied increasingly on their growing arsenals of nuclear weapons to protect their spheres of influence and to deter encroachment with the threat of mutual destruction. Tensions in the early Cold War were heightened when the United States refused to share atomic bomb technology with its erstwhile ally the Soviet Union. When the USSR successfully tested its own atomic bomb in 1949, the "arms race" between the superpowers to achieve nuclear superiority was on. In October 1952 the United States tested its first hydrogen bomb, at Eniwetok Atoll in the Pacific. By fusing light forms of hydrogen to create helium—the reaction responsible for the Sun's energy—the "H-



Two atomic bombs named "Fat Man" and "Little Boy." The latter was used to destroy Hiroshima, Japan, during World War II. (Library of Congress)

bomb" exploded with 1,000 times the power (the equivalent of some 10 million tons of TNT) of its fission predecessor. The USSR had the H-bomb by 1953. The U.S. then began developing new, solid-fuel intercontinental ballistic missles (ICBMs) to deliver bomb warheads from the U.S. western plains deep into the Soviet Union.

The post–World War II generation was the first to live under the shadow of total and instantaneous annihilation, a reality that—coming on the heels of the holocaust and the massive military horrors of World War II—gave rise to a new nihilism among some philosophers and social thinkers. Psychologists began to probe the possible mental health consequences of life in the atomic age, although many concluded that nuclear apocalyse was too vast and horrible to comprehend. And, not surprisingly, the atomic bomb began to find its way into everyday life and popular culture, sometimes in bizarre and even lighthearted ways.

Communities began to refashion their civil defense procedures to accommodate the bomb. In the United States, this meant designating bomb shelters—usually in the basements of schools and other public buildings, but also involving construction of individual shelters beneath the yards of single-family homes in the nation's burgeoning postwar suburbs. Stocked with canned foods and other provisions, these underground chambers typically were small and spartan, although some realtors seized the opportunity to offer luxury models with modern conveniences. While never ubiquitous, individual bomb shelters nevertheless became normalized, as reflected in a 1959 *Life* magazine story about a newlywed couple who spent two weeks of "unbroken togetherness" honeymooning in their bomb shelter.

The atomic bomb also became a key subject in 1950s American film, particularly in the science fiction genre. While some films spun out scenarios about how the nuclear powers might accidentally bring on Armageddon—as in *Fail-Safe* and Stanley Kubrick's classic tragicomedy *Dr. Strangelove*—most focused on the insidious and little-understood effects of radiation on human and animal life. However implausible the premises of *The H-Man, Attack of the Crab Monsters, The Incredible Shrinking Man, Attack of the Fifty-Foot Woman,* and their ilk, such fantasies resonated with widespread anxieties about genetic damage from atomic fallout.

Although government officials attempted to educate the public and military personnel about atomic civil defense, in retrospect these efforts seem hopelessly naive if not intentionally misleading. Army training films advised soldiers to keep their mouths closed while observing atomic test blasts in order to not inhale radioactive flying dirt. Civil defense films used a friendly animated turtle to teach schoolchildren to "duck and cover" during a nuclear attack—that is, duck under their desks and cover their heads. Such measures, of course, would have offered pitiful protection to those in the blast zone.

Opinion polls showed that American anxiety about the atomic bomb ebbed and flowed in response to geopolitical events. Concerns ran high in the late 1940s in the wake of the atomic bombings of Japan (many wondered whether the weapon would be used in all military conflicts), and as the Soviet Union undertook a crash program in rocketry and atomic-bomb development. These fears cooled temporarily in the early 1950s, particularly after the death of Soviet dictator Joseph Stalin in 1953 raised hopes for U.S.-Soviet rapprochement. But ICBM development and a wave of H-bomb testing in the Pacific in 1954 stirred up renewed public fears about fallout, especially after milk in heartland cities such as Chicago was found to contain elevated levels of isotopes. The national heart rate spiked during the tense days of the Cuban Missile Crisis in 1962, then slowed after the United States and the Soviet Union signed the Partial Test-Ban Treaty in 1963. Whereas 64 percent of Americans identified the threat of nuclear war as their leading concern in 1959, only 16 percent put the same concern first in 1964.

"ATOMS FOR PEACE": THE ORIGINS OF THE NUCLEAR POWER INDUSTRY

The Atomic Energy Act of 1946 represented the interests of American scientists who wished to see nuclear energy developed for nonmilitary purposes. It called for the establishment of a five-member civilian Atomic Energy Commission (AEC), which could deliver weapons to the military only on presidential order. But the military tensions of the early Cold War delayed civilian nuclear power development until 1948, at which time 80 percent of the AEC's budget went to military ends. In 1951, U.S. civilian nuclear power development consisted of only a small experimental government (liquid metal) reactor in Idaho.

Through the efforts of Captain Hyman G. Rickover, a naval engineering officer, the U.S. Navy made rapid strides in nuclear ship and submarine development. Garnering support from Edward Teller and other key figures outside the navy, Rickover brought in Westinghouse, General Electric, and the Electric Boat Company to construct the *Nautilus,* the world's first nuclear submarine. First tested in 1955, the *Nautilus* ran faster and ten times farther without surfacing than conventional submarines. Nuclear-powered aircraft carriers and other surface ships followed, and by the 1960s U.S. nuclear submarines were equipped with solid-fuel nuclear missiles. In contrast, the U.S. merchant marine did not emphasize use of nuclear power. Its demonstration vessel, the *Savannah,* operated by the U.S. Marine

Corporation beginning in 1959, used a pressurized-water reactor instead of a boiling-water reactor, and required a heavy government operating subsidy.

In 1953 the AEC began planning a full-scale nuclear plant at Shippingport, Pennsylvania. The plant was to be owned and operated by Duquesne Light Company, managed by Rickover, and equipped with a version of the navy's pressurized-water reactor. But rapid U.S. nuclear power development came only in the wake of the new Atomic Energy Act of 1954, which permitted private power companies to own reactors and to patent nuclear innovations; and the Price-Anderson Act of 1957, which limited liability for individual companies to $560 million and gave government subsidies for liability insurance. The first American plants went on line in the early 1960s.

The USSR operated the first nuclear power plant supplying a national grid at Obninsk, south of Moscow in 1954. This modest (5,000 kW) plant was the opening wedge in an aggressive Soviet drive for nuclear energy as reflected in the Bolshevik slogan "Communism equals Soviet power plus electrification of the entire country." Two years later, the United Kingdom operated a full-scale commercial nuclear plant at Calder Hall, a facility designed to produce plutonium for defense with electricity as a by-product. France ran an experimental reactor, at Marcoule, that year, but began its commercial program with the 'Electricite' de France Chinon reactor in 1962. By 1965 the United Kingdom led in nuclear power output with 3.4 million kWh, followed by the United States (1.2 million kWh), the USSR (895,700 kWh); Italy (622,000 kWh); and five other nations producing at least 500,000 kwh each. Within a few years the United States took the lead in technology and total output, its pressurized-water and boiling-water reactors supplied by enriched fuel from Manhattan Project plants.

These were heady times for the nuclear industry. Well-informed experts predicted that electricity soon would become "too cheap to meter." At the 1964–1965 New York World's Fair, General Electric's "Progressland" exhibit featured "the wonders of atomic energy"; and the company's Medallion City claimed to produce fusion energy (using a 0.000006 second, 100 million flash). Fusion energy, said General Eelectric, was going to supply a billion years of electrical energy.

Nuclear power developed unevenly across the globe. In 1987 the United States operated 110 of the world's 418 nuclear plants, the USSR 57, France 49, the United Kingdom 38, Japan 37, Canada 19, and Sweden 12, while some regions—the Arab states, Africa (except South Africa), and most of Central America—had few or none. Although the United States produces a third of the world's electricity, it derives only 15 percent of it from nuclear energy.

In each country, the pattern of nuclear development reflected national technical and economic resources, politics, and culture. In France, for example, national economic planners ensured the rapid growth of nuclear energy by limiting citizen participation and by standardizing reactor design. The Federal Republic of Germany followed a style closer to the United States, and thus sustained greater political challenges by its environmentally sensitive Green Party.

In balancing development with social safety, the Soviet Union was a grim outlier. Recent research has revealed a series of Soviet nuclear accidents, many of them concealed from the outside world, beginning with a dramatic waste dump explosion in 1957 near Kyshtym in the Urals that spread more than 2 million Ci over 20,000 sq miles. The explosion of Unit 4 at Chernobyl, north of Kiev, on April 26, 1986, was the worst recorded nuclear accident in history. It happened when operators were testing a voltage-regulating scheme on turbogenerator 8 when coolant pumps were slowed. Vigorous boiling led to excess steam, slow absorption of neutrons, and increased heat in a "positive feedback" cycle that resulted in meltdown and explosion. By early May, airborne contamination completely covered Europe. Thirty-one people were killed, 200 suffered radiation sickness, hundreds of thousands of people were confined indoors or evacuated, foods that were suspected of being contaminated were banned, and scientists projected some 1,000 extra cancer deaths over the next fifty years. Perhaps most troubling, follow-up investigation faulted Soviet plant design as much as or more than operator error. Following the accident, public opinion shifted sharply against nuclear energy in Europe, but in the United States, the tide already had turned.

NUCLEAR POWER UNDER SIEGE

In the 1950s, widespread anxieties about the atomic bomb in the 1950s spurred few into action. One notable exception in the United States was the Greater St. Louis Committee for Nuclear Information, founded by biologist Barry Commoner and other scientists at Washington University in 1958. This watchdog group published *Nuclear Information* (renamed *Science and the Citizen*) in 1964. For the most part, however, citizens of the two superpowers and their satellites or allies saw nuclear weapons as a necessary tool in the global contest between democracy and communism. Indeed, for many, atomic supremacy was the most accurate measure of national prowess.

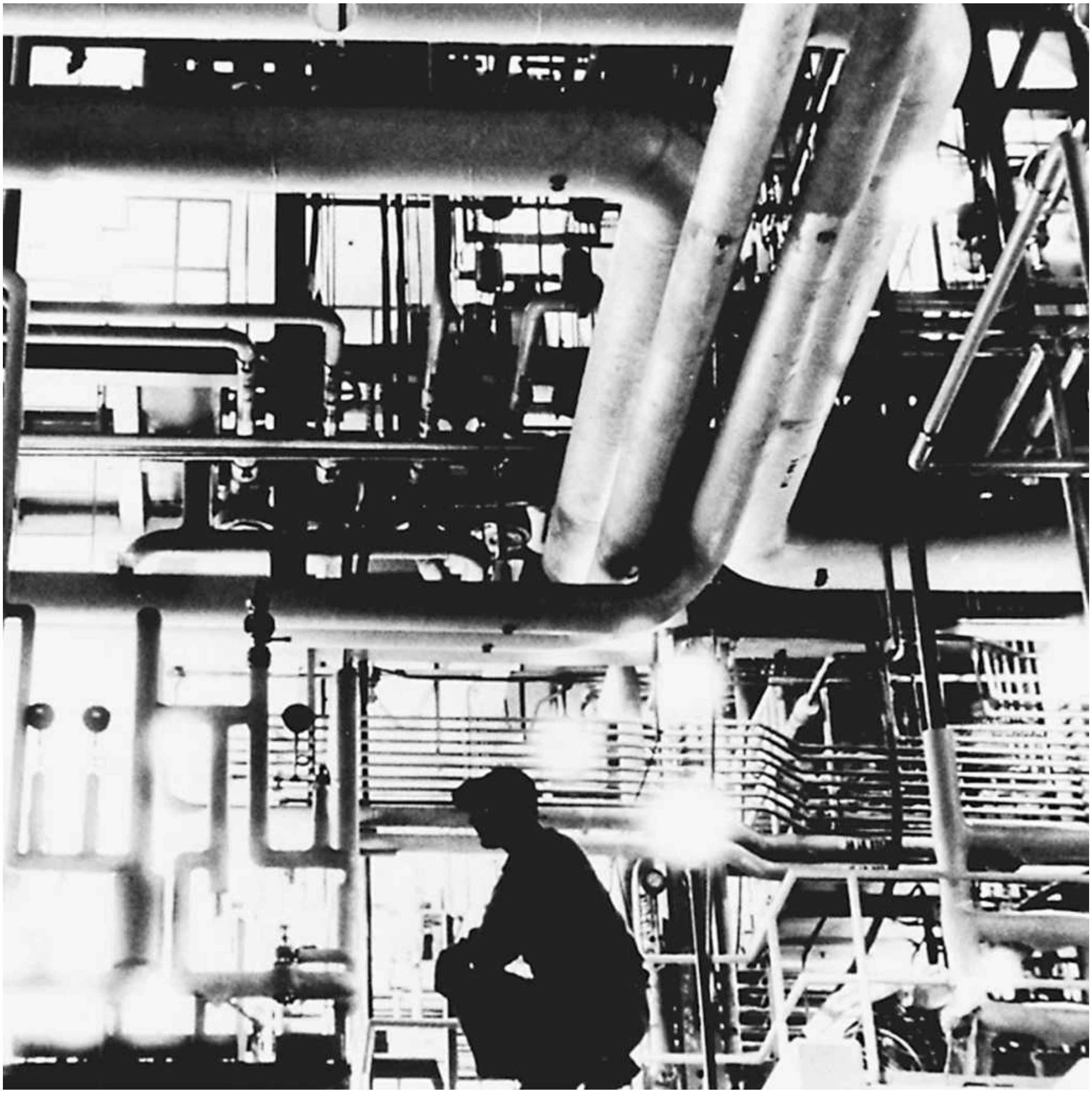
This changed dramatically in the 1960s and 1970s as antinuclear opinion switched from weapons to nuclear power and spread from activist groups into a large segment of the middle class. Antinuclear activism was strongest in the two nations with large nuclear power programs and the most open political systems—the United States and West Germany—although there were notable institutional and social differences between the two nations. And while antinuclear activism surely slowed the progress of nuclear power development in many countries, it was one of a cluster of economic, technological, political, and social forces constraining the industry.

Nuclear energy opponents initially targeted thermal pollution—the effect of nuclear power plants on local aquatic life by raising water temperature by several degrees. Attention soon shifted to the question of radioactive contamination of cooling water. When two scientists at Lawrence Livermore Laboratory in Berkeley, California—John Gofman and Arthur Tamplin—argued that the AEC acceptable level of 170 millirems per year would cause some 16,000 additional deaths (they later revised the figure to 74,000), an intense debate ensued that eventually led to a U.S. Senate committee and to a new standard of 25 millirems. Meanwhile, news of nuclear plant accidents in foreign countries reached the United States, which endured an accident of its own at Idaho Falls in 1961. New guidelines for remote nuclear plant citing followed, which resulted in scuttled plans for a Consolidated Edison plant in Queens, New York, and another in Los Angeles. Anti-nuclear activism was strongest in southern California, where protesters managed to stop the construction of a Pacific Gas & Electric plant at Bodega Bay in 1963 because it was near an earthquake fault.

The 1970s were hard times for the nuclear industry. The decade opened with the first Earth Day (April 22), which featured thousands of teaching events, many of them aimed at halting further nuclear power development, and ended with the accident at the Three-Mile Island nuclear plant in Pennsylvania. In between, the nuclear industry sustained increasingly sophisticated attacks from increasingly better organized opponents, such as the Environmental Defense Fund (1967); Ralph Nader's Critical Mass Project in Washington, D.C.; and the Union of Concerned Scientists, a consortium of nuclear scientists and engineers founded at MIT in 1971. In 1976, California passed a ballot initiative that halted nuclear plant construction until the federal government found a satisfactory way to dispose of radioactive wastes; while in New Hampshire the Clamshell Alliance was formed to oppose the construction of a nuclear plant at Seabrook, an effort that dragged on for years, involved tens of thousands of demonstrators, and helped force the plant's owner (Public Service Company of New Hampshire) to cancel plans for one of its two reactors and to declare bankruptcy in 1988—the first public utility to do so in American history.

In that case, protests caused delays that contributed to large cost overruns. But Seabrook was an exception; most nuclear utilities got into financial trouble with little help from protesters. Although oil prices rose dramatically in the 1970s—a spur to nuclear development—the "stagflation" of the times drove down demand for electricity from 7 percent to 2 percent per annum and drove up interest rates into the double digits. Between 1971 and 1978, nuclear capital costs rose 142 percent, making them more expensive to build per kilowatt-hour of capacity than new fossil fuel plants.

No case better illustrates the travails suffered by the nuclear power industry and its customers than the saga of the Washington Public Power Supply System (WPPSS). Thanks to abundant Columbia River hydropower, the Pacific Northwest enjoyed the nation's lowest electricity rates in the 1950s. By then the prime hydropower sites has been exploited, and demand for electricity continued to rise. In 1968 some one hundred utilities in the region, working through WPPSS, financed a ten-year, $7 billion Bonneville Power Administration plan to improve the region's hydropower and transmission and distribution assets as well as to build seven new thermal power plants, most of them nuclear. Construction costs skyrocketed, however; the second power plant (at Hanford, Washington), projected to cost $352 million in 1977, had consumed nearly $2 billion by 1980. Struggling to avoid bankruptcy, Bonneville boosted its wholesale rates 700 percent between 1979 and 1984. Irate ratepayers dubbed the project "WHOOPS."



Atomic electricity generating station in Shippingport, Louisiana. (Library of Congress)

Viewed in this context, the Three-Mile Island (TMI) accident was the *coup de grace* for an already foundering industry. In spite of the fact that the hydrogen gas bubble that accumulated in Reactor 2 did not explode, although some contaminated gas escaped; and that the commissions who investigated the accident faulted human error rather than equipment failure, TMI caused (as the *New York Times* put it) "a credibility meltdown" for the nuclear industry. (The release of a major motion picture with an eerily similar scenario, *The China Syndrome,* a few weeks before the real accident amplified the public relations crisis.) The TMI cleanup cost roughly $1 billion, less than a third covered by insurance. The Nuclear Regulatory Commission, a federal agency, began to enforce an informal moratorium on new plant licenses pending further investigation of TMI, and soon demanded expensive retrofitting for similar plants. But even without TMI, no new reactors have been ordered in the United States since 1978.

NUCLEAR ENERGY AND NUCLEAR DEFENSE FROM THE LATE 1970S TO THE PRESENT

As they did at the dawn of the atomic age, nuclear weapons have overshadowed nuclear energy for since the late 1970s. The shift began then, when the Carter administration switched its Soviet stance from détente to rearmament and ordered 200 new MX missiles. By then, the nonprofit organizations Union of Concerned Scientists—perhaps seeing civilian nuclear energy as moribund—shifted its attention to weapons. The Cambridge, Massachusetts–based Physicians for Social Responsibility, led by Helen Caldicott, began to publicize the horrors of nuclear war, garnered a large grant from the Rockefeller Foundation, and saw its membership surge from 500 in 1978 to 16,000 in 1980. At the same time, a movement favoring a moratorium on nuclear weapons began to take shape, with 1980 presidential hopefuls Governor Jerry Brown of California and Senator Edward Kennedy of Massachusetts joining the cause.

The election of Ronald Reagan to the White House in 1980 heightened nuclear tensions. Reagan dubbed the Soviet Union an "evil empire" and spoke of first-strike capabilities and strategies of limited nuclear war. Supporters of the 1981 Nuclear Weapons Freeze Campaign, in contrast, called for a verifiable treaty with the Soviet Union to halt nuclear proliferation. In 1983 Regan announced the extravagant Strategic Defense Initiative to shield the United States from nuclear attack using space-based missile interceptors and other yet-to-be-developed technologies. Endorsed by atomic bomb pioneer Edward Teller, the project was so expensive and technically speculative that it became popularly known as "Star Wars," after a Hollywood space fantasy. Although the president called it a purely defensive measure, many foreign-policy experts saw it as a form of destabilizing escalation. Others considered it a ploy to force the USSR into heavy military spending (to achieve parity with Star Wars) at a time when the Soviet empire was undergoing political and economic stress under the reformist regime of Mikhail Gorbachev.

The question of how much Star Wars may have contributed to the 1991 collapse of the Soviet Union continues to generate debate. But the breakup of the postwar nuclear superpower clearly ushered in a new era of nuclear quiescence. To be sure, the fate of the former Soviet Union's tens of thousands of warheads remains a vital concern, especially in light of the region's growing economic pressures. For example will unpaid military officers be tempted to sell nuclear weapons to well-funded terrorist groups? In the post-Soviet world, the threat of nuclear warfare no longer is seen as a matter of global annihilation bur rather as a risk in political hot-spots such as the Middle East, Korea, or—if Iraq's Saddam Hussein should acquire nuclear technology—the United States as a possible target. And dreams of a Star Wars-like defense live on; late in his final term, President Bill Clinton announced plans to consider the development of a nuclear defense system reminiscent of, though less elaborate than, the Reagan plan.

Although reactor-building continues steadily in France and Japan, and may take hold in parts of the less developed world, nuclear power will need a major technological breakthrough (or a fossil fuel energy crisis) to make a comeback in the United States and Germany. One possibility is the fusion reactor, which produces nuclear energy by combining two light elements (such as deuterium) into a single heavier one that weighs less than the sum of its parts, with the difference released in the form of energy. Fusion produces less radioactive waste than fission and faces no fuel constraints (deuterium is found in water). The problem is that scientists have not been able to combine light atoms through collusion—they bounce off each other—but rather do so by accelerating them with heat so intense it either consumes as much energy as the reaction produces, or (at much higher levels) it cannot be contained safely.

Since the earliest days of the atomic age, physicists and engineers have predicted the coming of practicable nuclear fusion within "ten years" or "a generation." History therefore offers many reasons to be skeptical about the promise of nuclear energy. At the same time, this unparalleled form of energy is not going to return to the Pandora's box pried open by the Manhattan Project more than a half century ago.

David B. Sicilia

See also: Einstein, Albert; Emission Control, Power Plant; Energy Management Control Systems; Environmental Problems and Energy Use; Ethical and Moral Aspects of Energy Use; Explosives and Propellants; Historical Perspectives and Social Consequences; Matter and Energy; Military Energy Use, Historical Aspects of; Nuclear Energy; Nuclear Fission; Nuclear Fission Fuel; Nuclear Fusion; Nuclear Waste.

BIBLIOGRAPHY

Boyer, P. S. (1985). *By the Bomb's Early Light.* New York: Pantheon.

Campbell, J. L. (1988). *Collapse of an Industry.* Ithaca NY: Cornell University Press.

Cantelon, P. L.; Hewlett, R. G.; and Williams, R. C., eds. (1991). *The American Atom.* Philadelphia: University of Pennsylvania Press.

Hughes, T. P. (1989). *American Genesis.* New York: Viking.

Joppke, C. (1993). *Mobilizing Against Nuclear Energy.* Berkeley: University of California Press.

Melosi, M. V. (1985). *Coping with Abundance.* Philadelphia: Temple University Press.

Mounfield, P. R. (1991). *World Nuclear Power.* London: Routledge.

Rhodes, R. (1986). *The Making of the Atomic Bomb.* New York: Simon & Schuster.

NUCLEAR FISSION

Nuclear fission is a process in which a heavy nucleus—usually one with a nucleon number of two hundred or more—separates into two nuclei. Usually the division liberates neutrons and electromagnetic radiation and releases a substantial amount of energy. The discovery of nuclear fission is credited to Otto Hahn and Fritz Strassman. In the process of bombarding uranium with neutrons in the late 1930s, they detected several nuclear products of significantly smaller mass than uranium, one of which was identified as ^{137}Ba. The theorectical underpinnings that exist to this day for nuclear fission were proposed by Lise Meitner and Otto Frisch. Shortly after Hahn and Strassman's discovery.

Some heavy nuclei will fission spontaneously. Others can be induced to fission through interaction with a neutron. In both spontaneous nuclear fission and induced nuclear fission the pool of neutrons and protons is conserved. For example, the nucleus ^{252}Cf (Californium) fissions spontaneously. The 98 protons and 154 neutrons in the nucleus of ^{252}Cf are reconfigured into other nuclei. Usually a few neutrons are released in the process. Pictorially, a typical spontaneous fission of ^{252}Cf producing two nuclei and three neutrons is shown in Figure 1.

The pictorial depiction of the spontaneous fission of ^{252}Cf can be summarized as an equation:

$$^{252}\mathrm{Cf} \rightarrow {}^{150}\mathrm{Ba} + {}^{99}\mathrm{Mo} + 3\mathrm{n} + \text{energy}$$

The nucleus ^{235}U (uranium) does not fission spontaneously, but it can be induced to fission through interaction with a neutron. Pictorially, a typical neutron-induced fission of ^{235}U producing two nuclei and three neutrons is depicted in Figure 2.

The pictorial depiction of the neutron-induced fission of ^{235}U can be summarized as an equation:

$$^{235}\mathrm{U} + \mathrm{n} \rightarrow {}^{144}\mathrm{Ba} + {}^{89}\mathrm{Kr} + 3\mathrm{n} + \text{energy}$$

Visually, nuclear fission is simply a rearrangement of neutrons and protons, and would seem to be possible for any nucleus. However, energy conservation principles limit significantly the number of possibilities for either spontaneous nuclear fission or induced nuclear fission. Total energy is conserved in a nuclear fission reaction as it is in all nuclear reactions. This means that the total energy before the fission equals the total energy of the fission products. For example, in the induced nuclear fission reaction discussed earlier, the combined energy of the neutron and ^{235}U nucleus equals the combined energy of the ^{144}Ba and ^{85}Kr nuclei and the three neutrons. The energy accounting includes kinetic energy as well as any potential energy associated with the nuclei and the neutrons. The energy associated with mass through the Einstein relation $E = mc^2$ is particularly relevant. The interaction of the neutron and the ^{235}U nucleus takes place when both are essentially at rest and, therefore, have negligible kinetic energy. Yet the reaction products—^{144}Ba, ^{85}Kr, and three neutrons—have significant kinetic energy. The kinetic energy of the products has its origin in the conversion of potential energy associated with the ^{235}U nucleus and the interacting neutron. The gain in kinetic energy is accompanied by a loss of mass in the interacting nuclei (neutron and ^{235}U nucleus). The kinetic energy of the reaction products (^{144}Ba, ^{85}Kr and 3 neutrons) is determined from $E = mc^2$, where m is the mass difference between the interacting nuclei and reaction products. The following example illustrates this principle. In atomic mass units the masses of the

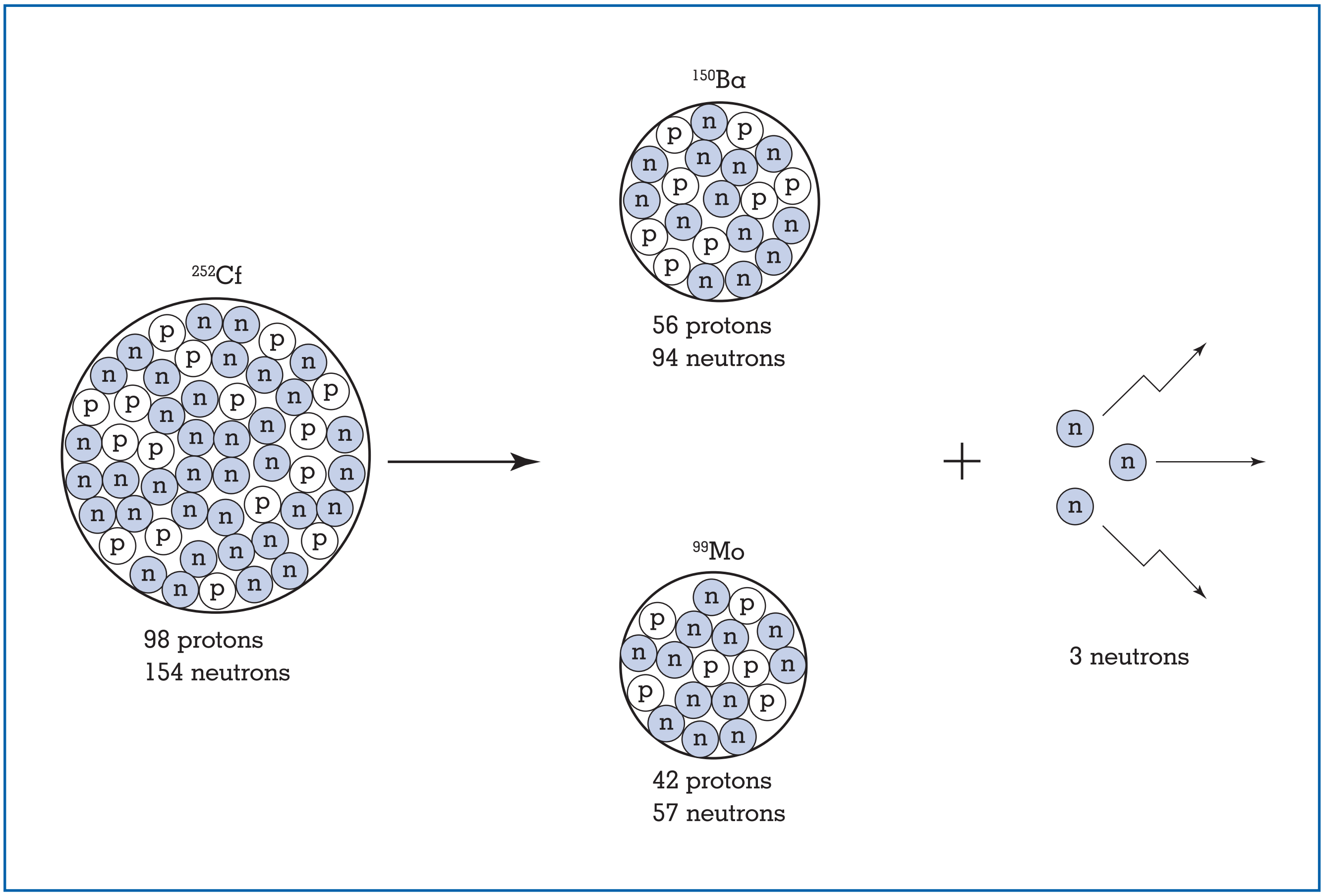


Figure 1.
Spontaneous fission process.

nuclei involved in the illustrative nuclear fission reaction are:

^{235}U	235.0439231 u	^{144}Ba	143.9229405 u
n	1.0086649 u	^{89}Kr	88.9176325 u
		3 n	3.0259947 u

Adding the left and the right columns yields 236.052588 u and 235.8665677 u, respectively. The reaction products have 0.1860 u less mass than the interacting nuclei. One atomic mass unit has an energy equivalent of 931.50 million electron volts (MeV). Therefore, 0.1860 u has an energy equivalent of 173.3 MeV. The energy acquired by the reaction products totals 173.3 MeV. This nuclear fission energy release is huge compared to that in a chemical reaction.

A typical chemical reaction such as that in burning a fossil fuel releases about 10 eV of energy. A release of 170 MeV in one nuclear fission reaction liberates the energy equivalent of 17 million chemical reactions in the burning of a fossil fuel.

A nuclear fission reaction will not occur unless the following occur: (1) the total mass of the reaction products is less than the total mass of the interacting nuclei, and (2) the sum of the neutrons and the sum of the protons in the interacting particles equals the sum of the neutrons and the sum of the protons in the products of the fission.

Any combination of reaction products consistent with these conservation principles is possible. For example, in the neutron-induced nuclear fission of ^{235}U it is possible to produce ^{140}Xe, ^{94}Sr, two neutrons, and 185 MeV of energy. The most likely reaction products are close in atomic number to xenon (Xe) and strontium (Sr), but the possibilities number in the hundreds.

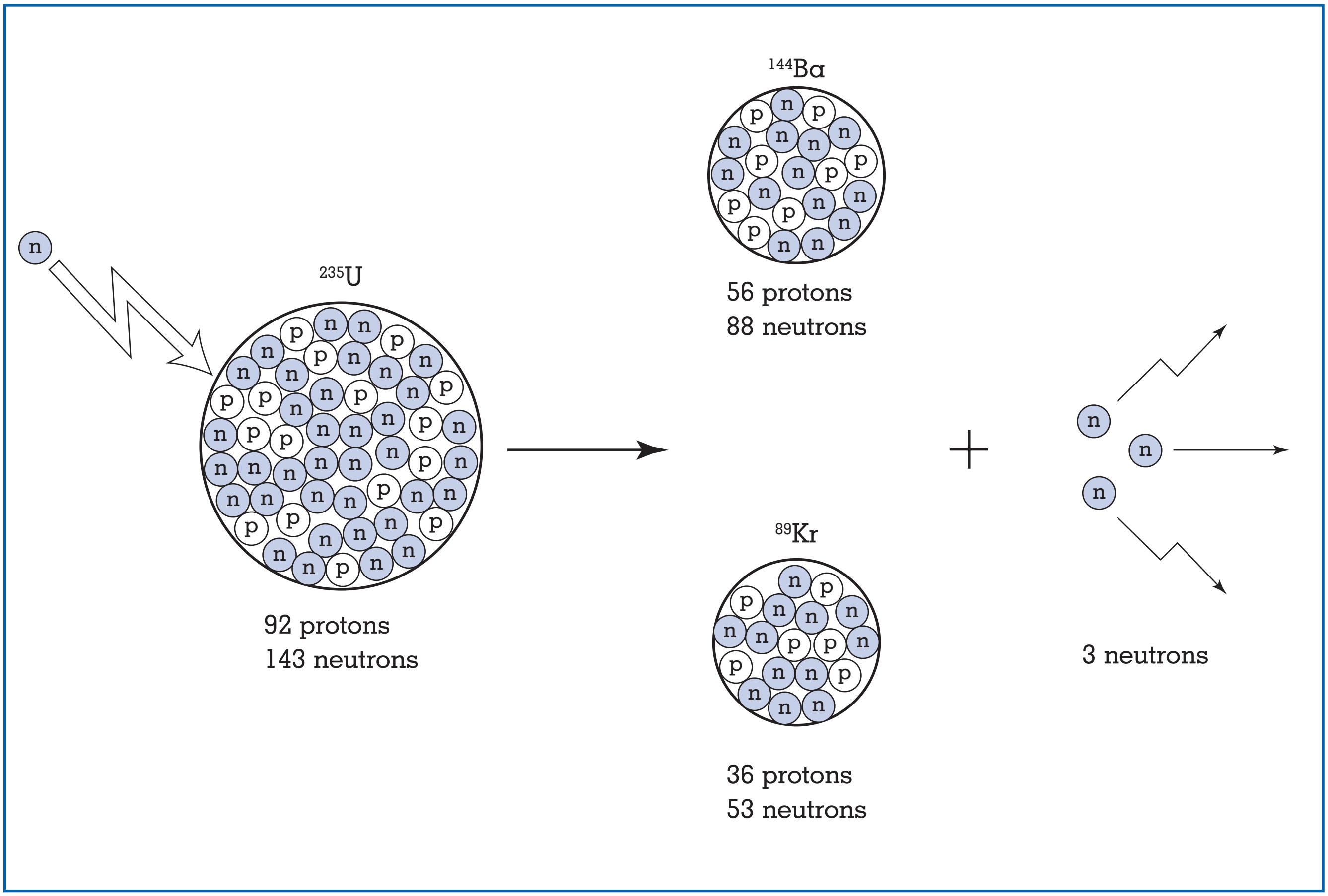


Figure 2.
Neutron-induced fission process.

Describing a neutron-induced nuclear fission reaction such as

$$^{235}\mathrm{U} + \mathrm{n} \rightarrow {}^{144}\mathrm{Ba} + {}^{89}\mathrm{Kr} + 3\mathrm{n} + \text{energy}$$

does not convey the likelihood of it actually happening. When a neutron encounters the ^{235}U nucleus, several nuclear events other than fission may occur. For example, the neutron may bounce off, changing the course of its path much like a marble bouncing off a bowling ball. Or the ^{235}U nucleus may capture the neutron and disintegrate in a way that does not involve nuclear fission. Whatever the event, there is only some chance that the event occurs when the neutron and the nucleus interact, and this probability depends strongly on the energy of the neutron. The fissioning of ^{235}U by a neutron is likely only for low-energy neutrons. Low in this sense means that the energy is of the order of the energy of a molecule in the air in a room.

A drop of water is a collection of molecules bound together by electric forces. Perturbing the drop in some way often causes the drop to separate into two smaller drops. A heavy nucleus such as ^{235}U has 92 protons and 143 neutrons and has features of a liquid drop that are useful for visualizing the dynamics of nuclear fission. The strong nuclear force that tends to bind the protons and the neutrons together competes with electric forces between protons that tend to cause the nucleus to break apart. Competition between these forces produces collective oscillations that are analogous to oscillations of a liquid drop. The energy added to the nucleus when it captures a neutron enhances the oscillations, and on some occasions the nucleus separates into two parts. The model is useful for visualizing the nuclear fission mechanism as well as for doing quantitative calculations involving the probability aspects of the process.

NUCLEAR FISSION CHAIN REACTION

A chain reaction results when some event triggers a sequence of identical events. A chain reaction on a crowded highway develops when a car runs into the rear of another car, the struck car collides with a car in front of it, and so on. An expanding chain reaction develops if a car hits two cars, each struck car moves on to strike two cars, and so on. A chain reaction is central to the operation of a nuclear fission reactor and a nuclear fission bomb. It begins with a neutron-inducing fission in a nucleus. Neutrons liberated in the fission in turn induce fission in other nuclei, and so on (see Figure 3). Each fission releases energy, and unless there is some purposeful control, the energy release proceeds uncontrolled. Controlled or uncontrolled, the chain reaction does not propagate unless (1) the fuel is properly arranged and (2) there is sufficient likelihood that a neutron induces fission.

This is tantamount to saying that a chain reaction of cars does not propagate unless (1) the cars are arranged properly and (2) there is a sufficient chance that a car will make other collisions.

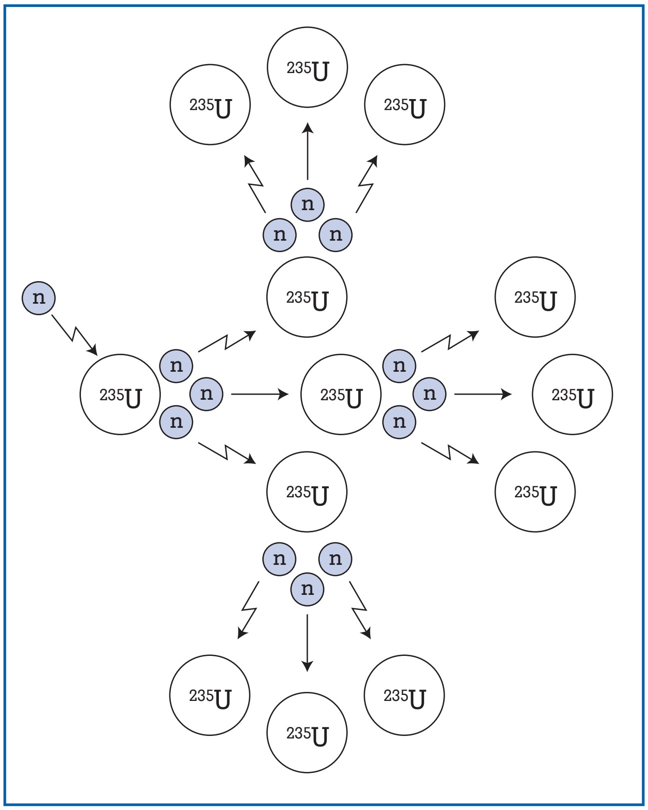


Figure 3.
A nuclear fission chain reaction.

NUCLEAR FUEL

Commercially, there are only two nuclear species that will function in a self-sustaining nuclear fission chain reaction; ^{235}U and ^{239}Pu (plutonium). Uranium occurs naturally in two predominant forms, ^{235}U and ^{238}U. The bulk, 99.3 percent, of natural uranium is ^{238}U. Even though ^{238}U differs from ^{235}U by having only three additional neutrons, the likelihood of neutron-induced fission of ^{238}U is so low that it cannot be used in a self-sustaining chain reaction. Thus the great bulk of natural uranium is not of a form that can be used in self-sustaining chain reactions. ^{239}Pu is not found naturally in appreciable amounts. Rather, ^{239}Pu is made through nuclear transmutations. The process begins by having ^{238}U capture a neutron to form ^{239}U. ^{239}U is unstable and undergoes radioactive decay, with the end product being ^{239}Pu.

NUCLEAR FISSION REACTOR PRINCIPLES

The fuel in a nuclear fission reactor is generally ^{235}U atoms arranged appropriately in a reactor vessel. Neutrons instigate fission of nuclei of ^{235}U atoms and liberate energy. The energy output may be controlled either by regulating the fuel and/or adjusting the neutron supply. Generally, the energy output is controlled by regulating the number of neutrons. Cylindrical control rods made of materials that readily absorb neutrons are moved up or down in strategically arranged spaces in the fuel assembly. Water circulating around the nuclear fuel removes heat produced by nuclear fission reactions and reduces moderates) the speed of neutrons liberated by fission. The main components of a nuclear fission reactor are shown in Figure 4.

Moderator

Neutrons produced in nuclear fission reactions are very energetic and are referred to as fast neutrons. To optimize the chain reaction, the speed of the reaction neutrons must be reduced. Water molecules surrounding the fuel help slow the neutrons through collisions with atoms. A proton is an optimum target for a neutron. Hydrogen atoms in water are a good source of protons because each water molecule has two hydrogen atoms and each atom has a single proton in its nucleus. A neutron produced in a nuclear fission reaction bounces around in a pool of water,

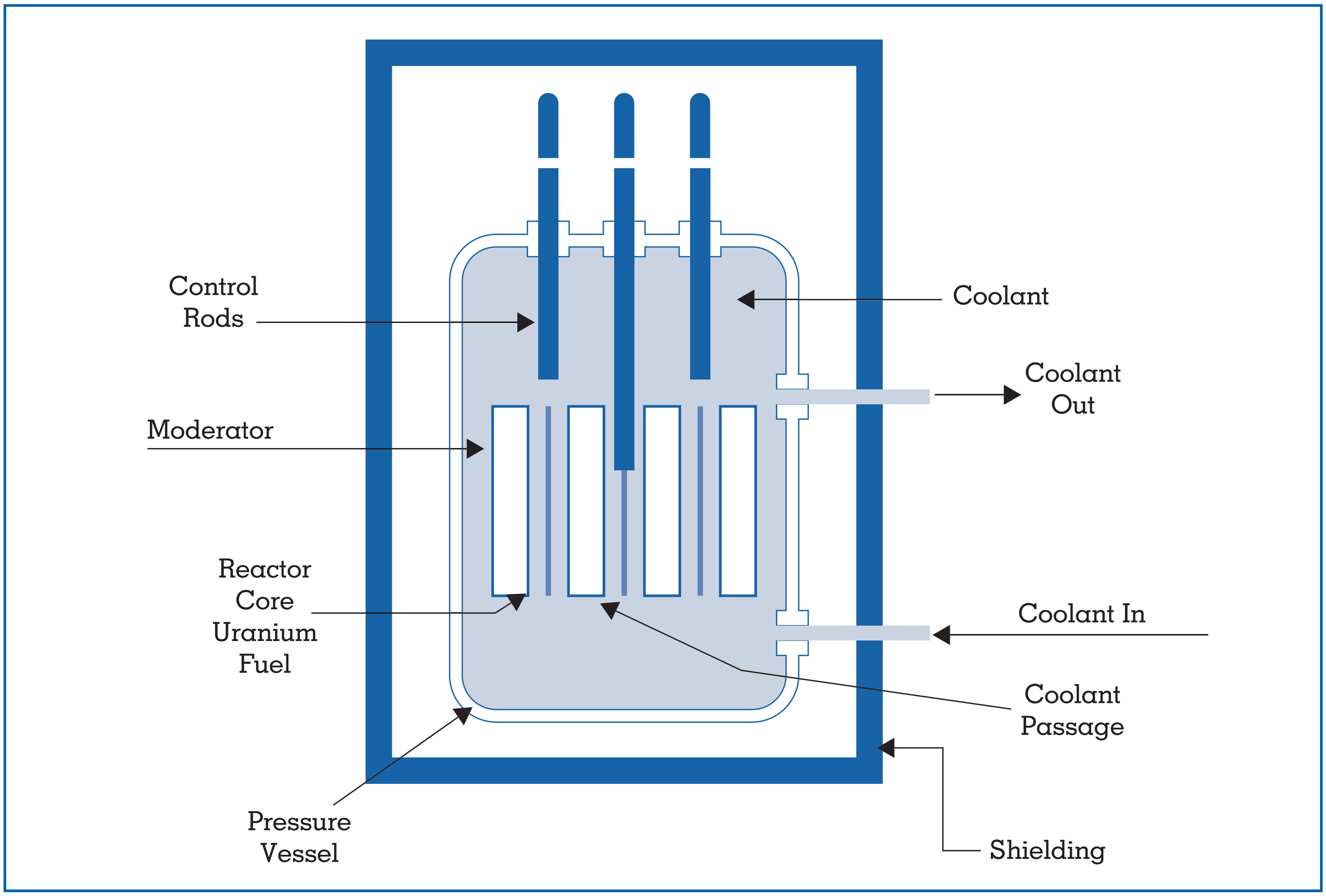


Figure 4.
Main components in a nuclear fission reactor.

losing energy with each collision with a proton. After its energy is reduced sufficiently, it may react with a ^{235}U nucleus to induce fission. Reducing the energy is referred to as moderation and the material used in the moderation is called a moderator. An object colliding head-on with a stationary object loses the greatest amount of kinetic energy when the masses of the two objects are the same. This is apparent when a billiard ball collides head-on with another billiard ball at rest. A hydrogen atom has roughly the same mass as a neutron. Accordingly, a neutron loses virtually all its kinetic energy in a head-on collision with a hydrogen atom. Because a water molecule (H_2O) has two hydrogen atoms, water is an efficient moderator for neutrons produced in nuclear fission reactions. Water is the moderator in nearly all reactors.

Control Rods

A nuclear chain reaction can be controlled by inserting materials within the fuel that deprive the fuel of neutrons. These materials are usually rod-shaped and are referred to as control rods. Control rods are made of materials that are more likely than the fuel to absorb neutrons. Boron and cadmium are favorite materials for control rods. To decrease the energy output, the control rods are inserted closer to the fuel. To increase the energy output, the control rods are pulled away from the fuel. Boric acid (H_3BO_3), a boron-containing liquid that strongly absorbs neutrons, is often included in the water moderator to provide further control of the energy output of the reactor.

Fuel

Uranium is a metal that is found naturally as a constituent of chemical compounds in minerals such as pitchblende. Uranium ore is mined much like coal: Open pits are used to mine shallow deposits, deeper deposits require shaft mining. Commercial ores yield 3 to 5 lb of uranium compounds per ton of ore. A material called yellow-cake is produced that is rich in the uranium compound U_3O_8. About 7 of every 1,000 uranium atoms in the yellow cake are of the variety ^{235}U.

Some nuclear fission reactors are designed to use natural uranium having 0.7 percent ^{235}U and 99.3 percent ^{238}U. CANDU reactors, manufactured in Canada, are examples. These reactors do not use ordinary water for the moderator. Most nuclear fission reactors use ordinary water for a moderator which requires that the fuel be about 3 percent ^{235}U and about 97 percent ^{238}U. Achieving this enrichment requires that the solid uranium compounds in the yellow cake be converted to gaseous uranium hexafluoride (UF_6). Following enrichment, gaseous UF_6 is converted to solid uranium oxide (UO_2) for fabrication of fuel elements for a nuclear reactor.

Uranium oxide fuel is formed into solid pellets about ⅜ in. in diameter and ½ in. long. Fuel pellets are stacked in tubes about 12 ft long. Neutrons must be able to penetrate the tube walls, called cladding, to interact with the uranium, and the walls must withstand the temperature required for heating water for a steam turbine. Stainless steel or a material called zircaloy is used for the cladding. Some 200 tubes are packed into a fuel bundle. About 175 appropriately arranged fuel bundles form the reactor core, where the energy is produced. A cylindrical steel vessel about 20 ft in diameter and about 40 ft high houses the core. Energy and neutrons are liberated from nuclear fission reactions within the fuel pellets. About 83 percent of the energy is in kinetic energy of the fission fragments. Fission fragments convert their kinetic energy to heat in the process of coming to rest in the fuel pellets. Heat is conducted from the pellets by water circulating around the fuel cladding. Water has two very important roles: conducting heat from the reactor core and moderating neutron speeds.

Energy Production

There are about 100 tons of uranium in the fuel bundles. With the control rods in their lowest position, there is no chain reaction, no energy produced, and no neutrons liberated from nuclear fission reactions. Energy production starts by raising the control rods and inserting a source of neutrons. Nuclear fission reactions begin producing energy and neutrons. Neutrons escape from the fuel elements into the water, where they lose speed and change direction with each collision with a proton. After losing sufficient speed, a neutron may reenter a fuel element and induce a ^{235}U nucleus to fission to help foster the chain reaction. If on the average one neutron from a nuclear fission reaction goes on to produce another nuclear fission reaction, the chain reaction is self-sustaining, and energy is produced at a steady rate. Raising the control rods allows more nuclear fission reactions and more energy production. Higher, but steady, energy production occurs at the new control rod position. Control of the chain reaction depends on

- the arrangement of the fuel elements
- the quality of the moderator
- the quantity of ^{235}U
- the neutron energy required for a high probability for inducing nuclear fission
- the material surrounding the reactor core that is used to minimize the escape of neutrons.

Although protons are very efficient neutron moderators, they also efficiently capture neutrons to form bound proton-neutron pairs called deuterons. Reactors using ordinary water for the moderator compensate for neutron capture by using fuel enriched to about 3 percent ^{235}U.

Usually atoms resulting from nuclear fission are radioactive. There are also radioactive atoms produced from neutron capture by both ^{235}U and ^{238}U. Both types of radioactive atoms remain in the nuclear fuel. It is these radioactive atoms that comprise the nuclear wastes that require disposal in an environmentally acceptable manner.

Boiling-Water Reactors and Pressurized-Water Reactors

Current light-water reactors have two basic designs. One is termed a boiling-water reactor (BWR), the other a pressurized-water reactor (PWR). The principles of the PWR design is illustrated in Figure 5. Water comes to a boil, producing steam inside the containment vessel of a boiling-water reactor. Water is kept under pressure and does not boil inside the containment vessel of a pressurized- water reactor. Steam for the turbine in a nuclear power plant using a pressurized water reactor comes from water brought to a boil in a steam generator garnering heat from water circulating around the reactor core. In a boiling- water reactor, steam is produced in the reactor vessel. In a pressurized-water reactor, steam comes from a system separated from the reactor vessel. Pressure and temperature in a boiling-water reactor are about 1,000 lb per sq in. and 545°F (285°C). These

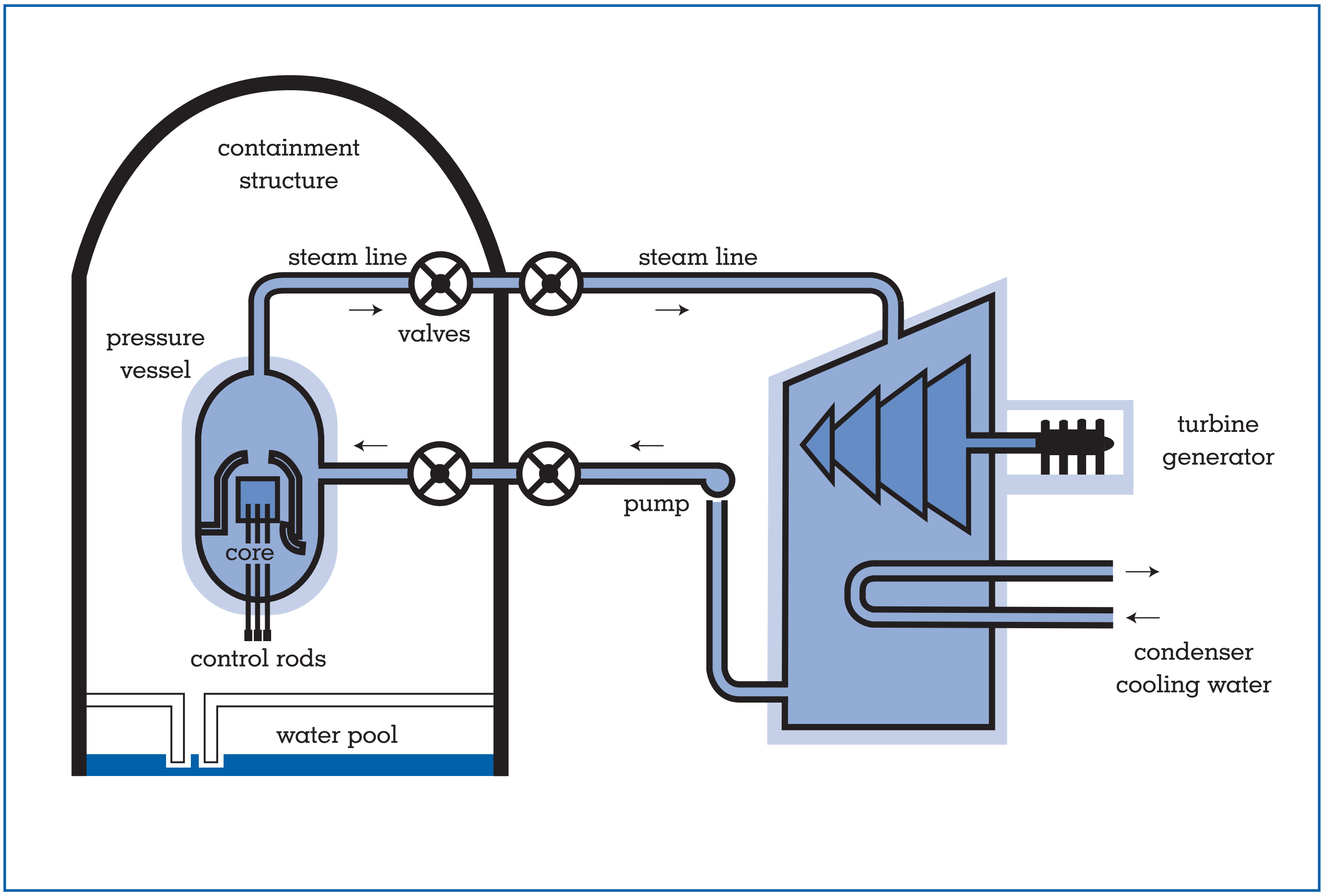


Figure 5.
The water-steam circuit in a pressurized water reactor.

numbers rise to about 2,250 lb per sq in. and 600°F (316°C) in a pressurized-water reactor. Roughly 70 percent of the reactors in the United States are of the PWR type. Essentially, the safety aspects of PWR and BWR reactors are equivalent.

At the beginning of the twenty-first century, 439 nuclear power reactors in 31 countries were generating 352 billion watts of electric power. More than 16 percent of the world's electricity was produced by these electric power plants. On average the power output of one of these plants is roughly 800 megawatts and is comparable to the power output of a large coal-burning electric power plant. Either one of these plants could supply the electricity needs of about a quarter of a million homes. The electric power outputs of solar and wind systems, for example, are usually factors of ten to a hundred times less than that of a large nuclear-fueled or coal-burning electric power plant.

REACTOR CONTROL

The products of nuclear fission reactions are radioactive and disintegrate according to their own time scales. Often disintegration leads to other radioactive products. A few of these secondary products emit neutrons that add to the pool of neutrons produced by nuclear fission. Very importantly, neutrons from nuclear fission occur before those from radioactive decay. The neutrons from nuclear fission are termed prompt. Those from radioactive decay are termed delayed. A nuclear bomb must function on only prompt neutrons and in so doing requires nearly 100 percent pure ^{235}U (or ^{239}Pu) fuel. Although reactor fuel is enriched above its natural 0.7 percent abundance, it is still some thirty times less concentrated than the fuel needed for bombs. A light-water nuclear reactor cannot achieve a self-sustaining condition using only prompt neutrons; delayed neutrons are needed as well. Control rods are slammed into position if a critical excursion from normal operation is sensed. The time involved in delayed neutron emission is crucial for success of the control operation.

A nuclear reactor and its auxiliaries involve a complex assembly of pumps, pipes, and valves for circulating water and steam. Minor malfunctions of the components are inevitable, and routine maintenance is inescapable. Distinct from the routine malfunctions is the possibility of loss of water in the main cooling of the reactor core. This bears the name "loss of cooling accident." Possible causes of such an accident include reactor operator errors, faulty construction, improper maintenance, natural disaster, and sabotage. A loss of cooling terminates the nuclear fission chain reaction because of the need for a moderator. However, radioactivity in the reactor core continues to produce heat at a level of 5 to 7 percent of its rated power. This percentage may seem low, but 5 percent of 3 billion watts of thermal power from normal operation amount to 150 million watts of thermal power that remains after the chain reaction stops. Many of the radioactive products are short-lived, and the power level drops accordingly. Still, there is sufficient heat to begin melting the core in about thirty seconds. Unless checked, this molten radioactive mass could melt its way through the bottom of the containment facilities. Emergency core-cooling systems must operate within this thirty-second time period. Although safety measures make the chance of an accident releasing radioactivity unlikely, it cannot be made zero.

THE NUCLEAR FISSION BREEDER REACTOR

Contemporary nuclear fission reactors are fueled with either ^{235}U or ^{239}Pu. In a nuclear fission reactor fueled with ^{235}U, some ^{238}U in the fuel is converted to ^{239}Pu through nuclear transmutations and becomes part of the fuel inventory. However, the reactor is not designed to optimize the production of ^{239}Pu. A nuclear breeder reactor is designed to both produce energy and convert ^{238}U to ^{239}Pu for fuel. The conversion process is termed breeding.

The fission of ^{235}U or ^{239}Pu liberates, on average, two to three neutrons. One neutron is required to sustain the nuclear fission chain reaction. In a nuclear breeder reactor, the extra neutrons are used to induce nuclear reactions that lead to the production of ^{239}Pu. The sequence begins by arranging for a neutron to be captured by a ^{238}U nucleus to form ^{239}U. ^{239}U is unstable and disintegrates producing a variety of neptunium, ^{239}Np. ^{239}Np is unstable and decays, forming ^{239}Pu. The breeding principle requires both fissionable fuel (^{235}U or ^{239}Pu) and a breeding material (^{238}U). Both resources are depleted as time passes.

The chance of inducing nuclear fission of ^{235}U is optimum for slow neutrons. Neutron capture by ^{238}U is significant only for energetic (or fast) neutrons. Accordingly, the nuclear fission reactions must be initiated by fast neutrons even though it is more difficult. Water cannot be used as a coolant, as in a light-water nuclear reactor, because of its moderating effect on neutrons. The alternative to water is usually liquid sodium. These reactors are termed liquid metal (for cooling) fast (for energetic neutrons) nuclear breeder reactors.

Nuclear Breeder Reactor Projects

An experimental nuclear breeder reactor has been operating at the Argonne National Laboratory near Idaho Falls, Idaho, since 1963. However, widespread commercial utilization of nuclear breeder reactors has not materialized. A commercial nuclear breeder reactor electric power plant was built in Monroe County, Michigan, in 1963, but was permanently terminated in 1966 after a blockage in a sodium cooling line produced a partial meltdown of its core. Plans were announced in February 1972 for a major nuclear fission breeder electric power plant to be built on the Clinch River in eastern Tennessee and planned for completion in 1980, but the project ultimately was not funded by the U.S. Congress. France and Japan have constructed major nuclear breeder reactor electric power plants, but at the turn of the century all had been closed. Reprocessing nuclear fuel to recover ^{239}Pu is necessity in nuclear fission breeder technology. This is a major stumbling block because of concerns for proliferation of ^{239}Pu for making nuclear bombs.

Joseph Priest

See also: Electric Power, Generation of; Environmental Problems and Energy Use; Explosives and Propellants; Meitner, Lise; Military Energy Use, Historical Aspects of; Molecular Energy; Nuclear Energy; Nuclear Energy, Historical Evolution of the Use of; Nuclear Fission Fuel; Nuclear Fusion; Nuclear Waste.

BIBLIOGRAPHY

Bodansky, D. (1996). *Nuclear Energy: Principles, Practices, and Prospects*. Woodbury, NY: AIP Press.

Clark, R. W. (1980). *The Greatest Power on Earth: The Story of Nuclear Fission*. London: Sidgewick & Jackson.

El Wakil, W. W. (1984). *Powerplant Technology*. New York: McGraw-Hill.

Forsberg, C. W., and Weinberg, A. (1990). "Advanced Reactors, Passive Safety, and Acceptance of Nuclear Energy." *Annual Review of Energy* 15: 133–152.

Golay, M. W., and Todreas, N. E. (1990). "Advanced Light Water Reactors." *Scientific American* 262(4):82–89.

Graetzer, H. G., comp. (1981). *The Discovery of Nuclear Fission*. New York: Arno Press.

Hafele, W. (1990). "Energy from Nuclear Power." *Scientific American* 263(8):136–143.

Krane, K. S. (1987). *Introductory Nuclear Physics*. New York: John Wiley & Sons.

Priest, J. (2000). "Nuclear Energy." In *Energy: Principles, Problems, Alternatives*, 5th ed., ed. J. Priest. Dubuque, IA: Kendall/Hunt.

Ruzic, D. N. (1998). "Light-Water Reactors and Their Advances." Urbana, IL: University of Illinois. <http://starfire.ne.uiuc.edu/ne201/course/topics/light_water_reactors/> August 28, 2000.

NUCLEAR FISSION FUEL

A nuclear power plant generates electricity in a manner similar to a fossil fuel plant. The fundamental difference is the source of heat to create the steam that turns the turbine-generator. A fossil plant relies on the combustion of natural resources (coal, oil) to create steam. A nuclear reactor creates steam with the heat produced from a controlled chain reaction of nuclear fission (the splitting of atoms).

Uranium is used as the primary source of nuclear energy in a nuclear reactor, although one-third to one-half of the power will be produced from plutonium before the power plant is refueled. Plutonium is created during the uranium fission cycle, and after being created will also fission, contributing heat to make steam in the nuclear power plant. These two nuclear fuels are discussed separately in order to explore their similarities and differences. Mixed oxide fuel, a combination of uranium and recovered plutonium, also has limited application in nuclear fuel, and will be briefly discussed.

URANIUM

Uranium (symbol: U; atomic number: 92) is the heaviest element to occur naturally on Earth. The most commonly occurring natural isotope of uranium, U-238, accounts for approximately 99.3 percent of the world's uranium. The isotope U-235, the second most abundant naturally occurring isotope, accounts for another 0.7 percent. A third isotope, U-234, also occurs naturally, but accounts for less than 0.01 percent of the total naturally occurring uranium. The isotope U-234 is actually a product of radioactive decay of U-238.

Nuclear power is now the only substantial use for uranium. But before uranium can be used in a nuclear reactor, it must undergo several processes. After uranium is mined from geological mineral deposits, it is purified and converted into uranium hexafluoride (UF_6). The UF_6 is next enriched, increasing the concentration of U-235 by separating out UF_6 made with U-238 atoms. The enriched UF_6 is then converted into uranium dioxide (UO_2), and pressed into fuel pellets for use in the nuclear reactor.

Uranium Mining and Milling

Uranium is found in most rock, in a concentration of two to four parts per million (ppm). Substantially greater average concentrations can be found in mineral deposits, as high as 10,000 ppm, or 10 percent. Most uranium deposits suitable for mining, however, contain an average of less than 1 percent uranium. Uranium is a metal, and thus its acquisition is not unlike the mining of any other metallic ore. Although uranium is found nearly everywhere on the earth, Canada leads the world in uranium production, mostly due to its heavy financial investment in uranium exploration, and to a few sizable deposits in the Saskatchewan territory. Table 1 depicts the total world uranium production in 1997.

Although the nucleus of the uranium atom is relatively stable, it is radioactive, and will remain that way for many years. The half-life of U-238 is over 4.5 billion years; the half-life of U-235 is over 700 million years. (Half-life refers to the amount of time it takes for one half of the radioactive material to undergo radioactive decay, turning into a more stable atom.) Because of uranium radiation, and to a lesser extent other radioactive elements such as radium and radon, uranium mineral deposits emit a finite quantity of radiation that require precautions to protect workers at the mining site. Gamma radiation is the

Country	*Tons of Uranium*
Canada	12,029
Australia	5520
Niger	3497
Namibia	2905
USA	2170
Russia	2000
Uzbekistan	1764
South Africa	1100
Kazakhstan	1000
France	748
Czech Republic	590
Gabon	472
China	500
Ukraine	500
others	897
Total world	**35,692**

Table 1.
1997 Uranium Mine Production
SOURCE: Uranium Institute, Core Issues 3/98; and Uranium Institute, 1998, Global Nuclear Fuel Market, Supply & Demand 1998–2020.

most prevalent, followed by beta and alpha radiation. Because airborne particulates can be ingested or inhaled, thus subjecting workers to internal doses of radiation, the most important step taken to protect mine workers is to minimize the amount of airborne particulate matter. In areas where airborne particulate is of great concern, the uranium ore can be wetted to prevent releasing dust to the air. In areas of even greater risk, workers can be outfitted with respiratory equipment. Radon gas is also emitted from uranium, but proper ventilation provides sufficient dilution of the gas so that workers are not affected.

In general, there are three methods available to mine uranium deposits: open-pit, underground, and in-situ leaching. Open-pit mining is used when the uranium deposit is relatively shallow. Underground mining is used when the deposit is relatively deep. But for underground mining to be cost-effective, the deposit must be relatively high-grade and abundant with uranium. For both of these mining methods, the uranium ore is crushed and ground into a fine powder, releasing the uranium from the material surrounding it in the mineral deposit. Next, the crushed uranium is leached from the ore with a solution designed to dissolve only the uranium. After the uranium is dissolved, the solution is separated. The uranium-rich liquid is run off into a collection facility, and the solid waste product is left behind. Note that this waste product, referred to as tailings, can be highly radioactive, and can accumulate in substantial quantity. Since the uranium is often found in deposits with other elements having substantially shorter half-lives, these short-lived isotopes are left behind in a concentrated waste mass. The waste water is evaporated from the sludge, and the remaining metals and minerals are stored, often in an open pit at the mill site, for the life of the facility.

A less environmentally invasive method of obtaining the uranium is through a process called in-situ leaching (ISL). Rather than crush the uranium mineral deposit in order to expose the uranium and then dissolve it, the dissolving solution is pumped directly into the uranium deposit. This way, the dissolved uranium solution is obtained without disturbing the minerals surrounding the uranium in the ore, and all of the waste solids are left behind. This assumes the mineral deposit is able to facilitate the leaching process. Uranium deposits must be permeable, like those found in sand and sandstone, to utilize ISL in order for the dissolving solution to pass easily into the deposit to extract the ore. Between 10 and 15 percent of the world uranium supply is mined via in-situ leaching.

Regardless of the method used to obtain the uranium—open pit, underground, or in-situ leaching—the result is a uranium-rich solution. The uranium is recovered from the solution through ion exchange or solvent extraction, and concentrated into a fine precipitate. The uranium-free solution can then be returned to the mine to extract more uranium. The uranium in the precipitate is a blend of uranium dioxide (UO_2) and uranium trioxide (UO_3), combined in a general uranium oxide (U_3O_8). The precipitate is usually dried, forming a yellow cake of uranium, simply called "yellow cake." This concentrated uranium is typically 70 to 90 percent uranium. In order to use the yellow cake as a nuclear fuel, the uranium must go through a process of conversion.

Uranium Conversion

Since the uranium from the milling process is still in an unusable form, the yellow cake is broken down once again. The uranium trioxide is reduced to uranium dioxide at very high temperatures. Refining of the product also takes place. Now the uranium product consists almost entirely of UO_2.

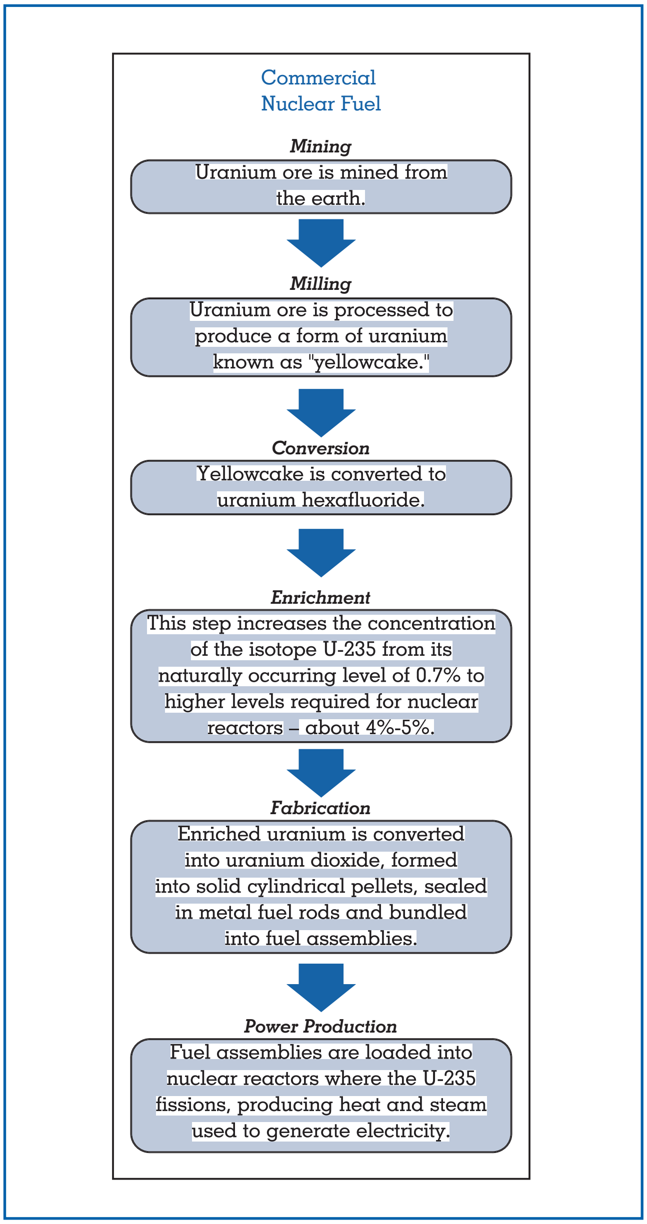


Figure 1.
Processes used in refining uranium for use in reactors
SOURCE: USEC, Inc.

During the conversion process, the object is to create uranium hexafluoride (UF_6), a highly corrosive substance that is gaseous at high temperatures, but is a white crystalline solid at lower temperatures. Uranium hexafluoride is easily transported in its crystalline form to an enrichment facility (the step taken after conversion), but the gaseous form is well suited for the enrichment process, itself. First, the UO_2 is combined with hydrogen fluoride (HF) to form uranium tetrafluoride (UF_4). This can be accomplished by reacting the UO_2 with either gaseous or liquid HF. The UF_4 is combined with gaseous fluorine to produce UF_6, and is shipped to a facility for enrichment.

Uranium Enrichment

Uranium-235 and U-238 behave differently in the presence of a controlled nuclear reaction. Uranium-235 is naturally fissile. A fissile element is one that splits when bombarded by a neutron during a controlled process of nuclear fission (like that which occurs in a nuclear reactor). Uranium-235 is the only naturally fissile isotope of uranium. Uranium-238 is fertile. A fertile element is one that is not itself fissile, but one that can produce a fissile element. When a U-238 atom is struck by a neutron, it likely will absorb the neutron to form U-239. Through spontaneous radioactive decay, the U-239 will turn into plutonium (Pu-239). This new isotope of plutonium is fissile, and if struck by a neutron, will likely split.

Nuclear fuel always contains a mixture of these uranium isotopes. The U-235, being naturally fissile, will sustain a controlled chain reaction in the nuclear reactor early in the fuel cycle. The U-238, being fertile, will eventually turn into plutonium, continuing to sustain the nuclear chain reaction late in the fuel cycle. The amount of U-235 initially blended into the nuclear fuel will dictate certain performance characteristics. Since naturally-occurring uranium is only approximately 0.7 percent U-235, any nuclear fuel containing more than 0.7 percent U-235 is said to be enriched. Light-water reactors, like those in the United States, use a fuel that has been enriched up to 5 percent. Some reactors are designed to use natural (unenriched) uranium, but they must rely on a medium other than light water to moderate the chain reaction. (These types of reactors typically use heavy water or graphite as a moderator for the nuclear reaction.) For these reactors, the U_3O_8 need only be refined and converted directly to UO_2 for use in fuel pellets, and enrichment is not necessary.

Uranium-235 is less massive than U-238, because each atom of U-235 contains three fewer neutrons than U-238. Although this difference appears insignificant, this 1 percent difference is precisely what allows the UF_6 to be divided into those molecules containing U-235 and those containing U-238. By separating the two quantities, enrichment plants can control the level of enrichment of the end product. Two processes for dividing the uranium isotopes are in large-scale commercial use. These are gaseous diffusion and centrifuge. A third enrichment process, laser enrichment, is being developed.

Gaseous Diffusion. Gaseous diffusion is responsible for producing over 90 percent of the world's enriched uranium. Gaseous diffusion has the ability to process large quantities of gaseous UF_6 at a time. Gaseous UF_6 is fed through a series of porous membranes that tend to filter out the heavier UF_6 molecules containing U-238 atoms. Gas that passes through the membrane will have a slightly higher percentage of U-235 molecules in the UF_6 than when it started, since some of the U-238 molecules did not pass through the membrane. The slightly enriched UF_6 is drawn off and forced through another series of membranes. Each iteration produces a slightly higher enrichment of U-235. After more than 1,500 iterations, the UF_6 is finally enriched to a usable form, at a U-235 concentration of 3 to 4 percent.

Centrifuge Enrichment. In a centrifuge, the UF_6 molecules are spun in a rotating cylinder. Because of the difference in mass between the UF_6 molecules containing U-235 and U-238, the lighter molecules remain at the center of the cylinder while the heavier molecules migrate toward the outside. The enriched UF_6 is drawn off the center of the centrifuge, and then centrifuged again. Similar to the diffusion process, each iteration produces a slightly higher enrichment of U-235. The centrifuge process is much more efficient than the diffusion process, producing a suitable enrichment after only ten to twenty iterations, and requiring significantly less energy to operate the enrichment plant than does a comparable diffusion plant. The capacity of a centrifuge facility is substantially less than the capacity of a diffusion plant, however, severely limiting its commercial application.

Laser Enrichment. Laser enrichment attempts to ionize gaseous U-235 atoms. By using a laser of a specific frequency, valence electrons of U-235 atoms can be removed without ionizing U-238 atoms. The positively charged U-235 atoms are attracted to a negatively charged metal plate, and liquid U-235 is collected. This process does not yet work with the UF_6 created from the existing uranium mining process, so its application is currently limited. Future laser enrichment processes will focus on using UF_6 in its enrichment technique.

Fuel Pellet Fabrication

Finally, the enriched uranium of converted back into UO_2. The UO_2 is pressed into small fuel pellets and packaged in a metal tube (made of a zirconium alloy) for use in a nuclear reactor.

Nuclear fuel can also be fabricated from existing stockpiles of fuel sources. Spent fuel can be reprocessed to extract the remaining U-238 for use in the next fuel cycle. Spent fuel also contains Pu-239, a direct replacement for the fissile U-235 in the enrichment process. Even military grade plutonium and highly enriched uranium can be reprocessed and diluted into a form usable in most nuclear reactors. The reprocessed isotopes are generally manufactured into mixed oxide (MOX) fuels and loaded into conventional nuclear reactors as fuel. The cost of increasing the fissile isotopes in MOX fuel represents a great cost savings over uranium enrichment, and thus reprocessing is increasingly becoming economically viable.

PLUTONIUM

Plutonium (symbol: Pu; atomic number: 93) is not a naturally occurring element. Plutonium is formed in a nuclear reaction from a fertile U-238 atom. Since U-238 is not fissile, it has a tendency to absorb a neutron in a reactor, rather than split apart into smaller fragments. By absorbing the extra neutron, U-238 becomes U-239. Uranium-239 is not very stable, and undergoes spontaneous radioactive decay to produce Pu-239.

Plutonium-239 is a fissile element, and will split into fragments when struck by a neutron in the nuclear reactor. This makes Pu-239 similar to U-235, able to produce heat and sustain a controlled nuclear reaction inside the nuclear reactor. Nuclear power plants derive over one-third of their power output from the fission of Pu-239. Most of the uranium inside nuclear fuel is U-238. Only a small fraction is the fissile U-235. Over the life cycle of the nuclear fuel, the U-238 changes into Pu-239, which continues to provide nuclear energy to generate electricity.

Not all of the Pu-239 will fission during the fuel cycle in a nuclear reactor. Some of the plutonium will not experience neutron bombardment sufficient to cause fission. Other plutonium atoms will absorb one or more neutrons and become higher numbered isotopes of plutonium, such as Pu-240, Pu-241, etc. Plutonium comprises just over 1 percent of nuclear reactor spent fuel—the fuel removed from the nuclear reactor after it has produced all the nuclear fission it can economically provide. Of the plutonium remaining in the spent fuel, slightly less than 60 percent will be fissionable Pu-239.

Because of the high concentration of Pu-240, the plutonium in spent fuel is unsuitable for military application (nuclear warheads). Plutonium-240 is a constant source of neutron decay and is highly unstable. Moreover, the Pu-240 cannot sustain a chain reaction. Military-grade plutonium consists of over 90 percent fissionable plutonium. Because it is impractical to separate the two isotopes of plutonium, special reactors are used to create Pu-239 with extremely small quantities of Pu-240. Even where nations are desperate for electric-generating capacity, the fear of rogue nations teaming to build and operate these special facilities to create military-grade plutonium is a main reason the International Atomic Energy Agency closely safeguards the transfer and use of nuclear energy technology.

Now that much of the world has agreed to nuclear disarmament, scientists and world leaders are searching for uses for the surplus weapons-grade plutonium. Instead of pursuing disposal options, one option is the use of mixed-oxide (MOX) nuclear fuel.

MIXED OXIDE FUEL

In order to deplete stockpiles of weapons-grade plutonium, it is possible to "burn" the fuel in a nuclear reactor. This option has two distinct advantages. First, the fissionable plutonium is a direct replacement for fissionable uranium in a nuclear reactor. Therefore, the plutonium will sustain a controlled chain reaction inside the reactor and provide the heat necessary to create steam for a power plant. Second, the plutonium not undergoing fission will become increasingly contaminated with other isotopes of plutonium (e.g., Pu-240) so that it will no longer be suitable for weapons applications. The longer the plutonium remains inside the nuclear reactor, the more Pu-240 is created as Pu-239 absorbs neutrons. The spent fuel from a reactor using MOX fuel will contain diminished amounts of Pu-239 and increased amounts of Pu-240, effectively removing weapons-grade plutonium from circulation.

Mixed oxide fuel is not appropriate for all nuclear reactors. Plutonium requires faster neutrons in order to operate in a sustained chain reaction. Light-water reactors operate in a highly moderated environment, and thus U-238 and Pu-239 do not make excellent fuel sources. Light-water reactors can handle up to approximately 30 percent MOX fuel. Heavy-water reactors, such as those in Canada, and other fast neutron reactors, can operate with 100 percent MOX fuel.

Other options for eliminating weapons-grade plutonium are to seal it permanently in solid radioactive waste and dispose of it in waste repositories, and to use the plutonium to fuel fast neutron reactors (without reprocessing the plutonium into a MOX fuel).

Plutonium has a much shorter half-life than uranium (24,000 years for Pu-239; 6,500 years for Pu-240). Plutonium is most toxic if it is inhaled. The radioactive decay that plutonium undergoes (alpha decay) is of little external consequence, since the alpha particles are blocked by human skin and travel only a few inches. If inhaled, however, the soft tissue of the lungs will suffer an internal dose of radiation. Particles may also enter the blood stream and irradiate other parts of the body. The safest way to handle plutonium is in its plutonium dioxide (PuO_2) form because PuO_2 is virtually insoluble inside the human body, greatly reducing the risk of internal contamination.

WASTE DISPOSAL

The main drawback to nuclear power is the production of radioactive waste. Spent fuel from a nuclear reactor is considered a high-level radioactive waste, and remains radioactive for a very long time. Spent fuel consists of fission products from the U-235 and Pu-239 fission process, and also from unspent U-238, Pu-240, and other heavy metals produced during the fuel cycle. That is why special programs exist for the handling and disposal of nuclear waste.

Brian F. Thumm

See also: Nuclear Fission; Nuclear Waste.

BIBLIOGRAPHY

Considine, D. M., ed. (1995). *Van Nostrand's Scientific Encyclopedia*, 8th ed. New York: Van Nostrand Reinhold.

Edwards, C. R. (1992). "Uranium Extraction Process Alternatives." *CIM Bulletin* 85(58):112–136.

Hotta, H. (1987). "Recovery of Uranium from Seawater." *Oceanus* (Spring):30.

Hunter, J. (1991). "Highland In-Situ Leach Mine." *Mining* (August):58–63.

Katz, J. J. (1951). *The Chemistry of Uranium*. New York: Dover.

Lide, D. R., ed. (1996). *CRC Handbook of Chemistry and Physics*, 77th ed. Boca Raton, FL: CRC Press.

Mann, C. A. (1979). *Nuclear Fuel.* East Sussex: Wayland.

Patton, F. S. (1963). *Enriched Uranium Processing.* New York: Oxford.

Sax, N. R., and Lewis, R. J., Sr. (1992). *Dangerous Properties of Industrial Materials*, 8th ed. New York: Van Nostrand Reinhold.

Seaborg, G. T. (1945). "The Chemical and Radioactive Properties of the Heavy Elements." *Chemical Engineering News* 23:2190–2193.

Singleton, A. L., Jr. (1968). *Sources of Nuclear Fuel.* Oakridge, TN: US Atomic Energy Commission, Division of Technical Information.

Uranium Information Centre, Ltd. (1997). "Plutonium." Melbourne: Author.

Uranium Information Centre, Ltd. (1997). "Uranium Enrichment." Melbourne: Author.

Uranium Information Centre, Ltd. (1998). "World Uranium Mining." Melbourne: Author.

The Uranium Institute. (1998). "Uranium: From Mine to Mill." London: Author.

NUCLEAR FUSION

The nuclear fusion of light elements powers the sun and other stars. The fusion process in stars and supernovas creates from primordial hydrogen heavier elements, including those needed for chemistry and life. Nuclear fusion is analogous to the chemical reaction of burning, such as the joining (fusing) of hydrogen and oxygen to produce water vapor and release energy. Both processes produce something new while releasing energy.

Our planet contains large quantities of light elements that are theoretically suitable for providing energy from fusion reactions in central power plants. The potential energy resource from fusion is vast—so vast that the world's energy needs could be met for billions of years at current rates of consumption. Laboratory experiments, as well as nuclear explosions, have released fusion energy on the Earth, and it may be feasible to provide useful energy from fusion in the future in a controlled and safe manner. While the fusion process is a nuclear reaction, the results of the reaction do not include many of the problems with fission: First, there are no products like long-lived fission products that cause many of the problems associated with nuclear waste from fission. Second safeguarding uranium and plutonium is not an issue with fusion because those materials are not used. Fusion does not release compounds such as carbon dioxide that contribute to global warming through the greenhouse effect. Third, starting and maintaining a fusion reaction in a controlled manner appropriate for producing useful energy is difficult and always involves a limited amount of fuel at any given moment so that an accidental nuclear runaway reaction in a large quantity of reactants is not possible.

Fusion energy comes from rearranging electrostatic and nuclear forces inside the nucleus, and in this respect fusion is like fission. The rearrangements of both fission and fusion release energy in proportion to the decrease of mass during the reaction according to the famous formula from Albert Einstein's relativity, $E = mc^2$, where E is the energy released, m is the mass lost, and c is the speed of light.

Chemical reactions, such as the burning of carbon in coal with oxygen, also get energy from rearranging electrostatic forces, but those forces are much smaller and consequently chemical energies released per atom are much smaller than nuclear energies released per nucleus.

Nuclei suitable for fusion must come near each other, where "near" means something like the nuclear radius of 10^{-12} cm. For positively charged nuclei to make such a close approach it requires large head-on velocities, and therefore multimillion-degree Celsius temperature. In contrast, fission can occur at normal temperatures, either spontaneously or triggered by a particle, particularly an uncharged neutron, coming near a fissionable nucleus.

Above a temperature of 10,000 degrees or so, matter starts to become ionized, meaning that electrons separate from atoms. Ionized matter, containing ions and electrons, is called "plasma"—a term not to be confused with the "plasma" of physiology and medicine, which means a liquid in which cells are suspended.

Stars are giant spheres of plasma. Stars fuse hydrogen into helium for their primary source of energy. Omitting details of the catalytic role of carbon in stars, four hydrogens combine to yield one helium, two neutrinos, and two positively charged electrons (positrons) with a total kinetic energy of approximately twenty-eight MeV. The unit "MeV," or million electron volts, is convenient for discussions of nuclear energy, as is the "eV" or electron volt (or calorie and Btu) for treatments of chemical reactions. A million

Particle beam fusion accelerator II (PBFA-II) was designed to deliver at least 100 trillion watts of power and was the first machine with the potential to ingnite a controlled laboratory fusion reaction. (U.S. Department of Energy)

electron volts is 1.6×10^{-13} joule, 3.8×10^{-14} calorie, 1.5×10^{-16} Btu, 4.5×10^{-20} kWh (thermal), which makes the MeV sound like a very small unit until one considers that an MeV of released energy is typically associated with approximately one atomic mass unit (AMU), of which there are 6×10^{23} in a gram.

The fusion reaction of four hydrogens to make a helium yields 185,000 kilowatt-hours (thermal) per gram of original hydrogen fuel—an enormous result for such a small mass. Fusion occurs mostly in the core of stars at a temperature near 15 million degrees and the energy released migrates to the much cooler 6,000 degree surface to produce the radiation emitted, some of which is visible light.

While stars shine from the fusion of plain hydrogen, that reaction is unsuitable for use on the earth. In substellar sizes the energy released leaves the reacting region as electromagnetic radiation much too quickly. Releasing fusion energy on a human scale requires fusion fuels that react with each other more rapidly than the slow-burning basic fuels of stars.

REACTIONS FOR USEFUL ENERGY

The most plausible fusion reaction for producing energy commercially involves two isotopes of hydrogen, deuterium (D) and tritium (T), or ${}^{1}H^{2}$ and ${}^{1}H^{3}$. Deuterium contains one proton and one neutron for an atomic number of two. Tritium contains one proton and two neutrons for an atomic number of three. The reaction is

$$D + T \rightarrow {}^{2}He^{4} + n$$

with the release of 3.5 and 14.1 MeV in the helium and neutron, respectively.

Schematic of a fusion reactor

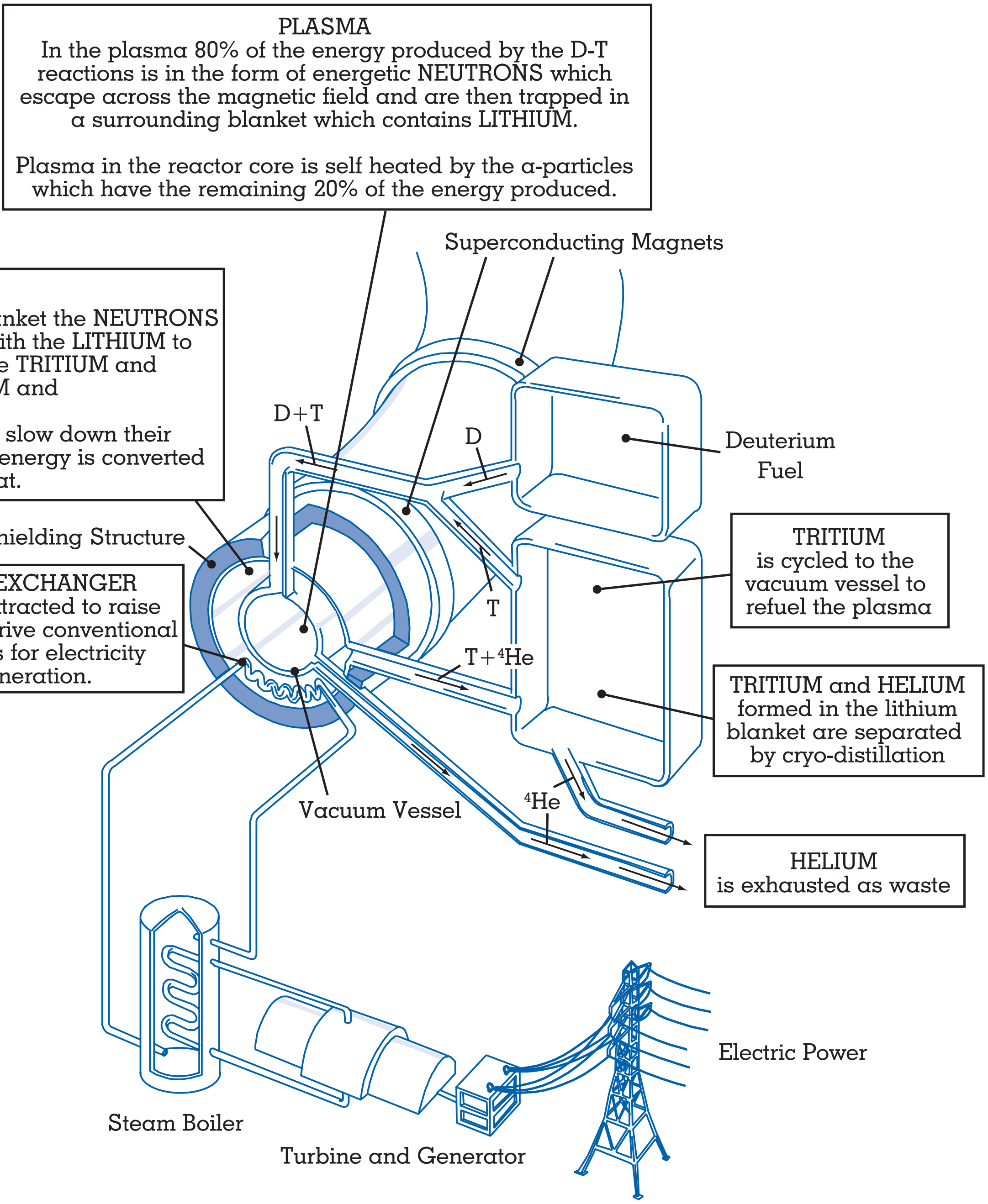


Figure 1.

Schematic of a fusion reactor, assuming a generally toroidal shape of the plasma and magnetic fusion. The principles emphasized are central hot core (red) at 100 million degrees, blanket and heat exchanger, shield, energy conversion, and the handling of D, T, and the "ash" He.

SOURCE: JET Joint Undertaking, Abingdon OX143EA, UK; Graphics Department; JG91.230.

Deuterium occurs naturally, mixed in with plain hydrogen in the tiny proportion of 0.015 percent; in other words, plain hydrogen is the more common isotope by a factor of 6,600. Tritium for fusion energy can be created from another nuclear process involving the interaction of the neutron (in the equation above) with lithium:

$$n + {}^{3}Li^{6} \rightarrow {}^{2}He^{4} + T$$
$$n + {}^{3}Li^{7} \rightarrow {}^{2}He^{4} + T + n$$

with the release of 4.8 MeV in the first reaction and the absorption of 2.5 MeV in the second reaction. The fuel resources for D-T fusion are, therefore, lithium and deuterium. Lithium occurs in brines, mines, and, at the high dilution of only 0.17 parts per million by weight, in sea salts. The resource of lithium in oceans dominates all other sources by a factor of approximately 10,000. And although lithium is plentiful, far fewer lithium atoms than deuterium atoms exist on the Earth making lithium the limiting the resource for D-T fusion.

An essential idea in fusion research is the following: The 3.5 MeV of energy carried by the ${}^{2}He^{4}$, which is also called an α-particle, stays in the fusing region to maintain the high temperature and keep the reaction going. The concept is called "ignition" by analogy to common chemical ignition of fuels in which a small initial spark starts a flame that continues on its own.

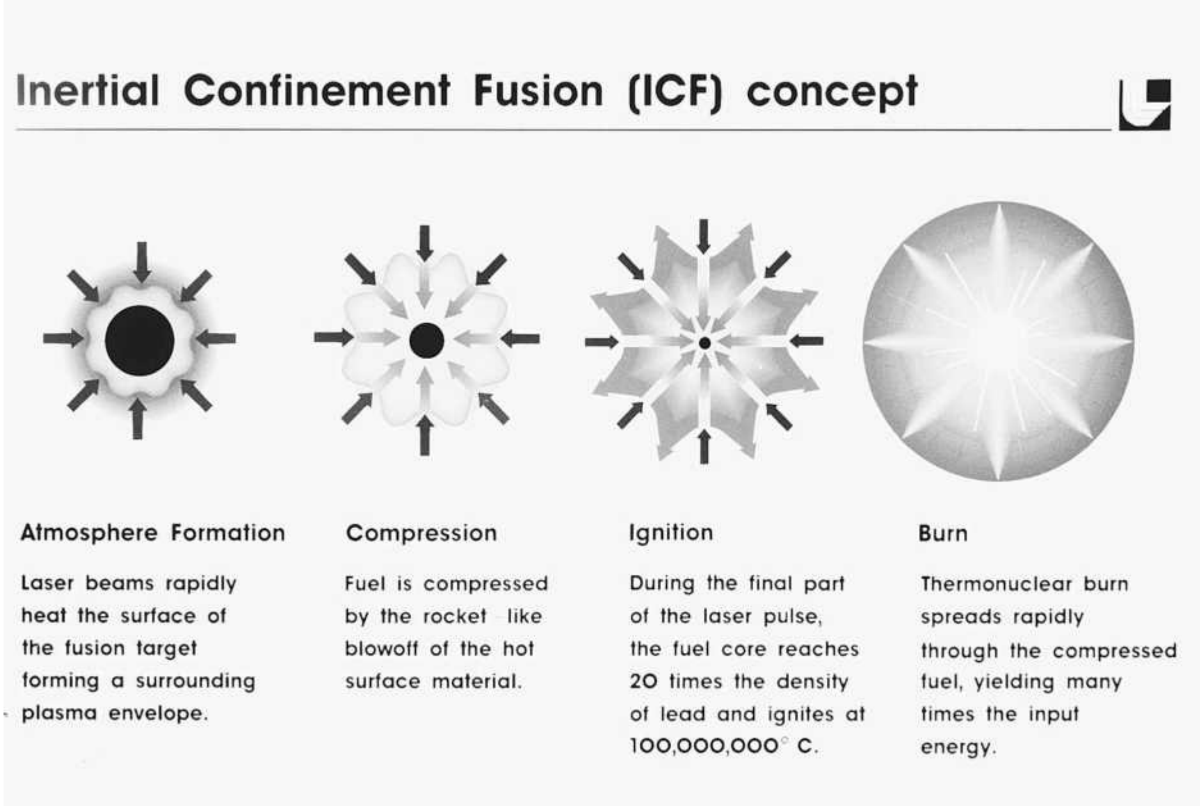
Other fusion reactions such as D plus ${}^{2}He^{3}$ and D plus D (not to mention the H plus H of stars) require far more difficult physical conditions than D plus T, but offer potential advantages in reduced neutron production, and even larger reserves of potential energy in the case of D-D.

CONDITIONS FOR USEFUL ENERGY

D-T fusion in substellar sizes on a human scale requires ultrahigh temperatures, approximately 100 million degrees, for the power produced to exceed the fundamental and irreducible cooling effect of power lost to electromagnetic radiation caused by the particles colliding with each other (bremsstrahlung, or braking radiation). Approximations to data on D-T reactivity lead to a simple and often quoted figure of merit for progress on fusion research, ntT, where n is the fuel particle density, t is an average time that energy stays in the fusion region, and T is the temperature. The best present experiments create ntT approaching the requirements for producing net energy, or 10^{21} m^{-3}-second-keV. (A keV is a kilo electron volt, which is approximately 11 million degrees.) An equivalent statement is that in some present experiments the ratio of fusion energy produced to the energy needed to assemble the hot particles, usually called Q, is near unity. A Q of unity is often called "scientific breakeven," to make clear the contrast with commercial feasibility which requires that fusion energy released be much more than the total energy to operate the machinery in the reactor power station.

The goal of research on nuclear fusion is to produce useful energy in the future as follows: A mixture of D and T is heated to ignition at 100 million degrees, and the fusion energy is released, largely carried by the 14 MeV neutrons produced, streams outwards to a sufficiently thick "blanket" containing lithium and a heat exchange medium, which could in fact be lithium or contain lithium. The neutrons interact with lithium to produce tritium that is retained at the reactor site for later use as fuel. The heat, which comes largely from absorbing the neutron energy in the blanket, is transferred to an external turbine engine to produce electricity with the standard efficiency of 35 to 40 percent. Low temperature heat— constituting of 60 to 65 percent—is not used for electricity. But it can be used for industrial processes, including space heating and refrigeration. The blanket absorbs most neutrons, and a shield outside the blanket absorbs any remaining neutrons and other escaping radiation. The region outside the blanket-shield is then largely free of radioactivity. The blanket, shield, and any material that must remain inside to operate the fusion reactor does become radioactive because of the interaction with the neutrons. The strength and nature of the radioactivity inside the outer surface of the shield has to do with the details of design and the materials used, but remote handling tools will be needed for maintenance. Equipment needed to create and maintain the reacting plasma is outside the shield: for example, electrical power supplies, lasers, ion beams, and superconducting magnets at a few degrees above absolute zero.

At the simplest level there is a remarkable feature of the fusion reactor: a 100 million degree reacting region is near the industrial heat source—the blanket, operating at several hundred degrees. Some kind of insulation must separate the reacting fusion region and the blanket. The two insulation options being explored are



Inertial Confinement Fusion (ICF) concept

Atmosphere Formation

Laser beams rapidly heat the surface of the fusion target forming a surrounding plasma envelope.

Compression

Fuel is compressed by the rocket like blowoff of the hot surface material.

Ignition

During the final part of the laser pulse, the fuel core reaches 20 times the density of lead and ignites at 100,000,000° C.

Burn

Thermonuclear burn spreads rapidly through the compressed fuel, yielding many times the input energy.

Representation of the inertial confinement fusion (ICF) concept. (U.S. Department of Energy)

magnetic fields for “magnetic fusion” or space for “inertial fusion.” The magnetic insulation is much like fiberglass in a walk or ceiling. Magnetic fields slow heat transfer from the motion of plasma particles, much as the fiberglass and trapped air slow the flow of heat through a wall. Space insulation is similar to that obtained in moving back from a blazing fire.

Enhancement of the fusion reaction has been explored in attempts to avoid the extreme requirements for temperature, plus the stringent conditions on density and time. An “atom” of a D or T nucleus and the short lived muon, which is 207 times more massive than the electron, has a much smaller size than a normal D or T atom, so much so that it might be possible to bring the D and T nuclei close enough for fusion at modest conditions. Certain metals can absorb hydrogen so densely that some spontaneous fusioning of D and T might occur. The latter idea is behind the announcement in 1989 of “cold fusion” in an electrolytic cell having platinum and palladium electrodes—an announcement that numerous investigations subsequently proved to be in error. Catalysis, or enhancement of fusion rates has not been pursued as an energy source in recent years.

INERTIAL APPROACH

Inertial confinement fusion has long succeeded in the context of military explosions—the hydrogen bomb. In the military application a fission bomb produces x-rays that drive an implosion of D-T fuel to enormous temperatures and densities such that fusion reactions occur during the short time that inertia keeps the fusing nuclei densely packed and hot.

For civilian purposes the inertial concept is to compress and heat a small sphere of D-T fuel with an external “driver” such as laser light, x-rays produced by lasers, or high energy ion beams. The resulting hot and extremely dense plasma burns the D-T mixture rapid-

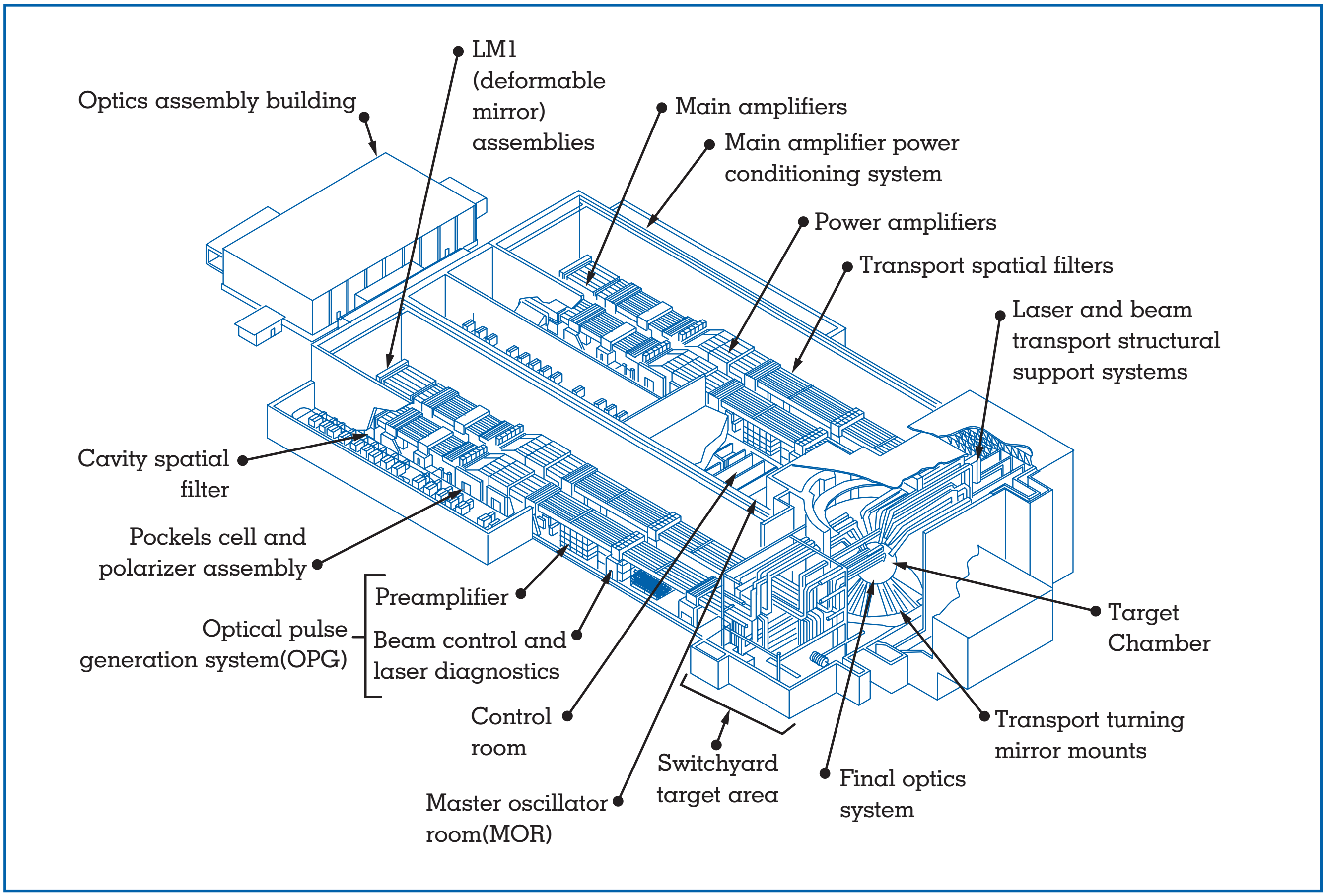


Figure 2.
Layout of the laser and target area building for the National Ignition Facility at the Lawrence Livermore National Laboratory in Livermore California.
SOURCE: LLNL ICF Annual Report 1997, p. 95.

ly before the microexplosion expands appreciably. Time scales are nanoseconds (billionth of a second) for each microexplosion, repeated steadily to produce power. Density scales are thousands times solid density. The mission of inertial fusion in the United States is primarily to address the science of high energy densities for the Defense Department program in stewardship of nuclear weapons. Applications to civilian energy needs are a secondary mission.

Much of the driver energy goes into ablation, or blowing-off the surface of the sphere of fuel to force the compression (implosion) by the rocket effect. As a result, the ntT needed for scientific break-even for inertial confinement is around twenty times higher than for magnetic confinement.

The largest inertial fusion program in the United States is the National Ignition Facility (NIF) now under construction at the Lawrence Livermore National Laboratory (LLNL) in Livermore, California (Figure 2). Completion of the stadium-sized construction project is planned for 2003 at an approximate total cost of $1 billion. The goal is to investigate the interaction between spherical targets and ultraviolet laser light producing x-rays in a “hohlraum”—cavity surrounding the target—at a power of 500 tera watts (trillion watts) for several nanoseconds for a total energy of 1.8 MJ (million Joules). Many other significant installations in the United States and around the world are investigating the potential of inertial fusion.

MAGNETIC APPROACH

Magnetic confinement has been pursued for civilian purposes since the 1950s. The idea is that a magnetic field, some of which might be generated by internal heating currents, slows the transfer of energy from the hot plasma core to the surroundings while extra heating power is applied from particle beams or radio frequency transmissions. Time scales are seconds for the energy replacement time, and indefinitely long for an energy-producing burn; density scales are ten thousands below atmospheric gaseous density.

In about 1968 the Soviet Union (now Russia) under the leadership of Academician Lev Artsimovitch took a large step forward in magnetic fusion with a concept called the "tokamak." The term is a Russian acronym for toroidal (inner-tube shaped) magnetic chamber. Since the mid 1970s the tokamak has dominated worldwide magnetic fusion research, culminating in highly successful D-T burning experiments in the United States with the Tokamak Fusion Test Reactor (TFTR) at Princeton University in New Jersey and the Joint European Torus (JET) in the United Kingdom. Numerous other tokamaks in many countries, including Japan (JT-60), France (Tore Supra) , Germany (ASDEX, TEXTOR), Italy (FTU), and the United States (D-III-D, Alcator-C), confirm and extend the progress while not using tritium.

Two essential ideas behind the tokamak are the externally supplied toroidal magnetic field that closes on itself repeatedly, and a slight twist to the field added by currents inside the reacting region. The closed magnetic field lines in the hot plasma do not strike a wall, so that energy is not instantly conducted to a wall. The twist is essential for good and stable confinement of energy. Prior to the emergence of the tokamak idea, the twist needed in closed, toroidal systems was provided often by external coils in the "stellarator" concept developed at Princeton University by astrophysicist Lyman Spitzer.

Many other magnetic fusion concepts have been pursued over the years: for example, the "magnetic mirror" with open field lines that always strike a wall somewhere but with good stability properties; linear and toroidal "pinches" featuring rapid compression and heating. Today the stellarator is enjoying a significant revival with large programs in

Mechanical Equilibrium and Absolute Trapping of Particles

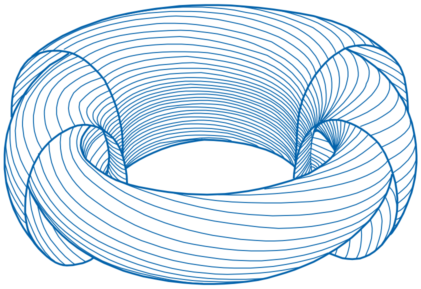
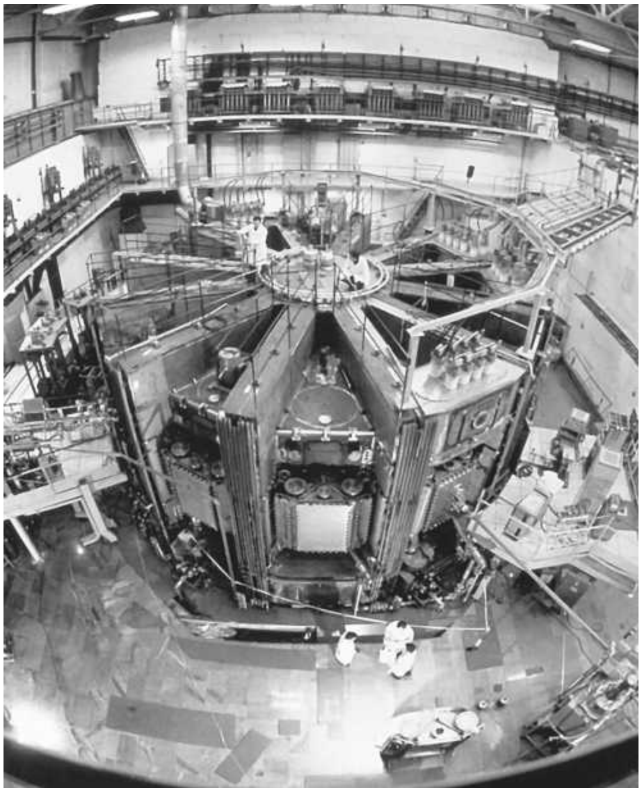


Figure 3.

Schematic representation of the magnetic structure of the Tokamak magnetic confinement device. The lines on the shells represent the direction of the total magnetic field, most of which comes from external coils. The portion that gives the twist, however, comes from current inside the hot plasma itself. The twisting is necessary for stable confinement.

SOURCE: *Transport and Structural Formation in Plasmas* by K. Itoh, S.-I. Itoh, and A. Fukuyama, Institute of Physics Publishing Ltd., 1999. p. 24.



Ring-shaped nuclear fusion research reactor Tokamak 15 at the Kurchatov Institute, Moscow, Russia. (Photo Researchers Inc.)

Japan (LHD) and in Germany (W-7-AS), plus smaller experiments elsewhere. At a basic research level, there is interest in a broad array of magnetic approaches.

Still, the tokamak is getting most of the attention, particularly regarding the matter of a large project to follow JET and TFTR and burn D-T so successfully that the plasma ignites—continues to burn on its own—thereby demonstrating in a convincing manner that with further work, the tokamak can be developed into a commercially viable source of energy. The working name for the ignition project is the International Thermonuclear Experimental Reactor (ITER). In 1998, they completed an integrated engineering design, supported by research and development on crucial components, and a reliable budget and schedule. The term "thermonuclear" refers to the multi-million degree temperatures needed to produce energy from nuclear fusion. The four international parties to ITER—Japan, the European Union, the United States, and the Confederation of Independent States (Russia)—had intended to proceed with cooperatively funded construction at a site to be chosen jointly. The cost estimate of approximately 10 billion and political factors have caused ITER to reexamine the design goals, particularly toward a less costly device, but one retaining the most basic and interesting scientific and technical goals. ITER management hopes to go ahead with construction after the year 2000, and operate ignited plasmas in a decade or so.

For both inertial or magnetic fusion energy the long range, but nonspecific, plan is the following: develop concepts to achieve ignition of D-T, whether in ITER, NIF, or other facilities, and proceed to a demonstration reactor (often called DEMO) and then to commercial implementation. DEMO and commercialization are decades away. Even though the ignition DEMO path is being followed by most people working on fusion, researchers generally agree on the need to continue developing the science of fusion and alternative approaches because uncertainties continue about what will eventually prove to be a successful approach.

SUMMARY

Creating useful energy from fusion on the earth has proved difficult: The forty year effort looks like it still has decades to go. The following inherent potential advantages tend to encourage continued effort: abundant fuel in lithium and virtually inexhaustible fuel in deuterium; absence of the greenhouse emissions of coal, oil, natural gas; no issues regarding the control of a nuclear chain reaction; radioactivity induced by fusion neutrons depends on the design of the device, not the fundamentals of nuclear physics.

If and when projects producing energy on an industrial scale begin to appear in the middle of the twenty-first century, the theoretical advantages will be subject to confirmation. Fusion, magnetic or inertial, might then join the mix of energy resources.

D. W. Ignat

BIBLIOGRAPHY

Bromberg, J. L. (1982). *Fusion: Science, Politics and the Invention of a New Energy Source.* Cambridge, MA: MIT Press.

Fowler, T. K. (1997). *The Fusion Quest.* Baltimore: Johns Hopkins University Press.

Häfele, W.; Holdren, J. P.; Kessler, G.; and Kulcinski, G. L. (1977). *Fusion and Fast Breeder Reactors.* Laxenburg, Austria: International Institute for Applied Systems Analysis.

Heppenheimer, T. A. (1984). *The Man Made Sun: The Quest for Fusion Power.* Boston: Little, Brown and Company.

Herman, R. (1990). *Fusion: The Search for Endless Energy.* New York: Cambridge University Press.

Huizenga, J. R. (1992). *Cold Fusion: The Scientific Fiasco of the Century.* Rochester: University of Rochester Press.

Lindl, J. D. (1998). *Inertial Confinement Fusion: The Quest for Ignition and Energy Gain Using Indirect Drive.* New York: Springer-Verlag.

McPhee, J. (1974). *The Curve of Binding Energy.* New York: Farrar, Straus and Giroux.

Niu, K. (1989). *Nuclear Fusion.* New York: Cambridge University Press.

Rhodes, R. (1995). *Dark Sun: The Making of the Hydrogen Bomb.* New York: Simon & Schuster.

Serber, R. (1992). *The Los Alamos Primer: The First Lectures on How to Build an Atomic Bomb.* Berkeley: University of California Press.

Sheffield, J. (1994). "The Physics of Magnetic Fusion Reactors." *Reviews of Modern Physics* 66:1015–1103.

NUCLEAR WASTE

KINDS OF NUCLEAR WASTE

Nuclear waste is radioactive material that has no immediate use. In the United States nuclear waste generally is divided into two main categories: high-level nuclear waste (HLW) and low-level radioactive waste (LLW). Two less common categories are tailings from uranium mining and milling, and a special category derived from particular aspects of nuclear weapons production and defense-related activities. This latter, less commonly discussed, category is called defense wastes and, because of its makeup, is sometimes called transuranic wastes; it makes up about half of the HLW in the United States. This article focuses on HLW and LLW from nuclear power reactors.

HLW comprises most of the radioactivity associated with nuclear waste. Because that designation can cover radioactive waste from more than one source, the term spent nuclear fuel (SNF) will be used to discuss HLW originating from commercial nuclear reactors. LLW comprises nearly 90 percent of the volume of nuclear waste but little of the radioactivity. Nuclear power reactors produce SNF and most of the nation's LLW, although there are approximately 20,000 different sources of LLW. The name SNF is a bit of a misnomer because it implies that there is no useful material left in the fuel, when in fact some fissionable material is left in it.

CONCERN FOR RADIOACTIVE WASTES

Nuclear waste is a concern due to the levels of radioactivity in it, especially in HLW. Relative to the wastes associated with other major methods of power production, the volumes are small. The concern is for the amount of radioactivity and the time of its duration. The time of radioactivity duration is measured in half-lives (the time required for half of the atoms of a given substance to disintegrate, at which time it becomes the new starting amount with which to begin the count again). Generally LLW not only has lower levels of radioactivity, it also has very low concentrations of long-lived radioactive substances. There are short-lived and long-lived substances in nuclear waste, with half-lives varying from seconds to thousands of years.

In the United States the federal government has taken responsibility for HLW, for both SNF and defense wastes, and for mill and mine tailings. Remediation of the effects of such tailings is well under way. Each state or groups of states (compacts) are responsible for the safe isolation of their own LLW. Three departments of the federal government have prime responsibility for matters related to nuclear waste: Department of Energy (DOE) and its predecessor agencies, which created most of the defense wastes; the Nuclear Regulatory Commission (NRC); and the Environmental Protection Agency (EPA). Regulations regarding care of nuclear wastes are also set by individual state agencies, especially for the LLW.

The federal government not only has responsibility for the safe isolation of SNF but also taxes on the generation of electric power to cover the cost of waste isolation at the rate of one mil/kWh ($0.001/kWh). Although the nuclear waste fund has been folded into the general federal budget, it comes from a special levy that can be changed to accommodate the needs of SNF handling. The office that manages this fund and that is responsible for SNF is in the U.S. DOE and is called the Office of Civilian Waste Management (OCRWM).

ORIGINS OF THE RADIOACTIVITY IN SPENT NUCLEAR FUEL

The nuclear waste generated from production arises when uranium atoms in the fuel split. Nuclear power reactors are fueled with assemblies of rods containing pellets, each about the size of a pencil eraser. About 3 percent of the uranium in these pellets is the fissionable isotope, or form, of uranium, U-235. This is the rarer of the two naturally occurring isotopes of uranium. In nature, the U-235 isotope is found in less than 1 percent of uranium ore—usually 0.5 to 0.7 percent of that ore. After uranium ore is mined and processed, the isotopes are separated, in a process called uranium enrichment. This is not a simple task and is mostly done by a gaseous diffusion plant. Usually the process is stopped when the mixture is about 3 to 3.5 percent of the fissionable form in the mix. (The majority uranium, U-238, is radioactive in its own right, emitting a short-range alpha particle, but it is not fissionable. When the fissionable isotope is extracted from uranium, the remaining material is called depleted uranium and has its own other uses, which will be addressed below.) The exact makeup of the reactor fuel varies some what but is mostly in the form of uranium dioxide, with the uranium being about 3 percent fissionable U-235.

There are three main products in the fissioning of uranium-235: tremendous amounts of energy, neutrons that perpetuate the chain reaction, and two new atoms. The splitting is such that eventually every element found on the periodic chart of the elements is made, especially in the flux of speeding neutrons that induced and perpetuated the fission. Every isotope of every element can be found in the waste. Many of the new isotopes produced are highly unstable—that is, they are radioactive. Many decay in seconds, but some exist in their unstable form much longer, eventually decaying into stable isotopes, though often with intermediates that are also radioactive. This results in radioactive decay chains. These chains of radioactive decay mean that SNF will be radioactive for a long time. It has been estimated that it will be approximately 7,000 years before the level of radioactivity in SNF drops to that of natural radioactivity, that of Earth itself. Since some parts of Earth will be radioactive forever, it can be said that SNF will be radioactive forever. In practice, one must decide what levels of radioactivity to call dangerous. That debate influences decisions about nuclear waste isolation. The debate about what is "safe" is not part of this particular topic. The amended version of the Nuclear Waste Policy Act (NWPA) calls for the safe isolation of HLW from the human environment for 10,000 years.

Besides fission products, the various forms of known but newly formed elements in the spent nuclear fuel, there is a small but significant amount of fissionable, or fissile, material in the SNF. This is quite important. There is some unused, unfissioned U-235 that has become too dilute to use. Like natural uranium ores in which chain reactions do not

Solid transuranic interim waste storage at the U.S. Department of Energy's Idaho National Engineering Laboratory in Idaho Falls, Idaho. (U.S. Department of Energy)

occur, the fuel will not sustain a chain reaction. But new fissile material also has been made in the intense neutron irradiation. Elements are made that are heavier than naturally occurring uranium, and because of their location on the Periodic Chart are referred to as transuranics. There are isotopes of plutonium and uranium made that can fission. None is present in concentrations that are great enough for a self-sustained chain reaction, but they do represent new fuel. The concentrations of U-239, Pu-239, Pu-240, Pu-241, and some short-lived intermediates are 1–2 percent of the SNF. Some of these are fissile material. Breeder nuclear reactors can be designed and run to produce significant amounts of these isotopes. "The SNF from a breeder reactor is rich in newly produced fissionable isotopes but it must undergo extensive reprocessing to become new reactor fuel. That reprocessing will adjust the concentration of fissionable isotopes and eliminate some of the fission products that tend to quench fission process."

HANDLING OF SNF

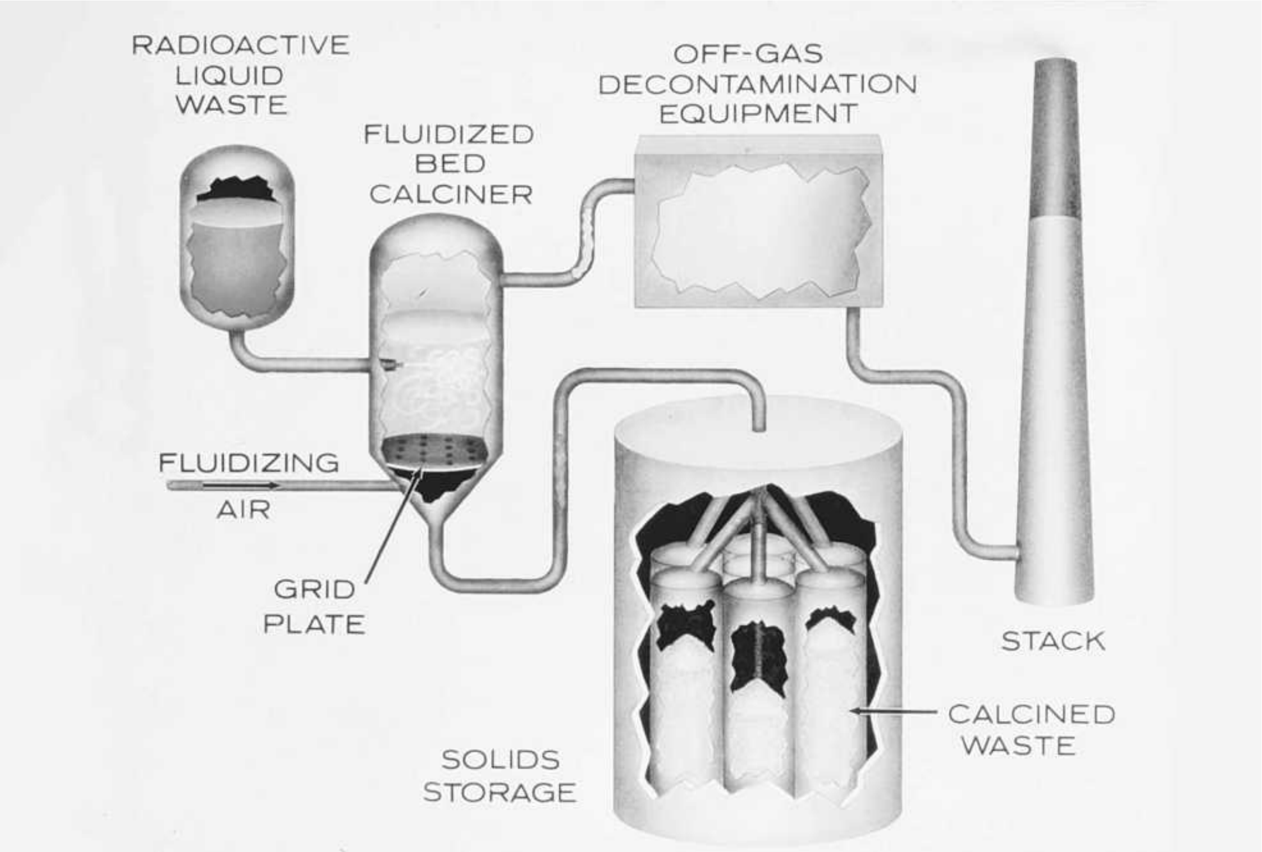
Workers wearing gloves can handle nuclear fuel that is being freshly installed into a nuclear reactor. This fuel assembly becomes dangerously and highly radioactive SNF after a short time in an operating nuclear reactor. Upon removal from a reactor (after about 18 months), a spent fuel assembly is now HLW and is far too "hot" thermally and radiologically to handle directly. It is removed remotely with a crane and stored in a bay of cooling water beside the reactor.

HLW still looks like a fuel assembly, a collection of long, skinny rods, each filled with fuel pellets, held in a rack that allows water to pass through and pick up thermal energy. The assembly is kept underwater to cool and also to shield the workers from the longer-range gamma radiation. The water also contains corrosion-inhibiting chemicals as well as chemicals to absorb the few neutrons emitted from lingering atoms of U-235 and other fissile material. Generally a ten-year cooling-off period allows that much of the heat to be emitted and the radioisotopes with short half-lives to have undergone decay. For power plants with limited pool storage there are two choices when the cooling pools are filled: rod consolidation and dry cask storage.

Rod consolidation can permit space saving by dismantling the fuel assembly racks. The rods are still maintained underwater. Rod consolidation is not a routine practice due to issues of heat dissipation and criticality, the possibility of a chain reaction continuing. Dry cask storage is the more common approach for the oldest fuel assemblies and calls for removal of the fuel assemblies from the cooling pool. Designs of the casks vary in detail, but they usually accommodate several spent fuel assemblies in a sealed steel container that is enclosed in a concrete box and/or metal canister. Usually these containers are stored and monitored at the plant site. As with all major operations at nuclear power plants in the United States, hearings before the NRC and the public are conducted on temporary SNF storage. Almost all dry cask storage occurs at the power plant site, where monitoring of the containers can be routine.

ULTIMATE ISOLATION OR REPROCESSING OF SNF

The ultimate disposition of HLW or SNF is a matter of significant importance and is controversial for some. In the United States all HLW, including SNF, has always been a federal responsibility, beginning with the Atoms for Peace Program in the 1950s. A



Cutaway view of the new waste calcining facility (NWCF) process. (U.S. Department of Energy)

study done by the National Academy of Sciences in 1957 indicated that nuclear waste disposal would not provide insurmountable technical problems and recommended deep geological isolation. Every country that has nuclear power plants has considered the safe isolation or disposal of HLW and has concluded that deep geologic isolation is the safest method. In the United States eight options, including deep-ocean-trench burial, space launches, Antarctica burial, and transformation into materials with shorter half-lives have all been studied. The media into which the HLW is to be isolated varies widely from country to country and include granite, salt beds, and tuffaceous material, welded volcanic ash, most commonly called tuff. United States attention is focused on Yucca Mountain, at the western edge of the Nevada Test Site, the location of above-and below-ground nuclear weapons testing in the past. It is a mountain of tuff.

YUCCA MOUNTAIN, NEVADA

The amended version of the NWPA has determined that only Yucca Mountain is to be characterized as a possible geologic repository for SNF and possibly some defense waste. It is approximately 100 miles northwest of Las Vegas, Nevada. The planned location of the SNF is approximately 1,000 feet above any groundwater and about the same distance below the surface. Characterization studies of the viability of Yucca Mountain are under way and have been since the early 1990s. The NWPA also prohibits studies of granite as a host rock. (Canada is considering SNF isolation in granite. Most of the eastern half of the United States is on granite.) Both private and governmental leaders—most concerned about tourism—in the state of Nevada are resisting the Yucca Mountain characterization studies. Furthermore, although the majority of scientists studying the mountain find no clear problems, a few find fault with the studies ad the site as a possible repository. The most difficult part of the studies involves predictions about the behavior of the tuff environment for 10,000 years and beyond. The fact that Yucca Mountain is in a region of ancient volcanic activity complicates the assessments. Most attention is devoted to water movement through the tuff and the vulnerability of the underlying earth to undergo any significant "stretching" in the foreseeable future. The Yucca Mountain project has a website listed in the bibliography for this article, with photos and links to scientific data being accumulated as well as links to opposition points of view.

Yucca Mountain, if it becomes the site for the isolation of SNF, will be laced with tunnels, waste in storage casks and monitoring equipment. A waiting period is planned while better isolation alternatives are sought. If Yucca Mountain is not used, it is to be refilled with the tuff material removed earlier. In the United States the SNF that would be isolated in Yucca Mountain would be waste that has not been reprocessed; it would be material that has come out of nuclear reactors and has been cooled at the plant site.

WASTE REPROCESSING

Since the amount of fissile material in the fuel assemblies is only about 3 percent of the uranium present, it is obvious that there cannot be a large amount of radioactive material in the SNF after fission. The neutron flux produces some newly radioactive material in the form of uranium and plutonium isotopes. The amount of this other newly radioactive material is small compared to the volume of the fuel assembly. These facts prompt some to argue that SNF should be chemically processed and the various components separated into nonradioactive material, material that will be radioactive for a long time, and material that could be refabricated into new reactor fuel. Reprocessing the fuel to isolate the plutonium is seen as a reason not to proceed with this technology in the United States.

Congress has decided that reprocessing will not be practiced in this country so that we will not be in the plutonium production business. This seems like a safe thing to do since this action will minimize terrorism threats. Reprocessing generates chemi cal wastes but greatly reduces the volume of the highly radioactive waste. It also isolates plutonium and unused fuel for possible use as new fuel. Reprocessing means that the volume of material calling for long-term isolation is reduced by one-third what it would have been before reprocessing.

THE VOLUME OF SNF

If one just concentrates on the radioactive material in SNF, the volume is very small, especially compared to waste from other power production practices. However, one can only discuss the separated radioactive material if it has undergone extensive reprocessing. If SNF is to be isolated, as in a place such as Yucca Mountain, with perhaps 70 miles of tunnels, the volume is that of the interior of this minor mountain. Isolation of up to 100,000 metric tons of SNF in Yucca Mountain means that for the United States, approximately all the SNF made to date and that expected in the operating lifetime of all current reactors can be put there. Approximately 2,000 metric tons of SNF are produced each year in the United States. Waste volume and placement depend on the amount of compaction and consolidation at the sites. The plans for the Yucca Mountain present a realistic and understandable picture of the volume of SNF.

A useful perspective may be seen about the amount of SNF by comparing it to the volume of other waste streams associated with power production. More than half of the electricity made in the United States comes from coal burning, where each large power plant generates bottom and fly ash in volumes measured in acre-feet annually. Each of these plants generates its own small mountain of ash. The gaseous wastes from coal burning and from methane and oil burning result in tons of carbon dioxide daily. This carbon dioxide is an infrared active gas and is thought by many to be contributing to global warming and climate change phenomena. The wastes associated with nuclear power are small in comparison, this is not surprising, considering the tremendous power in the nucleus of an atom.

LOW-LEVEL RADIOACTIVE WASTE

The other major category of nuclear waste, LLW, is generally that which is neither HLW nor the transuranic part of defense waste. Generally, LLW is generated in private and public labs, hospitals, and commercial enterprises and can involve lab clothing, paper, packing material and radioactive isotopes used in medical procedures, both diagnostic and therapeutic. Significant amounts of LLW are associated with defense wastes, but mostly are isolated at U.S. DOE facilities. Many pharmaceutical products require extensive use of radioactive tracers. Academic sources are minor. The major sources of LLW in volume and amount of radioactivity are nuclear power plants. Aside from the decommissioning of nuclear power plants, the volume of LLW is decreasing every year. This is because new technologies have become available to separate the radioactive material from that which is not radioactive. Also, generators of LLW are being more selective in the work that they do so that they generate less LLW each year. These changes have become about partly because the cost of isolating LLW has increased.

The LLW from nuclear power plants contains ion exchange resins as well as clothing, tools, and chemicals. Ion exchange resins, which comprise the majority of this LLW, are used to filter the water circulated in nuclear power plants. The ion exchange resins isolate and trap dissolved materials, much of which can be radioactive. Approximately three-fourths of commercially or privately generated LLW is in the form of the contaminated plastic beads that make up ion exchange resins.

The NRC categorizes LLW into Class A, B, or C. The wastes in Classes A and B contain materials that decay to safe levels in about 100 years. Class C wastes will require about 500 years to reach safe levels and contain mostly ion-exchange resins and filters. (There is a minor amount of LLW referred to as beyond class C material.) The majority of LLW is dry and is stored in cardboard or wooden crates. Some are in metal drums. LLW is most likely given shallow burial, compared to the deep geological isolation required for HLW. A difficult issues in LLW disposed involves a small amount of "mixed waste," where the radiological material is mixed with chemically or biologically hazardous substances. The contaminated chemicals may require special handling, since hazardous chemical wastes do not decay as do radioactive materials. In addition, a small amount of LLW consists of animal carcasses or waste and also needs special handling. Currently there are three commercial sites handling LLW for most of the states and compacts. The largest volume of LLW will come with the decommissioning and dismantling of nuclear power plants. In 1995 nearly 700,000 cubic feet of LLW were handled at commercial sites in the United States.

LLW is placed in sturdy, sealed containers. At the isolation site these containers are first put in concrete and/or metal vaults or bunkers and then buried in shallow trenches before being covered with backfill. Some of these are then paved over to prevent rainwater from entering the waste.

The LLW Forum coordinates information and regulations among and between the various state compacts and states. It maintains a website that is overseen by the Idaho Operations Office, which is part of the National Engineering Lab in Idaho. Although LLW is a responsibility of each compact, or of state for the LLW generated in that state, the regulations governing it are those of the U.S. DOE, the NRC, and the EPA. The regulations governing its isolation allow individual states to set additional requirements for handling the material. After being established, any particular compact can refuse to receive the LLW from other states.

OTHER RADIOACTIVE-WASTE SITES

The Waste Isolation Pilot Plant (WIPP) is in an excavated salt cavern in southern New Mexico, twenty-seven miles from Carlsbad. The WIPP site is 2,000 yards underground, and defense waste is being placed. There are plans to place there about 6 million cubic feet of material there containing fewer than five million curies of radio activity.

The most technically difficult category of HLW is that belonging to the U.S. DOE at its major plutonium production facilities in Hanford, Washington, and at the Savannah River facility near Aiken, South Carolina. Just sampling the material for characterization presents problems since much of it consists of radioactive acidic liquids. The rest of it in sludge form, and some is solid. A plant for the vitrification of that part of the material that can be made into glass bars is in operation at the Savannah River facility. The vitrification process encases the waste in glass "logs" to immobilize it and make it easier to handle. A conservative estimate is that it will take decades to clean up these facilities.

NUCLEAR WASTE IN OTHER COUNTRIES

The United States has the most radioactive nuclear waste and the most complicated array of waste types. Reprocessing of SNF is also practiced in some countries. Although costly, this practice reduces the volume of HLW requiring deep geologic disposal. (In the wide variety of elements in the fission products making up SNF, some are toxic heavy metals.) Reprocessing and the general lack of economy of scale in many other countries help to explain why there is less activity or progress related to HLW isolation elsewhere. Interim monitored retrievable storage is generally the focus of activity abroad. Different philosophies and cultures also mean that a wide variety of approaches to the problem are found. Some feel that the generation deriving the benefits of nuclear power is the one responsible for a solution. Others say that succeeding generations may have better technology or ideas about this problem than current generations: Current generations should not do anything permanent, committing future generations to the solutions seen by people living today. Among the technology that wold drastically reduce the cost of cleanup of waste sites robotics.

All the countries that produce nuclear waste have chosen the same alternative for the ultimate disposition of HLW, deep geological isolation, and they did so independently of one another. The United States has the most radioactive nuclear waste and the most complicated array of waste types of any "nuclear" country. Only in the United States can one find the same economy of scale for waste handling. Thus, it leads the world in most activities aimed at safe isolation. In France, Japan, and Great Britain, however, reprocessing is routinely practiced. Those countries reprocess HLW for many other countries. As mentioned above, reprocessing is not currently allowed in the United States.

France and Germany, local protests have dramatically slowed the choice of a final isolation site. As a result, interim storage is widely practiced in other countries. In Sweden, the waste is mixed in molten copper. The radioactive waste is immobilized in copper logs that are easily handled and stored in large underground caverns until a more permanent isolation is chosen. While most feel that those making the waste have the primary responsibility for its isolation, in Sweden they do not wish to commit future generations to solutions that might be vitrified; the glass "logs" are in concrete bunkers.

In the former U.S.S.R. vast areas of the country are contaminated by poor handling of nuclear waste, especially from that associated with the manufacturer of weapons. Some radioactive waste, especially from nuclear submarines has been isolated at the bottom of the Baltic Sea by the former Russian Navy.

CONCERNS

Transportation of HLW is among the most immediate concerns for the general public. Most experts agree that nuclear wastes, especially SNF, are being kept in places not intended for long-term storage and they acknowledge that the problem must be addressed. There is general agreement that SNF should be kept far from population centers (where the electricity is generated and used) and should be kept in a dry place, since water the most likely medium for its movement into the human environment. Movement of SNF is necessary, most likely by rail or roadway. Groups of U.S. DOE and state regulators are working together, continually revising transportation plans and designing and testing transportation casks. Some wish to see a combined transportation and burial cask, while others want very different qualities in each sort of cask. Planners build on past experience handling and transporting radiological materials. Extensive experience was gained when the first commercial nuclear power plant, at Shippingport, Pennsylvania, was decommissioned in 1989 and the reactor vessel was moved to Hanford, Washington, where it is buried.

Despite the challenges, many see nuclear waste issues as being mostly political and social. There is a growing awareness that technical answers alone will not solve the political and social concerns. Some consider nuclear waste issues small in comparison to the volume and challenges associated with other kinds of wastes, whether generated by power plants or other human activities.

Donald H. Williams

See also: Environmental Problems and Energy Use; Government Agencies; Nuclear Energy; Nuclear Energy, Historical Evolution of the Use of; Nuclear Fission; Nuclear Fission Fuel; Nuclear Fission.

BIBLIOGRAPHY

American Nuclear Society. <http://ans.org>.

American Nuclear Society. (annual). *High Level Radioactive Waste Management: Proceedings for the International Topical Meeting of the American Nuclear Society and the American Society of Civil Engineers.* Chicago: American Nuclear Society.

Murray, R. L. (1994). *Understanding Radioactive Waste*, 4th ed. Columbus, OH: Battelle Press.

Physics Today. (1998). 50(6).

Radwaste Magazine: Progress in Radioactive Waste Management and Facility Remediation. (monthly).

Savage, D., ed. (1995). *The Scientific and Regulatory Basis for the Geological Disposal of Radioactive Waste*. New York: Wiley & Sons.

Study Committee Home Page. <http://www.studycommittee.org>.

U.S. Department of Energy. (1996) "Integrated Data Base Report 1995: U.S. Spent Nuclear Fuel and Radioactive Waste Inventories, Projections and Characteristics, DOE/RW-0006, Revision 12." Washington, DC: Author.

U.S. Department of Energy: Office of Civilian Radioactive Waste Management. *The Yucca Mountain Project*. <http://www.ymp.gov>.

U.S. Environmental Protection Agency. *Yucca Mountain Home Page*. <http://www.epa.gov.radiation/yucca>.

O

OAK RIDGE NATIONAL LABORATORY

See: National Energy Laboratories

OCEAN DRILLING

See: Oil and Gas, Drilling for

OCEAN ENERGY SYSTEMS

Two-thirds of Earth's surface is covered by oceans. These bodies of water are vast reservoirs of renewable energy. In a four-day period, the planet's oceans absorb an amount of thermal energy from the sun and kinetic energy from the wind equivalent to the world's known oil reserves. Several technologies exist for harnessing these vast reserves of energy for useful purposes. The most promising are ocean thermal energy conversion (OTEC), wave power plants, and tidal power plants. All of these produce electricity from the oceans' reserves of renewable energy. Because the ultimate source of energy from the oceans is solar radiation (or the gravitational force of the sun and the moon in the case of tidal energy), ocean energy systems are renewable, have no fuel costs, and are relatively nonpolluting when compared to conventional sources of energy, such as fossil fuels. To date, the technologies for harnessing the oceans' energy on a large scale are still in the early stages of development and have high initial costs, making them more expensive than conventional alternatives.

OCEAN THERMAL ENERGY CONVERSION

Each day, tropical oceans absorb the energy equivalent to 250 billion barrels of oil. If less than one-tenth of 1 percent of this energy could be converted into electricity it would supply twenty times the amount of electricity, consumed daily in the United States. Unfortunately, this energy is spread out over 23 million square miles of ocean, providing a large volume of slightly heated water.

Ocean thermal energy conversion (OTEC) power plants generate electricity by exploiting the difference in temperature between warm water at the ocean surface and colder waters found at ocean depths. To effectively capture this solar energy, a temperature difference of 35°F or more between surface waters and water at depths of up to 3,000 feet is required. This situation can be found in most of the tropical and subtropical oceans around the world that are in latitudes between 20 degrees north and 20 degrees south.

The 35 degree temperature difference is necessary because extracting a little bit of heat from a large volume of water is inherently inefficient. In order to harness thermal energy, two reservoirs of heat at different temperatures are required. In such a "heat engine," heat from the high-temperature reservoir (high-grade energy) flows to the low-temperature reservoir (low-grade energy).

The maximum possible efficiency at which a heat engine can work is defined by the Carnot efficiency equation $E = (T2–T1)/T2$, where E is the efficiency of the heat engine, T1 is the temperature of the cold reservoir (Kelvin), and T2 is the temperature of the hot reservoir (Kelvin). Given tropical surface water temperatures of 80°F (300 K) and cold deep ocean temperatures of 40°F (278 K), this yields a maximum theoretical efficiency of about 7 percent. In comparison, the temperature differences found in conventional steam turbines result in maximum theoretical efficiencies of around 60 percent.

An OTEC facility, like any electric power plant, must use energy to run pumps and other electrically driven devices. It requires large amounts of electric energy to pump huge amounts of water from great depths. When the typical operating losses associated with a real-world OTEC power plant are taken into account, it is a challenge for engineers to design a facility that will be a net producer of electricity.

The concept of OTEC was first envisioned by the French physicist Jacques-Arsene d'Arsonval, in 1881. The first working OTEC system was built in Cuba in 1930, by the physicist Georges Claude, who also invented the neon lamp.

OTEC power plants can be located either onshore or at sea. The electricity generated can be transmitted to shore by electrical cables, or used on site for the manufacture of electricity-intensive products or fuels (such as hydrogen). For OTEC plants situated on shore to be economical, the floor of the ocean must drop off to great depths very quickly. This is necessary because a large portion of the electricity generated by an OTEC system is used internally to pump the cold water up from the depths of the ocean. The longer the cold water pipe, the more electricity it takes to pump the cold water to the OTEC facility, and the lower the net electrical output of the power plant.

There are three potential types of OTEC power plants: open-cycle, closed-cycle, and hybrid systems. Open-cycle OTEC systems exploit the fact that water boils at temperatures below its normal boiling point when it is under lower than normal pressures. Open-cycle systems convert warm surface water into steam in a partial vacuum, and then use this steam to drive a large turbine connected to an electrical generator. Cold water piped up from deep below the oceans surface condenses the steam. Unlike the initial ocean water, the condensed steam is desalinated (free of salt) and may be collected and used for drinking or irrigation.

Closed-cycle OTEC systems use warm surface waters passed through a heat exchanger to boil a working fluid that has a low boiling point, such as ammonia or a chlorofluorocarbon. The vapor given off is passed through a turbine/generator producing electricity. Then cold deep ocean water is used to condense the working fluid and it is returned to the heat exchanger to repeat the cycle. Hybrid OTEC systems combine both technologies and produce both electricity, with a closed-cycle system, and fresh water, with an open-cycle system.

Unlike electrical generation from most other forms of renewable energy that vary with weather and time of day, such as solar and wind energy, OTEC power plants can produce electricity 24 hours per day, 365 days per year. This capability makes OTEC an attractive alternative to conventional base-load electric power plants powered by fossil fuels or nuclear fission. Fresh water production is just one of the potentialy beneficial by-products of OTEC. The cold deep ocean water can also be used for aquaculture (fish farming), as it is pathogen free and nutrient rich. It can also be the source of air conditioning and refrigeration in nearby buildings.

OTEC power plants can have some negative impacts on the natural environment, but overall they are a relatively clean source of electricity compared to conventional options. Cold water released at the oceans surface releases trapped carbon dioxide, a greenhouse gas, but emissions are less than 10 percent of those from a fossil fuel power plant producing a similar amount of power. Also, discharging the cold water at the oceans surface could change local concentrations of nutrients and dissolved gases. This can be minimized by discharging the cold water at depths of greater than 200 feet and/or by using the cold water for air conditioning or refrigeration before it is released.

Despite the fact that OTEC systems have no fuel costs and can produce useful by-products, the initial high cost of building such power plants (up to $5,000 per kilowatt) currently makes OTEC generated electricity up to five times more expensive than conventional alternatives. As such, at the present time OTEC systems are largely restricted to experimental and demonstration units. One of the major facilities for OTEC research is the Natural Energy Laboratory of Hawaii Authority at Keahole Point on the island of Hawaii. An experimental OTEC facility located there has had a maximum *net* power production of 100 kilowatts, at the same time producing 5 gallons per minute of desalinated water.

The most promising markets for OTEC are more experience in building OTEC power plants and stan-

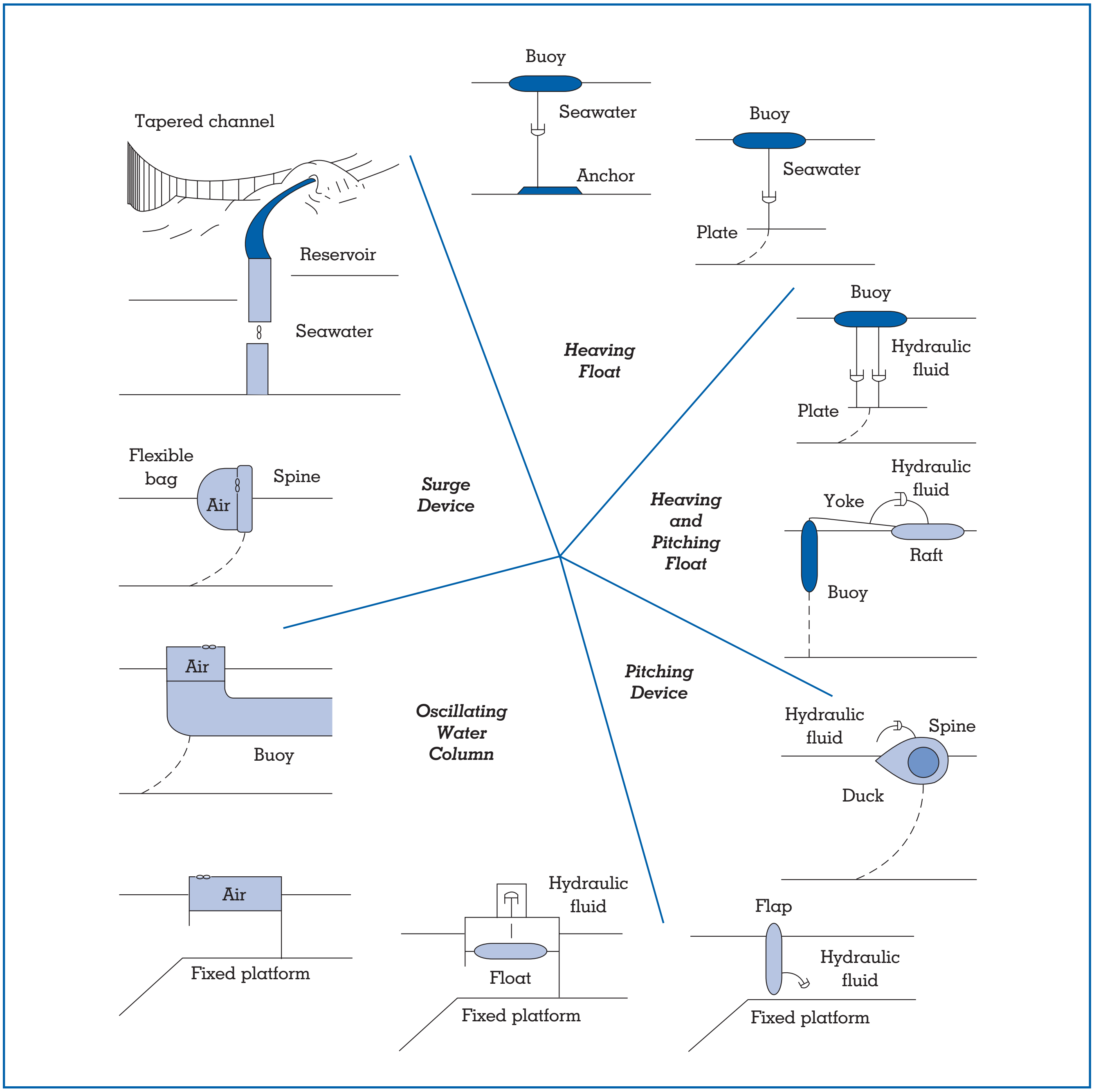


Figure 1.
Symbolic representation of various types of wave energy devices.

dardized plant designs could bring OTEC costs down in the future. However, the initial high costs, coupled with the risk of storm damage to expensive offshore or coastal OTEC facilities, may mean that OTEC electricity generation will never be competitive with conventional sources of electricity

WAVE POWER

Winds blowing across the surface of the world's oceans are converted into waves. The total amount of power released by waves breaking along the world's coastlines has been estimated to be 2 to 3 million megawatts, equivalent to the output of about 3,000 large power plants. Although this vast amount of energy is spread out along thousands of miles of coasts, in favorable locations, the energy density can average 65 megawatts per mile of coastline, an amount that can lead to economical wave-generated electricity. As of 1996, 685 kilowatts of grid-connected wave-generating capacity was operating worldwide.

Wave size is determined by wind speed and “fetch,” the distance over the oceans surface which the wind travels. Favorable wind energy sites are generally western coastlines facing the open ocean such as the Pacific Coast of North America and the Atlantic Coast of Northern Europe. Norway, Denmark, Japan, and the United Kingdom are the world leaders in wave energy technologies.

Although a wide variety of devices have been constructed to capture wave energy, commercial electrical generation from wave power is derived from one of three general types of devices: surface-following devises, oscillating water columns, and focusing, or surge devices.

Surface-following devices use a mechanical linkage between two floating objects or between a floating and a fixed object to produce useful mechanical power. This mechanical power can either be connected directly to a generator or transferred to a working fluid, such as water or air, which drives a turbine generator.

One high-efficiency-surface following device that has been tested is the “Salter duck,” named after its developer Steven Salter at the University of Edinburgh. It looks like a series of floating ducks that pivot about a stiff shaft; this pivoting motion drives a hydraulic fluid to produce electricity. Tests on the Salter ducks show that they are capable of capturing 80 percent of the energy carried by incoming waves.

Oscillating water columns (OWC) use the force of waves entering a fixed device to perform work or generate electricity. The simplest examples of these are navigational buoys; waves entering the anchored buoy compress air in a vertical pipe. This compressed air can be used to blow a warning whistle, or to drive a turbine generator producing electricity for a light. Japan has installed hundreds of OWC-powered navigational buoys along its coastlines since 1965. Larger OWC power plants include the United Kingdom’s experimental OSPREY (Ocean-Swell Powered Renewable Energy), a 2-megawatt offshore facility designed so that waves enter a fixed, submerged chamber, pushing air through a turbine with the rise and fall of each wave.

Focusing, or surge, devices are humanmade barriers that channel and concentrate large waves into a small area, drastically increasing their height. The elevated waves are channeled into an elevated reservoir. This water then passes through hydroelectric turbines on its way back to sea level, thus generating electricity. A 350-kilowatt grid-connected power plant has been operating on the North Sea coast of Norway since 1986. It uses a “tapered channel” design to focus the waves. The tapered channel system uses a narrowing concrete channel to funnel waves up into a reservoir located above sea level. This water then flows through a turbine on its way out, thereby generating electricity.

One of the problems associated with wave power plants is that, during severe storms, the energy unleashed by breaking waves can be over ten times that of the average wave for a coast. The original OSPREY power plant was destroyed by a storm shortly after its construction. The need to design wave power plants to survive such power adds a great deal to the cost of wave energy facilities. Although focusing devices are the most expensive of the three types of wave power plants to construct, their robust nature makes them less susceptible to damage by storms. The Norwegian Tapered Channel design has proven to be the most durable of any wave energy plants. In the late 1990s, this led to the production of two commercial multimegawatt power plants based on this design-one in Java, Indonesia, and one on King Island located between Tasmania and the Australian mainland.

Wave energy power plants consume no fuel during their operation and, as such, do not emit harmful pollutants. However, large-scale wave energy facilities can have an impact on nearby coastal environments. Offshore wave energy facilities such as surface followers or OWCs reduce the height of waves reaching the shore, potentially changing patterns of erosion and sedimentation. Focusing devices result in more erosion where the waves areconcentrated and more sedimentation in adjacent areas. The effect that these changes in erosion and sedimentation could potentially have on local ecosystems is variable and not fully understood by researchers.

TIDAL ENERGY

In coastal areas that have large tides, flowing tidal waters contain large amounts of potential energy.

The principle of harnessing the energy of tides dates back to eleventh-century England when tides were used to turn waterwheels, producing mechanical power. More recently, rising and falling tides have been used to generate electricity, in much the same manner as hydroelectric power plants.

Tides, the daily rise and fall of ocean levels relative to coastlines, are a result of the gravitational force of the moon and the sun as well as the revolution of Earth. The moon and the sun both exert a gravitational force of attraction on Earth. The magnitude of the gravitational attraction of an object is dependent upon the mass of an object and its distance. The moon exerts a larger gravitational force on Earth because, although it is much smaller in mass, it is a great deal closer to Earth than the sun. This force of attraction causes the oceans to bulge along an axis pointing toward the moon. Tides are produced by the rotation of Earth beneath this bulge in its watery coating, resulting in the rhythmic rise and fall of coastal ocean levels.

The gravitational attraction of the sun also affects the tides in a similar manner as the moon, but to a lesser degree. As well as bulging toward the moon, the oceans also bulge slightly toward the sun. When Earth, the moon and the sun are positioned in a straight line (called a full, or new, moon), the gravitational attractions are combined, resulting in very large "spring" tides. At half-moon, the sun and the moon are at right angles, resulting in lower tides called "neap" tides. Coastal areas experience two high and two low tides over a period of slightly longer than twenty-four hours. The friction of the bulging oceans acting on spinning Earth results in a very gradual slowing down of Earth's rotation, a phenomenon that will not have any significant effect on the planet for billions of years.

Certain coastal regions experience higher tides than others, which is a result of the amplification of tides caused by local geographical features such as bays and inlets. In order to produce practical amounts of electric power, a difference of at least 20 feet between high and low tides is required; the higher the tides, the more electricity can be generated from a given site, and the lower the cost of electricity produced. Worldwide, there are about forty sites with this magnitude of tidal range, and approximately 3000 gigawatts of power is continuously available from the action of tides. Due to the constraints of harvesting tidal energy, estimate that only 2 percent, or 60 gigawatts, can potentially be recovered for electricity generation.

The technology required to convert tidal energy into electricity is very similar to the technology used in traditional hydroelectric power plants. The first requirement is a dam or "barrage" across a tidal bay or estuary. Building dams is an expensive process, therefore the best tidal sites are those that exist where a bay has a narrow opening, thus reducing the length of dam which is required. At certain points along the dam, gates and turbines are installed. When there is an adequate difference in the elevation of the water on the different sides of the barrage, the gates are opened. This "hydrostatic head" that is created causes water to flow through the turbines, turning an electric generator to produce electricity.

Electricity can be generated by water flowing both into and out of a bay. Because there are two high and two low tides each day, electrical generation from tidal power plants is characterized by periods of maximum generation every twelve hours, with no electricity generation at the six-hour mark in between. Alternatively, the turbines can be used as pumps to pump extra water into the basin behind the barrage during periods of low electricity demand. This water can then be released when demand on the system its greatest, thus allowing the tidal plant to function with some of the characteristics of a "pumped storage" hydroelectric facility.

The demand for electricity on an electrical grid varies with the time of day, with demand highest during the day and lowest at night. Without pumped storage, the supply of electricity from a tidal power plant, which is dependent upon the slowly shifting times of the tides, can never match the demand on a system. But tidal power, although variable, is reliable and predictable and can make a valuable contribution to an electrical system that has a variety of sources. When tidal electricity is used to displace electricity that would otherwise be generated by fossil fuels or nuclear fission, it results in reduced emissions of greenhouse and acid gases, the risks associated with nuclear power.

Although the technology required to harness tidal energy is well established, because tidal power is expensive there is only one major tidal generating station in operation in the world today. This is a 240-megawatt power plant at the mouth of the La Rance River estuary on the northern coast of France, generating power roughly equal to the annual consumption of the nearby town of Rennes, which has a population of 200,000.

The Barrage de la Rance was the first hydroelectric power plant to generate energy using tidal power. (Corbis-Bettmann)

In operation since 1967, the La Rance generating station has been a very reliable source of electricity, producing electricity at a cost of 3.7 cents per kilowatt hour. Initially, La Rance was designed to be the first of many tidal power plants in France, but the country's nuclear program was greatly expanded in the late 1960s, thereby prohibiting the development of multiple tidal power plants. Elsewhere there is a 20-megawatt experimental facility at Annapolis Royal in Nova Scotia, and a 400-kilowatt tidal power plant near Murmansk in Russia.

Studies have been undertaken to examine the potential of several other tidal power sites worldwide. Scientists estimate that a barrage across the Severn River in western England could supply as much as 10 percent of the country's electricity needs (12 gigawatts). Similarly, several sites in the Bay of Fundy, in Cook Inlet, Alaska, and the White Sea in Russia have been found to have the potential to generate large amounts of electricity.

Like wave and OTEC power plants, one of the main barriers to the increased use of tidal energy is the initial cost of building tidal-generating stations. It has been estimated that the construction of the proposed facility on the Severn River in England would have a construction cost of $15 billion.

The major factors in determining the cost effectiveness of a tidal power site are the size (length and height) of the barrage required, and the difference in height between high and low tide. These factors can be expressed in what is called a site's "Gibrat" ratio. The Gibrat ratio is the ratio of the length of the barrage in meters to the annual energy production in kilowatt hours. The smaller the Gibrat site ratio, the more desirable the site. Examples of Gibrat ratios are La Rance at 0.36, Severn at 0.87, and Passamaquoddy in the Bay of Fundy at 0.92.

Tidal energy is a renewable source of electricity that does not result in the emission of gases responsible for global warming or acid rain, which are associated with fossil fuel-generated electricity. Use of tidal energy could also decrease the need for nuclear power. Changing tidal flows by damming a bay or an estuary could, however, result in negative impacts on aquatic and shoreline ecosystems, as well as on navigation and recreation.

OTHER OCEAN ENERGY TECHNOLOGIES

Theoretical concepts for generating electricity from ocean currents such, as the Gulf Stream, and salinity gradients (differences in salt content) are being investigated. More research and development is required before these concepts reach the stage of demonstration power plants.

Stuart E. Baird

BIBLIOGRAPHY

Avery, W. H., and Berl, W. G. (1997). “Solar Energy from the Tropical Oceans.” *Issues in Science and Technology* 14(2):41–42.

Bequette, F. (1998). “Harnessing Ocean Energy.” *UNESCO Courier* 51(7/8):33–35.

Bernshtein, L. B. (1995). “Tidal Power Development—a Realistic, Justifiable and Topical Problem of Today.” *IEEE Transactions on Energy Conversion* 10:591–599.

Bryden, I. G.; Naik, S.; and Fraenkel, P. (1998).”Matching Tidal Current Plants to Local Flow Conditions.” *Energy* 23:699–709.

Carless, J. (1993). *Renewable Energy: A Concise Guide to Green Alternatives*. New York: Walker and Company.

Cavanagh, J. E.; Clarke, J. H.; and Price, R. (1993). “Ocean Energy Systems.” In *Renewable Energy*, ed. L. Burnham. Washington, DC: Island Press.

Curran R.; Stewart, T. P.; and Whittaker T. J. T. (1997). “Design Synthesis of Oscillating Water Column Wave Energy Converters: Performance Matching.” *Journal of Power and Energy* 211:489–505.

D’Monte, D. (1997). “Construction Begins on India’s First Tidal-Energy Power-Plant Prototype.” *Solar Letter* 7(3):42–43.

DiChristina, M. (1995). “Sea Power.” *Popular Science* 246(5):70–74.

Edwards, R. (1998). “The Big Break.” *New Scientist* 160(2154):30–34.

Fraenkel, P. (1998). “Marine Current Power—An Emerging Energy Resource for the Millenium.” *Renewable Energy World* 1(2):64–69.

Golob, R., and Brus, E. (1993). *The Almanac of Renewable Energy*. New York: Henry Holt.

Grant, A. D. (1997). “Simplified Tidal Barrage for Small-Scale Applications.” *Journal of Energy Engineering* 123:11–19.

Pierce, F. (1998). “Catching the Tide.” *New Scientist* 158(2139):38–41.

Patterson, W. C. (1990). *The Energy Alternative*. London: Boxtree Ltd.

Sanders, M. M. (1991). “Energy from the oceans.” In *The Energy Sourcebook*, eds. Ruth Howes and Anthony Fainberg. New York: American Institute of Physics.

Sing, M. (1998). *The Timeless Energy of the Sun*. New York: Sierra Club Books/UNESCO.

OCTANE ADDITIVES

See: Gasoline and Additives

OCTANE NUMBER

See: Gasoline and Additives

OERSTED, HANS CHRISTIAN (1777–1851)

Hans Christian Oersted, the son of an impoverished pharmacist, made the great discovery that electricity and magnetism are related. Oersted was born on the small Danish island of Langeland, about halfway between Copenhagen and Hamburg. There was no school in Langeland, so he and his younger brother, Anders Sandoe, went to the homes of neighbors who taught the boys to read and write. Later the town surveyor taught them mathematics, and the mayor taught them English and French. When he was twelve, Hans began to help his father in the pharmacy, and the work stimulated his interest in science.

In 1794 Oersted and his brother Anders matriculated at the University of Copenhagen. Hans studied the sciences, and Anders, who eventually became a leading jurist and a minister of state, studied law. The

Hans Christian Oersted. (Library of Congress)

brothers were recipients of a small state scholarship, but largely supported themselves at the university by tutoring. They lived together, shared costs, and devoted themselves wholeheartedly to their studies. In 1797 Hans Christian Oersted was awarded a degree in pharmacy and, in 1799, received a Doctor of Philosophy degree.

After graduation, Oersted secured a position as a part-time lecturer at the university; he also managed a Copenhagen pharmacy. Word of Alessandro Volta's discovery of a way of producing a continuous electric current reached Copenhagen in 1800, and Oersted began experimenting with acids and alkalis using a voltaic pile. The following year, he left Copenhagen and visited a number of famous scientists during the traditional year of travel taken by European students following graduation. If all the scientists he met, Oersted was most influenced by Johann Wilhelm Ritter, an eccentric German physicist whom he visited for several weeks in Jena. Ritter had also begun his career as a pharmacist and had already discovered ultraviolet light, thermoelectric currents, and the process of electroplating. Oersted returned to Copenhagen in 1803 and applied to the university for a position as professor of physics, then called "natural philosophy," but was refused. He continued lecturing at the university in the schools of medicine and pharmaceuticals, and at the same time managed the pharmacy, carried on electrochemical experiments, and published his results. In 1806 he was finally made a professor of physics at the University, although he not become a full professor until 1817.

From 1803 to 1820, Oersted's life centered around the cultural and academic life of what was then the small city of Copenhagen. He took part in political and academic debates, participated in a royal geological expedition in Denmark, and became a popular public lecturer. He was knighted and achieved the position of secretary of the Royal Society of Copenhagen. In 1813 he again visited Germany and France and published a book about electrochemical forces. In it Oersted commented on magnetic forces and clearly stated that the connection between electricity and magnetism should be investigated. Before the publication of Oersted's book, it had indeed been suspected by many scientists that magnetism and electricity were somehow related. It was known that iron rods were magnetized by the action of the electric currents passing through them as the result of lightning strikes. However, no scientific verification or understanding of the relation existed.

Oersted's discovery of the relation between magnetism and electricity in 1820 is often described as the result of a lucky accident occurring during the course of a laboratory demonstration. However, Oersted declared that he had actually prepared the experiment before the demonstration and only carried it out during the demonstration to some advanced students because that was the first opportunity. What Oersted observed was that a wire carrying an electric current caused a nearby magnetic compass needle to assume a position perpendicular to the wire, and if the current were reversed, the needle would reverse position.

After this initial discovery, Oersted waited three months, apparently for the construction of a more powerful current source. He then carried out sixty experiments to show that the magnetic field due to the current in a wire is circular around the wire. He showed that the effect is independent of the type of wire, and that it is independent of any intervening common materials. Later, he proved that the effect is proportional to the current in the wire.

On July 21, 1820, Oersted published a four-page Latin monograph, “Experiments on the Effect of a Current of Electricity on the Magnetic Needle.” He distributed the monograph to leading scientists throughout Europe and, in the following months, the monograph was reprinted in translation in the most important scientific journals throughout Europe and Britain. A whole new field of investigation and technology was opened. Within a year the laws of electrodynamics were formulated, the electromagnet was invented, and the first primitive electric motor was demonstrated. The first electric telegraph and the first primitive electric generator would soon follow.

Oersted was named a fellow of several learned societies, presented with medals, and awarded cash prizes. At home, Oersted became Denmark’s leading citizen. He continued his research, but as an international figure he traveled extensively, became fluent in many languages, and met with the leading scientists of the time. He gave frequent public lectures and became a director of the Royal Polytechnic Institute of Copenhagen. He also had a lifelong interest in literature and, in 1829, he founded a literary journal to which he frequently contributed articles about science. In 1850, the fiftieth anniversary of his appointment at the university was celebrated as a national holiday, and he was given a country home by the government. When he died in 1851, more than 200,000 people joined the funeral procession.

Oersted had a kindly and sympathetic personality. He had a successful marriage and a large family. In 1819 he befriended a poor fourteen-year-old boy who over the years became virtually another member of the Oersted family. The boy, Hans Christian Andersen, was to become the great Danish storyteller. Andersen often referred to himself as “little Hans Christian” and to Oersted as “great Hans Christian.”

Leonard S. Taylor

See also: Electricity; Electricity, History of; Electric Motor Systems; Electric Power, Generation of; Magnetism and Magnets.

BIBLIOGRAPHY

Dibner, B. (1962). *Oersted.* New York: Blaisdell Publishing Company.

Meyer, H. W. (1971). *A History of Electricity and Magnetism.* Cambridge, MA: MIT Press.

OFFICE EQUIPMENT

TYPES OF EQUIPMENT

Energy use for office equipment ranks behind only lighting, heating, and air conditioning in the commercial sector. And because of growing use of office equipment, the hope is that more efficient equipment will offset the greater use.

Office equipment is an end use that generally is regarded to consist of five major products: personal computers, monitors, copiers, printers, and fax machines. This equipment can be defined by characteristics such as size and speed. Although office equipment is used in homes, this article focuses on office equipment energy use in actual offices.

Personal Computers (PCs)

PCs are desktop and deskside microcomputers that generally include a central processing unit (CPU), storage, and keyboard. PCs first appeared in office spaces during the 1980s. Prior to that time, many offices used simple terminals connected to mainframe computers. PC energy use assessments generally do not include laptops and PC servers—for example, new systems such as webservers—that are left on twenty-four hours a day.

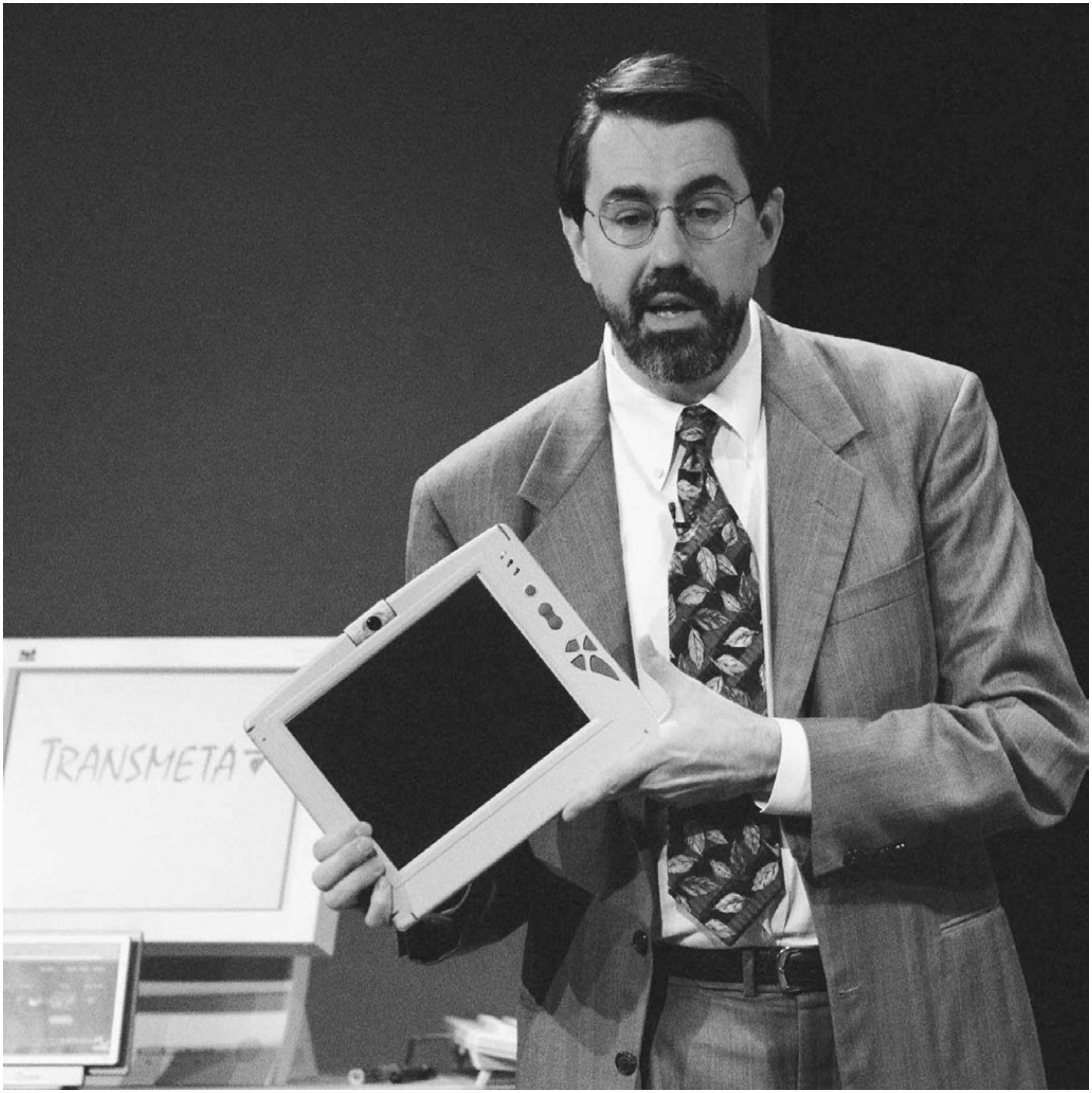
Monitors

Most monitors are display terminals that use cathode-ray tube (CRT) displays, which function by exciting a layer of phosphors with an electron gun. These devices include monitors used with PCs and terminals used with mainframes or minicomputers. Features such as color, resolution, and size influence power requirements. Most PC monitors are sold in screen sizes of fourteen inches, fifteen inches, seventeen inches, and twenty-one inches. In 1995, the typical monitor was fourteen inches and monochrome; by 2000, PC users were purchasing larger, higher-resolution color CRTs. Flat-panel displays such as those used in laptop computers are beginning to replace CRT monitors in desktop computers, though flat-panel models in 2000 cost considerably more than CRTs for the same display size. There are three types of flat-panel displays: liquid crystal (LCD), plasma, and electroluminescent emission (EL). Color LCDs are increasing their share of the monitor market because resolution is improving, costs are dropping, and they can be powered up and down more rapidly than most CRTs.

Copiers

The majority of copiers in the commercial sector use heat and pressure technologies to fix an image to paper. The basic principle consists of forming an image on a photosensitive drum with a laser or a lamp. The drum is then covered with toner, which is transferred to a sheet of paper and fixed to the paper by a fuser roll. The fuser is kept hot while the machine is in a standby mode, ready for the next copy. Some copiers have a low-power “energy-saver” mode in which the drum temperature is reduced; others have an “automatic off” feature, and some have both. Several factors influence energy use per page, such as whether duplexing (two-sided printing) is used and the size of the copy job. Single-page jobs are the most energy-intensive on a per-copy basis.

Copier characteristics vary with the speed of output, with faster machines adding more sophisticated handling of originals, duplexing, finishing of copies as with stapling and collating, and the ability to make more copies each month.



Transmeta Corp. CEO David R. Ditzel demonstrates the benefits of using the new "Crusoe" smart microprocessor and how it will revolutionize the field of mobile computing. (Corbis Corporation)

Printers

Since 1990 there has been rapid change from impact printing, such as daisy wheel and dot-matrix techniques, to nonimpact printing, dominated by energy-intensive laser printing. As is true for copiers, printer energy use is generally linked to print speed. Laser printing is relatively energy-intensive because of the heat and pressure requirements for the fuser. Most of the energy is consumed while the printer sits idle, keeping the fuser warm.

Inkjet printers are suitable in low-volume settings such as personal offices and the home. Inkjets provide low-cost color capability. Power use while the printer is idle is generally low, so that a distinct low-power mode is not needed.

Fax Machines

Facsimile machines, or fax machines, send and receive information from printed documents or electronic files over telephone lines. Three common types of fax machines to consider are direct thermal, laser, and inkjet. Direct thermal faxes apply heat to thermally sensitive paper, while inkjet and laser fax machines are similar to the printers discussed in the previous section. In 2000, laser fax machines had the broadest market share and are of greatest energy use. There were only a few fax machines in 1990 compared to ten years later. It is likely that the technology may change quickly again from 2000 to 2010. Fax cards used with computers suggest the possibility of a future with reduced paper communication by eliminating the need to print a document before sending it. Most contemporary fax machines reduce their power use when not active.

POWER REQUIREMENTS AND ENERGY STAR FEATURES

Office equipment energy use can be characterized by the power requirements in each "mode" of operation and by the number of hours per year for each mode. In 1993 the U.S. Environmental Protection Agency launched the Energy Star program, beginning with power management in PCs and monitors. This program now promotes the use of power management technology in all office equipment (see Table 1). The power modes in PCs and monitors can be described as follows:

Active mode: This is the power of the device when in operation.

Standby mode: This mode represents an intermediate state that attempts to conserve power with instant recovery. The system is idle.

Suspend mode: This mode, also known as sleep mode, has the lowest power level (without being off) but has a longer recovery time than for standby.

Off mode: The power in this mode is that drawn when the device is switched off (essentially zero), or for copiers when the device is unplugged.

Many office equipment models consume some electricity when nominally "off." This results in "leaking" or "standby" energy use. Usually this amounts to just 2 or 3 W, but in some copiers can be up to considerably more. Power management modes vary most widely in PCs, with the number of modes and sophistication of operation increasing. Power management is most transparent to the user with fax machines.

EQUIPMENT POWER LEVELS, OPERATING PATTERNS, AND ANNUAL ENERGY USE

Energy use of office equipment is primarily a function of three factors: the power characteristics of the device, the operation of power management, and manual turn-off patterns. Table 2 shows typical power levels for office equipment, though within each category there is a significant range. Printers and copiers are subdivided into categories based on their maximum speed of making images (in pages or copies per minute), with power levels and usage patterns varying across these categories.

In addition to the power levels above, extra energy is required for each image made by a device. While this varies considerably, a value of 1 Wh/image is common. For copiers, about 10 percent of annual energy use is used for making images, with the other 90 percent primarily used to maintain the copier in the "ready to copy" mode.

Power management capability has reached most office equipment, but on many devices the feature is not enabled or (for some PCs or monitors) is prevented from functioning. Equipment with functioning power management or that is turned off manually at night and on weekends use considerably less, and devices left fully on constantly use much more.

The Energy Star program has affected the office equipment market in several ways: increasing the portion of equipment capable of automatic power management; reducing low-power power levels; adding auto-off capability to many copiers; and increasing enabling rates of power management. All of this has saved energy for consumers with little or no additional manufacturing cost. Monitor power management has risen with larger screens becoming more common, though as LCD monitors replace CRTs the average should drop. The number of PCs and monitors that are properly enabled is likely to increase.

Another aspect of the Energy Star program is that "efficient" copy machines not only have the potential to lower direct use of electricity, they also should

Equipment Category	*Default Time to Low-Power State*	*Max. Power in Low-Power State*	*Date in Force*
PC (without monitor)[1]	**N.A.[2]**	**30 W**	**mid-1993**
Monitors[1]	**N.A.[2]**	**30 W**	**mid-1993**
Printers and printer/fax combos:			
1-7 pages per min.	15 min.	15 W	Oct 1, 1995
8-14 pages per min.	30 min.	30 W	Oct 1, 1995
Color and/or >14 pages per min.	60 min.	45 W	Oct 1, 1995
Fax machines:			
1-7 pages per min.	5 min.	15 W	July 1, 1995
8-14 pages per min.	5 min.	30 W	July 1, 1995
>14 pages per min.	15 min.	45 W	July 1, 1995
	Default time to Low/Off	*Max Power Low/Off*	
Copiers-Tier 1[3]:			
1-20 copies per min.	N.A./30 min.	N.A./5W	July 1, 1995
21-44 copies per min.	N.A./60 min.	N.A./40W	July 1, 1995
>44 copies per min.[4]	N.A./90 min.	N.A./40W	July 1, 1995
Copiers-Tier 2[5]:			
1-20 copies per min.	N.A./30 min.	N.A./5W	July 1, 1997
21-44 copies per min.	15 min./60 min.	(3.85xcpm + 5W)/10W	July 1, 1997
>44 copies per min.	15 min./90 min.	(3.85xcpm + 5W)/15W	July 1, 1997

Table 1.
EPA Energy Star Office Equipment Characteristics

NOTES: (1) PCs and monitors are to be shipped with the power saving features enabled and those features must be tested in a networked environment. Some PCs qualify for a second standard of 15 percent of the rating of the power supply, usually at least 30 W. For monitors, an initial low-power mode of 15 W is also required. (2) "NA" means "not applicable," which means that no requirement exists. (3) Additional Tier 2 requirement for copiers includes a required recovery time of 30 seconds for midspeed copiers (recommended for high speed copiers).

reduce paper use by encouraging double-sided copying, or duplexing. Paper costs in 2000 are about \$0.005/sheet. An average 40-cpm copier produces about 140,000 images per year. Approximately 124,600 pages are simplexed (copied on one side only) and 7,700 pages are duplexed (15,400 images). Doubling the duplex rate to handle 30,000 images per year saves about 15,000 pages per year or about \$75 per year, which is nearly twice the value of the energy savings.

In addition to the direct savings from reduced energy and paper, there are indirect savings from the energy embodied in the paper. This is about 16 Wh per sheet. Reducing annual paper use by 15,000 pages cuts embodied energy use by 300 kWh per year. By comparison, an Energy Star copier saves about 260 kWh per year directly.

Energy Star office equipment typically does not cost any more than the non–Energy Star devices, and power-management technologies have greatly improved since 1990. The first Energy Star PCs had slow recovery times from the low-power modes, which was one of the reasons power management was disabled by users. Some network technologies have not been compatible with PC power management. The PC operating system and hardware characteristics also influence the success of the power management. Monitor power management is gener-

Year	*Active (W)*	*Standby (W)*	*Suspend (W)*
Pre-Energy Star	70	70	70
Energy Star	44	25	25
Low Power	44	15	15

Monitor Equipment Power

Year	*Active (W)*	*Standby (W)*	*Suspend (W)*
Pre-Energy Star	60	60	60
Typical Energy Star	65	50	15
Low Power	65	30	3

Laser Printer Equipment Power

Year	*Active (W)*	*Standby (W)*	*Suspend (W)*
Pre-Energy Star	250	80	80
Typical Energy Star	250	80	25
Low Power	250	33	10

Copier Equipment Power

Year	*Active (W)*	*Standby (W)*	*Suspend (W)*	*Plug (W)*
Pre-Energy Star	220	190	190	10
Typical Energy Star	220	190	150	10
Low Power	220	190	150	2

Fax Equipment Power

Year	*Active (W)*	*Standby (W)*
Pre-Energy Star	175	35
Typical Energy Star	175	15
Low Power	175	7

Table 2.
Typical Power Requirements by Product Category

ally simpler to operate and saves more energy than PC power management.

Overall, office equipment currently uses about 7 percent of all commercial sector electricity, with that fraction projected to grow to about 8 percent by 2010. Total electricity used by office equipment would grow from 58 TWh in 1990 to 78 TWh in 2010 in the absence of Energy Star or any other government policies. While total energy use for office equipment has grown rapidly in recent years, this growth is likely to slow from 2000 to 2010 because the market is becoming saturated and because mainframe and minicomputer energy use is declining. Home office energy use is on the rise, but is still well below the energy use of office equipment in commercial buildings.

The likely energy and dollar savings in the commercial sector from the Energy Star program are significant on a national scale. Total electricity savings are estimated to range from 10 to 23 TWh/year in 2010, most likely about 17 TWh/year. Energy bill savings will exceed $1 billion per year after the year 2000. The cost of achieving Energy Star efficiency levels is estimated to be negligible, while the cumulative direct cost of funding the Energy Star program is on the order of a few million dollars. This policy therefore saves U.S. society large amounts of money with minimal expenditure of public funds.

Mary Ann Piette
Bruce Nordman

See also: Conservation of Energy; Consumption; Economically Efficient Energy Choices; Efficiency of Energy Use.

BIBLIOGRAPHY

Kawamoto, K.; Koomey, J. G.; Nordman, B.; Brown, R. E.; Piette, M. A.; and Meier, A. K. (2000). "A Simplified Analysis of Electricity Used by Office Equipment and Network Equipment in the U.S." In *Proceedings of the ACEEE 2000 Summer Study on Energy Efficiency in Buildings*. Washington, DC: American Council for an Energy-Efficient Economy.

Koomey, J. G.; Cramer, M.; Piette, M. A.; and Eto, J. H. (1995). "Efficiency Improvements in U.S. Office Equipment: Expected Policy Impacts and Uncertainties." Report No. LBNL-41332. Berkeley, CA: Lawrence Berkely National Laboratory.

Nordman, B.; Piette, M. A.; Kinney, K.; and Webber, C. (1997). "User Guide to Power Management for PCs and Monitors." Report No. LBNL-39466. Berkeley, CA: Lawrence Berkeley National Laboratory.

Nordman, B.; Meier, A. K.; and Piette, M. A. (2000). "PC and Monitor Night Status: Power Management Enabling and Manual Turn-Off." In *Proceedings of the ACEEE 2000 Summer Study on Energy Efficiency in Buildings*. Washington, DC: American Council for an Energy-Efficient Economy.

U.S. Environmental Protection Agency. (1999). *Energy Star Office Equipment Program.* <http://www.epa.gov/office/>.

OFF-SHORE DRILLING RIGHTS

See: Oil and Gas, Drilling for; Property Rights

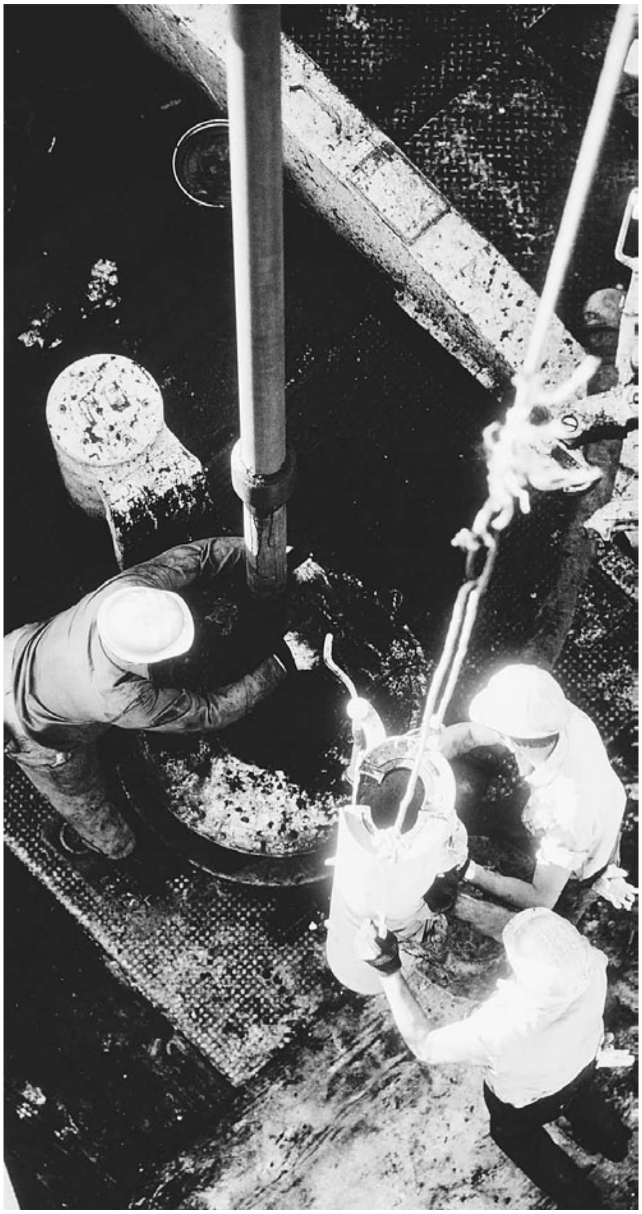
OIL AND GAS, DRILLING FOR

After all the exploratory analyses, drilling determines whether the exploration geophysicist has accurately located the reservoir (exploratory drilling) and whether the sites chosen for drilling into the same reservoir are optimal for efficient production (developmental drilling). When an exploratory hole produces neither oil nor gas, it is capped and abandoned. But if it does yield oil or gas, it is readied for production and becomes a completed well. To extract oil and gas requires drilling a well.

Drilling a well involves much more than making a hole. It entails the integration of complex technologies, requiring the driller to make individual decisions related to unexpected pressure regimes, practices, and rock formations. The resulting well is the sole conduit to move fluids from a reservoir to the surface—a conduit that must last at least fifty years and be flexible enough in design to allow for the application of future production technologies.

Drilling operators must confront and solve extremely difficult technical, safety, and control problems as they bore through layers of subsurface rock to access oil- or gas-bearing strata. Furthermore, drilling must protect the geologic formation, the ultimate productive capacity of the well, and the surface environment. Drilling problems must first be diagnosed using the information or data that is transmitted from the bottom of the well to the surface, where the information is collected on the rig floor.

Depending on the depth of the well, this time lag can consume valuable time needed to address the problem—either technical or geological—before it becomes worse and/or causes drilling operations to stop. Drilling a well involves all types of technical, geological, and economic risks. The greatest economic risk occurs when drilling operations must be halted after time and money have been invested. This is the nature of the challenge faced during the drilling process.



Oil rig workers begin drilling. (Corbis Corporation)

When a well has been drilled and lined with pipe, the connection between the geological formation and the well must be established. Well "completion" includes installing suitable metal pipe or casing, cementing this casing using rock section isolation devices, and perforating the casing to access the producing zones. In some reservoirs, the geological conditions dictate that "stimulation" processes be applied to improve reservoir permeability or fluid conductivity of the rock, thereby facilitating production through the well bore.

With an understanding of the nature of the rocks to be penetrated, the stresses in these rocks, and what it will take to drill into and control the underground reservoirs, the drilling engineer is entrusted to design the earth-penetration techniques and select the equipment. Substantial investments are at stake. Drilling and completion costs can exceed $400,000 per well, and offshore operations can easily escalate to more than $5 million per well. However, overall per-well costs (adjusted for inflation) have been dropping, falling more than 20 percent between 1980 and 2000. This technology-driven gain in extraction efficiency has more than made up for the higher costs associated with the increasingly more geologically complex operations and deeper depths faced in trying to extract the remaining oil and gas resources.

In the United States, 85 percent of wells are drilled by independent oil and gas operators, more than 90 percent of whom employ fewer than twenty people. Therefore, U.S. oil and gas production is dependent on the economic health of independent producers to offset the rising tide of imported crude oil.

Drilling is indeed a high-risk investment. Even with modern technology, drilling successful ex-ploratory wells can be elusive. In the United States, for exploratory wells drilled more than a mile from production, the chance of striking hydrocarbons are about one in ten, and exceeds one in forty for drilling in unproven frontier areas. Because of these daunting odds, the ever-increasing complexity of recovery, and the dwindling resource base, most energy experts in the late 1970s and early 1980s projected that the price of crude oil would double, triple, or quadruple by 2000. This did not happen. Profound advances in drilling technology as well as in other exploration and production technologies have allowed the inflation-adjusted price of crude oil to remain stable through the twentieth century except for short-term price distortions caused by geopolitical turmoil or misguided policies.

THE EARLY METHODS

On August 27, 1859, at a depth of only 69.5 feet, Francis Drake drilled and completed the first well at Titusville, Pennsylvania. By the end of the nineteenth century there were about 600,000 oil wells in more than 100 countries, with the United States and Russia dominating world production. In the United States, drilling was concentrated in the Appalachian oil fields until the Beaumont Texas, Spindletop Hill discovery on January 10, 1901, which then shifted exploration and production to the Southwest.

The world's largest petroleum reserves, in Turkey and the Middle East, were discovered just prior to World War I. Exploration and production continued to expand throughout the region in the 1920s, and culminated in the discovery of the vast oil resources of Saudi Arabia in 1938.

Early methods entailed cable-tool drilling. This crude impact-type drilling involved dropping a weighted bit by a cable and lever system to chisel away at the rock at the bottom of the hole. Periodically drilling had to be stopped to sharpen the bit and remove rock fragments and liquids from the bottom of the hole with a collecting device attached to the cable. Removal of liquids was necessary so the bit could more effectively chip away at rock. This dry method created the well-recognized "gusher," since a dry borehole allowed oil and gas to flow to the surface from the pressure of natural gas and water once the bit penetrated a producing reservoir. Usually considerable oil and "reservoir energy" were wasted until the well could be capped and controlled.

MODERN DRILLING

Rotary drilling became the preferred method of drilling oil and gas wells in the middle of the twentieth century (Figure 1). Using this method, the drill bit rotates as it pushes down at the bottom of the hole much like a hand-held power drill. Unlike in cable-tool drilling, the borehole is kept full of liquid during rotary drilling for two important reasons: The drilling mud circulates through the borehole to carry crushed rock to the surface to keep drilling continuous, and once the bit penetrates the reservoir, it does not create a gusher.

To drill a well, a site is selected and prepared. A drilling rig is transported to the site and set up. A surface hole is drilled, followed by drilling to the total planned depth of the well. The well is then tested, evaluated, and completed. Finally, production equipment is installed, and the well is put on production.

Drill-Site Selection and Preparation

Selection of the drill site is based largely on geological evidence indicating the possible accumulation of petroleum. The exploration drilling company wants to drill the well at the most advantageous location for the discovery of oil or gas. However, surface conditions also must be taken into consideration when selecting the drill site. There must be a nearly level area of sufficient size on which to set up the drilling rig, excavate reserve pits, and provide storage for all the materials and equipment that will be needed for the drilling program. All required legal matters must have been attended to, such as for acquiring a drilling permit and surveying of the drill site. When all of these matters have been resolved, work on site preparation will begin. Once the drill site has been selected and surveyed, a contractor or contractors will move in with equipment to prepare the location. If necessary, the site will be cleared and leveled. A large pit will be constructed to contain water for drilling operations and for disposal of drill cuttings and other waste. Many environmental regulations guide these practices. A small drilling rig, referred to as a dry-hole digger, will be used to start the main hole. A large-diameter hole will be drilled to a shallow depth and lined with conductor pipe. Sometimes a large, rectangular cellar is excavated around the main borehole and lined with wood. A smaller-diameter hole called a "rat hole" is drilled near the main borehole. The rat hole is lined with pipe and is used for temporary storage of a piece of drilling equipment called the "kelly." When all of this work has been completed, the drilling contractor will move in with the large drilling rig and all the equipment required to drill the well.

The Rig

The components of the drilling rig and all necessary equipment are moved onto the location with large, specially equipped trucks. The substructure of the rig is located and leveled over the main borehole. The mast or derrick is raised over the substructure, and other equipment such as engines, pumps, and rotating and hoisting equipment, are aligned and connected. The drill pipe and drill collars are laid out on racks convenient to the rig floor so that they may be hoisted when needed and connected to the drill bit or added to the drill strings. Water and fuel tanks are filled. Additives for the drilling fluid (drilling mud) are stored on location. When all these matters have been attended to, the drilling contractor is ready to begin drilling operations ("spud the well").

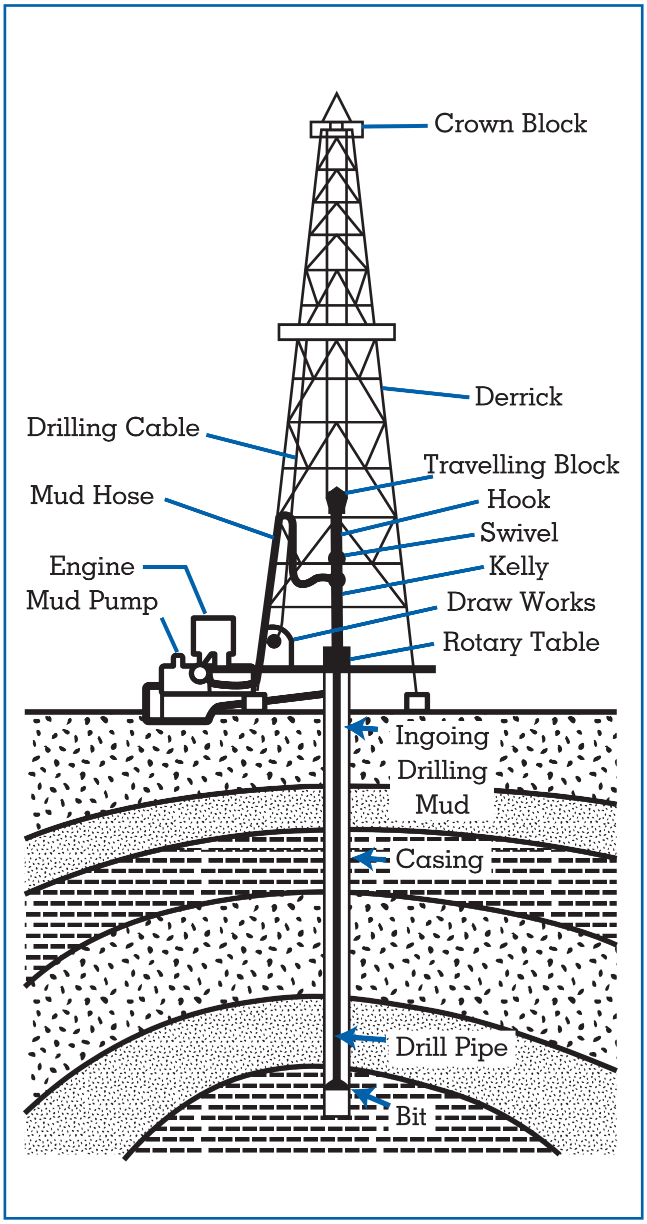


Figure 1.
A typical drilling rig operation. The string of pipe and drill bit rotate, with their weight adding to the cutting effectiveness of the bit. Borehole pressure is maintained by drilling mud, which also brings up the drill cuttings. Casing is used for weak formations or when there is risk of groundwater contamination.

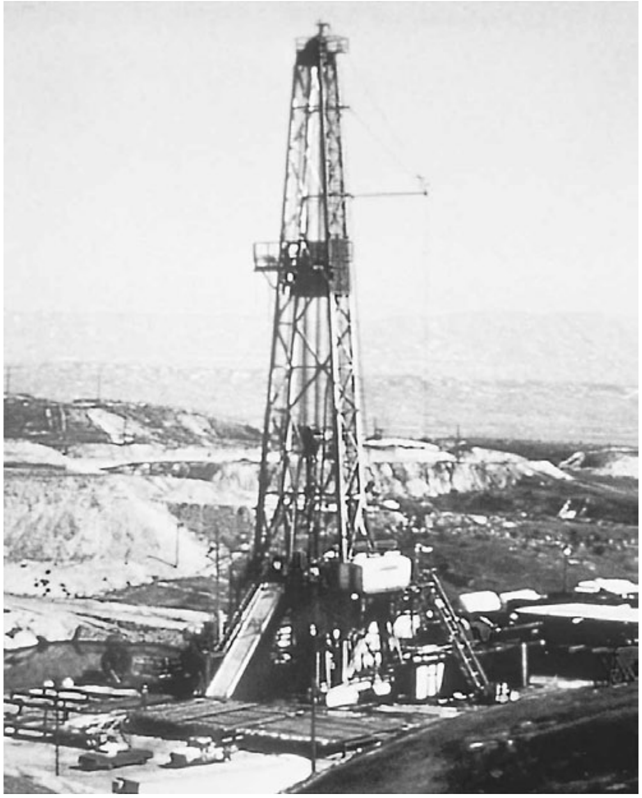
The drill string, consisting of a drill bit, drill collars, drill pipe, and kelly, is assembled and lowered into the conductor pipe. Drilling fluid, better known as drilling mud, is circulated through the kelly and the drill string by means of pipes and flexible hose connecting the drilling fluid or mud pumps and a swivel device attached to the upper end of the kelly. The swivel device enables drilling mud to be circulated while the kelly and the drill string are rotated. The mud pump draws fluid from mud tanks or pits located nearby. The drilling mud passes through the kelly, drill pipe, drill collars, and drill bit. The drilling mud is returned to the surface by means of the well bore and the conductor pipe where it is directed to a shale shaker, which separates the drill cuttings and solids from the drilling mud, which is returned to the mud tanks to be recirculated. As the drill string is rotated in the well bore, the drill bit cut into the rock. The drilling mud lubricates and cools the drill bit and the drill string and carries the drill cuttings to the surface.

Drilling the Surface Hole

When a well is spudded in, a large-diameter drill bit is used to drill to a predetermined depth to drill the surface hole, which is lined with casing. The casing protects aquifers that may contain freshwater, provides a mounting place for the blowout preventer, and serves as the support for the production casing that will be placed in the well bore if the drilling program is successful. The surface hole may be several hundred or several thousand feet deep. When the predetermined depth is reached, the drill string will be removed from the well bore. Steel casing of the proper diameter is inserted. Sufficient cement is pumped down the surface casing to fill the space between the outside of the casing and the well bore all the way to the surface. This is to ensure protection of freshwater aquifers and security of the surface casing. The casing and the cement are tested under pressure for twelve hours before drilling operations may be resumed. The blowout preventer is attached at the top of the surface casing. This device is required to control the well in the event that abnormal pressures are encountered in the borehole that cannot be controlled with drilling fluid. If high-pressure gas or liquid blows the drilling fluid out of the well bore, the blowout preventer can be closed to confine the gas and the fluid to the well bore.

Drilling to Total Depth

After the surface casing has been tested and the blowout preventer installed, drilling operations are resumed. They will continue until the well has been drilled to the total depth decided upon. Usually the only interruptions to drilling operations will be to remove the drill string from the well bore for the replacement of the drill bit (a procedure known as tripping) and for testing of formations for possible occurrences of oil or gas (known as drill-stem testing). Other interruptions may be due to problems incurred while drilling, such as the shearing off the drill string (known as "twisting off") and loss of drill-bit parts in the well bore (known as "junk in the hole").



Deep drilling oil rig at Naval Petroleum Reserve's Elk Hills site near Bakersfield, California. (U.S. Department of Energy)

As drilling operations continue, a geologist constantly examines drill cuttings for signs of oil and gas. Sometimes special equipment known as a mud logger is used to detect the presence of oil or gas in the drill cuttings or the drilling fluid. By examining the drill cuttings, a geologist determines the type of rock that the drill bit is penetrating and the geologic formation from which the cuttings are originating.

Today's conventional drill bit utilizes three revolving cones containing teeth or hardened inserts that cut into the rock as the bit is revolved. The teeth or inserts chip off fragments of the rock which are carried to the surface with the drilling fluid. The fragments or chips, while they are representative of the rock being drilled, do not present a clear and total picture of the formation being drilled or the characteristics of the rock being penetrated as to porosity and permeability. For this purpose a larger sample of the rock is required, and a special type of drill bit is used to collect the sample, called a core. The core is usually sent to a laboratory for analysis and testing.

If the geologist detects the presence of oil or gas in the drill cuttings, a drill-stem test is frequently performed to evaluate the formation or zone from which the oil show was observed. Drill-stem tests may also be performed when the driller observes a decrease in the time required to drill a foot of rock, known as a "drilling break." Since porous rock may be drilled easier than nonporous or less porous rock, a drilling break indicates the presence of the type of rock that usually contains oil or gas. A drill-stem test enables the exploration company to obtain a sample of the fluids and gases contained in the formation or interval being tested as well as pressure information. By testing the well during the drilling phase, a decision as to whether to complete it can be made. When the well has been drilled to the predetermined depth, the drill string will be removed from the well bore to allow insertion of tools that will further test the formation and particularly the well. These tools are suspended on a special cable. Specific properties of the formation are measured as the tools are retrieved. Signals detected by the tools are recorded in a truck at the surface by means of the electrical circuits contained in the cable.

Completing the Well

When drill-stem testing and well logging operations have been completed and the results have been analyzed, company management must decide whether to complete the well as a producing well or to plug it as a dry hole. If the evidence indicates that no oil or gas is present, or that they are not present in sufficient quantity to allow for the recovery of drilling, completion, and production costs and provide a profit on investment, the well probably will be plugged and abandoned as a dry hole. On the other hand, if evidence indicates the presence of oil or gas in sufficient quantity to allow for the recovery of the cost and provide a profit to the company, an attempt will be made to complete the well as a producing well. The casing is delivered and a cementing company is called in.

SlimDril drill bits are manufactured from polycrystalline with natural diamonds on the face of the bit for oil and natural gas drilling. (Corbis-Bettmann)

The well bore is filled with drilling fluid that contains additives to prevent corrosion of the casing and to prevent movement of the fluid from the well bore into the surrounding rock. The casing is threaded together and inserted into the well bore in much the same manner as the drill string. Casing may be inserted to the total depth of the hole, or a cement plug may have been set at a specific depth and the casing set on top of it. Cement is mixed at the surface, just as if the well were to be plugged. The cement is then pumped down the casing and displaced out of the bottom with drilling fluid or water. The cement then flows up and around the casing, filling the space between the casing and the well bore to a predetermined height. After cementing of the casing has been completed, the drilling rig, equipment, and materials are removed from the drill site. A smaller rig, known as a workover rig or completion rig, is moved over the well bore. The smaller rig is used for the remaining completion operations. A well-perforating company is then called to the well site. It is necessary to perforate holes in the casing at the proper depth to allow the oil and the gas to enter the well bore. The perforating tool is inserted into the casing and lowered to the desired depth on the end of a cable. Shaped charges are remotely fired from the control truck at the surface, and jets of high-temperature and high-velocity gas perforate the casing, the cement, and the surrounding rock for some distance away from the well bore.

A smaller-diameter pipe, called tubing, is then threaded together and inserted into the casing. If it is expected that oil or gas will flow to the surface via the pressure differential between the well bore and the formation, a wellhead is installed; it is equipped with valves to control the flow of oil or gas from the well. The wellhead is known as a "Christmas tree." If there is not sufficient pressure differential to cause the oil and the gas to flow naturally, pumping equipment is installed at the lower end of the tubing.

During well completion it is sometimes desirable or necessary to treat or "stimulate" the producing zone to improve permeability of the rock and to increase the flow of oil or gas into the casing. This may be accomplished by use of acid or by injection of fluid and sand under high pressure to fracture the rock. Such a treatment usually improves the ability of the rock to allow fluid to flow through it into the well bore. At this point the drilling and completion phases have ended.

The Drill Bits and Well Design

Designing more effective and durable drill bits is important because drilling depths of reservoirs can reach as deep as 5 mi (8,052 m). Worldwide use of the more expensive polycrystalline diamond compact (PDC) drill bit has slowly been gaining on the conventional roller cone bit because of faster rates of penetration and longer bit life. The 1998 PDC bit rotated 150 to 200 percent faster than similar bits a decade earlier, and averaged more than 4,200 ft (1,281 m) per bit as opposed to only 1,600 ft (488 m) per bit in 1988. This 260 percent improvement dramatically cuts time on site since the less often the drill bit needs changing, the less time and energy must be devoted to raising and lowering drill pipe. It also means that the 15,000-ft well in the 1970s that took around eighty days could be completed in less than forty days in 2000. Because wells can be drilled more quickly, more profitably and with less of an environmental impact, the total footage drilled by diamond bits went from about 1 percent in 1978 to 10 percent in 1985 and to 25 percent in 1997.

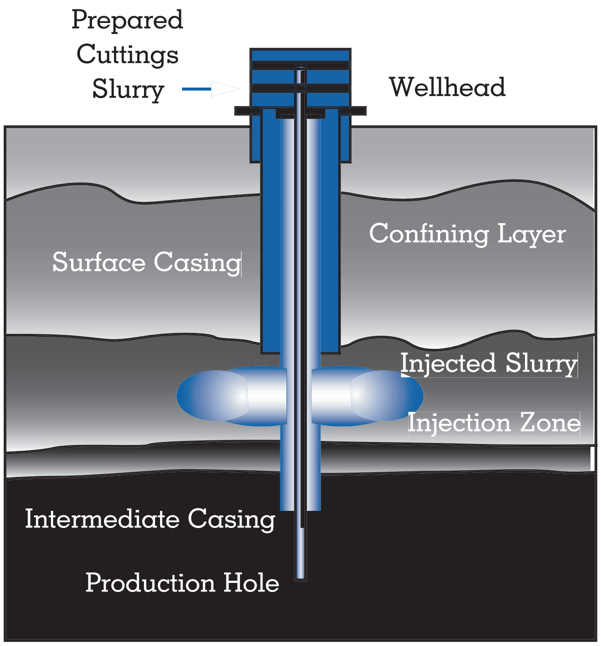


Figure 2.
To reduce the amount of waste and disposal costs, drill cuttings are increasingly being reinjected into the formations from which they came.

To reduce drilling and development costs, slimhole drilling is increasingly being used. Slimhole wells are considered wells in which at least 90 percent of the hole has been drilled with a bit fewer than six inches in diameter. For example, a typical rig uses a 8.5-in bit and a 5-in drill pipe, whereas a slimhole rig may use a 4-in bit and a 3.7-in drill pipe. Slimhole drilling is especially valuable in economically marginal fields and in environmentally sensitive areas, since the fuel consumption can be 75 percent less (mud pumps, drill power), the mud costs 80 percent less, the rig weight is 80 percent less, and the drill site is 75 percent smaller.

Another important advance for developmental drilling is coiled tubing technology. Often used in combination with slimhole drilling technology, coiled tubing is a continuous pipe and thus requires only about half the working space as a conventional drilling pipe operation. Moreover, coiled tubing

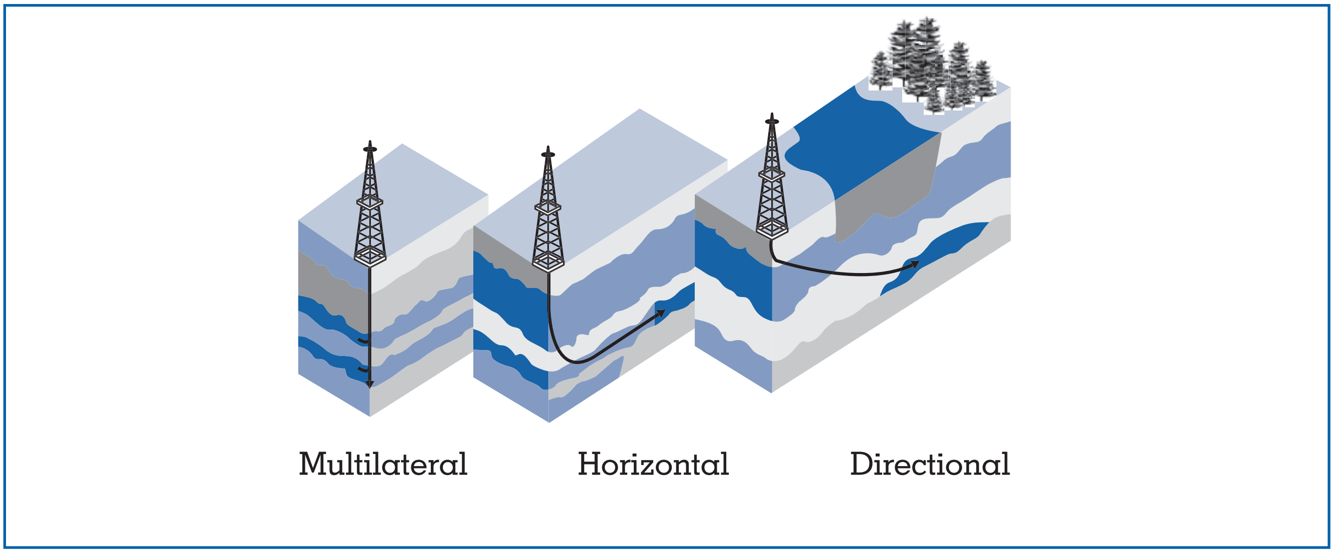


Figure 3.
Horizontal, Directional and Multilateral Drilling. Prior to 1988, fewer than 100 horizontal wells were drilled, but this total grew to more than 2,700, since horizontal drilling makes it possible to reach tight and thin reservoirs, and reservoirs inaccessible by vertical drilling. And because horizontal wells penetrate a greater cross section of any formation, more of the reserve can be extracted. Horizontal wells costs are 80 percent greater, yet production averages 400 percent more.

eliminates the costs of continuous jointing, reinstallation and removal of drill pipe. Since the diameter of coiled tubing is smaller than that of conventional pipe, coiled tubing also reduces operational energy use, noise, and the quantity of mud generated. The material of choice is currently steel, yet titantium is slowly replacing steel because titanium's greater strength promises a life cycle five to ten times longer than steel.

In weak formations or where groundwater needs to be protected, the borehole is lined with casing to prevent any transfer of fluids between the borehole and its surroundings. Strings of casing are progressively smaller than the casing strings above it, and the greater the depth of the well, the more casing sizes used. For very deep wells, up to five sizes of ever smaller diameter casings are used during the drilling process. This tiered casing approach has not only reduced the volume of wastes but also has been refined so that drilling cuttings can be disposed of by reinjecting them into the area around the borehole between the surface casing and the intermediate casing (Figure 2). In this way, wastes are returned to the geologic formations far below the surface, eliminating the need for surface disposal, including waste management facilities, drilling waste reserve pits, and off-site transit transportation.

Drilling is not a continuous operation. Periodically drilling must stop to check bottomhole conditions. To reduce the time lag between occurrence and surface assessment, measurement-while-drilling (MWD) systems are increasingly being used to measure downhole conditions for more efficient, safer, and accurate drilling. These systems transmit real-time data from the drill bit to the surface by encoding data as a series of pressure pulses in the mud column that is then decoded by surface sensors and computer systems. The accelerated feedback afforded by MWD systems helps keep the drill bit on course, speeds reaction to challenging drilling operations (high-pressure, high-temperature and underbalanced drilling), and provides valuable and continual clues for zeroing in on the reservoir's most productive zones. Since MWD systems provide a faster and more accurate picture of formation pore pressure and fracture pressure while the well is being drilled, it also substantially reduces the risks of life-threatening blowouts and fires.

HORIZONTAL, DIRECTIONAL AND MULTILATERAL DRILLING

Horizontal and directional drilling are non-vertical drilling that allow wells to deviate from strictly vertical by a few degrees, to horizontal, or even to invert toward the surface (Figure 3). Horizontal drilling is not new. It was first tried in the 1930s, but was abandoned in favor of investing in hydraulic fracturing technology (first introduced in 1947) that made vertical wells more productive. Rebirth of horizontal drillings began in the mid-1970s due to the combination of steerable downhole motor assemblies, MWD systems, and advances in radial drilling technology that made horizontal drilling investments more cost-effective. In 1996 more than 2,700 horizontal wells were dug, up from a very minimal number only a decade earlier.

In the simplest application, the borehole begins in a vertical direction and then is angled toward the intended target. The drill pipe still needs to move and rotate through the entire depth, so the deviation angle of the borehole needs to be gradual. A large deviation is made up of many smaller deviations, arcing to reach the intended target. The redirection of drill pipe is accomplished by use of an inclined plane at the bottom of the borehole. More recently, greater precision in directional drilling has been accomplished by using a mud-powered turbine at the bottom to drill the first few feet of the newly angled hole.

Horizontal and directional drilling serves four purposes. First, it is advantageous in situations where the derrick cannot be located directly above the reservoir, such as when the reservoir resides under cities, lakes, harbors, or environmentally sensitive areas such as wetlands and wildlife habitats. Second, it permits contact with more of the reservoir, which in turn means more of the resource can be recovered from the single well. Third, it makes possible multilateral drilling—multiple offshoots from a single borehole to contact reservoirs at different depths and directions. Finally, horizontal drilling improves the energy and environmental picture, since there is less of an environmental impact when fewer wells need to be drilled, and the energy future is brighter since more oil can be expected from any given well. Horizontal drilling has been credited with increasing the United States oil reserves by close to 2 percent or 10 billion barrels. Although horizontal drilling is more expensive than vertical drilling, the benefits of increased production usually outweigh the added costs.

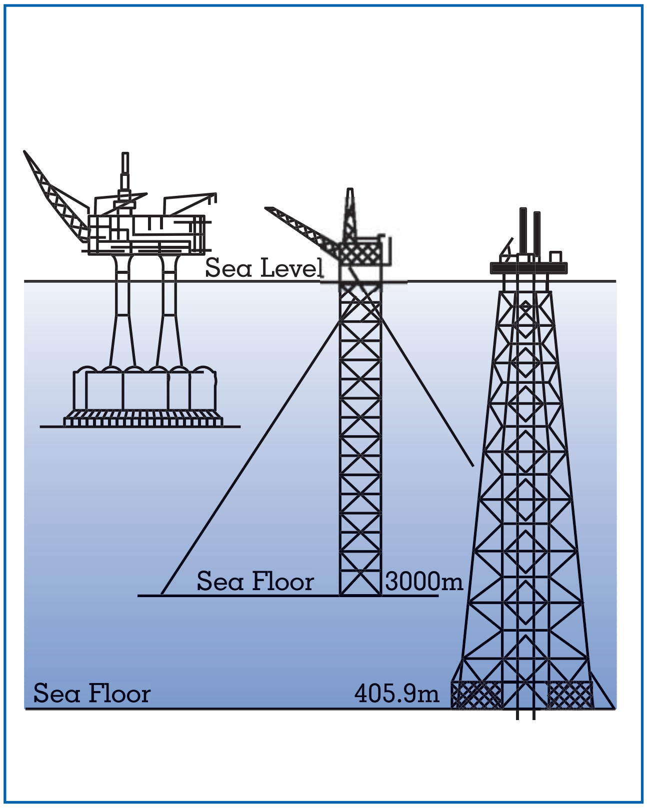


Figure 4.
Submersible base rigs. Flotational tanks built into the base are flooded so that the base rests on the ocean floor. The legs are constructed to the correct height for each site.

OFFSHORE DRILLING

Since oceans and seas cover more than two-thirds of the Earth, it came as no surprise when oil was discovered ten miles off the Louisiana coast in 1947. Fifty years later, offshore production had grown to approximately a third of all world output, most coming from the North Sea, the Persian Gulf, and the Gulf of Mexico. Technology has evolved from shallow-water drilling barges into fixed platforms and jack-up rigs, and finally into semisubmersible and floating drill ships (Figures 4–7). Some of these rigs are among the largest and most massive structures of modern civilization. And because drilling often takes place in waters as deep as 10,000 ft (more than 3,080 m) deep, as far as 200 mi (more than 300 km) from shore, and in unpredictable weather (it must be able to withstand hurricanes and uncontrollable releases of oil), the challenges facing designers, builders, and users of these rigs are daunting.

Offshore drilling operations entail much of the same technology as land-based drilling. As with onshore drilling, the offshore drill pipe must transmit rotary power and drilling mud to the bit for fixed rigs, floating rigs, and drill ships. Since mud also must be returned for recirculation, a riser (a flexible outer casing) extends from the sea floor to the platform. As a safety measure—to contend with the stress created from the motion and movement of the boat in relation to the sea bottom—risers are designed with material components of considerable elasticity.

Besides the much grander scale of offshore rigs, there are numerous additional technologies employed to contend with deep seas, winds, waves, and tides. The technical challenges include stability, buoyancy, positioning, mooring and anchoring, and rig designs and materials that can withstand extremely adverse conditions.

Shallow Water and Arctic Areas

For shallow-water drilling, usually water depths of 50 ft (15.4 m) or less, a drilling platform and other equipment is mounted on a barge, floated into position, and sunk to rest on the bottom. If conditions dictate, the drilling platform can be raised above the water on masts. Either way, drilling and production operations commence through an opening in the barge hull.

For drilling in shallow water Arctic environments, there is an added risk from the hazard of drifting ice. Artificial islands need to be constructed to protect the platform and equipment. Because of permafrost, drilling onshore in Arctic regions also necessitates building artificial islands. These usually are built from rock and gravel so that the ground does not become too unstable from permafrost melting around the drill site. To leave less of an environmental footprint, these artificial islands are increasingly being built from ice pads rather than gravel. Because the ice pads are insulated to prevent thawing, the drilling season in some Arctic areas has been extended to 205 days, and well operations to as long as 160 days. This heightens the chances of single-season completions that dramatically reduce costs, since equipment does not need to be removed to nontundra areas and brought back again.

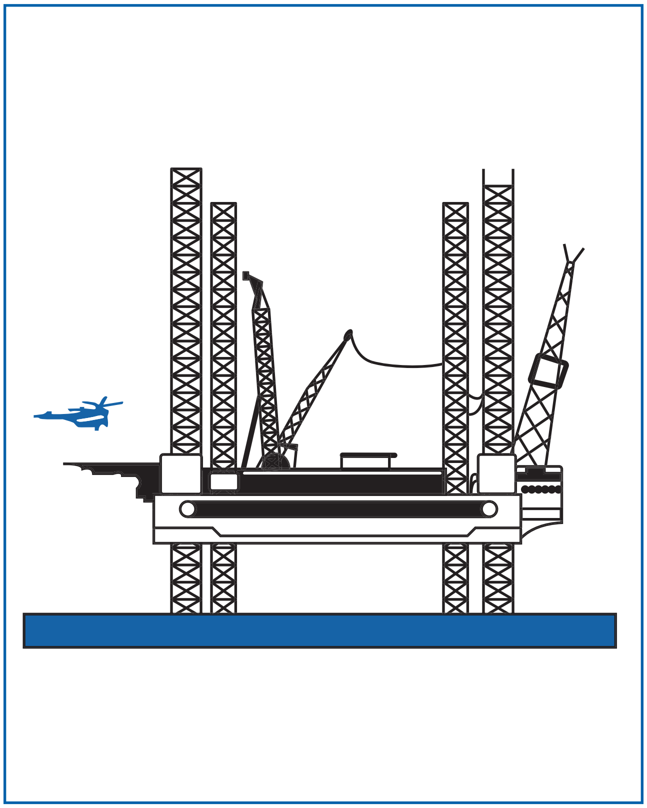


Figure 5.
Jack-up rigs. For drilling in water depths of up to 400 ft (122 m), the legs are cranked downward until they reach the sea floor and then jacked up so that the platform is 15 to 30 ft (4.6 m to 9.1 m) above the surface. Workers often reach the rig by helicopter (left).

Deep Water

For deeper drilling in open waters over the continental shelves, drilling takes place from drill ships, floating rigs and rigs that rest on the bottom. Bottom-resting rigs are mainly used for established developmental fields (up to 3,000 ft or 924 m); the semisubmersible and free-floating drill ships (up to 10,000 ft or 3,080 m) are the primary choice for exploratory drilling.

Where drilling and production are expected to last a long time, fixed rigs have been preferred. These rigs stand on stilt-like legs imbedded in the sea bottom and usually carry the drilling equipment, production

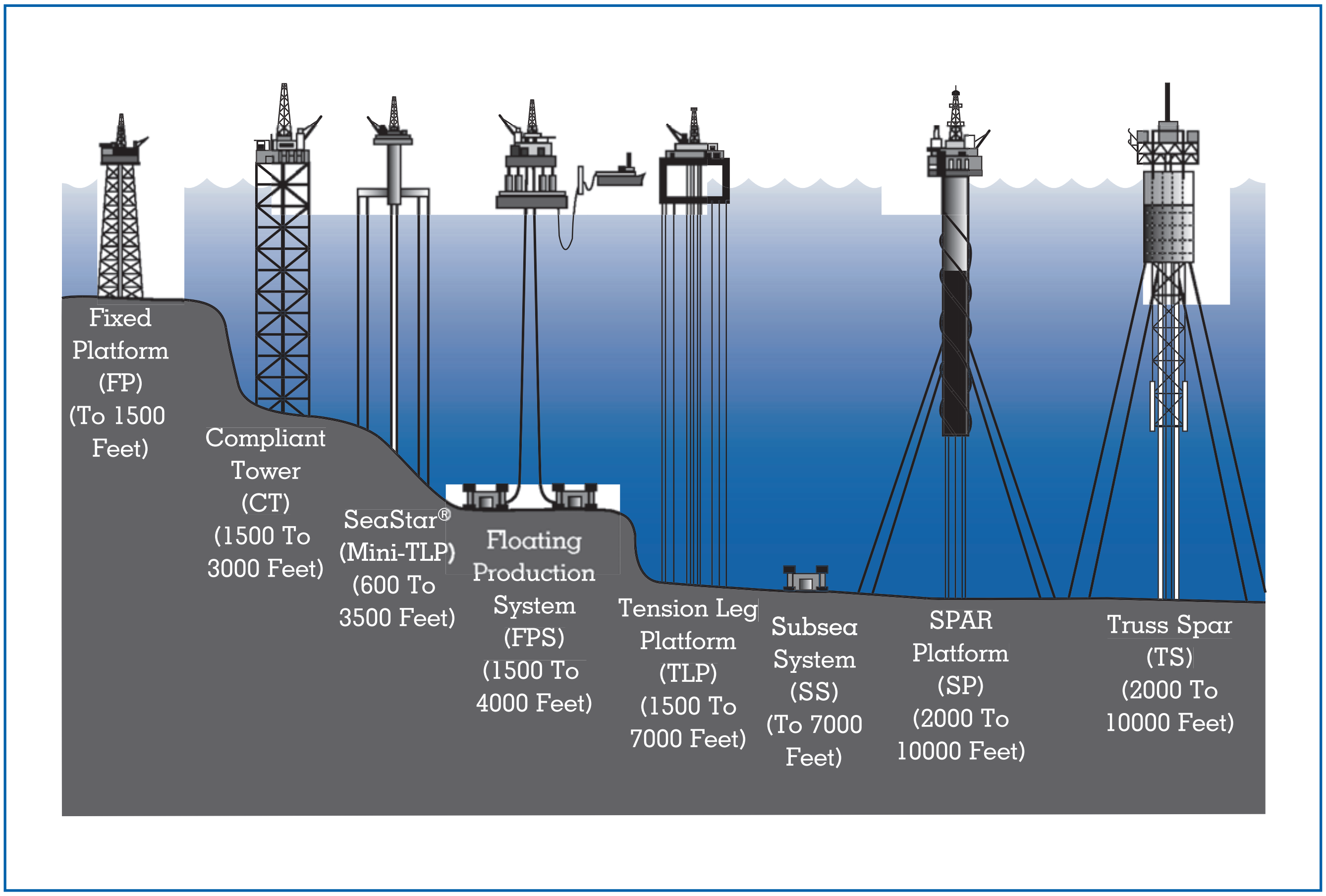


Figure 6.
Fixed rigs. Used in depths up to 1,500 ft (457 m), these rigs are imbedded in the ocean floor and tapered from bottom to top for stability. Compliant tower. Used in water depths from 1,500 to 3,000 ft (457 m to 915 m), these towers rely on mooring lines for stability.

equipment, and housing and storage areas for the work crews.

The jack-up platform, which is most prevalent, is towed to the site and the rack-and-pinion geared legs are cranked downward toward the bottom. Once the legs are stabilized on the bottom with pilings, the platform is raised 30 to 60 ft (9.2 to 18.5 m) above the surface. Another common fixed rig rests on rigid steel or concrete bases that are constructed onshore to the correct height. The rig is then towed to the drilling site, where the flotation tanks, built into the base, are flooded to sink the base to the ocean floor. For drilling in mild seas, a popular choice is a tender-assisted rig, where an anchored barge or tender is positioned alongside the rig to help with support. The hoisting equipment sits on the rig, with all the other associated support equipment (pumps, generators, cement units, and quarters) on the barge.

As more offshore exploratory operations move into deeper water—in 1997 there were thirty-one deep water rigs drilling in water depths greater than 1,000 ft (305 m) as opposed to only nine in 1990—there is sure to be more growth in the use of semi-submersible rigs and drill ships. In rough seas, semi-submersible drilling rigs are the preferred option since the hull is entirely underwater, giving the rig much more stability. The operational platform is held above the surface by masts extended upward from the deck. Cables are moored to the sea floor, with the buoyancy of the platform creating a tension in the cables that holds it in place.

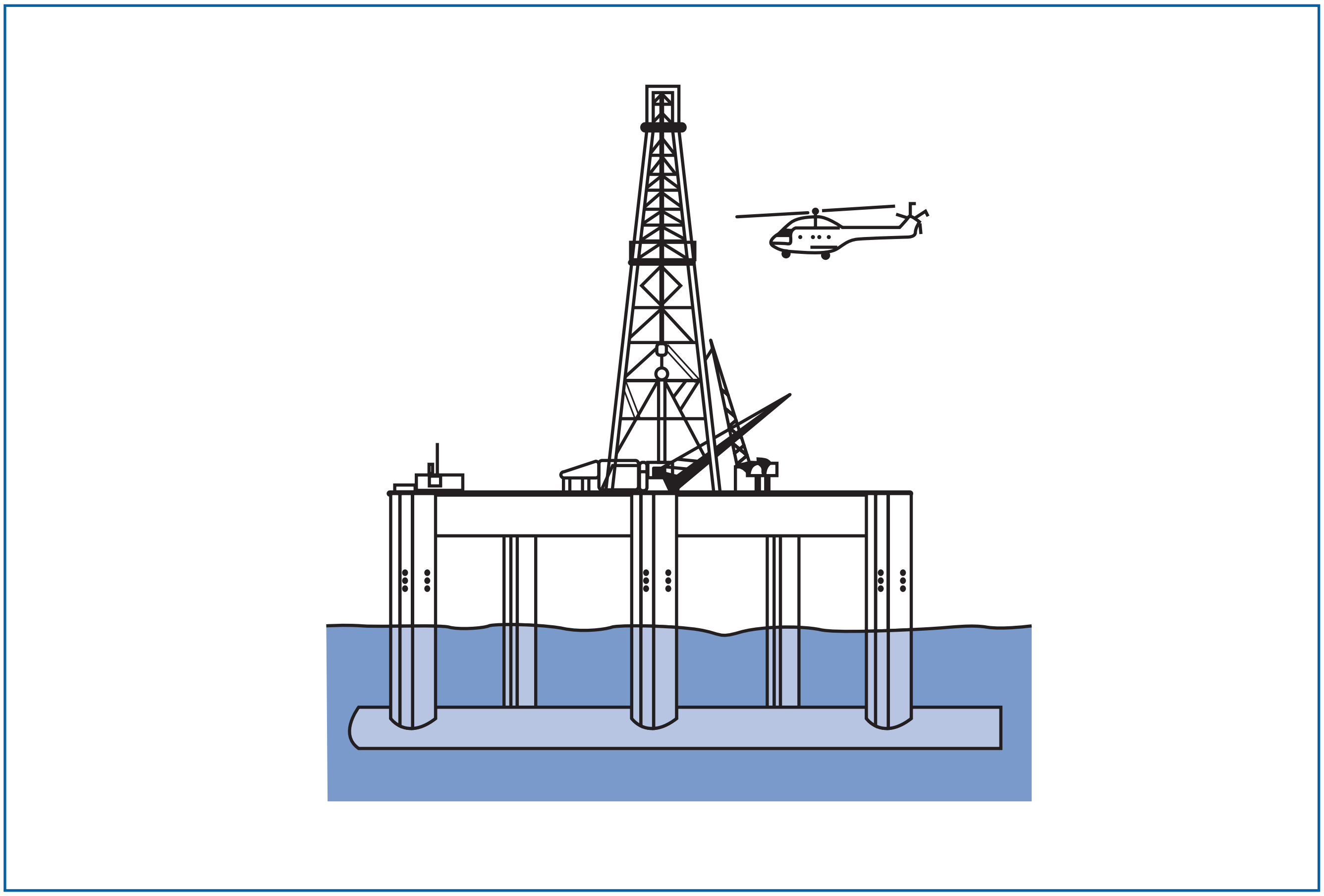


Figure 7.
Semisubmersible rigs. Huge cables and moorings anchor these floating rigs to the ocean bottom. The buoyancy of the rig creates the tension in the cables that keep the rig in place. When setting and removing mooring lines is excessively difficult or expensive, dynamic positioning systems are used instead.

When seas are less treacherous and there is greater need for mobility and flexibility, the drill ship is preferred over the semisubmersible. The drill ship has a derrick mounted in the middle over an opening for drilling operations, with several moorings used to hold the ship in position.

The use of dynamic positioning systems, instead of using moorings and cables, is becoming more popular in situations where it is expensive or extremely difficult to set and remove mooring lines. With the aid of advanced monitoring systems—gyrocompass wind sensors, global positioning systems, underwater sonar beacons, and hydroacoustic beacons—precise directional thrust propellers are used to automatically hold the vessel's orientation and position over the borehole. Despite the forces of wind, waves and ocean currents, at a water depth of 5,000 ft (1,526 m), a dynamic positioning system can reliably keep a drill ship within 50 ft (15.2 m) of the spot directly over the borehole.

Future Trends

As larger offshore discoveries become harder to find, more attention will be directed toward maximizing development from existing known reserves (workovers and recompletions), which seems to ensure a future for self-contained and jack-up rigs. There are certain to be reductions in size and weight of these new rigs. Self-elevating telescoping masts, which have improved dramatically in the 1990s, will become lighter and install more quickly. There also will be efficiency gains in equipment, and setup will be quicker, since modular units are becoming smaller, lighter, and less bulky.

The frontier remains deep water. Since the trend is toward longer-term operations, the design and materials of rigs that rest on the sea bottom will need to improve to handle the increased loads the extended reach needed in greater water depths. Productivity also will be increasing, since some of the latest ships can conduct simultaneous drilling operations using two drilling locations within a single derrick. Furthermore, because the costs of building and transporting massive drilling rigs can easily exceed several billion dollars, material research will continue to look for ways to prolong rig life cycles, and ways to better test structures regularly to avoid a catastrophic collapse. In particular, better welds, with fewer defects, will greatly extend the life cycle of offshore platforms from the continual forces of wind, waves, and ocean currents.

ANOTHER EQUIPMENT USE: CARBON DIOXIDE AND METHANE SEQUESTRATION

Since methane (CH_4) and carbon dioxide (CO_2) are greenhouse gases that governments of the world want reduced, the oil and gas industry—a major emitter of both—is looking for ways to effect such reductions. During drilling and production for oil and gas, significant amounts of CO_2 and CH_4 are produced. Where it is uneconomical to produce natural gas, this "stranded gas" is either vented as methane, flared (about 95% conversion to CO_2), converted to liquids (liquefied natural gas or methanol), or reinjected into the well. Over half of the oil exploration and production CO_2 emissions come from flaring, and almost all the methane emissions come from venting. The industry is trying to reduce emissions of both by better flare reduction practices, focusing on ways to convert more stranded gas to liquid fuels, and developing vapor recovery systems.

Another problem is when the carbon dioxide content of natural gas is too high and must be lowered to produce pipeline-quality gas. Although the current practice is to vent this CO_2, sequestration of CO_2 in underground geologic formations is being considered. Already, in the Norwegian sector of the North Sea, CO_2 has been injected into saline aquifers at a rate of 1 million tons a year to avoid paying the Norwegian carbon tax of $50 per ton of CO_2.

It may turn out that the same drilling technology that is being used to extract oil and gas, and that has been adapted for mining, geothermal, and water supply applications, will someday be equally useful in sequestering CO_2 in appropriate subsurface geologic formations.

John Zumerchik
Elena Melchert

See also: Climatic Effects; Fossil Fuels; Gasoline and Additives; Governmental Intervention in Energy Markets; Liquefied Petroleum Gas; Methane; Natural Gas, Processing and Conversion of; Natural Gas, Transportation, Distribution, and Storage of; Oil and Gas, Exploration for; Oil and Gas, Production of; Risk Assesment and Management.

BIBLIOGRAPHY

Backer, R. (1998). *A Primer of Offshore Operations*, 3rd ed. Austin, TX: University of Texas Press.

Benefits of Advanced Oil and Gas Exploration and Production Technology. (1999). Washington DC: United States Department of Energy.

Carter, G. (2000). "Offshore Platform Rigs Adapting to Weight-Space Restrictions for Floaters." *Offshore* 60(4).

Devereux, S. (2000). *Drilling Technology*. Tulsa, OK: PennWell.

Drilling for Oil, Unit 4: Activity 4-4/Teacher, pp. 4–48 to 4–53. Washington, DC: Teacher Training Program, National Petroleum Technology Office, Office of Fossil Energy, U.S. Department of Energy.

Hyne, N. J. (1995). *Nontechnical Guide to Petroleum Geology, Exploration, Drilling and Production*. Tulsa, OK: PennWell.

Pratt, J. A.; Priest, T.; and Castaneda, C. J. (1997). *Offshore Pioneers: Brown & Root and the History of Offshore Oil and Gas*. Houston, TX: Gulf Publishing Company.

U.S. Department of Energy. (1999). *Securing the U.S. Energy Environmental, and Economic Future*. Washington, DC: Author.

OIL AND GAS, EXPLORATION FOR

Exploration for oil and gas has evolved from a basic trial-and-error drilling to the application of sophisticated geophysical techniques to predict the best locations for drilling.

BASICS OF OIL AND GAS TRAPS

Oil and gas are usually associated with sedimentary rocks. The three basic types of sedimentary rocks are shales, sands, and carbonates. The shales are the sources of the hydrocarbons while the sands and carbonates act as the conduits and/or the containers.

Source Rocks

Shales often are deposited with some organic matter, for example, microscopic plant or animal matter, that become part of the rock. When the shales are squeezed under the pressure of overlying rocks and heated by the natural heat flowing from the earth, oil and gas may be formed depending on the type of organic matter involved and the temperature to which the source rock is raised. If a rock is capable of producing oil or gas, it is referred to as a source rock because it is a source for the hydrocarbons. An important element of modern exploration is making certain that a new area being explored has source rocks capable of generating hydrocarbons.

The petroleum industry relies on organic geochemists to analyze the source rocks in new exploration environments. The geochemists evaluate the potential of a source rock for producing oil and gas by determining the amount and type of organic matter as well as the amount of heating and pressure applied to the rock. The whole process of making hydrocarbons can be compared to cooking in the kitchen. The final product depends upon the ingredients and the temperature of the oven (or sedimentary basin in the case of a source rock). Geochemists perform studies of potential source rocks in order to evaluate both the thermal history and the basic ingredients.

The important ingredient in a source rock is the organic matter or kerogen (Figure 1). Kerogen originally meant "mother of oil" but is used to today to describe the organic component of sediment not solvable in common solvents. Kerogens consist of basic biochemicals including carbohydrates, proteins, polyphenols, and lipids. Oil and gas are derived from kerogens rich in lipids. These types of kerogens are often deposited in lakes or marine environments. Other types of kerogens derived primarily from land plants are more prone to producing only gas. The geochemist has to identify the type of kerogen in a source rock in order to predict the types of hydrocarbons generated.

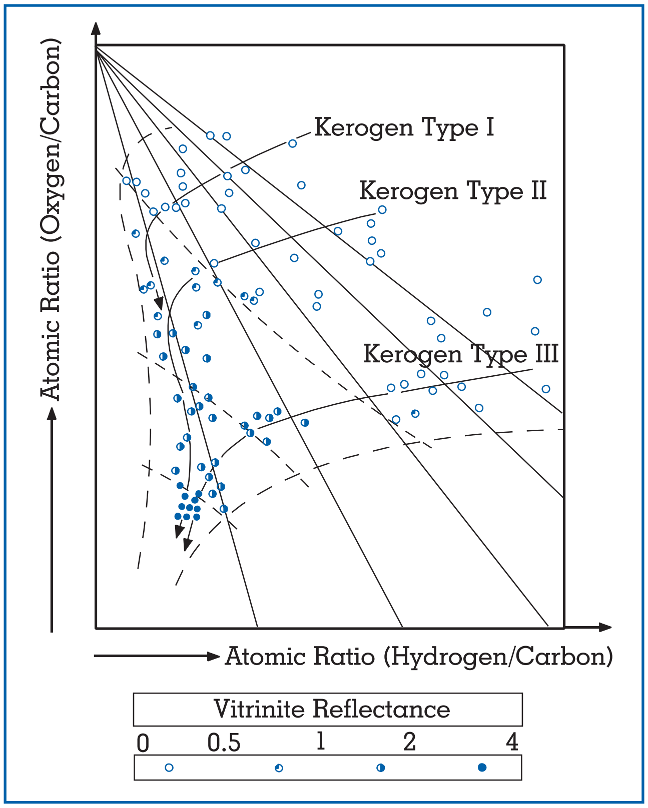


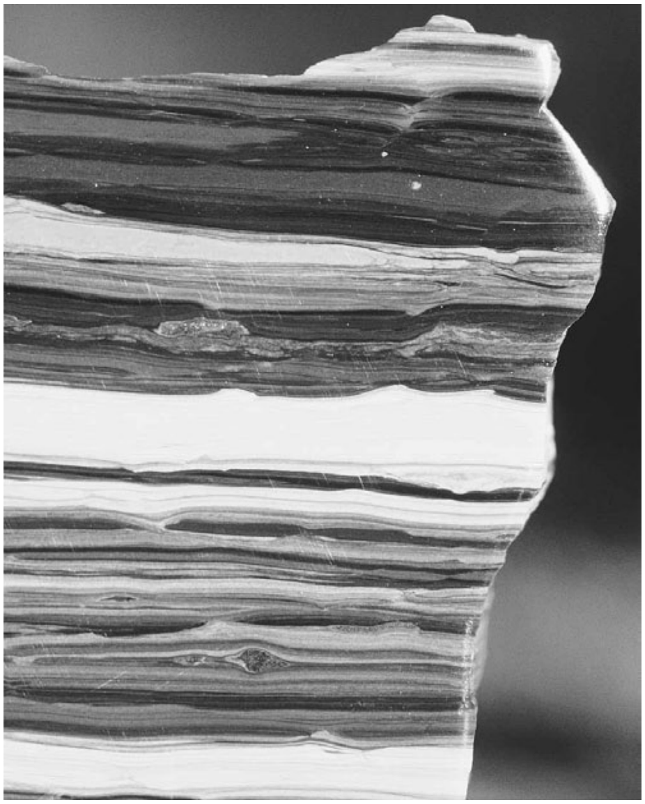
Figure 1.
General scheme of kerogen evolution from diagenesis to metagenesis in the van Krevelen diagram.

As the source rocks are buried and their temperature increases, the individual kerogens undergo chemical changes that are referred to as maturation. When the temperature reaches around 60–70° C and the rocks are buried approximately 1000 m below the surface, hydrocarbons begin to be released from the kerogen. At first oil is produced from the rock if the kerogen is the right type. Next, "wet" gas (or a gas with many different components) and finally "dry" gas (or a gas made up primarily of methane) is produced. As the rock continues to cook beyond the dry gas stage, the residual kerogen is of little use for producing oil and gas. Just as in the kitchen, too much heating can spoil even the best of ingredients.

Geochemists try to determine where hydrocarbons begin to be liberated and how their quantity and composition may vary with increasing maturity. This is equivalent to evaluating the amount and type of kerogen present in the rocks as well as the maturity (or amount of cooking in the analogy used above) for the potential source rocks in a region being explored. In addition, geochemists try to estimate when and in what directions the hydrocarbons migrated. This helps with the exploration process because some traps may have been formed too late in order to hold hydrocarbons.

One important measurement made by geochemists is the Total Organic Carbon (TOC) measurement. The results are expressed as a weight percentage. When the TOC is less than 0.5 percent, the rock is considered unlikely to have enough kerogen to produce oil or gas.

Other measurements made by geochemists try to determine the state or level of maturity of a rock. Just as with cooking, the color and appearance of kerogen changes as the source rock matures. As a result, there are several schemes for using color changes to measure the maturity of a source rock. One example is called the vitrinite reflectivity. Vitrinite is a main component of coal and is also found dispersed in source rocks. The reflectance of light from the vitrinite is measured electronically under a microscope and is denoted R_0 (the vitrinite reflectance). When R_0 is less than 0.5 to 0.7 percent, the rock is immature and still in the diagenesis stage of maturing. For R_0 greater than 0.5 to 0.7 percent to around 1.3 percent, the rock is mature and in the oil window of development. When R_0 is greater than around 1.3 percent and less than 2 percent, this stage of maturity implies the source rock is producing condensate and wet gas. When R_0 is greater than 2 percent and less than 4 percent methane is the primary hydrocarbon generated (dry gas zone). When the reflectance is greater than 4 percent, the source rock has been overcooked. Vitrinite reflectance R_0 combined with the measurement of the TOC and a determination of the type of kerogen can then be used to predict the amount and type of hydrocarbons expected from a source rock. The type or kerogen in a source rock is determined using laboratory techniques to determine the relative quantities of hydrogen and oxygen in the kerogen. For example, if the atomic ratio of hydrodren to carbon of the kerogen is plotted on a graph against the atomic ratio of oxygen to carbon, different types of kerogens can be identified. Three categories of kerogen can be identified using this type of plot. Type I kerogen is rich in lipids and is an excellent source of oil. Type II is still a good source rock but is not as rich in lipids and not as oil prone at the Type I kerogen. Type III kerogen is primarily a gas prone kerogen. Armed with results of this type of study, the geochemist is able to make predictions regarding the quality of the source rocks in an exploration area and the expected products (oil, gas, or both) expected from them. In addition, geochemists work with geologists to predict the timing and the direction of flow of hydrocarbons as they are expelled from the source rocks. This information is then used in exploration to evaluate the relative timing between trap formation and the creation of the oil. A potential oil trap might be capable of holding a great deal of oil and still not have one drop of oil because the trap was formed after the oil had migrated past the trap.



A piece of oil shale formed in the bed of Uintah Lake has been polished to reveal strata. (Corbis Corporation)

Reservoir Rocks

Once the potential for hydrocarbons has been verified in a region, explorationists look for reservoir rocks that potentially contain the hydrocarbons. Reservoir rocks are typically comprised of sands or carbonates that have space, called “porosity,” for holding the oil. Sands are present at most beaches, so it is not surprising that a great deal of petroleum has been found in sands that were once part of an ancient beach. Other environments responsible for sandstones include rivers, river deltas, and submarine fans. An example of a carbonate environment is a coral reef, much like the reef located off the coast of Florida. The sediment deposited on reefs consists of the skeletons of tiny sea creatures made up primarily of calcium carbonate. The sandstone and carbonate rocks become a type of underground pipeline for transporting and holding the oil and gas expelled from the source rocks. These reservoir rocks have pore spaces (porosity) and the ability to transmit fluids (permeability). A rock with a high porosity and high permeability is considered a desirable reservoir rock because it can not only store more petroleum, but the petroleum will flow easily when being produced from the reservoir. Finding reservoir rocks is not always easy and much of the effort of exploration is devoted to locating them. One approach to looking for reservoir rocks is to make maps over a region based on data from wells (Figure 2). Geologists can use these maps to predict the range of distribution for the reservoir rocks. For example, if the reservoir rocks were deposited by a river system, the map can show a narrow patch of sand following the meander of an ancient river channel. Because the speed of sound in sands and carbonates (reservoir rocks) is usually faster than in the surrounding shales (usually the source rock), the difference can be used to identify reservoir rocks by the reflection character and strength on the seismic data. One popular seismic method for examining the strength of the reflection coefficient as a function of source-receiver separation (referred to as "offset") is called "amplitude-versus-offset" or "AVO." Another way to identify reservoir rocks is their anisotropy when compared to the anisotropy of shales. Anisotropy is used here to describe the directional dependence of a material property (the velocity in this case). Shales have a natural anisotropy that causes the speed of sound to travel faster horizontally than vertically. Reservoir rocks tend to have a different type of anisotropy caused by fractures or no anisotropy at all (they are said to be "isotropic" in this case). When searching for reservoir rocks, the key is to separate the type of anisotropy caused by shales from the anisotropy (or lack of it) due to the reservoir rocks that are present. This approach can be used to seismically evaluate the potential for reservoir rocks in a region. Still another seismic approach to predicting the presence of reservoir rocks is based on a geological interpretation of the seismic data called "seismic stratigraphy." All of the above methods are used to identify the presence of reservoir rocks in an area being explored.

Traps

Once the petroleum has been generated and squeezed from a source rock into a neighboring sand or carbonate, the petroleum is free to move through the porous and permeable formations until it reaches the surface or meets a place beyond which the petroleum can no longer flow. When the petroleum flows to the surface, the hydrocarbons are referred to as seeps. When the hydrocarbons are prevented from reaching the surface, they are "trapped" (see Figure 3). Traps are divided into two basic categories: structural and stratigraphic. Structural traps are caused by a deformation of the earth that shapes the strata of reservoir rocks into a geometry that allows the hydrocarbons to flow into the structure but prevents their easy escape. The deformation can be caused by thrust faults such as those responsible for mountain building or normal faults such as those found in subsiding sedimentary basins such as the Gulf of Mexico. Stratigraphic traps are the result of lateral variations in layered sedimentary rocks so that porous rocks grade into nonporous rocks, that will not allow the petroleum to move any further. The easiest traps to find have been structural traps, because deformation usually affects all adjacent layers or strata. This means that the effects are easier to see on seismic data. The stratigraphic traps are the most difficult to find because they occur without affecting much of the surrounding geology. Many reservoirs are constrained by both stratigraphic and structural trapping.

GEOLOGICAL MAPPING FOR OIL AND GAS

Geologists can be thought of as the historians of the earth. History is important to exploration success. When a well is drilled, a number of devices are lowered down the well to log or identify the different formations that have been penetrated. The geologist uses the logs from many different wells to put together an interpretation of the geology between wells. A

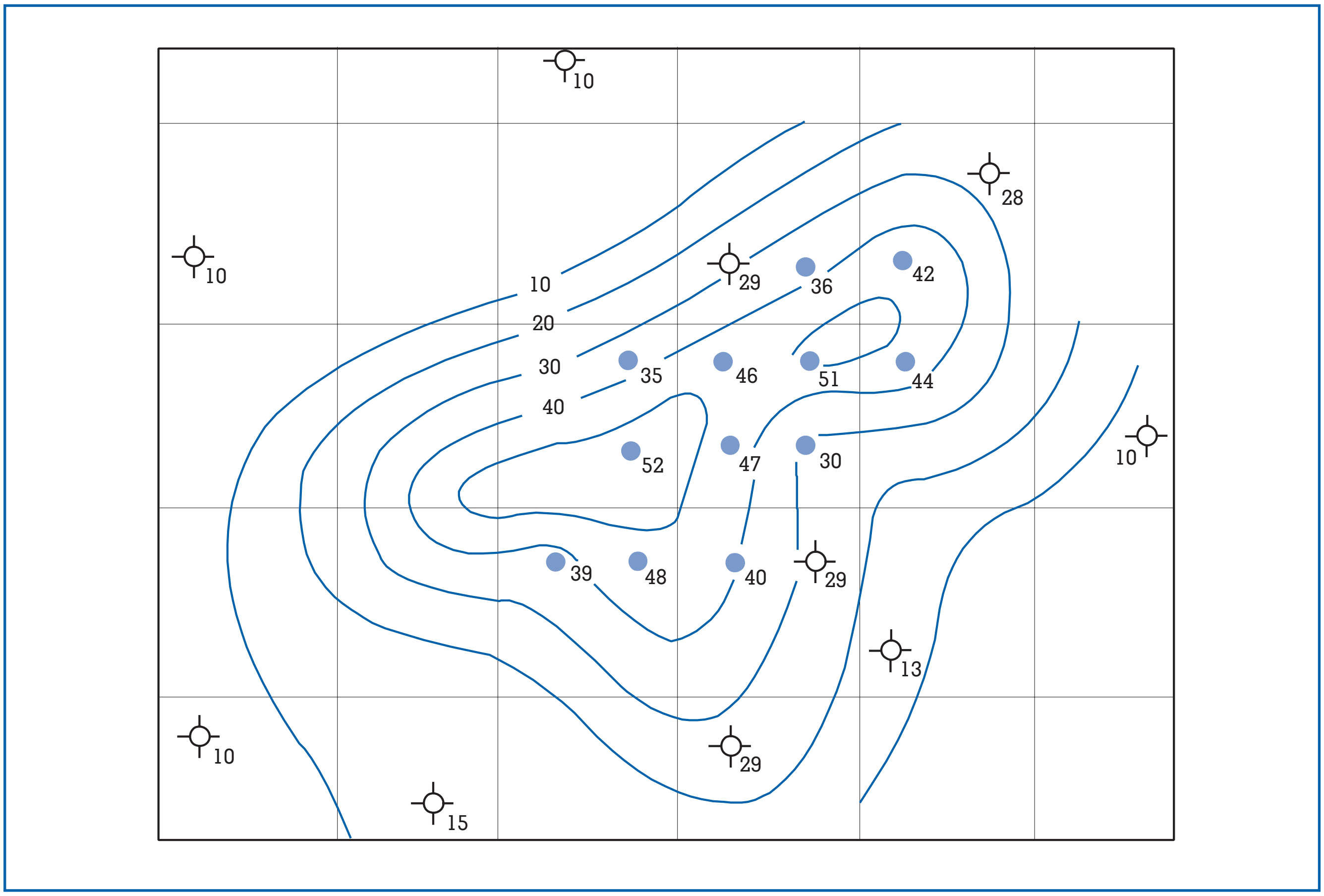


Figure 2.
Isopach maps depict thickness of a given interval.

part of the interpretation process involves making structure maps of the depth of sedimentary rock formations beneath the surface of Earth. In addition, geologists try to predict stratigraphic variations between wells using other mapping techniques. Besides being historians, geologists must solve geometric puzzles because they use the bits and pieces of information about rock layers observed from the well logs in order to form a complete picture of the earth. They can, for example compare well logs from two adjacent wells and determine if a fault cuts through one of the bore holes. The interpretation for modern geologists has been more challenging because many oil and gas wells purposely deviate from the vertical. Modem geologists employ numerous geometric tricks of the trade to assemble complicated 3-D interpretations of the earth and to identify where the reservoir rocks are located.

EARLY GEOPHYSICAL METHODS (1900s TO 1950s)

Geophysical methods are used to obtain information on the geology away from the wells where the location of oil- and gas-producing formations is known. In the early 1900s, the use of geophysics to find structural traps was accomplished by employing gravity and seismic methods. Gravity methods were the earliest geophysical method used for oil and gas exploration. The torsion balance gravimeter that was originally invented by Roland von Eotvos of Hungary (1890) later became an exploration tool. The first discovery by any geophysical method in the United States was made using a torsion balance in January 1924. Nash Dome was discovered in Brazoria County, Texas, by Rycade Oil Company, a subsidiary of Amerada. Sensitive gravity instruments measure the changes in the pull of Earth's

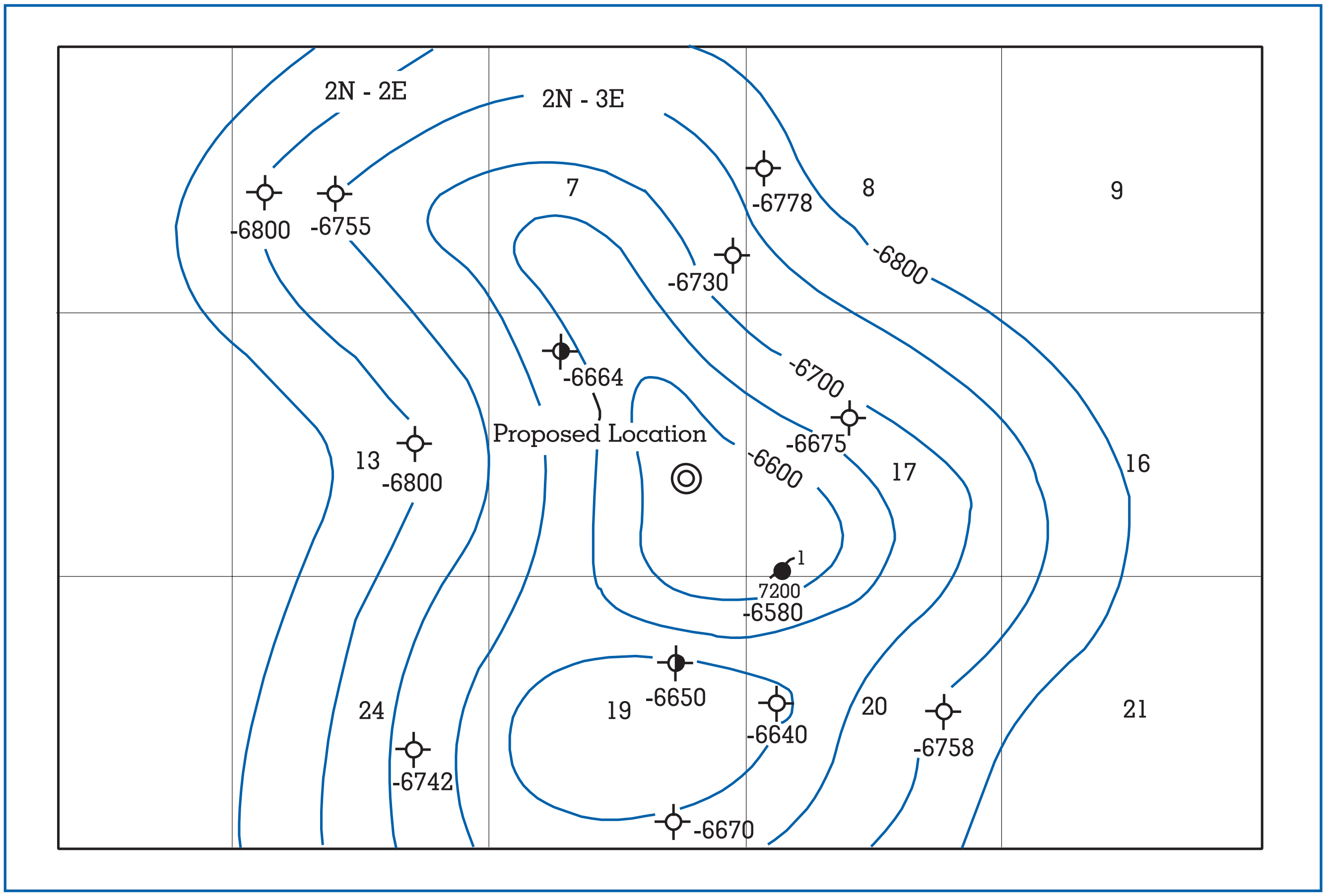


Figure 3.
Example of a structure map used to determine where oil is likely to be trapped.

gravity at different locations. The gravitational pull varies due to density variations in the different types of rocks. Because salt is lighter than most sedimentary rocks, gravity measurements were a popular approach to finding reservoirs associated with salt domes.

Sounding methods for spotting enemy artillery that were developed during World War I led to the major seismic exploration methods. On the German side, Ludger Mintrop developed the refraction exploration method that was most useful for early exploration of salt domes. Although Robert Mallet shot the first refraction survey in 1845, Mintrop recognized the commercial value of refraction surveying. He was able to verify the refraction method as an exploration tool by finding two salt domes in Germany in the period 1920–1921. In June 1924, shortly after the first torsion balance discovery, a Mintrop refraction crew (a company named "Seismos") discovered Orchard Dome in Fort Bend County, Texas. A patent filed in the United States in January 1917 by Reginald Fessenden, a Canadian by birth, laid down the fundamental ideas of using sound waves for exploration. The title of Fessenden's patent was "Methods and Apparatus for Locating Ore Bodies." In addition, the French, British, and U.S. wartime sound ranging efforts (for artillery) laid some of the foundation for the development of early seismic instrumentation and thinking developed by John Clarence Karcher. In 1921, Karcher used seismic reflections to map the depth to the top of the Viola limestone in Oklahoma. In 1927, Geophysical Research Corporation made the first seismic reflection discovery near Potawattamie County, Oklahoma. The successful well was drilled September 13, 1928. John Clarence Karcher and Everette Lee DeGolyer were the major contributors to this new exploration tool. Today the seismic

This is a team of trucks that create and record seismic vibrations, in an attempt to detect the presence of oil. (Corbis Corporation)

reflection method is the dominant form of exploration. The depth to the top of the oil and gas formations can be mapped with this technique, so the seismic reflection method can be used to directly identify structural traps.

Many other geophysical methods were attempted, but the gravity and seismic methods have survived the test of time. Gravity and seismic methods are often used in tandem because they give complementary information. This is helpful when exploring beneath salt or volcanic lava flows where seismic methods suffer difficulties. Some early work on magnetic measurements were accomplished, but magnetic surveys did not become popular until more sophisticated magnetometers were developed after World War II. When Conrad and Marcel Schlumberger joined forces with Henri Doll in the 1920s, they achieved the first geophysical logging of wells. The electric logs and other logs they developed played a major role in the early history of oil and gas exploration. The company they formed, Schlumberger Limited, has been an active contributor to numerous technological advances.

EXPLOSION OF TECHNOLOGY

In the 1950s the petroleum industry experienced an explosion of technology, a trend that continues today. One of the contributors in this effort was Harry Mayne, who developed common depth-point stacking, a method of adding seismic signals. The resulting stacked signals were then plotted in the form of a picture called a seismic section. The seismic sections, that appear as cross-sections of the earth, were initially used to find structural traps. The creation of the stacked sections benefited from digital recording that could be used to accurately record the amplitudes of seismic signals. True amplitude recording led to the discovery that the presence of hydrocarbons, especially gas, caused high-amplitude seismic reflections. These reflections, that appeared as bright spots on the seismic sections, enabled researchers to distinguish hydrocarbons from the surrounding rocks. Thus, in the 1970s bright-spot technology was developed by the petroleum industry to directly detect hydrocarbons from the surface. Shell Oil Company was the first company to use this new technology in the offshore region of the Gulf of Mexico, lowering the risk of falsely identifying oil and gas reservoirs. Unfortunately, other geological circumstances can create bright reflectors and some of these locations have been drilled.

As a result of these limitations in identifying stratigraphic traps, the industry sought a better understanding of stratigraphy. An important contribution to this understanding was the development of improved seismic methods to map stratigraphy. This improved method shaped the waveform leaving a seismic source into a shorter duration signal (the wave form is said to have been deconvolved). Enders Robinson, who had been a part of radar research efforts at MIT during World War II, introduced the deconvolution methods to the industry. Robinson applied some of the same radar technology to compress the seismic wavelets so that small changes in the stratigraphy could be mapped.

Seismic stratigraphy, originally developed by Peter Vail and his associates at Exxon, was another important contribution to the industry's ability to unravel the stratigraphy of the earth. As the industry pushed to find even more information about stratigraphic variations, methods were developed to use information about the shear-wave properties of rock as well as the compressional-wave velocities that had been the primary tool for seismic exploration. Shear waves or shear-related methods of exploration, such as AVO are used by the industry to identify changes in lithology (e.g., from sandstones to shales). These methods can also be used to evaluate bright spots or other types of reflection amplitude anomalies. For example, AVO methods are used to find rocks that have been fractured, thus making them into better reservoir rocks. In the early 1970s Jon Claerbout at Stanford University introduced an important seismic imaging principle that is still used today migrate seismic data accurately. "Migration" is a process used by geophysicists to plot seismic reflections in their true spatial positions. Today, many companies are using more sophisticated seismic imaging and analysis techniques called "inversion".

An important part of interpreting surface seismic data is the identification of the oil- and gas-bearing zones on the seismic data. Vertical seismic profile (VSPs) data are often recorded with a source on the surface and with receivers down the well to accomplish this task. In this way, the travel time to the reservoir can be measured along with other information relating the well data to the surface seismic data.

Another type of technology that has influenced oil and gas exploration is the acquisition of 3-D surface seismic data. This technology produces so much information that modern interpreters have been forced to give up looking at paper sections and use computers just to view their data. The 3-D seismic surveys are very much like a solid section of the earth. An interpreter can sit at a workstation and literally slice through the data viewing the seismic picture of the earth in almost every way possible. Some companies have assembled virtual reality rooms where the interpreter actually feels as if he or she is walking inside the earth to make the interpretation. Maps that might have taken months to produce in the early days of seismic mapping can now be accomplished in minutes. Computers are used not only for seismic processing and interpretation but for geological modeling and more accurate reservoir modeling. Integrated teams of scientists consisting of geologists, geophysicists, and petroleum engineers are pooling their talents to construct detailed 3-D models of reservoirs.

EXPLORATION TODAY

If the area being explored is a new one, the company has to make an assessment of the source rock potential. Next, they look for the reservoir rock and the trap. Sophisticated seismic methods including 3-D seismic surveys and VSPs are used to create a picture of where the oil and gas traps are located. They can also be used to determine when the traps were formed as well as the amount of oil and/or gas they contain. Seismic methods are often integrated with other geophysical measurements such as gravity, and magnetic or magnetotelluric measurements. In this way, the complementary data can be used to give a more complete picture of the geology.

EVALUATING RISK

Because financial risk is a major element of exploration, putting together a map that indicates where to drill is only part of the exploration process. Someone has to be convinced to put money into the effort to drill. Large companies go through multiple computer runs to evaluate the risks of drilling at a prospective site. Smaller companies tend to use a formula in which they multiply the probability of success times the current value of the oil that will be obtained if the well is successful. Next, they subtract from this product the probability of failure times the cost of drilling a dry well. The difference is called the "Expected Monetary Value" (EMV). If the EMV is a positive number, the well is judged to be a potentially profitable investment. Geophysical methods act to raise the probability of success for finding an economically successful well. However, the geophysical methods raise the cost of a dry well. Modern methods of geophysical exploration have raised the probability of success by as much as 30 percent in some instances over the original methods implemented in the 1920s. However, the question remains as to when the modern technology, such as 3-D seismic data, should be employed because it is more expensive. In offshore areas such as the North Sea, the deepwater Gulf of Mexico, and offshore Africa, there is the potential of finding large reserves. There is no question that 3-D data should be applied in these areas. The probability of drilling a successful well in these deep ocean areas is approximately the same as the shallow offshore areas, but what is driving deep ocean basin exploration is the potentially greater size of the reservoirs that remain to be found.

Raymon L. Brown

See also: Fossil Fuels; Oil and Gas, Drilling for; Oil and Gas, Production of.

BIBLIOGRAPHY

Allaud, L. A. (1977). *Schlumberger: The History of a Technique*. New York: John Wiley & Sons.

Curtis, D.; Dickerson, P. W.; Gray, D. M.; Klein, H. M.; and Moody, F. W. (1989). *How to Try Find an Oil Field*. Tulsa, OK: PennWell Books.

Mayne, H. W. (1989). *Fifty Years of Geophysical Ideas*. Tulsa, OK: Society for Exploration Geophysicists.

McCaslin, J. C. (1983). *Petroleum Exploration Worldwide: A History of Advances Since 1950 and a Look at Future Targets*. Tulsa, OK: PennWell Books.

Nettleton, L. L. (1940). *Geophysical Prospecting for Oil*. New York: McGraw-Hill.

Schlumberger, A. G. (1982). *The Schlumberger Adventure*. New York: Arco.

Sweet, G. E. (1966). *The History of Geophysical Prospecting*. Los Angeles: Science Press.

Tissot, B. P., and Welte, D. H. (1978). *Petroleum Formation and Occurrence*. New York: SpringerVerlag.

OIL AND GAS, PRODUCTION OF

The terms oil production and gas production refer to rates of extraction of liquid and gaseous hydrocarbon materials from natural underground deposits. Reserves and resources, on the other hand, refer to amounts of oil and gas that are present in the deposits, the difference between reserves and resources being whether or not the amounts can be economically recovered under current conditions. Supply refers to the amount of a product that becomes available for use in a given time, such as a year.

Worldwide production of crude oil increased through 1998 in response to increasing demand, although U.S. production was declining. The annual world supply of oil, which includes the annual production of crude oil and natural gas liquids grew from eighteen billion barrels in 1970 to twenty-seven billion barrels in 1998. A barrel is forty-two U.S. gallons. At the same time, the U.S. domestic supply declined from 4.3 billion barrels in 1970 to 3.5 billion barrels in 1998. So far, producing regions of the world have been able to meet the world demand without constantly escalating prices. However, the supply has become increasingly dependent on production in a few regions, notably the Middle East, where the remaining resources are concentrated. The rest of the world is becoming more dependent on these producers. Declining domestic production and growing demand have led to increasing imports of foreign oil into the United States. In 1997, one-half of U.S. consumption came from foreign imports.

Production of natural gas also has been increasing. World production of dry gas rose from sixty-six trillion cubic feet in 1987 to eighty-two trillion cubic feet in 1996. U.S. domestic dry gas production rose from seventeen trillion cubic feet in 1987 to nineteen trillion cubic feet in 1996, and the nation imported an additional three trillion cubic feet in 1996 to meet demand. Dry natural gas is produced from wellhead gas by removing most of the hydrocarbons heavier than methane. These heavy components, which tend to liquefy from the wellhead gas, are added as natural gas liquids to the oil supply and appear in the crude oil statistics.

Production of crude oil and natural gas involves technologies that have become increasingly complex as the remaining resources have become more difficult to locate and remove from their subsurface locations. Many new discoveries are made in sediments below the ocean floor in deep-water, and thus require removal of the oil and gas through long water columns. Other situations now require directional drilling of wells so that production involves transfers along wells that are far from vertical.

Most petroleum scientists believe that crude oil and natural gas formed over millions to tens of millions of years through the decomposition of organic matter buried by sediments. Generally, marine sediments have led to oil and gas, while freshwater

Gas production platform under construction in the Gulf of Thailand. (Corbis-Bettmann)

(swamp) sediments have produced coal and gas, although there are some exceptions. The liquid and gaseous hydrocarbons, in many cases, migrated through permeable subsurface rock, or along fractures, until they met an impermeable rock barrier, called a trap. There they collected to form an oil or gas reservoir. Oil and gas exploration teams using traditional and sophisticated geological techniques locate new reservoirs by studying the surface and subsurface geology in promising regions.

A reservoir is not a subterranean lake of pure oil or a cavity filled with gas. It is a porous and possibly fractured rock matrix whose pores contain oil, gas, and some water, or else, more rarely, it is a highly fractured rock, whose fractures contain the fluids. Such a reservoir is usually located in sandstone or carbonate rock. The rock matrix of an exploitable reservoir must be porous or fractured sufficiently to provide room for the hydrocarbons and water, and the pores and fractures must be connected to permit fluids to flow through the reservoir to the wells. In order to evaluate the production potential of a reservoir, a crew takes cores for analysis from appraisal wells drilled into the reservoir, and then examines properties of downhole formations using instruments lowered into the holes, a process called well logging.

When the exploration and appraisal phases at a reservoir are complete, a drilling crew drills production wells to extract the valuable hydrocarbons. Generally, a well is not left unlined. A pipe, called the casing, lines the well to protect the hole and the interior well structures, especially the tubing that transfers the oil or gas to the surface. In order to begin production, the well crew perforates the casing in the formations that contain the oil or gas and installs appropriate packing, plugs, valves, tubing, and any other necessary equipment required to remove the product.

A major factor in the production process is the driving force that moves the oil or gas to the borehole and lifts it to the surface. The fluid is, of course, under high pressure in the reservoir. If the pressure at the borehold exceeds the pressure of the rising column of oil or gas inside the tubing, it will flow out by itself. There are natural mechanisms that can provide the drive (energy) required to raise the product to the wellhead. One is the potential energy of the compressed oil or gas in the formation. The magnitude of this energy depends on the compressibility of the fluid or fluids involved. For gas, this mechanism is very effective because gas is quite compressible. As the formation gas expands into the borehole under the reduction of pressure at the borehole, it enters the tubing of the well and flows up to the surface. A gas well generally does not need to be pumped.

The situation is more complicated for oil because oil is relatively incompressible and is much denser. The natural driving pressure from the oil decompression is quickly reduced as oil flows into the borehole and, furthermore, the back-pressure produced by the heavy column of oil in the tubing is much greater than is the case for gas. Thus, one could expect this drive mechanism to be quickly lost. However, if the oil reservoir is in contact with either a large reservoir of water below it or a reservoir of gas above it, this contiguous fluid reservoir can supply the energy to maintain the drive for a much longer period. One says in such a case that the reservoir is water-driven or gas-cap-driven. There is, in fact, a third possibility called solution gas drive that originates from gas that is dissolved in the oil. When the formation pressure declines to a certain point, called the bubble point, the dissolved gas will begin to come out of solution. When this happens, the greater compressibility of the gas phase thus formed can contribute to an extended drive. However, this condition can be a mixed blessing because the gas can become the major product from the well, and it is less valuable than the oil.

The contiguous reservoirs of gas or water that contribute the drive pressures for an oil well can cause a problem. Gas and, in some cases, water, are more mobile than oil in an oil reservoir. As a result, during production the oil-gas or oil-water surface can move toward the region of reduced pressure around the borehole. If the interface reaches the borehole, the driver fluid will enter the well and be produced along with the oil. Since gas is not as valuable as oil, and water is unwanted, several measures are taken to limit ground water extraction. For example, in a gas-cap-driven reservoir, the production crew can locate the perforated section, where the oil enters the well, near the bottom of the pay zone so as to keep the collection point as far as possible from the oil-gas boundary. On the other hand, a water-driven well can be completed near the top of the zone to accomplish the same purpose. Reducing the rate of production from the well, and hence the drop in pressure around the borehole, is another way to avoid this problem.

The initial pressure in a reservoir is usually high enough to raise the oil to the surface. However, as production proceeds and the reservoir becomes depleted, the formation pressure falls and a new source of energy is required to maintain the flow of oil. A common practice is to inject either gas or water into the producing formation through special injection wells in order to maintain the pressure and move the oil toward the collection points in the producing wells. This procedure is called pressure maintenance or secondary recovery. Its success depends on maintaining a stable front between the oil and the injected fluid so that the oil continues to arrive at the collecting wells before the injected fluid. To the extent that the gas or water passes around the oil, arrives at the producing wells, and becomes part of the product coming up to the surface, the process fails to achieve its full objective. This undesirable outcome can occur, as noted above, when the mobility of the injected fluid exceeds that of the oil, as is the case for injected gas and for injected water with heavy oils. Gravity also can play a role in the success of secondary recovery depending on the geometry of the reservoir and the location of the injection and production wells.

In addition to pressure maintenance in a reservoir, there are other ways to maintain or improve well productivity as the formation pressure falls. A pump can be used to raise the oil, a process called artificial lift. One type is the familiar beam pump with its surface power unit driving up and down one end of a center-mounted beam while the other end executes the opposite down-up motion. The second end is attached to a string of “sucker rods” that extend down to the bottom of the well and operate a pump arrangement, consisting of a cylinder, plunger, and one-way valves. Other types of pumps in use eliminate the sucker rod and may have the power unit at the bottom of the well. A method for extending the period of natural flow is gas lift. In this procedure the well operators inject gas bubbles into the tubing to reduce the weight of the column of liquid that must be lifted to the surface. A third way to improve productivity is to "workover" an existing well. Mud, sand, or corrosion products may be plugging the perforations, or tubing, packing, pumps, or casing may be damaged and need repair. A workover crew will carry out repairs remotely from the surface or else remove interior components of the well for service.

Infill drilling is another possible way to extend production. For a variety of reasons, some oil may not be available to the original wells in the reservoir. Some wells may be spaced too far apart to capture the oil between them. Gas or water flooding may have bypassed some oil, or fractures or faults may block off certain parts of the reservoir from the rest so that they cannot be drained from existing wells. In these cases drilling new wells between existing ones can be an effective way to capture more of the resource.

Secondary recovery, infill drilling, various pumping techniques, and workover actions may still leave oil, sometimes the majority of the oil, in the reservoir. There are further applications of technology to extract the oil that can be utilized if the economics justifies them. These more elaborate procedures are called enhanced oil recovery. They fall into three general categories: thermal recovery, chemical processes, and miscible methods. All involve injections of some substance into the reservoir. Thermal recovery methods inject steam or hot water in order to improve the mobility of the oil. They work best for heavy oils. In one version the production crew maintains steam or hot water injection continuously in order to displace the oil toward the production wells. In another version, called steam soak or "huff and puff," the crew injects steam for a time into a production well and then lets it "soak" while the heat from the steam transfers to the reservoir. After a period of a week or more, the crew reopens the well and produces the heated oil. This sequence can be repeated as long as it is effective.

Chemical processes work either to change the mobility of a displacing fluid like water, or to reduce the capillary trapping of oil in the rock matrix pores. Reducing the mobility of water, for example by adding polymers, helps to prevent "fingering," in which the less viscous water bypasses the oil and reaches the production well. The alternate chemical approach is to inject surfactant solutions that change the capillary forces on the oil and release it from pores in which it is trapped. Another way to mobilize the oil is to inject a fluid with which the oil is miscible at reservoir conditions so that the oil trapped in the pores can dissolve in the fluid and thus be released. Carbon dioxide, nitrogen, and methane have been used for this purpose.

When oil or gas from a reservoir reaches the surface at the wellhead, it is processed to meet specifications for transportation and delivery to a customer such as a refinery or pipeline. The first step is usually to collect the output of a group of producing wells through a pipeline network into a gathering station where field processing takes place. In the case of crude oil, the objective is to produce a product relatively free of volatile gases, water, sediments, salt, and hydrogen sulfide, and fluid enough to transport to the refinery. In the case of wellhead gas, the objective is to produce a field-processed gas relatively free of condensable hydrocarbons, moisture, and various

Oil well pumps in Midway-Sunset Oil Field. (Corbis-Bettmann) gas contaminants such as hydrogen sulfide and carbon dioxide.

The field processing facilities separate the components of the wellhead liquids and gases using their differing physical and chemical properties. Tank and baffle arrangements that utilize density differences are effective for initial separation of the water, oil, and gas. Drying agents to take water out of the separated hydrocarbons, screens, or other means to remove mists of small droplets in the gas, and heat, de-emulsifying agents, or electrostatic means to eliminate emulsions can further refine the components. For wellhead gas, refrigeration of the raw gas will remove heavier molecular weight hydrocarbons, such as propane and butane, improving the quality of the purified gas and coincidentally producing valuable hydrocarbon liquids.

After field processing of wellhead products is complete, the oil and gas production phase of the industry passes into the refining stage, where the crude oil and field gas are further processed to make the products that ultimate users will purchase. The prices that refiners will pay determine the maximum allowable costs for oil and gas exploration and production. An oil or gas producer must be able to provide the product at a competitive price. For a vertically integrated oil company, no actual sale may take place, but the economics are much the same.

Oil and gas development in new fields is a somewhat risky business for several reasons. The future trend in the price of crude oil or natural gas is uncertain, especially since existing low-cost producers in the world can exert considerable monopoly power in the industry. Furthermore, the amount of product that a reservoir will produce at a given cost is uncertain, and there is a long delay between the time when the expenses of developing a new field are incurred and the time when the revenues come in from the sale of the products. These uncertainties affect the likelihood that an investment in a new prospective development will be worthwhile. Even the continuation of production in an existing field can pose uncertainty. For example, when should operators shut down operations in a declining oil field? Shutting down marginally profitable wells because the price of crude oil has dropped may be an irreversible action because of governmental decommissioning requirements, or even just the cost of restarting. Yet the price may go up again shortly so that profitable operations could resume. Thus, the business aspects of oil and gas production play a major role in decision-making in this extractive industry, and the players must be attuned to the risks involved.

George W. Hinman
Nancy W. Hinman

See also: Oil and Gas, Exploration for; Oil and Gas, Drilling for.

BIBLIOGRAPHY

Energy Information Administration. (1998). *Historical Natural Gas Annual 1930 Through 1997.* Washington, DC: U.S. Department of Energy.

Energy Information Administration. (1999). *International Petroleum Statistics Report.* Washington, DC: U.S. Department of Energy.

Guiliano, F. (1989). *Introduction to Oil and Gas Technology*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall.

Jahn, F.; Cook, M.; and Graham, M. (1998). *Hydrocarbon Exploration and Production.* Amsterdam, The Netherlands: Elsevier Science Publishers.

Royal Dutch Shell Group. (1983). *The Petroleum Handbook.* Amsterdam, The Netherlands: Elsevier Science Publishers B.V.

Tissot, B., and Welte, D. (1984). *Petroleum Formation and Occurrence*, 2nd ed. Berlin, Germany: Springer-Verlag.

OIL SHALE

See: Synthetic Fuel

ONSAGER, LARS (1903–1976)

Lars Onsager was a Norwegian-American chemist and physicist who received the 1968 Nobel Prize in Chemistry for “the discovery of the reciprocal relations bearing his name which are fundamental for the thermodynamics of irreversible processes.”

Onsager was born on November 27, 1903, to Ingrid and Erling Onsager. The family lived in Oslo (then Kristiania), Norway, where his father was a barrister. He grew up with two younger brothers, Per and Knut. Onsager's mathematical acumen was apparent early. At the age of fifteen he discovered for himself how to solve cubic equations. His school performance enabled him to skip one year at school, and at the age of sixteen he was ready for university studies. He entered the Norwegian Institute of Technology at Trondheim, where he graduated in 1925 with a chemical engineering degree.

As a freshman in Trondheim, he had studied the most recent theory, attributed to Peter Debye and Erich Hückel, of electrolytic solutions, a topic to which he returned continually throughout his life. His first contact with the scientific world was when he, at the age of twenty-two, entered Debye's office in Zürich, saying without introducing himself: "Good morning, Herr Professor, your theory of electrolytic conduction is incorrect" (Hemmer et al., 1996). Debye, impressed by Onsager's arguments, hired him as his assistant. Onsager's revision of the Debye-Hückel theory resulted in the so-called Onsager limiting law for electrolytic conduction, an important result in good agreement with experiments.

In 1928 Onsager emigrated to the United States, first to Johns Hopkins University, then in the same year to Brown University. Here he struggled with an attempt to base symmetry relations found in his work on electrolytes on general physical laws. Thus he considered more general irreversible processes than conduction and diffusion, the basic ingredients in electrolytic conduction. Irreversible transport processes are centered around fluxes of matter, energy, and electric charge and their dependence upon the forces that give rise to them. The processes are called irreversible because the fluxes are unidirectional and cannot run in the reverse direction. The main effects are well-known: diffusion (transport of matter) due to a concentration gradient or concentration difference (e.g., across a membrane), heat conduction (transport of energy) due to a temperature gradient or temperature difference, and electrical conduction (transport of electric charge) due to a potential difference.

However, a potential may give rise to more than one type of flux. There are cross-effects: A temperature difference can also result in diffusion, called thermal diffusion, and a concentration difference can result in a heat current. The general relation between fluxes J_i and the driving potentials X_i is of the form of linear relations

$$J_i = \Sigma_j L_{ij} X_j, \tag{1}$$

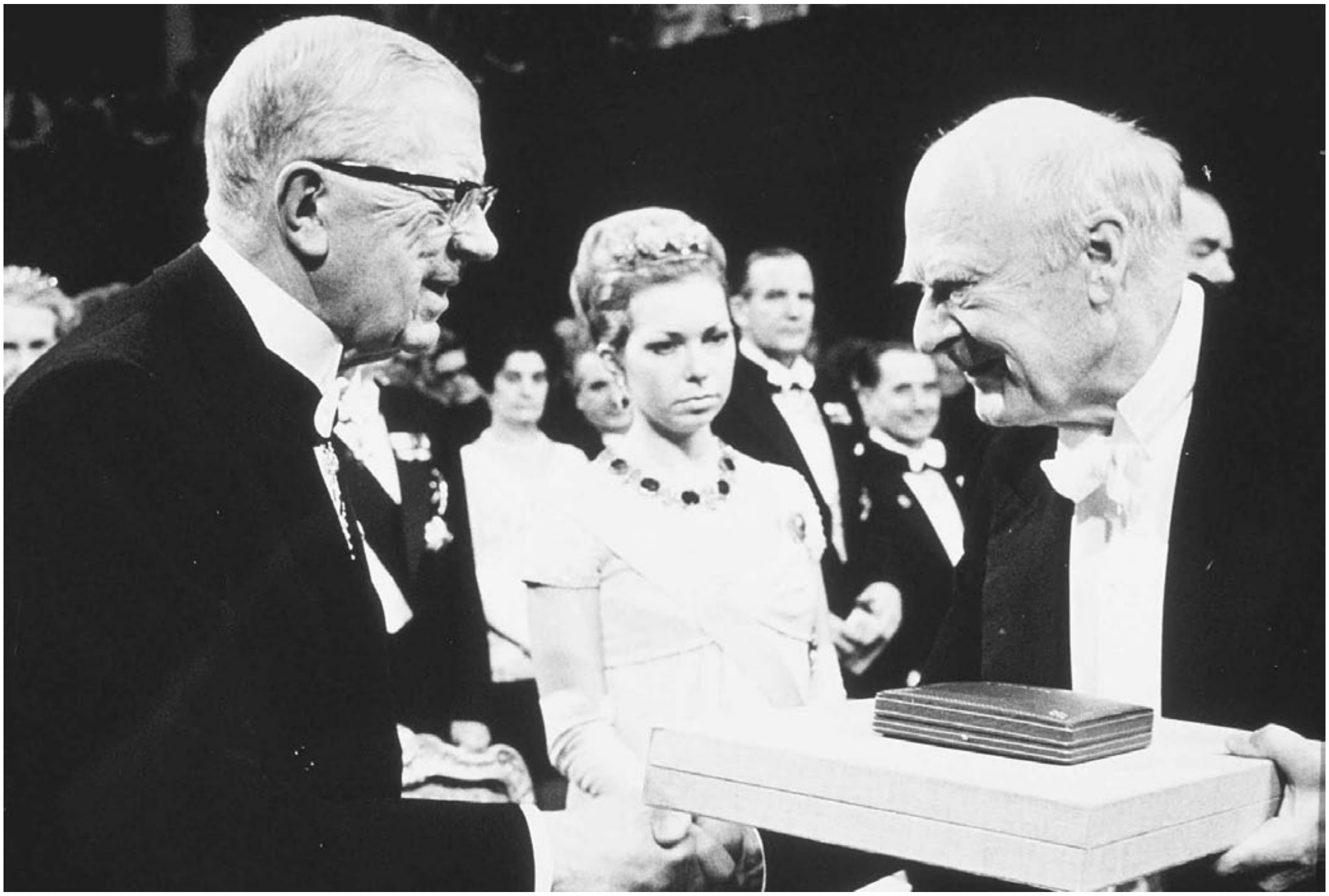
with coefficients L_{ij}, when the fluxes are not too great. Onsager showed that the coefficients of the cross-effects were equal:

$$L_{ij} = L_{ji}. \tag{2}$$

These are the famous Onsager reciprocal relations. Thus there is symmetry in the ability of a potential X_i to create a flux J_j, and of the ability of a potential X_j to create a flux J_i. The reciprocal relations are experimentally verifiable connections between effects which superficially might appear to be independent.

Such symmetries had been noticed in special cases already in the nineteenth century. One example was heat conduction in a crystal, where the X_i are the temperature gradients in the three spatial directions. Cross-effects between electric conduction and diffusion, and between heat conduction and diffusion, were also studied. Reciprocal relations of the form $L_{ij} = L_{ji}$ were confirmed experimentally for these cases, but a general principle, from which these would follow, was lacking. This situation changed dramatically in 1931 when Onsager, then at Brown University, derived the reciprocal relations. In the derivation, Onsager used the fact that the microscopic dynamics is symmetric in time, and he assumed that microscopic fluctuations on the average follow macroscopic laws when they relax towards equilibrium. The Onsager formulation was so elegant and so general that it has been referred to as the Fourth Law of Thermodynamics. It is surprising to note that the immediate impact of Onsager's work, which later earned him the Nobel Prize, was very minor, and that it was virtually ignored during the following decade.

Onsager spent the summer of 1933 in Europe. During a visit to the physical chemist Hans Falkenhagen, he met Falkenhagen's sister-in-law, Margrethe (Gretl) Arledter. They fell in love and married in September the same year. Another important event in 1933 was Onsager's move to Yale University, where he remained until his retirement. In the course of time the Onsager family grew to include one daughter and three sons. At Yale he was first a Stirling and Gibbs fellow, then assistant professor, associate professor, and finally in 1945 J. Willard Gibbs Professor of Theoretical Chemistry.



Lars Onsager (right), receiving the 1968 Nobel Prize in Chemistry from King Gustav Adolf of Sweden (left). (Archive Photos, Inc.)

Although Onsager's first appointment at Yale was a postdoctoral fellowship, he had no doctorate. It disturbed him that everybody called him "Dr. Onsager," and he decided to seek the Ph.D. from Yale. He was told that any of his published works would do for the thesis, but he felt he should write something new, and he quickly submitted a lengthy dissertation on Mathieu functions. Both the Department of Chemistry and the Department of Physics, found it difficult. The Department of Mathematics, however, was enthusiastic and was prepared to award the degree, whereupon the Department of Chemistry did not hesitate in accepting the thesis.

If Onsager's great achievement with the thermodynamics of irreversible processes met with initial indifference, Onsager's next feat created a sensation in the scientific world. In a discussion remark in 1942, he disclosed that he had solved exactly the two-dimensional Ising model, a model of a ferromagnet, and showed that it had a phase transition with a specific heat that rose to infinity at the transition point. The full paper appeared in 1944. For the first time the exact statistical mechanics of a realistic model of an interacting system became available. His solution was a tour de force, using mathematics almost unheard of in the theoretical physics of the day, and it initiated the modern developments in the theory of critical phenomena. Basing their decision especially on this work, the faculty at Cornell University nominated Onsager for the Nobel Prize in both physics and chemistry.

Although the reciprocal relations, electrolyte theory, and the solution of the Ising model were highpoints in Onsager's career, he had broader interests. From about 1940 Onsager became active in low-temperature physics. He suggested the existence of quantized vortices in superfluid helium, and provided the microscopic interpretation of the oscillatory diamagnetism in metals. In 1949, four years after he became an American citizen, he laid the foundation for a theory of liquid crystals. In his later years he studied the electrical properties of ice and took much interest in biophysics.

The importance of Onsager's work on irreversible processes was not recognized until the end of World War II, more than a decade after its publication. From then on irreversible thermodynamics gained momentum steadily, and the reciprocal relations have been applied to many transport processes in physics, chemistry, technology, and biology. The law of reciprocal relations was eventually recognized as an enormous advance in theoretical chemistry, earning Onsager the Nobel Prize in 1968. Thus more than a third of a century passed before Onsager's work was suitably recognized. In his presentation speech at the Nobel ceremonies S. Claesson said: "Here we thus have a case to which a special rule of the Nobel Foundation is of more than usual applicability. It reads: 'Work done in the past may be selected for the award only on the supposition that its significance has until recently not been fully appreciated.'"

Onsager was reluctant to publish. Many of his original discoveries appeared first in the discussion periods at scientific meetings and were not published for years, if at all. He was very much an individualist. Although he had students and coworkers, especially in his later years, he preferred to work alone. Therefore, he never created a school around him. As a person, Onsager was modest and self-effacing, with a wry sense of humor. He had an awesome memory and was at ease with history, language, and the literature of the West. Games of all kinds appealed to him.

After retirement he went to the Centre for Theoretical Studies at Coral Gables, Florida, where one of his main interests was the question of the origin of life. He died at Coral Gables on October 5, 1976.

P. C. Hemmer

See also: Thermodynamics.

BIBLIOGRAPHY

Hemmer, P.C., Holden, H., and Ratkje, S. Kjelstrup, eds. (1996) *The Collected Works of Lars Onsager*. Singapore: World Scientific.

Onsager, L. (1931) "Reciprocal Relations in Irreversible Processes." *Physical Review* 37:405–426 and 38:2265–2279.

OPTOELECTRONICS

See: Lighting

OTTO, NIKOLAUS AUGUST

(1832–1891)

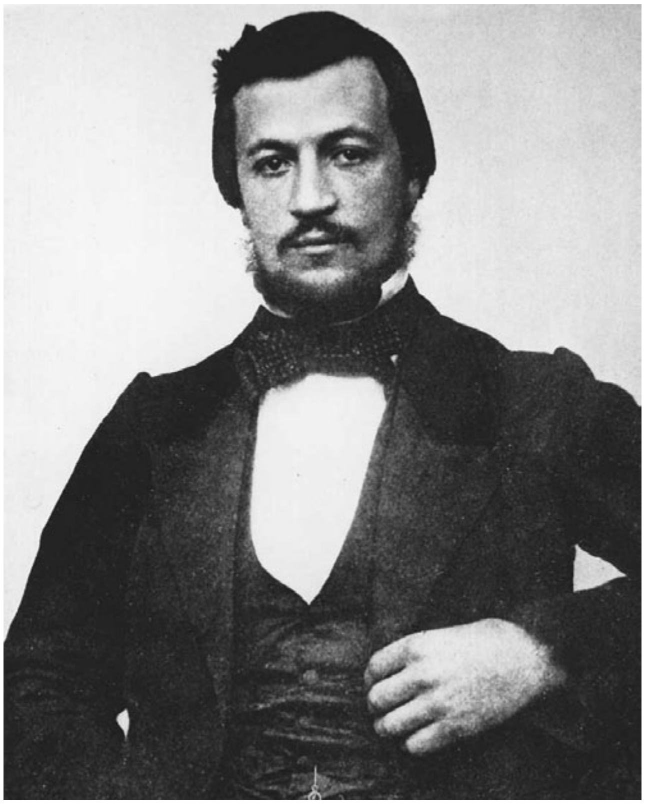
Otto was born in the village of Holzhausen, near Frankfurt. His father, an innkeeper, died when Otto was a small boy. This was a period of declining prosperity in Germany, although Otto's mother had hoped her son might receive a technical education, Otto was forced to leave school and take up employment in a Frankfurt grocery store. His brother Wilhelm, who had a textile business in Cologne, then secured him a position as a salesman in tea, sugar, and kitchen ware to grocery stores. Otto's travels took him along the German border with France and Belgium as far as Cologne where in 1859 he met Anna Gossi and began a nine-year courtship.

In 1860 Jean Lenoir built the world's first mass production internal combustion engine, a noncompression double-acting gas engine. Otto saw the potential of the Lenoir engine and perceived that the future development of the internal combustion engine would be dependant on its fuel source. In an attempt to make the engine mobile he devised an alcohol-air carburetor, and with his brother filed for a patent in 1861. Hampered by a lack of funds and a technical education, Otto spent all his leisure time in developing an engine that would surpass that of Lenoir. In 1861 Otto commissioned a Cologne instrument maker, Michael Zons, to build an experimental model Lenoir engine. Otto carried out a series of experiments with the purpose of lessening piston shock loading at the start of combustion. In one series of experiments he turned the engine over drawing in the explosive mixture, the piston was then returned to top center compressing the mixture that was then ignited. Otto found that ignition took place with great violence. This undoubtedly set the scene for the development of Otto's compression engine. In 1862

Zons built for Otto a flat four-cylinder engine. Each cylinder contained a free "cushion" piston, this engine, however, was a failure. In 1863 Otto had Zons built a one-half horsepower atmospheric engine.

In 1864 Otto's luck took a turn for the better when his engine came to the attention of Eugen Langen. Langen was a partner in a sugar refining business that he had expanded into refining sugar beets as well as cane sugar, using processing equipment of his own design. Although Langen could see the deficiencies in Otto's design he quickly made up his mind to invest in the engine. With the help of his father, Langen raised sufficient capital to form a company. In March 1864 Otto and Langen signed an agreement forming N.A. Otto & Cie, the world's first company to be set up to manufacture internal combustion engines. By 1867 the engine design was perfected and a patent applied for. The improved engine was exhibited at the 1867 Paris Exposition. After a series of tests that examined the engine efficiency and performance, the Otto and Langen engine was found to consume less than half the gas of the best noncompression engine and was awarded a gold medal. Although orders now flooded in, the partners still lacked sufficient funds. New partners were sought and Otto, whose stock holding, fell to zero, accepted a long-term employment contract. In 1872 Gasmotoren-Fabric Deutz AG came into being with Langen in control and Otto as Technical Director. Gottlieb Daimler joined the firm as production manager bringing with him Wilhelm Maybach to head the design department. Improvements were made to the engine probably by Maybach, but Daimler demanded they be patented in his name, an action that led to disputes between him and Langen. Langen also had to arbitrate numerous disagreements between the autocratic Daimler and the oversensitive Otto. Through the use of licenses and subsidiaries, Otto and Langen established an international industry with engines being built in most industrialized nations.

Around 1872 Otto returned to the problem of shockless combustion. It occurred to him, if only air was brought into the cylinder first and then the gas/air mixture, the charge would become progressively richer towards the source of ignition. Otto built a small hand-cranked model with glass cylinder to study his stratification idea by using smoke. A prototype four-stroke cycle engine was constructed in 1876. Although, at the time, Daimler thought Otto's new ideas would prove a waste of time, Langen assigned Otto an engineer and let him continue his research. Otto's ideas concerning stratification were patented in 1877. The new engine had a single horizontal cylinder and bore some similarity to the Lenoir engine. The important difference was in the admission of gas and air, also gas flame ignition. Otto seemingly had more faith in ignition by flame rather than electrical spark. The basic elements of the modern four-stroke engine are to be found in this experimental engine. A slide valve controlled air intake with a second cam operated slide valve controlling gas admission. As the piston began its outward stroke, the air valve opened and the gas valve remained closed until the piston had completed half its travel. At the end of the stroke, the valves closed and the piston began the return stoke compressing the air gas mixture. At the end of the return stroke, the gas air mixture is ignited and the piston commences its second outward stroke, the power stroke. At the end of this stroke the exhaust valve opens, and the piston returns exhausting the burnt gasses. In Otto's new design there was one power stroke for every four strokes of the piston or two revolutions of the flywheel.


Nikolaus August Otto. (Granger Collection, Ltd.)

Wilhelm Maybach redesigned Otto's prototype engine for mass production. The "Otto Silent" engine rapidly established the gas engine as a practical power source and Deutz jealously protected their position as the world's sole supplier of Otto's new engine. Through litigation and threats of action Deutz pursued any manufactures who attempted to infringe on the Otto patent. These actions, while effectively emasculating other engine builders, acted as a stimulus for basic research into the correctness of Otto's ideas on stratified charge and in turn the validity of his patent. Deutz's manufacturing competitors in endeavoring to discredit the stratified charge theories discovered an obscure patent filed by Frenchman A. B. de Rochas in 1862. Although de Rochas's patent had never been published, through a failure to pay the required fees, nor was there any evidence he had produced a working engine, his patent was used to challenge Deutz's monopoly. During litigation Otto strove to establish prior invention, but suffered, as have many other inventors, from inadequately documenting his research work. Although Otto must be considered the true inventor of the four-stroke cycle, patent law is more a matter of establishing priority of ideas than production of working machines. The Otto patents were overturned in Germany in 1886. In Britain, Otto's designs were licensed by Crossley who successfully defended their position by having the de Rochas patent ruled inadmissible.

Although his ideas were vindicated in Britain, the loss of his German patents were a blow to Otto's pride from which he apparently never recovered. Otto died five years later in 1891. Otto's only son Gustov became an engineer and aircraft designer in Munich, and in 1901 founded an airframe company that produced aircraft for World War I.

Robert Sier

See also: Engines; Gasoline Engines.

BIBLIOGRAPHY

Barlow, K. A. (1994–1995). "Nikolaus August Otto and the Four-Stroke Engine." *Transactions of the Newcomen Society* 66: suppl. no. 1.

Clerk, D. (1896). *The Gas and Oil Engine*, 6th ed. London: Longman, Green and Co.

Cummins, C. L. (1976). *Internal Fire*. Oregon: Carnot Press.

Donkin, B. (1896). *Gas, Oil and Air engines*, 2nd ed. London: Griffin and Co.

P

PARSONS, CHARLES ALGERNON (1854–1931)

Charles Algernon Parsons was born in London on June 13, 1854, the son of a wealthy, aristocratic Anglo-Irish family that was scientifically very distinguished. His father, William Parsons (third Earl of Rosse), a member of parliament, was an engineer, astronomer, and telescope-maker who had built the largest telescope in the world. His mother, Mary Countess of Rosse, best remembered as a photographer, was adept at architectural design and cast-iron foundry work.

The greater part of Parsons' childhood was spent at the family's castle and estate at Parsonstown (now Birr), County Offaly, Ireland. The workshops there were extensive, and he grew up surrounded by tools and machinery. He was privately educated by, among others, Sir Robert Ball, a man who encouraged him to spend time not just in the classroom, but also in the workshops. With these facilities and the guidance of his parents and tutor, he developed an exceptional talent for engineering.

The key to Parsons' success lies in the nature of his early training in science and engineering. In a letter concerning the design of yachts, he wrote that "they seem to design by rule of thumb in England—they should rely on model experiments, which would put them right." This is illustrative of an approach to engineering that was highly practical. He solved problems by building experiments and then meticulously observing and analyzing the results. It was an aspect of his character that was formed by the time he left Birr for his university education.

In July 1871, Parsons entered Trinity College Dublin, where he studied mathematics. In 1873, he entered St. John's College Cambridge and took his B.A. degree in mathematics in 1877. Thus, at the age of 23 he graduated from the university armed with a battery of practical and theoretical skills that allowed him to gain unique insights into the engineering problems of the time.

Having spent five years at the university honing his theoretical skills, Parsons then returned to the practical by entering the engineering firm of W. G. Armstrong & Co. in Newcastle-upon-Tyne. As a premium apprentice he was allowed to spend some of his time constructing his own experiments, provided that he paid the material and labor costs himself. Using this time, he developed a high-speed, four-cylinder epicycloidal steam engine.

The Armstrong company was not interested in producing the engine; therefore in 1881, Parsons moved to Kitson's of Leeds who took it into production. At Kitson's he occupied himself with experiments in rocket powered torpedoes that, although unsuccessful, provided useful background for his next project, the steam turbine.

Parsons left Kitson's and joined Clarke, Chapman & Co. of Gateshead as a junior partner. This was the turning point in his career, his previous working life having been devoted to learning and developing his skills. In 1884, he decided "to attack the problem of the steam turbine and of a very high speed dynamo." During the next ten years of his life, he made huge strides in the construction of dynamos and steam turbines.

In the late 1880s the Industrial Revolution was demanding ever-increasing amounts of energy. Electrical power was an obvious solution to this demand; however, it was not yet possible to generate very high power. Parsons realized that high-speed

Charles Algernon Parsons. (Library of Congress)

dynamos were the answer, and that the best source of rotational energy would be a steam turbine, whose nonreciprocating nature makes it capable of far higher speeds than conventional reciprocating engines.

Turbine designers had been experimenting with machines in which the blades were driven by a high-velocity steam jet that expanded through them in a single stage. It had been shown that steam-jet velocities of the order of 1000 m/s were required.

For a steam turbine to work efficiently, it is necessary that the velocity of the tips of the turbine blades be proportional to the velocity of the input steam jet. If the velocity of the input steam jet is high, it is necessary to increase the radius of the blades so that their tip velocity becomes correspondingly high. These long rotor blades subject the entire assembly to unmanageable mechanical stresses at high speed.

Parsons realized that he could, by using multiple rings of turbine blades on the same shaft, allow the steam to expand in stages through the turbine, and therefore extract energy more efficiently from the steam jet. This allowed him to utilize much lower jet velocities, of the order of 100 m/s, with correspondingly shorter turbine blades.

Parsons attacked the problem with characteristic gusto, so much so that by 1885 his first prototype was running successfully, giving 4.5 kW at 18,000 rpm. In 1889, frustrated by what he perceived as a lack of progress, he broke his partnership with Clarke Chapman and set up on his own in Heaton. He lost all his patent rights to the steam turbine; however, he was not daunted by this and simply developed a radial-flow machine that circumvented his own patents. In a radial-flow turbine, the steam expands outwards at right-angles to the rotational axis in contrast to the axial-flow machine where the steam expands along the length of the rotational axis.

Parsons realized that his invention was ideal for powering ships. Having received an unenthusiastic response from the admiralty, he set about proving his point in a characteristically irreverent manner. He built the *Turbinia*, a vessel powered by steam turbine and screw propellers. On the occasion of her diamond jubilee in 1897, Queen Victoria was reviewing the royal navy fleet at Spithead when Parsons appeared in the *Turbinia* and raced through the fleet at the then unbelievable speed of 30 knots. The point was well made, and on November 21, 1899, HMS *Viper* had her trials, reaching a speed of 32 knots.

Parsons was, first and foremost, a compulsive inventor. He spent his days inventing everything from children's toys to the Auxetophone, a mechanical amplifier for stringed musical instruments. His success as an inventor lies in his inquisitive nature and the fact that he was equally comfortable with theory and practice.

Parsons loved travel and the sea. It was on a cruise to the West Indies that he died aboard ship on February 11, 1931, in Kingston Harbor, Jamaica.

Bob Strunz

BIBLIOGRAPHY

Appleyard, R. (1933). *Charles Parsons*. London: Constable & Company Ltd.

Parsons, C. A. (1911). *The Steam Turbine*, Cambridge, England: Cambridge University Press.

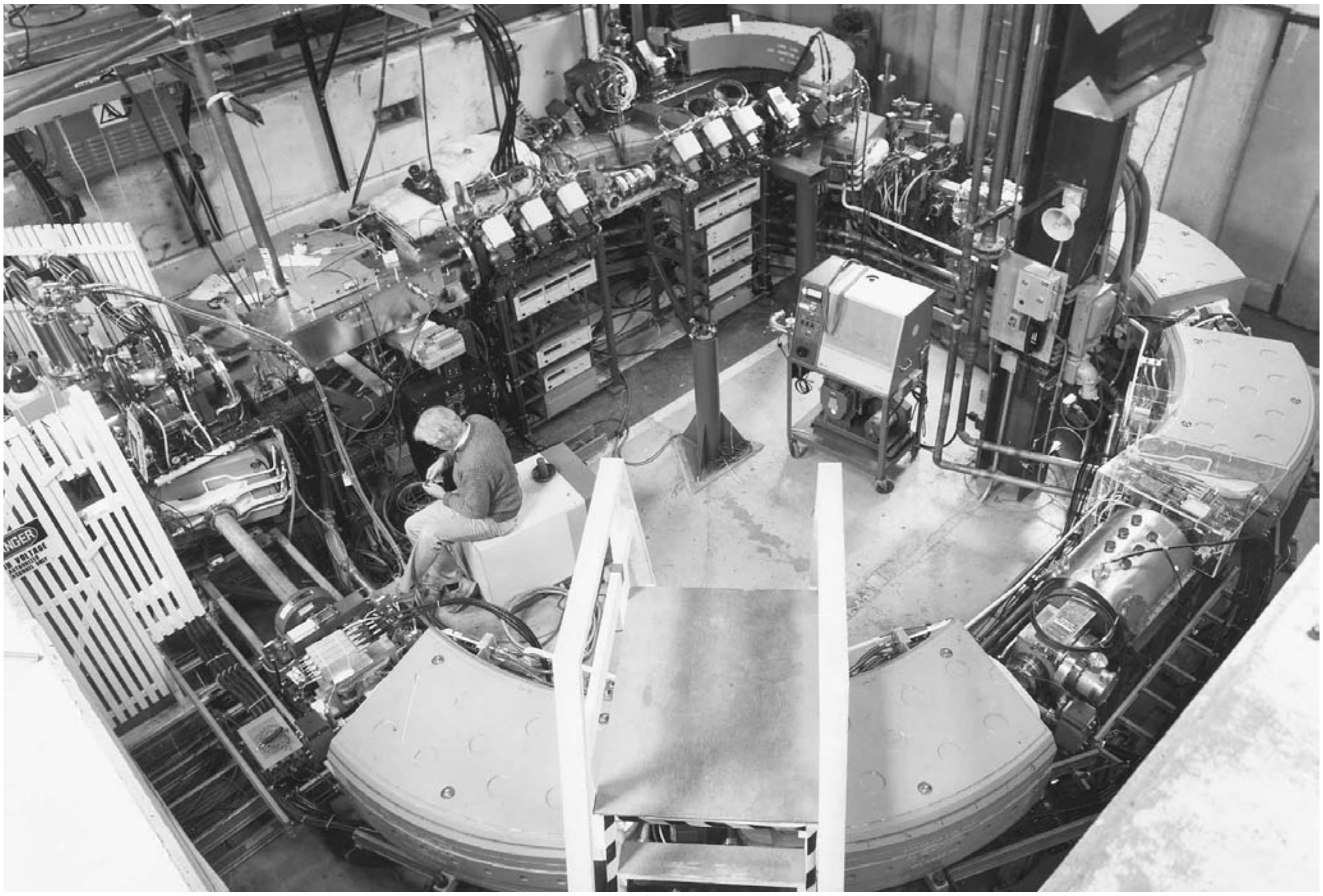
Richardson, A. (1911). *The Evolution of the Parsons Turbine*. London: Offices of Engineering.

Scaife, W. G. (1992). "Charles Parsons—Manufacturer." *Journal of Materials Processing Technology* 33:323–331.1

PARTICLE ACCELERATORS

A particle accelerator is a device that accelerates charged particles—such as atomic nuclei and electrons—to high speeds, giving them a large amount of kinetic energy. The most powerful accelerators give particles speeds almost as great as the speed of light in a vacuum (300,000 kilometers per second), which is a limiting speed that electrons and nuclei can approach but cannot reach.

If a particle is moving much slower than the speed of light, for example, one-tenth the speed of light, then its kinetic energy is proportional to the square of its speed, in accordance with Newtonian mechanics. For example, if an accelerator quadruples the kinetic energy of a slowly moving particle, its speed doubles. However, if the particle is already traveling fast, for example, at 90 percent of the speed of light, and the accelerator quadruples the kinetic energy of the particle, then the speed cannot double, or the particle would go faster than the speed of light. In this case, one has to use Einstein's more general formula for the relation between speed and kinetic energy, according to which the speed goes from 90 percent of the speed of light to a little under 99 percent. No matter how much the kinetic energy of the particle is increased by the accelerator, the particle still will travel slower than light. If the particle is traveling at 99 percent of the speed of light, no matter how much energy it is given, its speed can only increase by about 1 percent. For this reason, powerful accelerators are really particle energy boosters and not speed boosters.



Fermilab designed and built the proton therapy accelerator as the first proton accelerator specifically for the treatment of cancer. (U.S. Department of Energy)

ACCELERATOR PHYSICS

The basic unit of energy used in accelerator physics is the electron volt (eV), which is the energy acquired by an electron when accelerated through a potential difference of one volt. An electron volt is a very small unit compared to an energy unit such as a food calorie (kilocalorie). A kilocalorie is about 26 billion trillion times as large as an eV. Common multiples of eV are MeV (million eV), GeV (billion eV), and TeV (trillion eV).

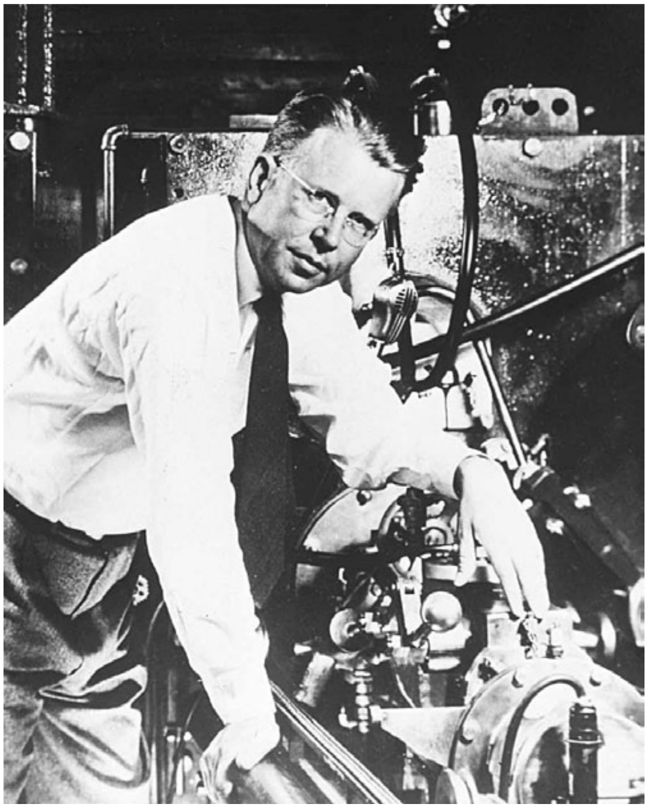
It is instructive to compare the kinetic energy acquired by a particle to its rest energy. If an airplane travels at 1000 kilometers per hour, its rest energy is more than 2 million million (2 (10^{12})) times its kinetic energy. In contrast, the highest energy electron accelerator gives an electron a kinetic energy almost 200,000 times its rest energy. The speed of such an electron is very nearly equal to the speed of light. If a burst of light and a 100 GeV electron leave a distant star at the same time, and it takes the light 2,500 years to reach the Earth, the electron will reach the earth about one second later.

The main reason particle accelerators were invented and developed is so that scientists can observe what happens when beams of particles strike ordinary matter. However, particle accelerators are also used in a wide variety of applications, such as irradiating cancers in people.

HISTORY

The development of particle accelerators grew out of the discovery of radioactivity in uranium by Henri Becquerel in Paris in 1896. Some years later, due to the work of Ernest Rutherford and others, it was found that the radioactivity discovered by Becquerel was the emission of particles with kinetic energies of several MeV from uranium nuclei. Research using the emitted particles began shortly thereafter. It was soon realized that if scientists were to learn more about the properties of subatomic particles, they had to be accelerated to energies greater than those attained in natural radioactivity.

The first particle accelerator was built in 1930 by John Cockroft and Ernest Walton. With it, they succeeded in accelerating protons (hydrogen nuclei) by means of a high-voltage source. The protons reached an energy of 300,000 eV. The following year Robert Van de Graaff succeeded in obtaining a high voltage by means of a charged moving belt. Subsequently, accelerators of the Van de Graaff type accelerated particles to more than 10 MeV.



Ernest Lawrence with cyclotron. (Corbis Corporation)

The Cockroft–Walton and Van de Graaff accelerators are linear; that is, they accelerate particles in a straight line. A short time later Ernest Lawrence got the idea to build a circular accelerator, called a cyclotron, and with the help of M. Stanley Livingston he constructed it in 1932. This first cyclotron accelerated protons to about 4 MeV. Since then, many other cyclotrons have been built, and they have been used to accelerate particles to more than 50 times as much kinetic energy as the original one. Also, other kinds of circular accelerators, such as synchrotrons, have been constructed.

OPERATION

The principle of a circular accelerator is that forces from properly arranged electromagnets cause the charged particles of the beam to move in circles, while properly arranged electrical forces boost the energy of the particles each time they go around. The radius of the circle depends on the mass and speed of

1. The first accelerator, built by J.D. Cockroft and E.T. Walton in 1930 at the Cavendish Laboratory of the University of Cambridge, England, accelerated protons to an energy of 300,000 electron volts (eV). At this energy, the protons have a speed of about 7,600 kilometers per second, and are traveling about 2.5% of the speed of light. The first circular accelerator, or cyclotron, was built by Ernest Lawrence with the help of M. Stanley Livingston in 1932 at the University of California in Berkeley.
2. The highest energy accelerator in existence today, the Tevatron, is at the Fermi National Accelerator Laboratory, located west of Chicago. It accelerates both protons and antiprotons to an energy of about 1 trillion electron volts (TeV). At this energy, the protons and antiprotons have a speed almost as great as the speed of light, which is 300,000 kilometers per second. The difference in speed is only about one part in 2 million.
3. The Large Hadron Collider is an accelerator scheduled for completion in the year 2005 at the CERN laboratory near Geneva Switzerland. It will accelerate protons to an energy of 8 TeV. At that energy, the protons will have a speed only about one part in 130 million slower than the speed of light.
4. Accelerators are responsible for many fundamental discoveries as well as many practical applications. An example of a fundamental discovery was the observation of the W and Z particles that carry the weak force that is responsible, along with the strong and electromagnetic forces, for the fact that the sun shines. An example of a practical application is the use of accelerator beams to kill cancerous tumors in patients.
5. In the coming decades, by colliding particles at ever greater energies, physicists hope to discover what causes mass to exist.

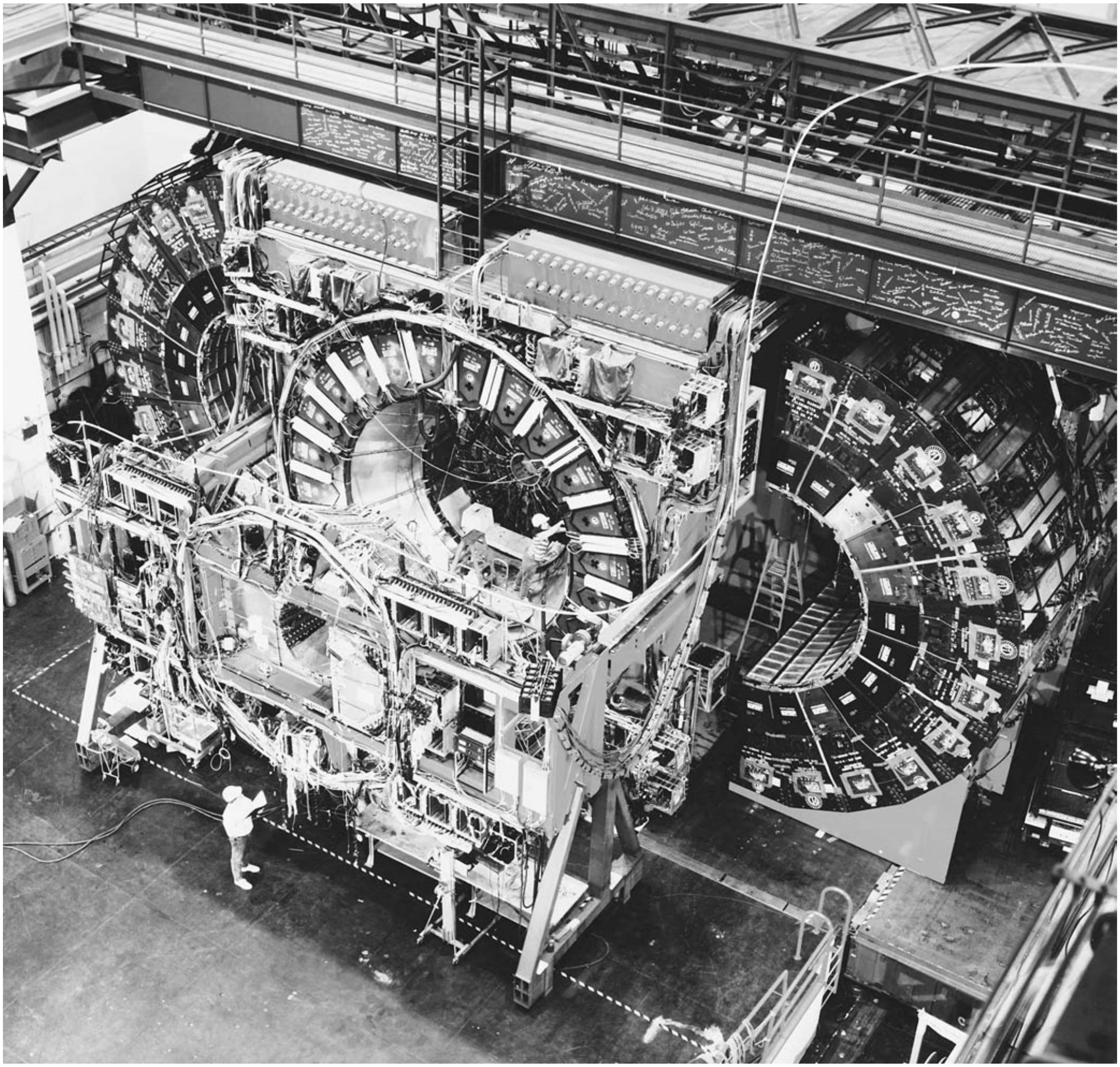
the particle and on the strength of the magnetic forces. In a cyclotron, the magnetic forces are kept constant, and as the particles gain energy they spiral outward into larger and larger circles. In a synchrotron, as the particles gain energy the magnetic forces are increased enough to keep the particles going around in the same circle.

After a beam of particles is accelerated to its final energy, it can be made to strike atoms in a target fixed in space or else collide with particles in a high–energy beam traveling in the opposite direction. Accelerators with two beams going in opposite directions are called colliders. In all particle accelerators, the particles move in an enclosed region from which nearly all the air has been removed. Otherwise, too many of the particles in the beam would collide with the molecules of air rather than with the target.

If subatomic particles moving at speeds close to the speed of light collide with nuclei and electrons, new phenomena take place that do not occur in collisions of these particles at slow speeds. For example, in a collision some of the kinetic energy of the moving particles can create new particles that are not contained in ordinary matter. Some of these created particles, such as antiparticles of the proton and electron, can themselves be accelerated and made to collide with ordinary matter.

If one wants to achieve the highest possible energy, the accelerator of choice is a collider. If two particles moving in opposite directions with the same kinetic energy collide with each other, the total kinetic energy of both particles is available to create new particles. On the other hand, when a beam strikes a stationary target, the law of conservation of momentum requires that the struck particle and any created particles carry off the same total momentum as the particle in the beam initially had. This fact requires the final particles to have significant kinetic energy—energy that cannot be used to create additional particles.

The highest energy accelerators in existence are synchrotron colliders. The one with the highest energy is called the Tevatron, and is located at the Fermi Accelerator Laboratory near Chicago. Each beam has an energy of about a TeV (hence the name), about 250,000 times the energy achieved in Lawrence's first cyclotron. The Tevatron accelerates protons in one beam and antiprotons (antiparticles of the proton) in the other beam. The antiprotons are accelerated after being created in collisions of protons with a target. The Tevatron is about six kilometers in circumfer-



The Collider Detector at Fermi National Accelerator Laboratory, home of the Tevatron. (Corbis Corporation)

ence and gives both protons and antiprotons kinetic energies of about 1,000 times their rest energies.

When an electrically charged particle is accelerated, it radiates electromagnetic energy. The higher the energy of the particle, the more energy it radiates, and, for a given energy, the smaller the mass of the particle and the smaller the radius of the circle in which it goes, the more energy it radiates. In a synchrotron, the particles go around the accelerator thousands of times, each time radiating electromagnetic energy, appropriately called synchrotron radiation.

Synchrotron radiation is not a serious problem for proton synchrotrons, but it is for high energy electron synchrotons because the mass of the electron is only about one two-thousandth as large as the mass of the proton. For this reason, a high-energy electron synchrotron has to have a large circumference or it wil radiate too much power to be economically useful. The largest electron synchrotron is located at the European Laboratory for Nuclear Research (CERN) near Geneva, Switzerland, and partly in France. Sited in an underground tunnel, it has a circumference of 27 kilometers. It gives electrons kinetic energies of about 100 GeV or 200,000 times their rest energy. After attaining this energy, the beam of electrons is made to collide with an equally energetic beam of positrons (antiparticles of electrons) going in the opposite direction.

SYNCHROTON RADIATION

If the object of a synchrotron is to accelerate electrons to the highest possible energy, synchrotron radiation is a serious obstacle that limits the energy attainable. On the other hand, the electromagnetic radiation from a synchrotron can be useful for experiments on the properties of solids and for other purposes. For this reason, some electron synchrotrons are built primarily for the synchrotron radiation they emit.

To avoid the problem of synchrotron radiation, linear electron accelerators have been built. The principle of acceleration is not a static voltage, as in the Cockroft–Walton and Van de Graaff accelerators, but a traveling electromagnetic wave that has the right phase to accelerate the electrons as they travel inside a long straight tube. The longest linear accelerator is located in California at the Stanford Linear Accelerator Center. It is two miles long and accelerates electrons and positrons to an energy of about 50 GeV, after which they are made to collide.

The 27-kilometer tunnel of the CERN laboratory also houses a proton collider that is scheduled for operation beginning in the year 2005. This collider will be the highest energy accelerator in the world, with the protons in each beam attaining energies of about 8 TeV, which is about 8000 times the rest energies of the protons. Although so far, electron synchrotrons have achieved electron kinetic energies that are the highest multiple of their rest energies, proton synchrotons, which are less limited by synchroton radiation, have achieved the highest actual particle energies.

Don Lichtenberg

See also: Matter and Energy.

BIBLIOGRAPHY

Lederman, L. (1993). *The God Particle.* New York: Houghton Mifflin.

Newton, D. (1989). *Particle Accelerators.* New York: Franklin Watts.

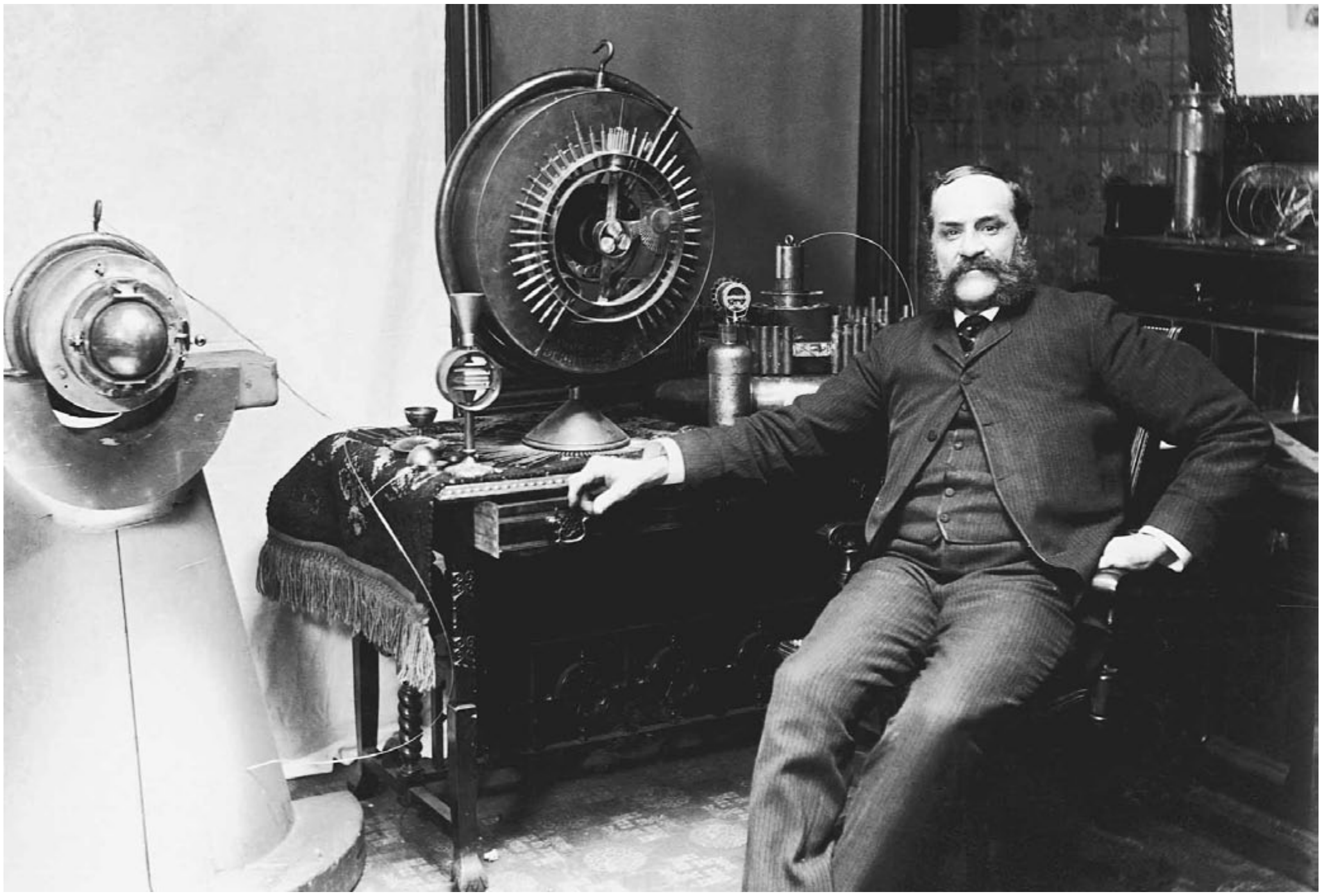
PEAK LOAD MANAGEMENT

See: Electric Power, Generation of

PERPETUAL MOTION

The idea of perpetual motion, which has been around for centuries, is to make a device that will produce more work output than the energy input—in short, to get something for nothing. Robert Fludd in 1618 was one of the first to discover it is impossible to get something for nothing. He designed a pump to drive some of the output from a water wheel to recirculate water upstream, which would then run over the wheel again. The remaining portion of the output could then be used to operate a flour mill. The only problem with this device is that it took more energy to pump the water than the entire energy output of the water wheel. Friction will always cause such a water wheel to "grind" to a halt even in the absence of doing useful work.

Would-be inventors frequently employ magnetic or electrostatic interactions because these forces are less understood (by them). For instance, a weight can be lifted by a magnetic field, perhaps produced by a superconducting magnet, with no power loss. Then, when the magnetic field is turned off or taken away, the falling weight can be harnessed to do useful work. The magnet can continually lift and drop the weight. The details can be quite complicated, but when any design is correctly analyzed it is always found that the energy required to turn the magnetic or electric field on and off, or to move the field from place, to place exceeds the work obtained by the falling weight. The first perpetual motion device using magnetic forces



John W. Keely (1827–1898) seated next to what he claimed was the first perpetual motion machine. It was proved to be a fake, but not until after he died. (Corbis Corporation)

was proposed by John Wilkins in the 1670s. Wilkins proposed using the magnetic material lodestone, not a superconducting magnet.

To prove that a particular design of a perpetual motion machine will not work can be very time consuming, and the predictable negative result has never been worth the effort. Therefore, the U.S. Patent Office has a policy to not examine applications covering perpetual motion machines unless the applicant furnishes a working model.

The nonexistence of perpetual motion machines, despite centuries of effort to design them, has been used to support the law of conservation of energy. This law is based, however, not on this negative result, but on all the experiments performed to date in which energy is carefully accounted for. It has never been observed to fail. This law is, therefore, a good basis from which to analyze perpetual motion machines. It clearly states that the goal of getting more energy output than the energy input is impossible. It also gives a basis for considering another lesser goal of perpetual motion, which is to produce a device that will run forever with no further external inputs.

There are many systems in nature that for practical purposes are perpetual. The rotation of the Earth does not change perceptibly in any person's lifetime. Very careful measurements can detect such changes, and they are predictable based on the tidal interaction with the sun and the moon. At the current rate of decrease, the Earth's rotation relative to the Sun would stop in 5.4 billion years. Systems that are unchanging for practical purposes, but which are running down ever so slowly, do not count as perpetual motion.

There are clocks that do not need winding or battery replacement. They run by taking advantage of small changes in atmospheric pressure or the movement, or by utilizing solar cells. These do not count as perpetual motion either because they use the flow of energy from other sources to maintain their motion.

The electrons of atoms in their ground state are in perpetual motion. Also, in a gas, each atom has an average kinetic energy (motion) depending on the temperature of the gas. These motions don't count either because humans demand something on their own scale that they can see or take advantage of before they consider it as true perpetual motion.

Perpetual motion itself would not defy the conservation of energy. A pendulum swinging forever at the same amplitude does not change its energy. However, energy comes in several different forms and it is impossible to keep a system in a single energy mode because of all the possible interactions with the environment.

The pendulum, for instance, will encounter air resistance, transforming its kinetic energy into random motion or heat in the air. If the pendulum were mounted in a vacuum, it would swing for a longer period of time, but any friction in the support would eventually bring it to rest. Suppose that the support could be made absolutely frictionless, would that do it? Again, no, because as the pendulum moves back and forth in the light of the room, the light pressure is greater as the pendulum moves toward the light than it is when it moves away. The light itself will slow and eventually stop the pendulum. Suppose the frictionless pendulum were mounted in a dark vacuum? That still would not do it, because the heat (infrared) radiation from the walls of the chamber will slow the pendulum the same way as the light does. The frictionless pendulum mounted in a dark vacuum chamber at absolute zero might do it if the small tidal effects gravitational interaction of the moving pendulum with the surroundings (including the Earth) could be neglected, which they cannot. Without gravitational interaction, the pendulum would not swing downward. If true perpetual motion is pursued, no interaction, no matter how small, can be neglected.

The consideration of the simple pendulum illustrates the basic problem behind devising a perpetual motion machine. The problem is the fact that energy exists in several forms and is transformed from one form to the other, especially when motion is involved. Even if friction is eliminated, there are still the electromagnetic radiation and gravitational interactions which, given the present understanding of physics, would be impossible to eliminate.

Since transformation of energy from form to form is inevitable, what about devising a system that will just pass the energy around and thus keep going forever? The problem with such a device is more subtle. It involves the second law of thermodynamics. Once the energy of a pendulum swinging in air, for example, is transferred as heat to the air, there is no way to transfer all of the heat energy back to the kinetic energy of the pendulum. According to the second law, some energy could be returned to the pendulum, but not all of it. Gradually, even with a perfect heat engine (Carnot cycle or Stirling cycle) operating to restore energy to the motion of the pendulum, its swinging would eventually die away.

There would remain some very small residual motion of the pendulum due to the air molecules striking it at random (Brownian motion), but that does not count in the "game" of perpetual motion. In the condition of residual motion, the pendulum is just another (big) molecule sharing equally in the average kinetic energy of all the individual air molecules. In other words, the pendulum eventually comes to thermal equilibrium with the air.

One statement of the second law of thermodynamics is that there is a tendency in nature for an isolated system (no external inputs) to move toward conditions that are statistically more probable (higher entropy). For the swinging pendulum, given a certain energy, there is only one amplitude at which it can swing. That is, there is only one way the pendulum can have this energy. When this energy is transformed as heat to the air, there are almost an infinite number of ways for the air molecules to contain the same energy. The swinging pendulum mode is just one mode out of many for the energy to exist in the system. Thermodynamics does not predict that all the energy can never return to the swinging mode; it just states that in the foreseeable duration of the universe it is not likely to happen because there are so many other possibilities.

Similar considerations apply to all other schemes to produce perpetual motion. No matter how it is designed, an isolated system will always tend toward a condition of thermal equilibrium, which never involves motion of objects on a human scale, and is thus not "perpetual motion."

Don C. Hopkins

BIBLIOGRAPHY

Feynman, R. P.; Leighton, R. B.; and Sands, M. (1963). *Lectures on Physics, Vol. I*, Chapters 39–46. Reading, MA: Addison-Wesley.

Ord-Hume, A. (1977). *Perpetual Motion: The History of an Obsession*. New York: St. Martin's Press.

PETROLEUM CONSUMPTION

NATURE OF PETROLEUM

Crude oil generally appears as a dark-colored liquid. It is found underground under pressure and therefore wells must be drilled in order to bring it to the surface. Part of the crude oil is in the form of a gas. The latter separates out from the oil upon reaching the lower surface pressures, and is commonly referred to as natural gas.

Structurally, oil is made up of hydrocarbon molecules containing various combinations of carbon and hydrogen. The configurations of these molecules depend on the number and arrangement of the carbon atoms. These may be linked in straight chains, branched chains, and circular, or ring formation. With respect to molecular weight, the lighter molecules (fewer than 4 carbon atoms) tend to be gases; the medium weight molecules compose liquids; and the heavier (15 or more carbon atoms) are heavy liquids and solids.

Hydrocarbons are segmented into a variety of categories. Each category possesses a distinct molecular profile and, in turn, set of chemical and physical properties. Each class of hydrocarbons therefore has historically served different markets. Crude petroleum is composed of four major hydrocarbon groups: paraffins, olefins, naphthenes, and aromatics.

The paraffins are saturated hydrocarbons with formula $C^nH^{(zn+z)}$. All possible positions on the carbon atoms that can combine with other atoms are occupied by hydrogen, and paraffins are chemically inactive under a wide range of conditions.

There are two major types of paraffin molecules found in crude oil: "normal" paraffins are those with straight or linear carbon chains; "iso-" paraffins are branched (i.e., portions of the molecule that are attached at angles to the main, linear chain). The former tend to lower the octane number of gasoline resulting in engine knock. Because of their superior burning properties, they are desirable components of diesel fuel, kerosene, and fuel oils. Compared to the normal paraffins, the iso-paraffins are selected for use in automotive and aviation fuel due to the high octane numbers they impart to gasolines.

In contrast to the paraffin group, the olefins are unsaturated compounds. Simple olefins (mono-olefins) contain a single double bond [-c=c-] with formula, C^nH^{zn}. Thus, all available positions around the carbon atoms are not occupied by hydrogen, and olefins are far more reactive than the paraffins. The "di-" olefins $C^nH^{(zn-z)}$ contain two double bonds between carbon atoms. These bonds break apart readily resulting in a highly unstable intermediate that can combine easily with other molecular types. Diolefins have a low octane number and tend to produce gummy materials that affect combustion and foul surfaces. In contrast, the mono-olefins have a high octane number and, with only a single double bond in a molecule, are stable enough to prevent the rapid buildup of gum-like residues.

The naphthenes and aromatics both have cyclic (or ring-like) molecular structures and both possess high octane numbers. Napthenes are saturated and aromatics contain alternate double bonds on their ring. They are typically found in gasoline. The naphthenes also are an important part of kerosene.

In addition to the above hydrocarbon groups, crude oil also contains a number of inorganics, including sulfur, nitrogen, oxygen, metals, and mineral salts. Not all crudes are the same. They differ in average molecular weight (boiling point, viscosity) and composition according to the geographical locations at which they are found.

PRODUCTS AND MARKETS OF CRUDE OIL

Crude oil is the source for over 3,000 petroleum-based products for both industrial and consumer applications. The technique of distillation, the first stage processing of petroleum, exploits the different boiling points of the various petroleum fractions to separate out and isolate for use the different portions of the crude. The type and proportions of hydrocarbons present in each fraction depends upon the type of crude oil used and the range of temperatures employed. The major products produced directly from crude oil from distillation processes include gasoline, kerosene, lubricants, and waxes.

These various fractions are processed further into additional products. These value-added operations generally involve chemical transformations often using catalysts. They include cracking, hydrogenation, reforming, isomerization, and polymerization. The main output from these processes is fuels and petrochemicals.

Gasoline has been the primary product extracted from petroleum since the 1920s. The major portion of the gasoline group used for automotive and aviation purposes is cracked gasoline obtained through the thermal or catalytic cracking of the heavier oil fractions (e.g., gas oil). A wide variety of gasoline types are made by mixing straight run gasoline (gasoline obtained through distillation of crude oil), cracked gasoline, reformed and polymerized gasolines with additives, such as tetraethyl lead, alcohols, and ethers.

Kerosene is heavier than gasoline and lighter than gas oil. The lighter portion of kerosene is most suitable as an illuminant for lamps. The heavier portions of kerosene traditionally have been used as stove oil. Since the 1950s, kerosene has been used as a major component in jet fuel.

Gas oil, which is heavier than kerosene, is the raw material of choice in cracking and other refinery operations. Cracking of gas oil produces a variety of fuels for automotive, industrial, and domestic (furnace) use.

The heaviest products obtained directly from oil are lubricants, waxes, asphalt, and coke. These products have both domestic and industrial uses. Lubricants, for example, are applied in the operation and maintenance of industrial equipment and machinery. Asphalt, because it is not reactive to chemicals in the environment, is a superb material of construction in the building of roads and in roofing. It is also used in the waterproofing of concrete, the manufacture of black paints, and as a material for tire threads, battery housing, electrical insulation, and other applications. The heaviest of all the petroleum products, coke, is used extensively as a major component of industrial electrodes and as a commercial fuel.

PRODUCTION PATTERNS

United States

Between the 1860s and the mid-1880s, the bulk of petroleum crude came form Pennsylvania, particularly the area near Pittsburgh. This oil, which had a high paraffin content and low level of impurities, could be readily distilled into illuminant products and lubricants.

After 1885, oil production moved into the Midwest, chiefly Ohio and Indiana. While these fields held large amounts of oil, compared to Pennsylvania, the crude was of an inferior grade. It had both a high asphalt and sulfur content. The former reduced its use as an illuminant; and the high sulfur content created a foul smell that made it unsuitable for domestic consumption. At the turn of the century, the Standard Oil Trust, adopting a new type of separations technology, was able to remove the sulfur from "sour" crude and produce commercially usable "sweet" fuel products, including kerosene.

By 1900, other products from petroleum began to take on importance. Lubricants especially became prominent. This was due to the growth of industrialization in the United States, a shortage of naturally occurring lubricants (e.g., vegetable and whale oils), and an intense and creative marketing effort on the part of Rockefeller and his Standard Oil Trust. By 1910, Standard Oil Trust was also marketing coke and asphalt to a variety of manufacturers as well as the construction industry.

During the period 1900 and 1919, there was tremendous growth in U.S. oil production. Geographically, oil production also underwent important changes. In 1900, total annual U.S. oil production reached 63.6 million barrels. By this time, major production had shifted from the Northeast to the Appalachian region and the Midwest, and in particular the famous Lima-Indiana fields. In 1900, these areas together produced over 91 percent of total U.S. oil.

By 1920, total U.S. oil production had grown seven-fold, to 442.9 million barrels. The Appalachian and Lima-Indiana fields no longer dominated oil production: in 1920, in aggregate, they accounted for less than 8 percent of total production. Now the mid-continental fields of Louisiana and Oklahoma represented over 56 percent of U.S. output. California also had grown rapidly as an oil producing region accounting for nearly a quarter of all U.S. oil production. With the opening of the Spindle Top fields in Texas, this period also saw the beginning of oil producing dominance in the Gulf states. In 1920, this region already drilled more than 6 percent of all U.S. oil. The Rocky Mountain region and Illinois also increased their status as oil producing areas; together they also accounted for over 6 percent of total U.S. oil output.

Over the next two decades, oil production within the United States continued a generally upward path as producers discovered new fields and worked established fields more intensely. By 1929, total U.S. oil output exceeded the one billion barrel per year level. Despite the Depression, and in part due to the still growing demand for automotive gasoline, oil production grew over 39 percent over the following twelve years, topping off at over 1.4 billion barrels in 1941.

At this time, the four important producing areas were the mid-continental states, The Gulf region, California, and Illinois. The Mid-Continental states continued to generate over 50 percent of the nation's oil. But its share had fallen from 1920 levels, from over 56 percent to about 51 percent. California too had seen a decline in its share, from about 24 percent to 16 percent. The fastest growing area was the Gulf region: between 1920 and 1941, its share of U.S. oil production had climbed from 6 percent to 16 percent.

Annual domestic crude oil production expanded over 50 percent between 1945 and 1959, from 1.7 billion barrels annually to almost 2.6 billion, induced by a rise in domestic oil prices. During this period, Texas was the largest single producing state followed by California, and Louisiana. These states combined accounted for 60 to 80 percent of total domestic production through the period. There was also a resurgence of production in Illinois and Kansas, and the emergence of new fields in Colorado, New Mexico, and Wyoming.

Overall, U.S. oil production continued to climb through the 1960s as the Gulf, Mid-Continental, and West Coast fields were more actively exploited. In particular, drilling went deeper and off-shore sites were mined to a greater extent than previously. By 1973, U.S. producers were shipping about 3.4 billion barrels of oil annually. A major shift in oil production came in the 1970s with the construction of the Alaskan pipeline. By the late 1980s Alaska accounted for well over 10 percent of U.S.-produced crude. By the end of the 1990s, this figure had risen to close to 19 percent, or about one-fifth of total U.S. oil production.

But overall there was a steady contraction in domestic production during the 1970s and 1980s. This unprecedented decline was the result of increasing imports of foreign oil, especially from the Middle East, Venezuela, and Mexico, and fluctuations in the price of crude worldwide, largely due to OPEC policies. By 1983, U.S. production had declined to 3.2 billion barrels per year.

The tail end of the century has seen a continued decline in U.S. oil production as a result of the precipitous fall in prices. In 1997, U.S. production stood at only 2.3 billion barrels, or less than 1959 production volume. This declining petroleum activity has continued from 1997 to early 1999, during which period the number of rotary oil drillingrigs fell from 371 to 111.

The fact remains however that the United States still has considerable untapped supplies of petroleum. It is estimated that total remaining recoverable U.S. oil from all sources may exceed 200 billion barrels, or about 70 years worth of energy, assuming the current rate of U.S. consumption. During the first decade of the twenty-first century, producers will be extracting more oil from such regions as the Gulf of Mexico and other redundant off-shore sites. Employing new types of alloys in their equipment, they will also be drilling deeper into the ground to reach more corrosive high sulfur crudes and other previously untouched reserves.

International

As the automobile came to control the market for fossil fuel, the United States increasingly dominated world oil production. From the end of the nineteenth century to the World War I period, U.S. share of world oil production grew from around 40 percent to over 70 percent. By the early 1920s, there were other oil-producing powers, notably Mexico and Russia, although they could not compete with the United States in the production and processing of petroleum.

Over the next twenty years, world production rose from nearly 1.4 billion barrels per year to over 2 billion barrels. Foreign countries greatly increased their production and refining capability. Between 1929 and 1941, U.S. exports of petroleum abroad declined (although Canada remained the leading foreign consumer of American oil). During this period, foreign production increased its world share from 34 percent to 41 percent.

This growth resulted from additional countries entering as major producers of petroleum. By 1941, Venezuela was producing over one-half of the crude oil extracted in the Western Hemisphere. During this period as well, the Middle and Near East first began to flex its muscles as an oil producing region. Iran, Iraq, the Bahrain Islands, and Saudi Arabia were the major oil-producing countries in the region. As a whole, by the end of World War II, the region supplied about 13 percent of the total world production. Middle East oil wells tend to be much more economical than those in North America. There are fewer dry holes, and each mideast well produces on average ten times the volume of a U.S. well.

Following the war, there was a steady increase in imports into the United States relative to total supply. During this period, the U.S. share of world oil production declined from 60 percent to 40 percent (1959). Crude oil imports as a percentage of total crude oil going to refineries increased from about 5 percent to 12 percent and the U.S. industry for the first time shifted its position from a net exporter to a net importer of mineral oils. At the same time, U.S. companies were establishing a strong presence abroad. Indeed, much of the oil produced abroad and imported into the United States at that time was owned by American-based companies.

Petroleum imports into the United States have continued to increase through the 1970s, 1980s, and 1990s. In terms of energy content, between 1970 and 1996, the energy contained in petroleum imported into the United States almost tripled, from 7.47 to 20.1 quadrillion Btus. Between 1970 and 1997, the Middle East countries in particular have greatly expanded their imports into the United States, from 222 million barrels to 1,349 million barrels. By the latter half of the 1990s, imports as a percentage of total domestic petroleum deliveries had reached the 50 to 60 percent level.

It is believed that world reserves may total 2 trillion barrels. This represents sufficient energy to last at the least to the end of the twenty-first century. Oil production worldwide is on the rise. Both Canada and Mexico have been expanding production. Brazil is becoming an important oil producing center. In Eastern Europe, and especially Romania, new oil is being drilled in and around the Black Sea. The former Soviet Union, especially in and around the Caspian Sea, has access to one of the world's richest oil reserves. Southeast Asia, especially Indonesia and Malaysia, are extending their petroleum production and processing capabilities. Western Europe is working regions of the North Sea. Other promising oil-producing countries include Australia, India, Pakistan, China (off-shore), and parts of the West Coast of Africa.

U.S. CONSUMPTION PATTERNS

The Nineteenth Century

From the late 1850s to the turn of the century, the major interest in petroleum was as a source of kerosene, as both a fuel and illuminant. As early as the 1820s, it was generally known from chemical experiments that illuminants, heating fuels, and lubricants could be obtained relatively cheaply from the distillation of crude oil.

In the 1840s, America's first petroleum reservoir was discovered in Tarentum, Pennsylvania. By the late 1850s, lubricants and kerosene were being extracted commercially from crude oil.

At this time, the petroleum industry was still a small scale affair and largely decentralized: production, transportation, refining, and distribution of refined products were undertaken by separate companies. Periods of overcapacity caused price declines that cut deep into the profits of producers. In the last two decades of the nineteenth century, John D. Rockefeller imposed order on the industry by building a fully integrated corporation: The Standard Oil Trust.

The Early Twentieth Century: 1900–1920

The period 1900 to 1919 set the stage for the takeoff of oil as a major energy source. The four main products for consumption extracted from petroleum were naphtha and gasoline, illuminating oil, lubricating oil, and fuel oil. During these years demand for these products increased in the wake of significant increases in the auto population, especially in the urban areas, growing industrialization, and an expanding national income.

Thus we find that between 1900 and 1919, total energy consumption more than doubled from 7,529 trillion Btus to 18,709 Btus and the percentage of total energy consumed in the United States in the form of oil increased from less than 5 percent to about 12 percent. On the other hand, the trend for coal was downward as its percentage of total energy dropped from over 89 percent to 78 percent.

Use for the various petroleum products grew at different rates. The expansion in consumption of fuel oil production was an important trend. Increasingly, fuel oil, due to its lower price, more complete consumption, and relative ease of transport, became the fuel of choice over coal in industrial, commercial, and residential facilities as well as a fuel in rail and ship transport. During these years, fuel oil output showed the greatest increase, nearly 2,387 percent, from 7.3 million barrels to 180.3 million barrels.

Even more important was the growing importance of gasoline. At the turn of the century, virtually all of the gasoline produced (6.2 million barrels) was used as a solvent by industry (including chemical and metallurgical plants and dry cleaning establishments). But by 1919, 85 percent of the 87.5 million barrels of gasoline produced went into the internal combustion engine to power automobiles, trucks, tractors, and motor boats. During this period, production of gasolines grew from 6.7 million barrels to 97.8 million barrels.

At the same time gasoline was finding its first large markets, kerosene was in the fight of its life with the burgeoning electrical power industry. (The temporary exception occurred in the rural areas without access to electricity.) Not surprisingly, gasoline and kerosene exhibited a widening price differential: between 1899 and 1919, the per gallon price for gasoline exceeded that for kerosene and the price differential grew over 86 percent, from 1.3 cents to 12.3 cents. Overall, production of Illuminating oil increased from 29.5 million barrels to 56.7 million barrels.

With respect to the various byproducts of petroleum, including paraffin wax and candles, lubricating jellies, asphalt, and coke, total value of this group grew 543 percent, from $12.3 million to $79.1 million. But this good performance would be the last major growth for the group. The one exception was asphalt which would continue in demand as a material of choice for the highway and roofing construction industry. By 1920, more than 1.3 million tons of the substance entered the market, up from the 44,000 tons produced annually at the turn of the century.

The 1920s and 1930s

Over the next ten years (1919–1929), U.S. annual energy demand grew from nearly 18.7 to 25.4 quadrillions Btus. The percentage of energy represented by petroleum increased from a little over 12 percent to 25 percent. Natural gas also nearly doubled its share, from 4.3 perent to 8 percent. Coal continued to lose ground as a source of energy, from 78 percent to 63 percent of total energy consumption.

With respect to petroleum products, during the 1920s, the amount of kerosene placed on the market remained unchanged at close to 56 million barrels annually for the period, and lubricant output increased somewhat, from 20.2 to 35.6 million barrels. Production of both gasoline and fuel oil outpaced all other petroleum products in growth. For example, output of fuel oil increased two and one half times, from near 182 million barrels to close to 449 million barrels.

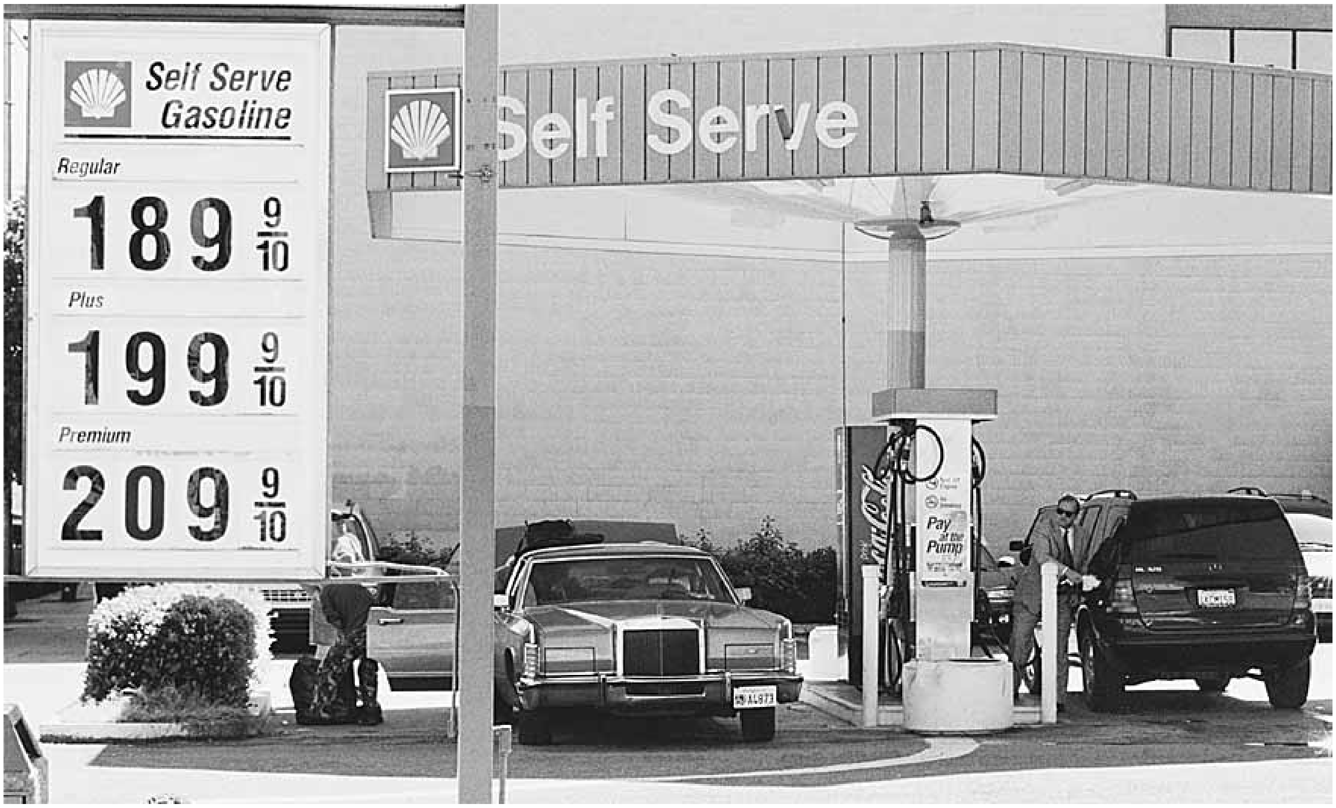
But these years put the spotlight on gasoline above all. The years following World War I saw a rapid growth in the number of automobiles on the road and in miles of highway constructed. The number of registered vehicles increased from 7.6 million to over 26.5 million and the length of surface highways expanded from 350,000 to over 694,000 mi. As a result, the average distance traveled per car grew from 4,500 mi to 7,500 mi and the average annual consumption of gasoline per car rose from 400 gal to over 590 gal. In response, during the 1920s, gasoline output expanded fourfold from 94.2 million barrels to over 435 million barrels. By 1930, gasoline accounted for almost 48 percent of all petroleum products produced in the United States.

From the onset of the Depression (1929) to the beginning of World War II (1941), petroleum's share of total U.S. energy consumption continued to expand, from 24 percent to more than 34 percent (while natural gas increased its share from 8 to 11 percent). Coal continued to lose share in the nation's energy output, from 62 percent to 54 percent. Over the same period, output of motor fuel, while not matching the growth of the previous decade, nevertheless continued its upward trend, from 256.7 million barrels to 291.5 million barrels.

It was during the 1930s that aviation solidified its position as a major and expanding gasoline market. Through the 1930s, consumption of aviation fuel increased over eightfold, from only 753 thousand barrels in 1929 to over 6.4 million barrels at the start of World War II. The other petroleum products—kerosene, lubricating oils, asphalt—saw growth as well, but behind gasoline and refinery products as a whole.

The Rise of Petrochemicals and World War II

By the start of World War II, fossil fuel products were being used increasingly to make advanced organic synthetics. The petrochemicals industry began in the years immediately following World War I. In the 1920s and 1930s, petrochemicals—chemicals made from either petroleum or natural gas—focused



The San Francisco metropolitan area had the nation's highest average price per gallon of gas as consumer costs escalated during 2000. (Corbis Corporation)

on making and utilizing for chemical synthesis the basic olefin compounds: ethylene, propylene, and the butylenes. These were generated in increasing volumes from the refinery plants as off-gases or byproducts of the cracking process. From these building blocks issued a range of chemical products and intermediates. Compared to coal, the traditional raw material, petroleum offered companies such as Union Carbide and Dow Chemical, the efficient, large-scale production of a great variety of synthetics.

Even with direct access to the basic raw materials, the refiners were slow to enter the field. Eventually, in the years leading up to World War II, the refiners began to perceive how their refining operations could be supplemented by petrochemical manufacture. By the start of world War II, they were beginning to compete in earnest with the chemical industry in petrochemical synthetics markets. Between 1929 and 1941, the byproduct refinery gases consumed by both the chemical and petroleum industries for the purpose of manufacturing chemicals more than doubled, from 38.6 million barrels to 83.4 million barrels.

This activity expanded greatly during World War II. Refiners, such as Jersey Standard, Sun Oil, Shell, and Sacony-Vacuum, pioneered the mass production of advanced and strategically critical petrochemicals including butadiene for use as raw-material for synthetic rubber, toluene for in advanced explosives, as well as high-octane motor fuel and aviation gasoline.

The Postwar Period: 1945–2000

The postwar period witnessed increased use of petroleum both for chemicals and energy. Petrochemicals enabled the production of synthetic fibers and high-performance plastics as well as synthetic resins and rubber. By the early 1950s, 60 percent of organics were produced from petroleum. By the early 1970s, over 2,000 chemical products, having a combined weight of over 25 million tons, were derived from petroleum within the United States By the 1990s, petrochemicals represented over 90 percent of all man-made commercial organics.

Petroleum made further gains as a source of energy in the United States in the form of gasoline and liquid gas. The portion of total energy consumption in the United States represented by petroleum and petroleum liquid gas rose from over 34 percent to nearly 46 percent. During these years, the production of gasoline increased from 695 million barrels to 1.5 billion barrels.

If chemicals and gasoline represent the high-growth areas for petroleum, the other petroleum products, for the most part, did not perform quite as well. The exception was kerosene. Output of kerosene rose from 100 million barrels in 1949 to over 200 million barrels in 1959, a result of increased demand in jet fuel. At the same time, demand for fuel oil declined as customers switched to natural gas.

From the 1970s through the 1990s, consumption of petroleum fluctuated closely with prices that to a large extent depended on OPEC supply policies. The OPEC oil crisis of 1973, in particular, forced oil prices upward and, in turn, caused a sharp reduction in petroleum consumption. Between 1973 to 1975, for example, in the face of high oil prices, U.S. consumption of petroleum declined to a level of 16 million barrels per day (bpd). In the 1975 to 1979 period, prices declined and consumption increased to a level near 19 million bpd. The period 1984 to 1994 was one of price reduction and growing consumption. During these years, crude oil prices declined regularly from $36/barrel to a low of $12/barrel in 1994. Then, between 1994 to 1996, with tightening supplies worldwide, prices increased to $18.50/barrel. There was another reversal during the late 1990s as prices steeply contracted, in some areas of the United States to below $10/barrel. By 1997, consumption of petroleum reached 18.6 million bpd, or the level close to that achieved by the 1978 record consumption. Overall, petroleum consumption in absolute terms in the United States has increased at an average of 0.5 percent between 1972 and 1997.

Even so, the last three decades of the twentieth century have been a period of a declining presence for petroleum in the U.S. energy picture. Between 1970 to 1996, energy consumption represented by petroleum has declined from 44.4 percent to 38.1 percent. This contraction is the result of increasing efficiency in the use of energy and growth of such alternate energy sources as nuclear, natural gas, and, more recently, coal. However, a fall in petroleum prices and the decommissioning of nuclear power plants will result in a resurgence of petroleum usage. By the year 2020 government predictions are that petroleum will comprise over 40 percent of the nation's total consumed energy.

THE TRANSPORT AND STORAGE OF PETROLEUM

The petroleum transportation and distribution network constitutes a major portion of the industry's total infrastructure and represents a large capital expenditure for producers. This means that production often outpaces demand.

Economies of scale dictate full-capacity use as companies must keep the infrastructure working at full throttle in order to maintain their investment. This is especially true of the pipeline that must be run at a throughput for which it was designed because the costs per barrel involved in operating pipelines varies inversely with level of throughput. Over the last century, the transportation and distribution network has expanded greatly all the while undergoing significant innovation.

Pipelines

The U.S. pipeline system began in the 1870s and is today composed of a maze of trunk lines and side pipelines that split off in numerous directions and over many miles. The pipeline grew over the years with the volume of oil production. Since the turn of the century, the U.S. petroleum pipeline has expanded from less than 7,000 mi to nearly 170,000 mi. The greatest growth of the pipeline system occurred in the period between the wars in the Mid-Continent and Gulf regions. This network expansion was required not simply to carry larger volumes but also because fields increasingly were further removed from population centers (and thus markets) as well as refining and distribution points.

The carrying capacity of the pipeline did not depend only on the laying down of more pipe. A number of technical improvements expanded the capacity of a given mile of pipe. Pipe making technology has extended pipe size from only 5 in. at the turn of the century up to 12 in. by the 1960s. Larger pipes carrying oil longer distances required greater pumping force. At first, this was provided by steam-powered engines at pumping stations. By the 1920s, these were replaced, at first by diesel engines, and soon thereafter electric power.

From the 1930s and well into the postwar period, pipeline technology advanced through the application of numerous improvements that increased the economies of laying down pipe and facilitated oil transport over long distances and in a variety of climatic conditions. These included new methods of welding pipe joints, the introduction of seamless tubing and pipe, new types of pipe linings and materials of construction to fight against corrosion and advance streamline oil flow, and development of automatically controlled central pipeline pumping stations and remote controlled line stations.

Shipping

Prior to 1900, water transport of oil was carried out in makeshift vessels, such as barges and converted freighters and ore-carrying ships. These were often contracted out (in contrast to pipelines that more often were owned by the producers). There was an increase in water transport of oil during the first two decades of the century.

The larger companies led the way in acquiring water transport fleets. Increasingly after World War I, producers, to be sure of sufficient capacity when required and to avoid charter prices, invested in their own tankers. In this period, about one-third of all U.S. oil was shipped by barge or tanker at some point in its trip to the refiner or market. This figure has increased to over 60 percent through the post-World-War-II period. In the years leading up to World War II, approximately 400 U.S. vessels were involved in oil transport. It is estimated that in the postwar period, the number of oil carrying vessels has fluctuated between 200 and 600 vessels.

Oil tankers today can be either single- or double-hulled. The capacity of tankers usually lies between 1 to 2 million barrels. They are often designed for special purposes. For example, in the 1970s, specially constructed tankers were used in the transport of Alaskan oil. A new generation of oil tankers is being designed as the fleet built during the 1970s and 1980s ages. Designs of the newer double-hulled tankers are most concerned with preventing oil spills, minimizing corrosion, improving maneuverability, advancing engine power and efficiency, strengthening hull structures, and installing vapor-control systems.

Storage

Oil has been stored in tanks since the late nineteenth century. Stored oil is essentially inventory waiting to be distributed to markets. Storage tanks are used to hold both crude oil and the variety of distilled and refined products.

There are two types of storage tanks: underground and aboveground. The former are typically found at service stations. Aboveground tanks are usually used to store crude oil and various refined products. They are often installed at marketing terminals, at various points along a pipeline, and at refineries. In addition, both types of storage tanks are utilized by a wide range of industrial facilities to hold oil used in fueling heir operations. It is estimated that there are 3 to 4 million tanks store oil in the United States.

Technology has improved storage facilities so that oil can be held for longer periods of time with minimal degradation of the liquid, internal corrosion and structural damage, evaporation loss, and leakage. Storage of crude is an expensive operation. Evaporation, corrosion, and leakage increase the costs of storing oil, as do insurance expenses as well as opportunity costs involved in keeping inactive large amounts of petroleum. Accordingly, producers attempt to reduce their inventories of crude and so minimize their reliance on storage. This means the maintaining of continuous flow conditions from well to refinery.

PETROLEUM AND THE REGULATORY CLIMATE

Since the 1970s, the petroleum industry has been increasingly affected by the regulatory push at the international as well as federal and state levels. Oil spills, such as the one involving the Exxon Valdez in Alaska in the 1980s, put the regulatory spotlight on the environmental dangers inherent in moving oil by marine tanker. Since the 1970s, The International Maritime Organization (IMO), in conjunction with the U.S. Department of Transportation (DoT) and the U.S. Environmental Protection Agency (EPA), has stepped up its efforts to implement measures to increase protection of the environment on U.S. and international waters.

The pipelines and storage facilities also present environmental hazards, mainly from leakage and subsequent seepage of oil into the soil and groundwater. Pipeline regulations are the responsibility of the EPA and the DoT. DoT's Office of Pipeline Safety in particular is responsible for regulating interstate hazardous liquid pipeline systems. With respect to storage tanks, both the under- and above-ground types are regulated by the EPA under the Resource Conservation and Recovery Act (RCRA). The major regulation affecting oil-bearing underground storage tanks (USTs), effective on December 22, 1998, required the upgrade of tanks including improved leak detection systems as well as measures to prevent future corrosion of tanks.

Sanford L. Moskowitz

See also: Diesel Fuel; Fossil Fuels; Gasoline and Additives; Kerosene; Liquified Petroleum Gas; Oil and Gas, Drilling for; Oil and Gas, Exploration for; Refining, History of; Residual Fuels.

BIBLIOGRAPHY

Amuzegar, J. (1999). *Managing the Oil Wealth: OPEC's Windfalls and Pitfalls*. New York: I. B. Tauris and Co. Ltd.

Bromley, S. (1991). *American Hegemony and World Oil: The Industry, the State System, and the World Economy*. University Park: Pennsylvania State University Press.

Business Trend Analysts, Inc. (1998). "The Consumption of Specialty Steel in U.S. and International Petroleum and Petrochemical Markets." In *The U.S. Stainless Steel Industry*. Commack, NY: Business Trend Analysts, Inc.

Department of Commerce. (1998). *Statistical Abstracts of the United States*. Washington, DC: U.S. Government Printing Office.

Economides, M., et al. (2000). *The Color of Oil: The History, the Money and the Politics of the World's Biggest Business*. Houston: Round Oak Publishing Company.

Enos, J. (1962). *Petroleum, Progress and Profits: A History of Process Innovation*. Cambridge, MA: MIT Press.

Giebelhaus, A. W. (1980). *Business and Government in the Oil Industry: A Case Study of Sun Oil, 1876–1945*. Greenwich, CT: JAI Press.

Hartshorn, J. (1993) *Oil Trade: Politics and Prospects*. New York: Cambridge University Press.

International Energy Association. (1996). *Global Offshore Oil Production: Prospects to 2000*. Paris: Author.

Larson, H. M., et al. (1971). *History of Standard Oil Company (New Jersey): New Horizons, 1927–1950*. New York: Harper and Row.

Moskowitz, S. L. (1999). "Science, Engineering and the American Technological Climate: The Extent and Limits of Technological Momentum in the Development of the U.S. Vapor-Phase Catalytic Reactor, 1916–1950." Diss. Columbia University.

Spitz, P. (1988). *Petrochemicals: The Rise of an Industry*. New York: John Wiley and Sons.

Waddams, A. L. (1980). *Chemicals from Petroleum*, 4th ed. Houston: Gulf Publishing Co.

Williamson, H. F., et al. (1959). *The American Petroleum Industry, Vol. I: The Age of Illumination, 1859–1899*. Evanston, IL: Northwestern University Press.

Williamson, H. F., et al. (1963). *The American Petroleum Industry, Vol. II: The Age of Energy, 1899–1959*. Evanston, IL: Northwestern University Press.

PHANTOM LOAD

See: Electricity

PHOTOSYNTHESIS

See: Biological Energy Use, Cellular Processes of

PIEZOELECTRIC ENERGY

Piezoelectric energy is a form of electric energy produced by certain solid materials when they are deformed. (The word *piezo* has its roots in the Greek word *piezein* meaning "to press.") Discovery of the piezoelectric effect is credited to Pierre and Jacques Curie who observed in 1880 that certain quartz crystals produced electricity when put under pressure.

Quartz is a crystalline piezoelectric material composed of silicon and oxygen atoms and is illustrative of a piezoelectric material. If no forces are applied to a quartz crystal the distribution of charges is symmetric at the sites of the atoms and there is no net internal electric field. Squeezing or stretching a quartz crystal displaces atoms from their normal positions producing a separation of positive charges and negative charges (electrons) that give rise to a net internal electric field. A voltage develops across the whole crystal making it, in a sense, a battery. As with a battery, an electric current is produced by the voltage when something is connected to the crystal.

The piezoelectric effect will not be used to produce energy to energize a light bulb, for example, yet there are many applications. One common usage is as an igniter in an outdoor gas-fired grill used for cooking. A spring-loaded plunger released by a button on the front of the grill gives a sharp blow to a piezoelectric crystal. The voltage produced generates a spark between two separated metal contacts near the incoming gas, causing the gas to ignite.

Another important but little-known piezoelectric effect is found in some electronic systems. Speaking produces pressure variations that propagate through the air. Forces are produced on anything in contact with this vibrating air so that when contact is with a piezoelectric crystal, tiny voltage variations are produced. A crystal microphone is designed to make use of this piezoelectric effect.

Pressure gauges used to monitor pressures in the fluids pumped throughout an automobile engine are also often based on the piezoelectric effect. An experimenter in a laboratory can apply a known pressure to a piezoelectric crystal and measure the voltage produced. Once the relationship between voltage and pressure is known, a pressure can be inferred from a measure of the voltage. In this manner, the piezoelectric material functions as a pressure gauge. A variety of gases and oils that rely on piezoelectric gauges are pumped throughout the working components of an automobile engine.

Interestingly, a reverse piezoelectric effect is the basis for a number of applications. In this reverse process, a piezoelectric crystal vibrates when an electric current is provided from an external source such as a battery. The precise frequency of these vibrations is used to great advantage in watches and other timepieces. The most common piezoelectric crystal used is quartz. Quartz occurs naturally, but quartz crystals can also be grown in the laboratory. The most common crystals in timepieces are tiny U-shaped tuning forks that vibrate 32,768 times per second. An electronic circuit in the timepiece records time by counting the vibrations. If you peek inside a "quartz" watch you will not see wheels and a spring as in a mechanical watch. Rather, you will see a battery that provides electric current for the crystal and an electronic circuit. You might also see the U-shaped tuning fork.

Since discovering and making use of the piezoelectric effect in naturally occurring crystals such as quartz and Rochelle salts, scientists have produced a wide range of piezoelectric materials in the laboratory. An early example is barium titanate, used in an electrical component called a capacitor. Currently, most piezoelectric materials are oxide materials based on lead oxide, zirconate oxide, and titanium. These very hard piezoelectric materials are termed piezoceramics.

Piezoelectricity is also a natural by occurring phenomenon in the human body. Studies have shown that piezoelectricity causes electric potentials in dry bones of humans and animals. Whether or not this is the cause of the electric potentials that occur in wet, living bone is debatable. More than likely, any piezoelectric effect occurring in moist bone is overshadowed by other sources of electric potentials.

Joseph Priest

See also: Capacitors and Ultracapacitors.

BIBLIOGRAPHY

Ikeda, T. (1990). *Fundamentals of Piezoelectricity*. New York: Oxford University Press.

PISTONS

See: Engines

PLUTONIUM

See: Nuclear Energy; Nuclear Fission

POTENTIAL ENERGY

Potential energy is the energy that something has because of its position or because of the arrangement of its parts. A baseball in flight has potential energy because of its position above the ground. A carbohydrate molecule has potential energy because of the arrangement of the atoms in the molecule.

A diver standing on a platform above the water has potential energy because of a capacity for doing work by jumping off. The higher the diver is above the water and the greater the diver's mass, the greater the potential energy—that is, the capacity for doing work. It is the gravitational force that pulls the diver downward and into the water. For this reason, the potential energy of the diver is called "gravitational potential energy." As an equation, gravitational

The bow stores elastic potential energy as the hunter pulls on it. The energy is transferred to the arrow as kinetic energy upon release. (Corbis Corporation)

potential energy (U) measured in joules (J) is given by $U = mgh$, where m is the mass in kilograms (kg), g is the acceleration due to the gravitational force (9.80 meters/second^2, and h is the height in meters (m). An 80-kg diver 3 m above the water would have 2,350 J of gravitational potential energy.

Any type of potential energy is associated with a force and involves some favorable position. A compressed spring in a toy gun does work on a projectile when the spring is released. Before release, the spring had potential energy that we label "elastic potential energy." The force involved is that of the spring, and the favorable position is the amount of compression. Other examples of elastic potential energy are involved in flexing the bow of a bow and arrow, flexing a vaulter's pole, compressing the suspension springs of an automobile, and stretching a rubber band.

A loss in potential energy by a system is accompanied by a gain of energy in some other form. All the potential energy of a mass held above a spring is converted to kinetic energy just before the mass hits the spring. The mass loses kinetic energy as it pushes against the spring, but the spring gains potential energy as a result of being compressed. Water atop a dam in a hydroelectric power plant has gravitational potential energy. In falling, it loses potential energy and gains kinetic energy. When the water impinges on the blades of a water turbine, the rotor of the turbine gains rotational kinetic energy at the expense of the water losing kinetic energy.

Molecules have potential energy associated with electric forces that bind the atoms together. In a chemical reaction liberating energy—heat, for example—the atoms are rearranged into lower potential energy configurations. The loss of potential energy is accompanied by an increase in other forms of energy.

Protons and neutrons in the nucleus of an atom have potential energy associated with nuclear forces. In a nuclear reaction liberating energy, the nuclei are rearranged into lower potential energy structures. The loss of potential energy is accompanied by an increase in other forms of energy. A single nuclear reaction in a nuclear power plant liberates nearly 10 million times the energy of a single chemical reaction in a coal-burning power plant.

Joseph Priest

See also: Conservation of Energy; Gravitational Energy; Kinetic Energy; Nuclear Energy.

BIBLIOGRAPHY

Hobson, A. (1995). *Physics: Concepts and Connections.* Englewood Cliffs, NJ: Prentice-Hall.

Priest, J. (2000). *Energy: Principles, Problems, Alternatives*, 5th ed. Dubuque, IA: Kendall/Hunt Publishing Co.

Serway, R. A. (1998). *Principles of Physics*, 2nd ed. Forth Worth, TX: Saunders College Publishing.

POWER

Power is the rate of doing work or the rate of converting energy. A unit of work or energy is called a joule. A person doing 4,000 joules of work in four seconds would be doing work at a rate of 1,000 joules per second. A joule per second is called a watt (W) in honor of James Watt, inventor of the steam engine. Hence, the person developed a power of 1,000 watts. When a hundred watt light bulb is lit, electric energy is converted to heat and light at a rate of one hundred joules per second. In the United States, the power developed by engines and motors is usually expressed in horsepower (hp). The horsepower unit was coined by Watt who estimated the rate at which a typical work horse could do work: One horsepower equals 746 watts. In an effort to use metric units for all physical quantities, there is a trend in Europe to rate engines and motors in watts.

In terms of the rate at which work is done or energy is converted, consider that

- the human heart in a person at rest does work at a rate of about 0.01 watt,
- a flashlight converts energy at a rate of about four watts,
- a student in daily activity does work at a rate of about ninety watts or about 0.1 hp,
- an automobile engine develops about 40,000 watts, a large electric power plant produces about 1,000,000,000 watts of electric power, and
- the solar input to the earth (at the top of the atmosphere) amounts to about hundred million billion watts.

When the power is a large number, as in the case of an electric power plant, it is convenient to express the power in megawatts (MW) where one megawatt equals one million watts. An electric power of 1,000,000,000 watts would be expressed as 1,000 MW. A large coal-burning or nuclear power plant produces about 1,000 MW of electric power. The sum total of the electric power produced by all electric power plants is expressed in units of gigawatts (GW). One gigawatt equals one billion watts. An electric power of 1,000,000,000,000 watts would be expressed as 1,000 GW.

A unit of energy divided by a unit of time is a unit of power. If you are burning wood at a rate of 10,000 Btu per hour then Btu per hour denotes power. Similarly, a unit of power multiplied by a unit of time is a unit of energy. Electric power in a home is usually measured in kilowatts (kW) (i.e., thousands of watts, and time in hours [h]). Therefore, kilowatts multiplied by hours (kWh) is a measure of energy. A light bulb rated at one hundred watts (0.1 kW) converts electric energy at a rate of one hundred joules per second. If the bulb is on for two hours then the electric energy provided by the power company is 0.1 kW × 2 h = 0.2 kWh. The power company would be paid for 0.2 kWh of electric energy. Typically, the cost is about 10¢/kWh, so the amount due the company is two cents. It is interesting that electric utilities are referred to as power companies when we pay them for energy, not power.


Each cyclist must convert the energy stored in his or her muscles into torque to move the pedals. To go uphill, the necessary rate of work (power) increases as does the required torque. (Corbis Corporation)

A bicyclist has to expend energy to cycle up a steep hill. The energy the cyclist expends is determined by the product of the power the cyclist develops and the time the cyclist is pumping (energy = power × time). The cyclist may choose to develop a relatively large amount of power and pump for a short period of time. Or he may prefer to take it easy developing a relatively small amount of power and pump for a longer period of time. The energy would be the same in both cases, but the effort would be different. Even after shifting to the lowest gear, cyclists often find that it is still easier to make their way up a hill by traversing back and forth across the road. The traversing takes less effort (less power), but requires more time. This is also why roads up a mountain take winding paths rather than trying to go directly up—vehicles with less reserve capacity power can still climb the mountain.

By definition, a force (F) acting on some object as it moves some distance (x) along a line does work (W):

$$W = Fx.$$

This idea says nothing about the time involved in doing the work. Asking for the rate at which work was done (i.e. the power) and how fast was the object moving, we find power (P) equals force times velocity (v):

$$P = Fv.$$

The power developed by a person pushing a box on a floor with a force of 300 newton (N) at a speed of 1 m/s is 300 W or about 0.4 hp.

A force is required to rotate an object. The response to the force depends not only the size of the force, but also on the manner in which the force is applied. A bicyclist must push down on a pedal to cause the sprocket to rotate. But if the shaft to which the pedal is attached is vertical, no rotation results. The greatest response occurs when the bicyclist pushes down on the pedal when the shaft is horizontal. The concept of torque is used to describe rotational motion. The bicyclist pushing down on the pedal when the shaft is vertical produces zero torque. The maximum torque is produced when the shaft is horizontal. In this case the torque is the product of the force and the length of the shaft. For any other position between vertical and horizontal, the torque is the product of the force, shaft length, and sine of the angle made by the shaft and direction of the force. A force develops power when linear motion is involved and a torque develops power when rotational motion is involved. The power developed by a force is the product of force and linear velocity ($P = Fv$) and the power developed by a torque is the product of torque and angular velocity ($P = T\omega$).

Power and torque are directly related. An automobile engine develops power when producing a torque on the drivetrain. Generally, the owner of an automobile is more interested in power than torque. However, technicians in the automobile industry usually evaluate the performance in terms of torque rather than power. The power developed in the drive chain of an automobile is essentially constant. But the rotational speed of the drivetrain depends on the torque developed. The larger the torque, the smaller the rotational speed.

Joseph Priest

See also: Watt, James; Work and Energy

BIBLIOGRAPHY

Hobson, A. (1999). *Physics: Concepts and Connections.* Englewood Cliffs, NJ: Prentice-Hall.

Priest, J. (2000). *Energy: Principles, Problems, Alternatives*, 5th ed. Dubuque, IA: Kendall/Hunt Publishing.

Serway, R. A. (1997). *Physics for Scientists and Engineers*, 4th ed. Fort Worth, TX: Saunders College Publishing.

Serway, R. A., and Faughn, J. (1999). *College Physics*, 5th ed. Fort Worth, TX: Saunders College Publishing.

POWER PLANTS

See: Electric Power, Generation of

PRESSURE

Pressure is force divided by the area over which the force acts:

$$\text{pressure} = \text{force}/(\text{area of contact})$$

or

$$P = F/A.$$

In the metric system, pressure has a unit of newtons per square meter, which is called a pascal (Pa). Although the pascal is the scientific unit and is preferred, pounds per square inch (lbs/in^2) is common in the United States. For example, in most of Europe, tire pressure is recorded in pascals (typically 220,000 Pa), whereas tire pressure in American cars is measured in pounds per square inch (typically 32 lbs/in^2). As a point of reference, the pressure that the earth's atmosphere exerts on anything at the earth's surface is roughly 100,000 Pa or 15 lbs/in^2.

A large pressure does not necessarily mean a large force because a large pressure can be accomplished by making the contact area small. For example, a 160-pound man having fifty square inches in his shoe soles produces a pressure of 3.2 lbs/in^2 on the ground when he stands flatfooted. But the same person wearing ice skates having 2.5 square inches in the runners of the blades produces a pressure of 64 lbs/in^2 when standing on ice. When a person pushes down on the head of a thumbtack, the pressure at the point is quite large because the contact area of the sharp thumbtack is small.

Blaise Pascal, the French scientist for whom the pascal pressure unit is named, was the first to discover that a pressure applied to an enclosed fluid is transmitted undiminished to every point of the fluid and the walls of the containing vessel. The earth's atmosphere exerts a pressure on the surface of water in a swimming pool (a container) and this pressure is felt equally throughout the pool including the walls. This principle is employed in a hydraulic press that has many applications. The pressure due to the force f on the piston of area a is P = f/a. The piston of area A feels a force F creating a pressure P = F/A. According to Pascal's principle, the two pressures are equal so that F/A = f/a or F = (A/a)f. If the area A is larger than a, then the force F is larger than f. Therefore, a small force on the left piston causes a larger force on the right piston. Lifting an object requires an upward force larger than the weight of the object. This is why rather massive objects such as automobiles, trucks, and elevators can be lifted with a hydraulic press. For example, a car weighing 10,000 N (about 2,000 pounds) can be raised with a hundred N force if the ratio of the areas of the pistons (A/a) is one hundred.

Work is done by both forces when the pistons move and the larger force cannot do more work (force × displacement) than the smaller force. Accordingly, the larger piston moves a much smaller distance than the smaller piston. Theoretically, if the ratio of the areas is one hundred then the smaller piston must move one hundred times further than the larger piston.

The mechanism for rotating the wheels in a car or truck has its origin in pistons in the engine that are forced to move by expanding gases in the cylinders. The expanding gas is a result of igniting a mixture of gasoline vapor and air. An expanding gas exerts a force on a piston doing work as the piston moves. In principle, it is no different than a person doing work by pushing a box on the floor of a hallway. The energy for the work comes from the energy of the gas produced by the ignition of the gasoline vapor and air mixture. As in the calculation of work done by a constant force the work done by the gas is the product of the force (F) and displacement (d, the distance the piston moves). Because the force on the piston is equal to the pressure (P) times the area (A) of the piston the work (W) can be written W = PAd. The quantity Ad is just the change in volume (ΔV) of the gas so that the work can also be written as $W = P\Delta V$. This is the basic idea in determining the performance of engines, which are used to power cars, trucks, and all sorts of other equipment.

The air pressure created by the clown's lungs on the inner walls of the balloon force it to expand. (Corbis Corporation)

The expansion of a balloon when it is inflated is evidence of the pressure exerted on the interior walls by molecules blown into the balloon. If you tie off the open end of the balloon, you can squeeze the balloon with your hands and actually feel the result of the pressure in the balloon. If you were to cool the balloon you find that the pressure goes down. Warming causes the pressure to increase. This happens also with the pressure in an automobile tire after it has warmed by running on a road. When there is no change in the pressure, temperature, and volume, we say the gas is in equilibrium. The pressure, volume, and temperature of a gas in equilibrium and how they are related is an important feature for describing the behavior of a gas when the gas does work. The relationship for all possible variations of P, V, and T is very complicated, but for moderate temperatures (around 300 K or 70°F) the equilibrium state of all gases is given by $PV = NkT$ where N is the number of atoms (or molecules) and k is the Boltzmann constant, 1.38×10^{-23} joules/kelvin. A gas is said to be ideal if it subscribes to the relationship $PV = NkT$ and the equation is called the ideal gas equation.

Joseph Priest

BIBLIOGRAPHY

Hobson, A. (1999). *Physics: Concepts and Connections.* Englewood Cliffs, NJ: Prentice-Hall.

Priest, J. (2000). *Energy: Priciples, Problems, Alternatives*, 5th ed. Dubuque, IA: Kendal/Hunt Publishing.

Serway, R. A. (1997). *Physics for Scientists and Engineers*, 4th ed. Fort Worth, TX: Saunders College Publishing.

Serway, R. A., and Faughn, J. (1999). *College Physics*, 5th ed. Fort Worth: Saunders College Publishing.

PROPANE

See: Liquified Petroleum Gas

PROPELLERS

A propeller converts through helical motion the energy supplied by a power source into thrust, a force that moves a vehicle forward in a fluid medium. They are used primarily for marine and aerial propulsion, but they are found on other technologies such as hovercraft and wind turbines as well. Propellers, which are essentially a series of twisted wings, or blades, connected to a central hub, are efficient energy transmission devices for those applications. The blades strike the air or water at a certain angle, called the pitch, and create an area of low pressure in front of the propeller. As a result, the blades generate thrust through either fluid or aerodynamic means by pushing forward through the low-pressure area. Slip, or the energy lost as the propeller rotates, offsets the full output of thrust. The effectiveness of the propeller is measured by propulsive efficiency, the ratio of engine power to the actual thrust produced minus slip, during one complete revolution of the propeller. Because this process resembles the twisting movement of a carpenter's screw as it advances through wood, marine propellers are often called "screws" and aerial propellers are called "airscrews."

There exists a wide range of propellers for different applications. Fixed pitch propellers, most commonly found on ships and small aircraft, are simple in operation and efficient for one operating regime. Transport aircraft use variable-pitch propellers that can alter their pitch (sometimes automatically) to perform efficiently over a variety of conditions. Counter-rotating propellers are two propellers placed in tandem and operating from the same power source. They produce thrust efficiently, but they are also very complicated in design. Feathering propellers can set their blades parallel to the movement of either water or air in case of engine failure, which increases passenger safety. Reversible pitch propellers decrease landing distances for aircraft and aid in the maneuvering of ships by changing the direction of thrust. Depending on their application and operating regime, propellers can be made from wood, metal, or composite materials.

The concept of the propeller originates from three important antecedents. First, the Greek mathematician Archimedes developed a method of transporting water uphill through helical motion during the third century B.C.E. Second, the emergence of windmills in western Europe in the twelfth century C.E. indicated an inverse understanding of the principles and capabilities of thrust. Finally, Leonardo da Vinci's adaptation of the Archimedean helical screw to the idea of aerial propulsion during the fifteenth century inspired later propeller designers. These ideas influenced the development of both marine and aerial propellers but were mainly conceptual. It was not until the nineteenth and twentieth centuries that the availability of adequate power sources such as steam and internal combustion engines would make propellers feasible power transmission devices.

MARINE PROPELLERS

Propellers for the marine environment appeared first in the eighteenth century. The French mathematician and founder of hydrodynamics, Daniel Bernoulli, proposed steam propulsion with screw propellers as early as 1752. However, the first application of the marine propeller was the hand-cranked screw on American inventor David Bushnell's submarine, *Turtle* in 1776. Also, many experimenters, such as steamboat inventor Robert Fulton, incorporated marine propellers into their designs.

The nineteenth century was a period of innovation for marine propellers. Reflecting the growing importance of the steam engine and the replacement of sails and side paddle wheels for propellers on ocean-going vessels, many individuals in Europe and America patented screw propeller designs. Two individuals are credited with inventing the modern marine propeller: English engineer Francis P. Smith and American engineer John Ericcson. Both placed important patents in 1836 in England that would be the point of departure for future designs. Smith's aptly named ship *Archimedes* featured a large Archimedean screw and demonstrated the merits of the new system. Swedish-born Ericcson employed a six-bladed design that resembled the sails on a windmill. He immigrated to the United States in 1839 and pioneered the use of screw propellers in the United States navy. The first American propeller-driven warship, the U.S.S. *Princeton*, began operations in 1843. Other notable screw propulsion designs included British engineer Isambard Kingdom Brunel's 1843 iron-hull steamship, the *Great Britain*, which was the first large vessel with screw propellers to cross the Atlantic. In 1845, the British Admiralty sponsored a historic "tug-of-war" between the paddle wheel-driven H.M.S. *Alecto* and screw propeller-driven H.M.S. *Rattler*. The victory of the *Rattler* indicated the superiority of propellers to both sail and paddle wheel technology. However, naval vessels did not rely exclusively on screw propulsion until the 1870s. The early twentieth century witnessed the universal adoption of screws for oceangoing ships, which include the majority of vessels from the largest battleships to the smallest merchant marine vessels.

A worker walks underneath a ship's propeller in a dry dock in Canada. (Corbis Corporation)

Typical marine propellers are fixed pitch and small in diameter with very thin, but broad, blade sections. They are made from either cast metal, corrosion-resistant metal alloys such as copper, or composite materials. Marine propellers normally operate at 60 percent efficiency due to the proximity of the ship's hull, which limits the overall diameter of the propeller and disturbs the efficient flow of water through the blades. As a result, the blades have to be very wide to produce adequate thrust. Marine propeller designers use innovations such as overlapping blades and wheel vanes to offset those problems and improve efficiency.

Another important consideration for marine propeller design is cavitation, the rapid formation and then collapse of vacuum pockets on the blade surface at high speed, and its contributions to losses in propulsive efficiency. The phenomenon can cause serious damage to the propeller by eroding the blade surface and creating high frequency underwater noise. Cavitation first became a serious problem in the late nineteenth and early twentieth centuries when innovations in steam and diesel propulsion technology drove propellers at unprecedented speeds. The introduction of gearing to make propellers rotate more efficiently at high revolutions and blades designed to resist cavitation alleviated the problem. However, research in the 1950s found that cavitation was a desirable trait for many high-speed marine propeller designs such as hydrofoils.

AERIAL PROPELLERS

As early as the eighteenth century, flight enthusiasts gradually began to consider the aerial propeller a practical form of transmitting propulsion. French mathematician J. P. Paucton revived the idea of the aerial propeller in Europe in his 1768 text, "Theorie de la vis d'Archimede." His *Pterophore* design used propellers for propulsion as well as for overall lift. Pioneers of lighter-than-air flight such as the French aeronauts Jean-Pierre Blanchard and Jean Baptiste Meusnier, used propellers on their balloons and airships in the 1780s. The increased attention given to the development of a practical heavier-than-air flying machine during the late nineteenth century ensured that experimenters rejected power transmission devices such as flapping wings, oars, sails, and paddle wheels, and incorporated propellers into their designs. Even though the propeller became a common component of proto-aeroplanes such as wings, aeronautical enthusiasts did not recognize the importance of propeller efficiency.

Wilbur and Orville Wright first addressed the aerial propeller from a theoretical and overall original standpoint during the successful development of their 1903 *Flyer*. They conceptualized the aerial propeller as a rotary wing, or airfoil, that generated aerodynamic thrust to achieve propulsion. They determined that the same physics that allowed an airfoil to create an upward lifting force when placed in an airstream would produce horizontal aerodynamic thrust when the airfoil was positioned vertically and rotated to create airflow. As a result, the Wrights created the world's first efficient aerial propeller and the aerodynamic theory to calculate its performance that would be the basis for all propeller research and development that followed. Used in conjunction with the reciprocating internal combustion piston engine, the aerial propeller was the main form of propulsion for the first fifty years of heavier-than-air flight.

Others built upon the achievement of the Wrights by improving the overall efficiency of the propeller. American engineer William F. Durand developed the standard table of propeller design coefficients in his landmark 1917 National Advisory Committee for Aeronautics study, "Experimental Research on Air Propellers." Besides the need for an aerodynamically efficient blade shape, propellers needed to be efficient over a variety of flight regimes. Developed simultaneously by Frank W. Caldwell in America, H. S. Hele-Shaw and T. E. Beacham in Great Britain, and Archibald Turnbull in Canada in the 1920s and becoming widely adopted in the 1930s, the variable-pitch mechanism dramatically improved the performance of the airplane. It linked aeronautical innovations such as streamline design, cantilever monoplane wings, and retractable landing gear with the increase in power brought by sophisticated engines, fuels, and supercharging. The properly designed variable-pitch propeller has been critical to the economic success of commercial and military air transport operations since the 1930s.

After World War II, the variable-pitch popeller was combined with a gas turbine to create the turbine propeller, or turboprop. The use of a gas turbine to drive the propeller increased propulsive efficiency, fuel economy, and generated less noise than conventional piston engine propeller aircraft. Turboprop airliners, first developed in Great Britain, began commercial operations in the early 1950s and were considered an economical alternative to turbojet airliners. Propellers are the most efficient form of aerial propulsion because they move a larger mass of air at a lower velocity (i.e., less waste) than turbojets and rockets. That efficiency is only present at speeds up to 500 mph (800 km/h). Beyond that, the tips of the rotating propeller suffer from near-sonic shockwaves that degrade its aerodynamic efficiency to the point where it loses power and cannot go any faster. Given that limitation and the higher efficiencies of turbojets at speeds above 500 mph for long distance flights, the propeller appeared to be obsolete as a viable energy conversion device for air transport. What occurred was that each propulsion technology proved the most efficient in different applications. The high efficiency of the aerial propeller, between 85 to 90 percent, and its lower operating costs created a strong economic impetus for air carriers to encourage the improvement of propeller-driven aircraft for the rapidly expanding short haul commuter market in the 1970s.

Further attempts to improve the efficiency of airplanes resulted in advances in propulsion technology based on the aerial propeller. Introduced in the 1960s, the turbofan, a large, enclosed multiblade fan driven by a turbojet core, harnessed the efficiency of the propeller while developing the thrust of the turbojet. NASA's long-range aircraft energy efficiency program of the 1980s developed the propfan, which is an advanced turboprop that employs multiple scimitar-shaped blades with swept-back leading edges designed for high speed. Propfans are capable of operating at speeds comparable to those generated by turbofan and turbojet-powered aircraft at a 25 percent savings in fuel.

The extreme climatic and physical environments in which propellers operate test the limits of aerodynamics, mechanical engineering, and structural theory. Depending upon the size of the power source, aerial propellers can be made from wood, metal, or composite materials and feature from two to six long and slender blades. The blades must be able to withstand exposure to harsh weather and stand up to aerodynamic loading, engine vibration, and structural bending while efficiently creating thrust.

A significant characteristic of the aerial propeller since its invention has been noise. Aircraft propellers, especially those made of metal, often produce an extremely loud "slapping, beating" sound when operating at high speeds. The increased use of propeller-driven commuter aircraft in the 1970s contributed to the growing noise pollution around busy urban airports. Measures to quiet aircraft included propeller synchronization to prevent high frequency vibration on multiengine aircraft and the use of elliptical blades, thinner airfoil sections, smaller propeller diameters, and lower rotational speeds.

Technologies other than conventional aircraft use aerial propellers for propulsion. Helicopters and vertical and short takeoff and landing aircraft (V/STOL) use rotor blades shaped to produce aerodynamic lift much like a propeller produces thrust. Hovercraft, or air-cushion vehicles, use propellers designed to be efficient at slower speeds for both movement and maneuverability when mounted on swiveling pylons. The search for an alternative energy source to fossil fuels in the 1970s encouraged many experimenters to use propellers to catch the wind and generate electricity through a turbine. Wind turbines benefit from advances in propeller design to create energy rather than converting the energy of a power source into forward movement.

Jeremy R. Kinney

See also: Aircraft; Ships; Turbines, Wind.

BIBLIOGRAPHY

Anderson, J. D., Jr. (1989). *Introduction to Flight.* New York: McGraw-Hill.

Anderson, J. D., Jr. (1997). *A History of Aerodynamics and Its Impact on Flying Machines.* New York: Cambridge University Press.

Anderson, J. D., Jr. (1999). *Aircraft Design and Performance.* New York: McGraw-Hill.

Breslin, J. P., and Andersen, Poul. (1994). *Hydrodynamics of Ship Propellers.* New York: Cambridge University Press.

Durand, W. F. (1917). Technical Report No. 14, "Experimental Research on Air Propellers." *Annual Report of the National Advisory Committee for Aeronautics, 1917.* Washington, DC: U.S. Government Printing Office.

Gilfillan, S. C. (1935). *Inventing the Ship*. Chicago: Follett Publishing.

Gillmer, T. C. (1986). *Modern Ship Design*. Annapolis, MD: Naval Institute Press.

Gillmer, T. C., and Johnson, B. (1982). *Introduction to Naval Architecture*. Annapolis, MD: Naval Institute Press.

Jakab, P. L. (1990). *Visions of a Flying Machine: The Wright Brothers and the Process of Invention*. Shrewsbury, England: Airlife.

Kerwin, J. E. (1986). "Marine Propellers." *Annual Review of Fluid Dynamics* 18:367–403.

Metzger, F. B. (1995). *An Assessment of Propeller Aircraft Noise Reduction Technology*. Hampton, VA: Langley Research Center, National Aeronautics and Space Administration.

Miller, R., and Sawers, D. (1968). *The Technical Development of Modern Aviation*. London: Routledge and Keegan Paul.

Rosen, G. (1984). *Thrusting Forward: A History of the Propeller*. Windsor Locks, CT: United Technologies Corporation.

Smith, E. C. (1937). *A Short History of Naval and Marine Engineering*. Cambridge, England: Babcock and Wilcox.

Sparenberg, J. A. (1984). *Elements of Hydrodynamic Propulsion*. Boston: Martinus Nijhoff.

Weick, F. E. (1930). *Aircraft Propeller Design*. New York: McGraw-Hill.

PROPERTY RIGHTS

In many situations, property rights are easily determined. Boundaries to stationary, observable assets such as land can be defined and enforced clearly. Property rights to crude oil and natural gas are far more troublesome, for two reasons. Both resources lie below the surface, making it difficult to locate the exact limits of the deposits, and they migrate within subsurface reservoirs. Some reservoirs cover many square miles, including hundreds of property owners, each with a competing claim to the oil and gas. Moreover, other reservoirs straddle national boundaries, encouraging international competition for the oil and gas. For these reasons energy production from crude oil and natural gas can involve serious common-pool losses. Although there is much to be gained from cooperation among property owners to avoid these losses, reaching a cooperative solution can be very difficult.

The problems of the common pool were most dramatically illustrated in Garrett Hardin's famous 1968 article "The Tragedy of the Commons." Exploitation of valuable resources when property rights are absent or poorly defined can be extremely wasteful. Because the parties must capture the resource before it is appropriated by another, time horizons become distorted, with undue emphasis on short-term production and neglect of long-term investment. Even if the trend of market prices indicates that resource values would be greater from future, rather than current, use, production will not be postponed. The losses are compounded by excessive capital investment in extraction and storage.

Common-pool problems characterize many resources where it is difficult to define property rights to restrain access and use. In North America, common-pool conditions in oil and natural gas production are created when firms compete for hydrocarbons in the same formation under the common-law rule of capture (which also governs fisheries and many other natural resources).

Under the rule of capture, ownership is secured only upon extraction. In the United States, production rights are granted to firms through leases from those who hold the mineral rights, typically surface land owners. Each of the producing firms has an incentive to maximize the economic value of its leases rather than that of the reservoir as a whole. Firms competitively drill and drain, including the oil of their neighbors, to increase their private returns, even though these actions can involve excessive investment in wells, pipelines, and surface storage, higher production costs, and lower overall oil recovery.

Oil production requires pressure from compressed gas or water to expel oil to the surface. There are three main types of reservoir drives to flush oil to wells: dissolved gas drive, gas-cap drive, and water drive. With a gas drive, the oil in the reservoir is saturated with dissolved gas. As pressures fall with oil production, the gas escapes from solution, expands, and propels oil to the surface. Hence it is important to control gas production so it remains available to remove the oil. With a gas-cap drive, the upper part of the reservoir is filled with gas, and oil lies beneath it. As oil is withdrawn, the compressed gas expands downward, pushing oil to the well bore. As with a dissolved gas drive, gas production from the gas cap should be restricted to maintain reservoir pressure to expel the oil. Finally, with a water drive, the oil lies above a layer of water. The compressed water migrates into the oil zone as the oil moves to the well and helps to push it from the rock formation. High oil recovery rates require that subterranean water pressures be maintained.

Maximizing the value of the reservoir requires that full reservoir dynamics be considered in drilling wells and in extracting oil. Gas and water must be recycled through the strategic placement of injection wells; wells with high gas-oil or water-oil ratios must be closed or not drilled; and the rate of oil production must be controlled to maintain underground pressure.

Each reservoir generally has a dominant drive, an optimal pattern of well locations, and a maximum efficient rate of production (MER), which, if exceeded, would lead to an avoidable loss of ultimate oil recovery. Unfortunately, oil, gas, and water are not evenly distributed within the reservoir. With multiple leases above the reservoir, some lease owners will have more oil, gas, or water than will others, and coordination among competing firms in well placement and in controlling production rates is difficult. Efficient production of the reservoir suggests that some leases not be produced at all. Further, since each firm's production inflicts external costs on the other firms on the formation, some mechanism must be found to internalize those costs in production decisions.

Maintaining subsurface pressures; effectively placing wells; coordinating timely investment; and controlling oil production across leases while protecting the interests of the various lease owners under conditions of geologic and market uncertainty are formidable challenges. Yet, if they are not adopted, common-pool losses are likely.

Common-pool problems in oil and gas have been observed since commercial production began in the United States in 1859. The solution, unitization, also has been understood for a long time. With unitization, a single firm, often with the largest leased area, is designated as the unit operator to develop the reservoir, ignoring lease boundaries. With unitization, optimal well placement and production are possible as is coordination in other activities to increase overall recovery from the reservoir; These activities include pressure maintenance; secondary recovery; and enhanced oil recovery (EOR), whereby heat, carbon dioxide, or other chemicals are injected into the reservoir.

All lease owners share in the net returns of unitized production, based on negotiated formulas. As residual profit claimants, the lease owners now have incentives to develop the reservoir (rather than the lease) in a manner that maximizes its economic value over time. When producers expect unitization to occur, exploration is encouraged because greater recovery rates and reduced costs are anticipated. Bonuses and royalties to landowners are higher because the present value of the oil and gas resource is greater with unitization.

Despite its attractions for reducing common-pool losses, use of unitization has been more limited than one would expect. Joe Bain (1947, p. 29) noted: "It is difficult to understand why in the United States, even admitting all obstacles of law and tradition, not more than a dozen pools are 100 percent unitized (out of some 3,000) and only 185 have even partial unitization." Similarly, in 1985 Wiggins and Libecap reported that as late as 1975, only 38 percent of Oklahoma production and 20 percent of Texas production came from field-wide units.

To be successful, a unit agreement must align the incentives of the oil-producing firms over the life of the contract to maximize the economic value of the reservoir without repeated renegotiation. Unit contracts involve a number of difficult issues that have to be addressed by negotiators. Because remaining production often lasts twenty years or more, unit agreements must be long-term and be responsive to considerable uncertainty over future market and geological conditions. They must allocate unit production and costs among the many firms that otherwise would be producing from the reservoir. Additionally, they must authorize investments that may be made later to expand reservoir production, and distribute the ensuing costs among the individual parties.

These are formidable tasks for unit negotiators. The negotiating parties must agree to a sharing rule or participation formula for the allocation of costs and revenues from production. There may be different sharing rules for different phases of unit production, such as primary and secondary production, and the rules should apply to all firms on the reservoir. This arrangement is termed a single participating area. There should not be separate participating areas for oil and gas; otherwise, different incentives for oil and gas production will emerge. Development, capital, and operating costs must be allocated in proportion to revenue shares in an effort to align all interests to maximize the economic value of the reservoir. In

STAKING A CLAIM: THE HOMESTEADING ALTERNATIVE

Property rights assignment by homesteading is based on a theory of philosopher John Locke. Locke believed “as much land as a man tills, plants, improves, cultivated and can use the product of, so much is his property.” Such labor, Locke continues, “enclose(s) it from the common.”

In a Lockean world, mineral rights do not accompany surface rights in either original or transferred ownership. Minerals would not be owned until homesteaded by the acts of discovery and intent to possess. In the case of oil and gas, initial ownership would occur when the oil or gas entered the well bore and was legally claimed by the driller. The reservoir would then turn from a “state of nature” into owned property. The homesteader (discoverer) could be the property owner directly overhead, another land owner, or a lessee of either.

The oil and gas lease under homestead law would simply be the right to conduct drilling operations on a person’s surface property. The lease would not constitute a claim to minerals found under a particular surface area as under the rule of capture rule. Since the reservoir could be reached from different surface locations with slant drilling, the economic rent of a homestead lease would be far less than the value of a capture-rule lease. The difference in value would accrue to the driller-finder, thereby encouraging production by making drilling more efficient and profitable.

Under a homestead theory of subsurface rights, the first finder of a mineral area would have claim to the entire recognized deposit. In the case of oil and gas, the geologic unit is the entire reservoir, however shaped, as long as it can be proven to be contiguous. Separate and distinct reservoirs in the same general area, whether vertical or horizontal, would require separate and distinct discovery and claim. Alien wells draining a claimed reservoir would be liable for trespass if the homesteader can reasonably prove invasion (whether by well distance, well depth, crude type, geological formation, reservoir pressure, etc.). The pool owner would also be able to use his discovery as chosen. The pool could be left untapped, depleted, or used for storage. The reservoir space, in addition to the virgin contents of the reservoir, is newfound property.

The operation of homestead property rights system can be facilitated by pragmatic rules established either by law or judicial opinion. When a new oil or gas field is discovered, the surrounding acreage could be declared off-limits for other producers until the discoverer has a “reasonable” opportunity to drill frontier wells to delineate the field. An exception would be made if another operator had begun to drill before the nearby discovery was made so long as opportunistic slant drilling did not occur. If the discoverer does not drill development wells to delineate the field, then other operators may drill and become co-owners of the same contiguous reservoir upon discovery, requiring cooperation to minimize drilling costs.

While the homestead theory was never seriously considered in the United States experience, it has been recommended as part of a subsoil privatization program for Latin America. The major attraction is increased economic efficiency: Production and management of a contiguous oil and gas reservoir is most efficiently done by one operator, and homesteading creates one owner in place of multiple owners under the rule of capture.

Robert L. Bradley

BIBLIOGRAPHY

Bradley, R. L., Jr. (1996). *Oil, Gas, and Government: The U.S. Experience*. Lanhan, MD: Rowman & Littlefield.

Locke, J. (1947). “An Essay Concerning the True Original Extend and End of the Civil Government.” In *Social Contract*, ed. E. Barker. London: Oxford University Press.

that case, each party will share in the net returns from production and hence will have an incentive to minimize development and production costs and to maximize revenues. Further, a single unit operator must be selected to develop the field. Multiple unit operators lead to conflicting objectives and hinder the coordinated production practices necessary to maximize the value of the reservoir. Supervision of the unit operator by the other interests must be determined, with voting based on unit participation.

All of these issues can be contentious, but reaching agreement on the revenue sharing or participation formula is the most difficult. Shares are based on estimates of each firm's contribution to the unit. Those firms with leases that reflect a natural structural advantage (e.g., more resources or easier access to resources) seek to retain this advantage in the unitization formula. Such firms are unlikely to agree to a unitization agreement that does not give them at least as much oil or gas as they would receive by not unitizing. Even if the increase in ultimate recovery from unitization is so great that these parties will receive more from unit operations than from individual development, they have a much stronger bargaining position in negotiations than less-favored tract owners. The former can hold out for the most favorable allocation formula, secure in the knowledge that the regional migration of oil will continue toward their tracts during any delay in negotiations. Indeed, holding out may increase the value of a structurally advantageous location. If the firms form a subunit without the participation of the owners of better-located tracts, the pressure maintenance operations of the unit may increase the amount of oil migration toward the unsigned tracts. The holdouts, then, benefit from the unit without incurring any costs of the pressure maintenance activity.

The information available to the negotiating parties for determining lease values depends upon the stage of production. During exploration, all leases are relatively homogeneous, and unitization agreements can be comparatively easy to reach using simple allocation formulas, often based on surface acreage. Information problems and distributional concerns, however, arise with development as lease differences emerge. Because reservoirs are not uniform, the information released from a well is descriptive of only the immediate vicinity. There will be disagreements over how to evaluate this information in setting lease values. As a result of disagreements over subsurface parameters, unit negotiations often must focus on a small set of variables that can be objectively measured, such as cumulative output or wells per acre. These objective measures, however, may be poor indicators of lease value.

Conflicts over lease values and unit shares generally continue until late in the life of a reservoir. As primary production nears zero, accumulated information causes public and private lease value estimates to converge, so that a consensus on share values is possible. Without unit agreement, secondary production from all leases may be quite limited. The timing of a consensus on lease values, however, suggests that voluntary unit agreements are more likely to be reached only after most of the common-pool losses have been inflicted on the reservoir.

Reaching agreement on these issues is a complex process. In a detailed examination of seven units in Texas, Wiggins and Libecap showed in 1985 that negotiations took four to nine years before agreements could be reached. Moreover, in five of the seven cases, the area in the final unit was less than that involved in early negotiations. As some firms became frustrated, they dropped out to form subunits, which led to a partitioning of the field and the drilling of additional wells but generally did not minimize common-pool losses. For example, 28 subunits, ranging from 80 to 4,918 acres, were established on the 71,000-acre Slaughter field in West Texas. To prevent migration of oil across subunit boundaries, some 427 offsetting water injection wells were sunk along each subunit boundary, adding capital costs of $156 million.

Other costs from incomplete unitization are shown on Prudhoe Bay, North America's largest oil and gas field, which was first unitized in 1977. Two unit operators, separate sharing formulas for oil and gas interests, and a disparity between revenue sharing and cost sharing among the working interests have resulted in protracted and costly conflicts. The parties on the field do not have joint incentives to maximize the economic value of the reservoir, but rather seek to maximize the values of their individual oil or gas holdings. Because ownership of oil and gas in Prudhoe Bay is very skewed, with some firms holding mostly oil in the oil rim and others holding mostly natural gas in the gas cap, the firms have been unable to agree to a complete, reservoirwide agreement. Disputes have centered on investment and whether natural gas should be reinjected to raise oil production, or liquefied and sold as natural-gas liquids. In 1999 Libecap and Smith showed that skewed ownership in the presence of gas

Prudhoe Bay Oil Field. (Corbis Corporation)

caps characterizes other incomplete unit agreements elsewhere in North America.

Most states, as well as the U.S. government, have some type of compulsory unitization rule to limit the ability of a minority of holdouts to block a unit. Mandatory unitization began in the United States in Louisiana, in 1940. Due to political opposition by small firms that receive regulatory-related benefits, Texas, surprisingly, has not had a compulsory unitization law. Texas court rulings have tended to restrict unitization.

Although unitization is most problematical in North America, where there generally are more operators on a given oil field, the problems of coordinated production to avoid common-pool losses exist elsewhere. In the North Sea, Argentina, Ecuador, and the Caspian Sea, lease and national boundaries often do not coincide with reservoir boundaries. Under these conditions a nation may engage in competitive drilling to drain the oil and gas before its neighbor does. This incentive appears to explain the initial pattern of Norwegian drilling only along its boundaries with the United Kingdom (Hollick and Cooper, 1997).

Gary D. Libecap

See also: Governmental Approaches to Energy Regulation; Government Intervention in Energy Markets; Oil and Gas, Drilling for; Oil and Gas, Exploration for; Oil and Gas, Production of; Pipelines.

BIBLIOGRAPHY

Bain, J. S. (1947). *The Economics of the Pacific Coast Petroleum Industry, Part III: Public Policy Toward Competition and Prices.* Berkeley: University of California Press.

Hardin, G. (1968). “The Tragedy of the Commons.” *Science* 162:1243–1248.

Hollick, A. L., and Cooper, R. N. (1997). “Global Commons: Can They Be Managed?” In *The Economics of Transnational Commons*, ed. P. Dasgupta, K-G. Mäler, and A. Vercelli. Oxford: Clarendon Press.

Libecap, G. D. (1989). *Contracting for Property Rights*. New York: Cambridge University Press.

Libecap, G. D., and Smith, J. L. (1999). "The Self-Enforcing Provisions of Oil and Gas Unit Operating Agreements: Theory and Evidence." *Journal of Law, Economics, and Organization* 15(2):526–48.

Libecap, G. D., and Wiggins, S. N. (1984). "Contractual Responses to the Common Pool: Prorationing of Crude Oil Production." *American Economic Review* 74:87–98.

Libecap, G. D., and Wiggins, S. N. (1985). "The Influence of Private Contractual Failure on Regulation: The Case of Oil Field Unitization." *Journal of Political Economy* 93:690–714.

McDonald, S. L. (1971). *Petroleum Conservation in the United States*. Baltimore: Johns Hopkins University Press.

Smith, J. L. (1987). "The Common Pool, Bargaining, and the Rule of Capture." *Economic Inquiry* 25:631–644.

Sullivan, R. E., ed. (1960). *Conservation of Oil and Gas: A Legal History, 1948–1958*. Chicago: American Bar Association.

Weaver, J. L. (1986). *Unitization of Oil and Gas Fields in Texas: A Study of Legislative, Administrative, and Judicial Policies*. Washington, DC: Resources for the Future.

Wiggins, S. N., and Libecap, G. D. (1985). "Oil Field Unitization: Contractual Failure in the Presence of Imperfect Information." *American Economic Review* 75:376–385.

PROPULSION

Whenever anything is set in motion, there must be some type of propulsive force that moves it. Propulsion is a key element of many activities, including athletic events, recreation, transportation, weapons, and space exploration. This article explains the basic principles involved in any propulsion system, differentiates the types of propulsion systems, and discusses some practical aspects of propulsion.

BASIC PHYSICS

Momentum/Impulse

To change an object's speed it is necessary to exert a net force on it for some duration of time. The product of the force and the length of time it is applied is called the impulse and is the same as the change of momentum. Any increase in speed is directly proportional to the impulse and inversely proportional to its mass:

$$\begin{aligned}\text{Change of Speed} &= \text{Force} \times \text{Time/Mass}\\ &= \text{Impulse/Mass}\end{aligned}$$

A 1,000 kg (metric ton) car accelerating from 0–30 m/s (68 mph) in 6 seconds, thus requires an average force of 5,000 newtons. This is a force approximately half the weight of the car. It is likely that the car would "burn rubber" as the drive wheels spun on the road during this acceleration. To accelerate to the same speed in 12 seconds would require a force of only 2,500 newtons.

Energy/Work

An increase of speed corresponds to an increase of kinetic energy. This increase of kinetic energy is a result of the net force acting in the direction of motion through some distance. Net force times distance is the work done on the object:

$$\begin{aligned}\text{Change of Kinetic Energy} &= \text{Force} \times \text{Distance}\\ &= \text{Work}\end{aligned}$$

Kinetic energy equals 1/2 mass × speed^2. Therefore, a formula for the change of the square of the speed is

$$\begin{aligned}\text{Change of speed}^2 &= 2 \times \text{Force} \times \text{Distance/Mass}\\ &= 2 \times \text{Work/Mass}\end{aligned}$$

The car used in the example above would travel 90 meters during its acceleration (assumed constant) to 30 m/s in 6 seconds. The increase in kinetic energy would be 450,000 joules, which exactly equals the work input of 5,000 newtons × 90 meters.

In the case of maintaining the speed of an object against opposing forces such as friction or air resistance, the energy analysis is more appropriate than a momentum analysis. For friction forces, the conservation of energy requires that work by propelling force is equal to the energy dissipated by friction, or, in the case of lifting an object at constant speed through a gravitational field, that the work by propelling force is equal to the gain in potential energy.

In designing a rocket with sufficient propulsive power to launch a satellite into orbit, all three energy considerations are involved. The satellite gains speed (more kinetic energy), it is lifted upward against gravity (more potential energy), and it encounters air resistance (energy dissipated by friction).

Whatever the type of propulsion, it is necessary to exert a force on the object for a period of time or through a distance.

Newton's Third Law

The force of propulsion must be exerted by some object other than the one being propelled. Forces always appear as pairs. As expressed by Newton's third law, the object being propelled and the propelling agent interact by equal, opposite and simultaneous forces. For instance a runner pushes against the starting blocks and the starting blocks push against the runner, impelling him forward. A ship's screw pushes against the water propelling it backward, and the water pushes against the ship's screw, propelling the ship forward. The rocket exhaust thrusts against the rocket, propelling it forward as the rocket thrusts the exhaust backward. And the wind pushes against the sail, propelling the boat forward as the sail pushes against the wind, slowing it down or deflecting its motion.

Efficiency

There are numerous ways to look at the efficiency of propulsion systems, depending on the goal of the system. One way might be to define it in terms of an energy ratio—the ratio of the energy required to achieve the desired motion to the energy consumed by the propelling agent. For instance, consider accelerating the one-ton car to 30 m/s, giving it a kinetic energy of 450,000 joules. Suppose during the acceleration the engine uses 1/2 cup of gasoline with a heat of combustion of 4,200,000 joules, then the energy efficiency would be 450,000/4,200,000 = 10.7 percent.

Other definitions of efficiency arise where energy considerations are irrelevant. In rocket propulsion, for instance, the goal is to give a certain mass of payload a very large speed. A more useful measure here might be the ratio of the speed boost given to the payload mass compared to the total mass of fuel, including its container and the engine.

The most well-known propulsion efficiency is the amount of energy (fuel) required to move a vehicle a certain distance against opposing forces. Efficiencies of road vehicles, for instance, are usually measured in miles per gallon.

TYPES OF PROPULSION

Propulsion can be classified according to the type of object with which the propulsive device interacts. It might interact with (A) a massive object, (B) a fluid medium in which it is immersed, (C) material that it carries along to be ejected, or (D) a flowing stream.

Thrust Against a Massive Object

The object interacts directly with another relatively massive object. For example, with every fall of a foot, a runner interacts ultimately with the earth, as do automobiles, trains and bicycles. A pitcher pushes against a mound of dirt as he throws a ball, and bullets interact with the gun.

Thrust Against the Ambient Medium

The object interacts with a relatively lightweight medium such as air or water that it encounters along its path. A ship takes some of the water in which it floats, propels it backward, and the ship is in turn propelled forward. Other examples of this type are propeller airplanes and jet planes. The jet plane takes in air (23 percent oxygen by weight) through which it is flying, adds some fuel to it, and the chemical reaction with the oxygen (burning) results in a high-velocity exhaust. The mutual interaction between the exhaust (which, in the case of the fan jet, mixes in even more air) and the airplane propel it forward. Unlike a thrust against a massive object, which barely moves, the fluid carries away significant amounts of kinetic energy during the interaction.

Rockets

The object to be propelled carries the material to be pushed against, and usually carries the energy source as well (fuel and oxidant). Some examples of this type are: rockets, inflated balloons released to exhaust their air, and ion engines. The inflated balloon would be propelled by interaction with its released air even if it were released in space. It does not push against the air around itself. In fact, the ambient air only tends to impede its motion.

Flowing Streams

The object interacts with a flowing medium. Sails interact with the wind, rafts float downstream, a spacecraft is propelled by the stream of photons (light) from the sun. This is more of a "channeling" type of propulsion in that the propelled object deflects the flowing stream in such a way that it is forced to move in desired direction.

COMPARISON

Theoretically, the first type of propulsion, where the interaction is with a relatively immovable object, is the most energy efficient. In the absence of friction, it is theoretically possible to convert 100 percent of the available energy into kinetic energy, the energy of motion. This is because the more massive immovable object takes away a very small amount of kinetic energy even though it shares equally (but oppositely) in the impulse or change of momentum.

A common energy source for this type of propulsion is an internal combustion engine, in which fuel is burned and expanding gasses convert the energy into mechanical motion. This process is governed by the laws of thermodynamics, which limit the efficiency to much less than 100 percent. Further losses are encountered through transmission and through rolling friction. The efficiencies are in the 20 percent range for automobiles.

Human propulsion, which depends on muscle power, obtains its energy through the metabolism of food. This process has an efficiency of about 30 percent for a long-distance runner during a race. A cyclist who pedals at a tenth horsepower (climbing a 5 percent grade at 5 miles per hour) for eight hours, and who eats a diet of 3,000 calories, has an efficiency of 17 percent for the day, assuming no additional work is done.

The efficiency of propulsion for a given energy source will always be less when the interaction is with a fluid medium such as air or water. For example, the kinetic energy of the wind generated by the propellers of an airplane is energy that is dissipated and does not end up as part of the airplane's kinetic energy.

In this second type of propulsion the efficiency depends on the speed of the object relative to the speed at which the medium is being pushed backward. The greatest theoretical efficiency occurs when the propeller of the ship, for example, pushes a large quantity of water backward at a relative speed just slightly greater than the speed of the ship moving forward. Practical limits on the size of the screw determine that the water is usually pushed backward much faster than the ship's speed in order to get adequate thrust. The efficiency turns out to be very small when the screw is pushing water backwards at high rates of speed.

The jet engine also interacts with the ambient medium as it goes along, so it is basically the same type of propulsion as with the propeller. The decision of whether to use a propeller or a jet engine for an airplane depends on the design speed. The dividing line is about 300 miles per hour. The propeller is more efficient at pushing large quantities of air at relatively low speeds, so it is the choice for low-speed flight. Since it is always necessary to push the air backward at a relative speed greater than that at which it is desired to fly, a jet engine is necessary for high speed flight.

When the efficiency of transportation, by whatever propulsion method, is measured on a distance per amount of energy basis (miles per gallon), it is found that two very important factors are speed and aerodynamic design. At normal road speeds, and for subsonic (most commercial) aircraft, the wind resistance for a given size and shape depends on the square of the speed. The fuel consumption for a given vehicle is influenced directly by the friction forces, including wind resistance. An aircraft flying the same distance will need four times the energy to overcome wind resistance flying at 400 miles per hour as it uses flying at 200 miles per hour. The actual fuel consumption difference is less than this because of constants such as the amount of fuel needed to take off and to gain altitude.

The same rule is true of automobile travel. An automobile traveling at 70 miles per hour will need nearly twice the energy to overcome wind resistance as if it made the same trip at 50 miles per hour. Again, the difference in fuel consumption will be less due to the effect of rolling friction (nearly independent of speed), the fuel used to accelerate up to speed after every stop, and the energy used to operate lights, air conditioner, etc.

The third type of propulsion is best typified by a rocket engine. The rocket must carry its energy supply and ejecta material as well as the object to be launched. The energy supply is usually two chemicals that, when they combine, give off a large amount of heat, vaporizing the reaction products. This hot gas is ejected at high velocity, propelling the rocket. The two chemicals may be hydrogen and oxygen, which combine to produce very hot water (steam). The fuel might also be a solid mixture of chemicals that are designed to react at a controlled rate.

To achieve the goal of the greatest speed with a given mass of fuel, it is necessary that the ejecta come out of the rocket as fast as possible. Therefore, the efficiency (or effectiveness) of rocket fuels is better measured in terms of the speed of the ejected material than by energy efficiency.

One advantage of the rocket is that the payload can reach speeds in excess of the relative speed of the ejecta, which is not the case with a jet engine. The thrust of the jet engine goes to zero when its speed through the air reaches the relative speed of the ejecta, whereas the thrust of the rocket stays the same no matter what the speed. The jet engine has the advantage that it uses the air it encounters along the way as both the oxidant for the fuel and most of the ejecta. It, therefore, does not need to carry nearly as much mass as the rocket to accomplish the same impulse.

An ion rocket engine does not need to carry its energy supply like the chemical rocket does. It can use solar cells to collect energy from the sun to provide an electric field to accelerate ions (charged xenon atoms, for instance) that provide a thrust when they are ejected at high speed. Since the energy is "free," the energy efficiency is not relevant. The relative speed of the ejected ions from NASA's Deep Space 1 ion propulsion system is about 62,000 miles per hour, which is ten times that from conventional chemical rockets. Therefore, for a tenth of the "fuel," it can achieve the same boost in speed as a conventional rocket. By this measure the ion rocket engine is ten times more efficient than the chemical rocket.

The ion engine produces a very tiny thrust, about the weight of a sheet of typing paper, but exerts it continuously for a long period of time (several months). This very long time multiplied by the relatively small thrust (force) still results in large impulse to the space probe. Because it is not possible to push more xenon ions through the engine in a given time only low thrust is possible. Greater thrust would require a much larger engine and a more powerful source of energy.

Nuclear energy could be a compact source for application in an ion engine, but no practical devices have yet been constructed. Nuclear energy would be necessary for the ion drive of an interstellar mission for which solar energy would not be available.

Launching a satellite from earth could not be done with the low thrusts of currently conceived ion rockets. Chemical rockets, which can generate large amounts of thrust, will be needed for the foreseeable future.

The fourth type of propulsion makes use of a flowing stream. In this case, as in the ion engine, the energy efficiency is irrelevant because the supply of energy is unlimited for practical purposes.

The sailboat being propelled downwind is easy to understand. Also it is not too hard to see that turning the sail at an angle to bounce the wind off to the side propels the boat at some angle to the wind. What is amazing is that it is possible for the sailboat to be propelled faster than the wind, and to make headway upwind. These feats are possible because it is easier to move a boat in the direction of its keel (lengthwise) than it is to move it abeam (crosswise).

Another flowing stream that might be used for propulsion in the future consists of photons (light particles) from the sun. These photons have momentum and thus exert a small pressure on everything they strike. If reflected from a mirror, the mirror would act exactly like a sail in the wind. The pressure of the light would always be at right angles to the plane of the mirror. Therefore, it would be possible to use a large, lightweight reflecting surface as a sail to navigate anywhere in the solar system. The solar sail, like the ion engine, would generate a very small thrust. At the earth it would be only 7 newtons per square kilometer (4 pounds per square mile) of reflecting sail.

Don C. Hopkins

BIBLIOGRAPHY

Reynolds, W. C., and Perkins, H. C. (1977). *Engineering Thermodynamics.* New York: McGraw-Hill.

Harrison, J. P. (1996). "The Jet Age." In *Mastering the Sky.* New York: Sarpedon.

Zumerchik, J. (1997). "Sailing." In *Encyclopedia of Sports Science, Vol. 1.* New York: Macmillan.

Leary, W. E. (1998). "Ion Propulsion of Science Fiction Comes to Life on New Spececraft." *The New York Times,* Oct. 6.

PROSPECTING, GEOPHYSICAL

See: Oil and Gas, Exploration for

PULLEY

See: Mechanical Transmission of Energy

Q–R

QUANTUM MECHANICS

See: Molecular Energy

RADIANT ENERGY

See: Heat Transfer

RADIOACTIVE WASTES

See: Nuclear Waste

RAILWAY FREIGHT SERVICE

See: Freight Movement

RAILWAY PASSENGER SERVICE

Travelers usually have a choice among several transportation alternatives, most prominent being automobile, bus, airplane and railway. Rail was the preferred means of travel in the mid-nineteenth century and continuing until the 1920s, but began to decline afterward because it could not compete with the greater mobility of the automobile and the greater speed of air travel. Despite this decline, rail passenger service may yet find a new role where its potential to be more fuel-flexible, environmentally friendly, and energy efficient can be realized. Moreover it has the flexibility to avoid the congestion delays of air and roadway traffic that is likely to only worsen.

About 60 percent of transport energy in the United States is used for passenger transport, almost entirely by autos, light trucks and aviation. In fact, rail passenger services in the U.S. carry only about 2 percent of total passenger-kilometers (a passenger-kilometer is one passenger moved one kilometer), with the remainder carried by auto, air and bus. In other countries the rail role is larger, ranging from 6 to 10 percent of passenger-kilometers in many European countries to about 20 percent in India and as high as 34 percent of passenger-kilometers in Japan. Depending on the country, energy in transport could be saved by making rail passenger services more energy-efficient (not so important in the United States but much more so in Europe, India and Japan) or, probably more important, in getting passengers to switch from less-energy-efficient modes to potentially more efficient modes of rail passenger service.

Rail passenger service offers many alternatives, with different characteristics and customers. In urban areas, slow (50 km/hr), frequent-stop trolleys ("streetcars") have long operated in many cities, often in the same right-of-way where autos and buses drive. There are "light rail trains" (LRT), midway in speed (up to about 80 km/hr) and capacity between trolleys and traditional subways, operating in a number of large metropolises. "Heavy rail" metros (up to 110 km/hr) form the heart of the transport network in many of the world's largest cities. Many cities supplement their metros with longer-range, higher-speed (up to 140 km/hr) suburban rail services, and some have mixtures of all of these (e.g., Cairo and or Moscow).

In the intercity market there are glamorous high-speed trains (200 to 300 km/hr) such as the Japanese Shinkansen, the French TGV, and Amtrak's Metroliner. Such trains carry enormous numbers of passengers in Japan (almost 300 million per year), France (about 65 million), and Germany (more than 20 million). These trains provide high-quality, city-center-to-city-center passenger service in competition with air (below 500 km or so) and auto (over 150 km or so). Though they are important, only rarely do these high-speed trains carry a sizeable percentage of the country's total rail passengers.

The workhorses of most rail passenger systems are ordinary passenger trains operating at 70 to 160 km/hr. Table 1 shows the rail passenger traffic carried by most of the world's railways. The top three railway systems account for more than half of the world's passenger services, and the top six account for over two-thirds. Table 1 demonstrates another important point: the *entire* developed world accounts for only 30 percent of rail passenger transport. North American, European, and Japanese experiences are not representative of the vast bulk of the world's rail passengers.

	Passenger-Kilometers (000,000)	*Percent of Passenger-Kilometers*
India	357,013	20.1
China	354,261	19.9
Japan	248,993	14.0
Russia	168,679	9.5
Rest of W. Europe	68,107	3.8
Ukraine	63,752	3.6
Germany	60,514	3.4
France	55,311	3.1
Egypt	52,406	2.9
Italy	49,700	2.8
Rest of dev. Asia	42,849	2.4
Republic of Korea	29,292	1.6
United Kingdom	28,656	1.6
Rest of CEE	27,963	1.6
Poland	20,960	1.2
Kazakhstan	20,507	1.2
Pakistan	19,100	1.1
Romania	18,355	1.0
Rest of Middle East	17,803	1.0
Latin America	17,204	1.0
Rest of CIS	16,079	0.9
All of Africa	14,242	0.8
US Commuter	11,135	0.6
Amtrak	8,314	0.5
Australia	4,904	0.3
Canada VIA	1,341	0.1
World Total	**1,777,440**	**100.0**
Developing Countries		69.8
Developed Countries		30.2

Table 1.
The World's Rail Passenger Traffic

ENERGY USE BY RAIL PASSENGER TRAINS

Rail passenger energy efficiency depends on many variables, and a complete analysis would consume volumes of argument. In large part, though, an analysis of energy consumption in passenger transport begins with four basic factors: surface contact friction (rolling wheels), air drag (wind resistance), mass and weight (which influence acceleration/deceleration requirements and contact friction), and load factor (the percentage of occupied seats or space).

Surface friction is a source of rail's advantages in transport energy efficiency. Under similar conditions, steel wheels on steel rail generate only about 20 to 30 percent of the rolling friction that rubber wheels on pavement generate, both because rails are much smoother than pavement and because steel wheels are much more rigid than rubber tires, so steel wheels deform much less at the point of contact with the ground. Each rail wheel has only about 0.3 sq in of surface in contact with the rail, whereas an automobile tire can have 20 to 30 sq in of rubber in contact with the pavement. The greater rubber tire deformation takes energy that is wasted as heat in the tread, and the lower the pressure, the greater the deformation. Slightly offsetting this difference is the friction of the rail wheel flange, which keeps the train on the track.

Air drag is minimal at very low speeds, but it rises rapidly with increasing speed. In fact, drag is related to speed squared, and the power required to overcome drag is related to speed cubed, and it affects all vehicles. With good design of rolling bearings, rolling friction does not increase as rapidly with speed as does air drag, so in most cases air drag begins to exceed rolling friction at speeds above 60 to 100 km/hr in rail passenger vehicles, and it dominates energy requirements at speeds above about 150 Km/Hr. Streamlining can reduce air drag for any vehicle, but rail again has an inherent advantage because, for the same number of passengers, trains can be longer and thinner than buses or airplanes, and the thinness of the form affects air drag significantly.

Mass affects energy consumption because it takes energy to accelerate the vehicle and its passengers, and most of this energy is subsequently wasted as heat in the braking system when the vehicle slows. Weight also increases rolling friction. Surprisingly, rail passenger vehicles are relatively heavy. A fully loaded five-passenger automobile will have a gross weight of no more than 800 pounds per occupant and a bus slightly less, whereas a fully loaded rail passenger coach will have a gross weight of between 2,000 and 3,000 pounds per occupant. The average train, including locomotive, diner, and sleeping cars, can average 4,000 pounds or more per passenger.

Engineering models of tractive effort calculation derive from a basic model, generally called the Davis formula (after its initial formulator, W. J. Davis). The Davis formula calculates the resistance (Rt, in pounds per ton of weight being pulled) as:

$$Rt = (1.3 + 29/w + bV + CAV^2/wn)/wn$$

In this formula, w is the weight in tons per axle; b is a coefficient of flange friction (0.03 for passenger cars); V is the speed in miles per hour; C is the air drag coefficient (0.0017 for locomotives, 0.00034 for trailing passenger cars); and A is the cross-sectional area of locomotives and cars (120 sq ft for locomotives, 110 sq ft for passenger cars).

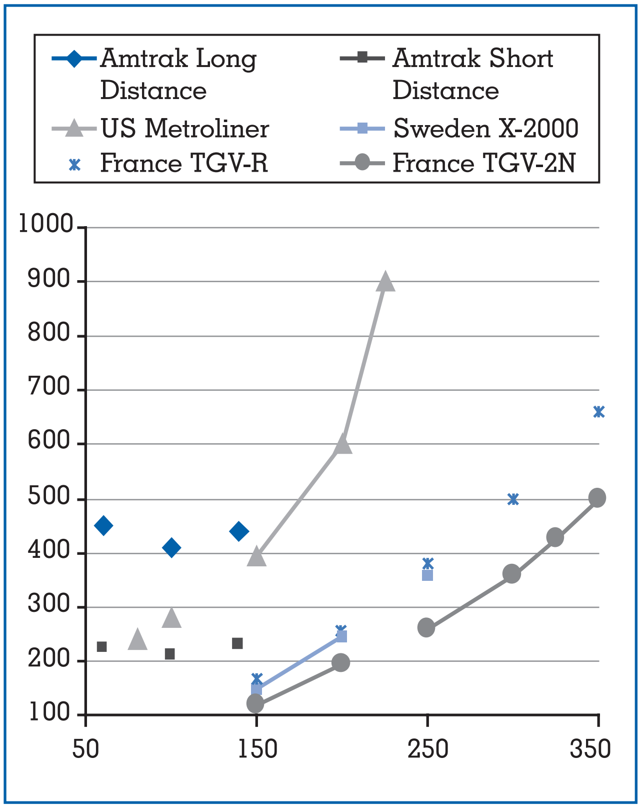


Figure 1.
Rail Passenger Fuel Consumption at Various Speeds (KJ/P-km for full trains).

The coefficients shown for the formula are examples; they would differ with changing designs of freight and passenger rolling stock and are specific to each train. The purpose here is to show the form of the relationship, not the exact values. Figure 1 displays energy consumption on various full-passenger trains over a range of speeds. Energy consumption rises rapidly with speed—and the energy consumed by the highly streamlined TGV or Swedish X-2000 is far less than the energy consumed by the boxy Metroliner.

The figure leaves out another important determinant of energy efficiency: *load factor*. Roughly speaking, a half-empty vehicle consumes twice as much energy per passenger-kilometer as a full vehicle, and a one-fourth-full vehicle four times as much. No matter what the potential energy advantages might be, empty trains waste energy, and full autos can be highly efficient. Unless modes are compared while operating at the appropriate load factor, the energy efficiency conclusions reached can be seriously flawed—and it is not necessarily valid to assume the same load factor for all modes. Indeed, the annual average load factor for Amtrak in the United States is only 46 percent, while the average load factor for U.S. airlines is 65 percent. Commuter trains do run full (or more than full) in the loaded direction during rush hour, but they often run nearly empty in the outbound direction and during midday, and actually average less than 35 percent load factor.

Even after the readily quantifiable engineering and operating issues are argued, the picture is still incomplete. Energy consumption in passenger transport is also affected by a large number of less predictable, real-world factors. Mountainous countries, for example, cause reduced railway efficiency because railways cannot climb steep grades (the downside of reduced rolling friction), so railway lines usually have to be 30 to 50 percent longer than highways between the same two end points in mountainous terrain. Bad maintenance practices and older equipment can severely reduce energy efficiency and increase pollution. These factors, and others such as driver expertise, can easily dominate the engineering and operating factors that can be readily measured.

Taking all these factors into account, well-calibrated computer models show that U.S. freight trains of various types could theoretically operate at energy consumption rates of 75 to 140 kJ per ton-km (kJ/t-km) of cargo depending on the type of train, with low-speed, heavy coal trains being the most efficient and high-speed, high-drag, double-stack container trains being the most energy-consuming. In 1997, the actual rail industry average of 250 kJ/t-km was nearly twice the theoretical upper bound of 140 kJ/t-km, reflecting the running of partially empty trains, wagon switching, and locomotive idling practices, among other factors. Though the technology is effectively the same, average energy consumption in Kenya Railways freight traffic is more than 1,000 kJ/t-km as a result of far less than optimum operating and maintenance practices.

Similar variations occur in reported cases of rail passenger service where engineering models show trains potentially operating at 200 to 400 kJ per passenger-kilometer. In practice, though, passenger trains are rarely full, and some trains, such as many long haul Amtrak trains, carry sleepers and diners which add significantly to the weight of the train but do not add many passenger seats. Commuter trains do not usually carry diners, of course, but they often operate essentially empty on their midday trips or on trips against the flow of rush-hour traffic. Thus the theoretically high rail passenger energy efficiency gets transmuted into an Amtrak system-wide average of 1,610 kJ per passenger-kilometer and an average for all United States commuter railways of 2,131 kJ per passenger-kilometer. Because of the effect of actual practices, the optimal efficiency from the engineering model is, in practice, a factor of four to eight better than what is experienced.

In general, the range of reported experience supports the conventional wisdom. Rail is potentially the most efficient method of motorized passenger travel; but bus is almost the equal of rail. Motorcycles are also quite efficient (and this assumes only the driver—motorcycles with passengers are the most energy-efficient vehicles of all). Automobiles and airlines do typically consume more energy per passenger-kilometer than rail.

More important, though, is the fact that there are significant overlaps in the energy consumption ranges, depending on the factors discussed above. While rail and bus are generally a toss-up, there are conditions in which a fully loaded Boeing 747-400 can be more energy-fficient than a partially loaded Amtrak Metroliner. A fully loaded automobile can well be a better energy and environmental choice than an empty bus or an empty train. There is no single answer, and actual conditions are very important.

Though only partly energy-related, passenger trains have additional features that deserve highlighting, such as the potential for electrification, space efficiency, and impact on urban form.

Electrification

Because trains travel path is well defined, where train traffic is dense it can be economically feasible to construct either overhead electric power supply wires (the “catenary” system) or ground-level electric supply systems (usually called the “third rail”), which permit trains to be powered directly by electric traction motors. Though a few cities have electric trolley buses, the cost of building and operating a system to supply electricity to buses, autos, and trucks is normally so high as to restrict road travel to use of fossil fuels, though advances in battery technology and hydrogen fuels may eventually reduce this dependence.

Electric traction power in railways offers three significant advantages: diversity of fuels, easier control of pollution, and high tractive effort per unit of train weight. Because electricity can be generated by a wide range of fossil fuels such as coal or natural gas, electric trains are not solely dependent on petroleum products for their operation. In fact, hydroelectric or nuclear power uses no fossil fuels at all (though each has its own environmental implications), and some countries, such as Switzerland, can use hydropower to drive almost all of their passenger and freight trains. Even when the electricity used by trains is generated by fossil fuels, the pollution is far more easily controlled from a single electric power station than from thousands of cars and buses, and the pollution and engine noise will be emitted at the power station and not in the city centers.

Because it takes time for an electric motor to overheat when exposed to high power currents, properly designed electric motors can operate for limited periods of time at power ratings as much as twice the level at which they can operate continuously. This means that electric motors can deliver short bursts of acceleration that require power well beyond the power needed for cruising. Diesels, by contrast, are limited by the power of the engine, and cannot exceed this rating even for short periods of time. In addition, in the diesel-electric system, which almost all diesel trains use, the diesel engine drives a generator, which in turn feeds electric traction motors, which drive the train; this involves extra weight, so that diesel systems are heavier than all-electric systems. As a result, electric traction has advantages where high acceleration is required, or where (as in high-speed trains) the weight of the train must be controlled to reduce the forces the train exerts on the track. Offsetting the performance advantages of electric traction is the added cost of the catenary or third rail and transformers needed to feed the train with electricity.

Most urban rail passenger systems are electrically operated because electric power permits high acceleration and thus closer spacing of trains. Most high-speed trains are electrically driven because diesel engines of the high power required are too heavy (and damage the track), and gas turbines do not handle acceleration well because they are not energy-efficient at speeds other than the optimum speed for which the turbine is designed. Ordinary, longer-haul passenger trains are a mix of diesel and electric traction: Diesel traction is far cheaper on light-density lines, and high acceleration and light weight are not so important on slower trains.

Because railway electric motors can be made to slow a train as well as drive it, the energy of braking can be regenerated onboard and put back into the catenary or third rail for use by other trains. Regeneration has not been prevalent in railways because of the complexity and weight of the onboard equipment involved, and regeneration has generally not been economically feasible where the catenary is providing alternating current (ac) traction current because of the difficulty of matching the frequency and phase of the regenerated power to the power in the catenary. Technology is changing in this respect, and ac regeneration is emerging as a possibility where train traffic is dense enough to support the added investment.

Railway electric traction systems use either ac or direct current (dc). Third rail systems operate on dc at 600 to 700 V. Overhead systems can be found with either 1,500 or 3,000 V dc, or using ac at a range of voltages (from 11,000 to 50,000) and frequencies (16 2/3 to 60 Hz). The modern standard for overhead catenary systems is usually 25,000 V and either 50 or 60 Hz depending on the industrial standard for the country (50 Hz for most of Europe, 60 Hz for the United States and Japan).

The total energy efficiency of a diesel system is not much different from an electrically driven system—a point that many energy comparisons neglect. The percentage of the energy in diesel fuel that is actually translated into tractive effort seems low, about 26 percent, in comparison with electric drive trains, which convert 90 percent of the power received from the wire or third rail into traction. However, the percentage of the fuel consumed in a power plant that is converted into electric energy is only about 40 percent (slightly higher in modern, compound cycle plants), and electric systems lose energy due to resistance in the transmission lines and transformers. Measured by the percentage of the initial fuel's energy that is actually translated to tractive effort on the train, there is not a great deal of difference between diesel at about 26 percent efficiency and electric traction systems that deliver about 28 percent of the initial fuel energy into tractive effort at the rail.

Space Efficiency

The land area required for a double-track railway (a minimum right-of-way approximately 44 ft wide) is less than that required for a four-lane highway (at least 50 ft wide at an absolute minimum, and usually much more). Depending on assumptions about load factors, a double-track railway can carry 20,000 to 50,000 passengers in rush hour in each direction, about twice to four times the peak loading of a four-lane highway (2,200 vehicles per lane per hour). Putting a double-track railway underground is also much cheaper than the equivalent capacity by highway because rail tunnels are smaller and require much less ventilation. Overall, high-density railways can carry three to eight times the amount of traffic a highway can carry per unit of land area used, and can do so with limited and controlled environmental impact.

Urban Form

Because of their ability to handle massive numbers of people effectively, using little space and emitting little or no pollution (in the urban area, at least), railways can support the functioning of much denser urban areas than auto-based urban sprawl; in fact, the term "subway" was coined to describe the underground railway that consumes almost no surface space. Where urban congestion is important (in Bangkok there have been estimates that traffic congestion lowers the gross domestic product of Thailand by several percentage points), various rail options can have a high payoff in moving people effectively. In densely populated countries, as in most of Europe and Japan, longer-haul rail passenger service plays a similar function of moving people with minimum burden on land space and overloaded highway and air transport facilities.

Technology in transport is also on the move. Automobiles have seen advances in engine design (reducing energy and pollution), body design (reducing air drag), radial tires and tread design (reducing rolling friction), and replacement of steel with plastics and aluminum (reducing weight). As a result, auto energy mileage per gallon of fuel, which was about 13.5 miles per gallon in the United States in 1970, has improved to about 21.5 miles per gallon in 1997 (a 37% improvement), and there are vehicles now available that can produce 50 to 70 miles per gallon. Diesel engine technology has improved in parallel, as has vehicle design for both rail and bus vehicles, yielding roughly a 46 percent improvement in rail efficiency since 1970. Advances in aircraft design and size, along with improved engines, have kept pace with the surface modes. Since the basic technology in engine, air drag, and vehicle frame weight management is similar and is available to all modes, there is reason to believe that all modes will (or could) improve, but this is no convincing basis for arguing that any of the modes will dramatically improve its energy efficiency relative to the others.

CONCLUSIONS

The energy/transport relationship is complex and resists easy generalization. Against this backdrop, what useful conclusions can be drawn about the role of passenger rail services in saving energy or reducing environmental impacts from transport?

In the urban arena, energy-efficient rail passenger services can reduce air pollution and help manage urban congestion. Rail passenger services can be vital to the form and function of large cities by putting large movements of people underground or overhead without consuming undue space. This potential cannot be reached, however, unless rail services are carefully planned and managed so that they operate where needed and at high load factors. Rail's effects are likely to be localized, though, and the potential will be limited if competitive modes are not fully charged for the congestion and pollution they cause. Urban rail probably does not offer much to the effort to control greenhouse gas emissions, because of the relatively small amounts of energy involved and because of urban rail's inherently low load factors.

Rail's contribution in the intercity passenger area is less clear. Where population density is high and travel distances short, and especially where fuel prices and tolls are high and airline travel expensive, there is a need for rail passenger service that can operate efficiently. This describes conditions in Japan, and parts of Western Europe and possibly the northeast corridor in the United States, where high-density and high-speed rail services do exist. But with rail passenger services carrying less than 2 percent of the intercity traffic in the United States, the contribution of this traffic to energy-efficiency objectives is probably minimal. Where populations are high and extremely poor, and where rail tariffs are kept low, there is also a significant demand for rail passenger services, as in India, China, the CIS countries, and Egypt. It is possible that rail passenger service in these countries is making a contribution to the reduction of energy consumption and thus to control of CO_2 emissions. In the "middle" countries, including the United States and Canada (outside a few dense urban corridors), which have long distances and low population densities, it seems doubtful that rail passenger services can make a measurable contribution to energy efficiency and CO_2 reduction. In all countries it is unlikely that intercity rail passenger services will be useful in reducing localized urban air pollution.

Louis S. Thompson

See also: Air Travel; Automobile Performance; Diesel Cycle Engines; Diesel Fuel; Electric Motor Systems; Electric Vehicles; Engines; Freight Movement; Hybrid Vehicles; Locomotive Technology; Magnetic Levitation; Mass Transit; Power; Steam Engines; Tires; Traffic Flow Management; Transportation, Evolution of Energy Use and.

BIBLIOGRAPHY

Association of American Railroads. (1997). *Railroad Facts*. Washington, DC: Author.

Archondo-Callao, R., and Faiz, A. (1992). “Free Flow Vehicle Operating Costs.” Washington, DC: World Bank.

Anderson, D., and Mattoon, D. (1993). *Train Operation and Energy Simulator (TOES) Technical Manual*. Chicago: Association of American Railroads.

Anderson, D.; Mattoon, D.; and Singh, S. (1992). *Revenue Service Validation of Train Operation and Energy Simulator (TOES)*, Version 2.0. Chicago: Association of American Railroads.

Association of American Railroads. (1990). *Locomotive Improvement Program, Eleventh Research Phase Final Report*. Washington, DC: Author.

Chan, Lit-Mian, and Weaver, C. (1994). “Motorcycle Emission Standards and Emission Control Technology.” Washington, DC: World Bank.

Chester, A., and Harrison, R. (1987). *Vehicle Operating Costs: Evidence from Developing Countries*. Baltimore: Johns Hopkins University Press.

Commission of the European Communities. (1992). *Green Paper on the Impact of Transport on the Environment: A Community Strategy for Sustainable Mobility*. Brussels: Author.

Commission of the European Communities. (1996). *White Paper: A Strategy for Revitalizing the Community’s Railways*. Brussels: Author.

Davis, S. C. (1999). *Transportation Energy Data Book*, Edition 19 of ORNL-6958. Washington, DC: U.S. Department of Energy.

DeCicco, J., and Ross, M. (1994). “Improving Automotive Efficiency.” *Scientific American* 271(6):52-57.

Drish, W. F., and Singh, S. (1992). *Train Energy Model Validation Using Revenue Service Mixed Intermodal Train Data*. Chicago: Association of American Railroads.

Faiz, A.; Weaver, C. S.; and Walsh, M. P. (1996), *Air Pollution from Motor Vehicles*. Washington, DC: World Bank.

Gordon, D. (1991). *Steering a New Course: Transportation, Energy and the Environment*. Washington, DC: Island Press.

Hay, W. W. (1961). *An Introduction to Transportation Engineering*. New York: John Wiley & Sons.

Heierli, U. (1993). *Environmental Limits to Motorization Non-Motorized Transport in Developed and Developing Countries*. St. Gallen, Switzerland: SKAT, Swiss Centre for Development Cooperation in Technology and Management.

Imran, M., and Barnes, P. (1990). “Energy Demand in the Developing Countries: Prospects for the Future.” Washington, DC: World Bank.

Ishiguro, M., and Akiyama, T. (1995). *Energy Demand in Five Major Asian Developing Countries*. Washington, DC: World Bank.

Levine, M. D., et al. (1991). *Energy Efficiency, Developing Nations, and Eastern Europe*. A Report to the U.S. Working Group on Global Energy Efficiency.

Martin, D. J., and Michaelis. (1992). *Research and Technology Strategy to Help Overcome the Environmental Problems in Relation to Transport: Global Pollution Study*. Luxembourg: EEC.

Noll, S. A. (1982). *Transportation and Energy Conservation in Developing Countries*. Washington, DC: Resources for the Future.

Organisation for Economic Cooperation and Development. (1985). *Energy Savings and Road Traffic Management*. Paris: Author.

Redsell, M.; Lucas, G.; and Ashford, N. (1988). *Comparison of on-road Fuel Consumption for Diesel and Petrol Cars*. Berkshire, Eng.: Transport and Road Research Laboratory.

Rose, A. B. (1979). *Energy Intensity and Related Parameters of Selected Transportation Modes: Passenger Movement*. Oak Ridge, TN: Oak Ridge National Laboratory.

Schipper, L., and Meyers, S. (1992). *Energy Efficiency and Human Activity: Past Trends, Future Prospects*. New York: Cambridge University Press.

Schipper, L., and Marie-Lilliu, C. (1998), *Transportation and CO2 Emissions: Flexing the Link*. Washington, DC: World Bank.

Thompson, L. S., and Fraser, J. (1995). *Energy Use in the Transport Sector*. Washington, DC: World Bank.

United Nations. Economic Commission for Europe. (1983). *An Efficient Energy Future: Prospect for Europe and North America*. Boston: Author.

U.S. Congress, Office of Technology Assessment. (1991). *Improving Automobile Fuel Economy: New Standards and Approaches*. OTA-E-504. Washington, DC: U.S. Government Printing Office.

U.S. Department of Transportation. (1990). *Compendium of National Urban Mass Transportation Statistics from the 1987 Section 15 Report*. Washington, DC: Urban Mass Transportation Administration.

U.S. Department of Transportation. (1973). *The Effect of Speed on Automobile Gasoline Consumption Rates*. Washington, DC: Federal Highway Administration.

U.S. Department of Transportation. (1993). *Transportation Implications of Telecommuting*. Washington, DC: Author.

U.S. Department of Transportation. (1997). *National Transport Statistics*. Washington, DC: Bureau of Transportation Statistics.

U.S. Department of Transportation. (1991). *Rail vs. Truck Efficiency: The Relative Fuel Efficiency of Truck Competitive Rail Freight and Truck Operations Compared in a Range of Corridors*. Washington, DC: Federal Railway Administration.

Wagner, F. A., and Institute of Transportation Engineers. (1980). *Energy Impacts of Urban Transportation Improvements*. Washington, DC: Institute of Transportation Engineers.

World Bank. (1988). *Road Deterioration in Developing Countries: Causes and Remedies*. Washington, DC: World Bank.

World Bank. (1993). *Energy Efficiency and Conservation in the Developing World*. Washington, DC: World Bank.

World Bank. (1996). *Sustainable Transport: Priorities for Policy Reform*. Washington, DC: World Bank.

RANKINE, WILLIAM JOHN MACQUORN (1820–1872)

William Rankine has been credited with many things derived from his brilliant career, with perhaps the most unique being the transition of his empirical work into scientific theories published for the benefit of engineering students. He is considered the author of the modern philosophy of the steam engine and also the greatest among all founders of and contributors to the science of thermodynamics.

Born in Edinburgh, Scotland, on July 5, 1820, Rankine received most of his education from his father, David Rankine, a civil engineer, plus various private tutors. The elder Rankine worked as superintendent for the Edinburgh and Dalkeith Railway, imparting to his son a love of steam engines. On top of this fine education came two inspiring years at the University of Edinburgh, which helped launch his career in civil engineering, even though he left without earning a degree. He spent a year assisting his father, then worked in Ireland for four years on projects including railroads and hydraulics. Returning to Scotland, he apprenticed under Sir John Benjamin MacNeil, a respected civil engineer of his time. Rankine remained in this profession until the late 1840s before switching to the practice of mathematical physics.

At the age of thirty-five, Rankine was appointed by the Queen's Commission to the Chair of Civil Engineering and Mechanics at the University of Glasgow in Scotland. This Regius Chair was established in 1855 with great distinction at the fourth oldest university in Great Britain. Despite the great achievements of British inventors, most self–educated, they were often deprived of an appropriate claim to true professional recognition in the areas of architecture and engineering. As a professor, Rankine argued passionately with his University commissioners for the establishment of a diploma in engineering science, later adding a Bsc in science. These were just a few of the achievements that earned him accolades as a pioneer in engineering education.

Rankine's most famous textbook, *A Manual of the Steam Engine and Other Prime Movers* (1859), set an engineering standard for years to come, delivering systematic instruction to upcoming engineers. This publication was credited with affording inventors a new basis beyond their experimentation to advance the development of the steam engine. Other books Rankine authored include *A Manual of Applied Mechanics* (1858); *A Manual of Civil Engineering* (1862); and *A Manual of Machinery and Millwork* (1869). He also published numerous practical manuals on civil and mechanical engineering and scientific tables for construction, carpentry, architecture and surveying. Somehow in his spare time, he found time to write for entertainment, including *Songs and Fables* (1874) and *A Memoir of John Elder* (1871). His research led to such renowned legacies as the Rankine cycle, for ideal operation of a steam engine, as well as his Rankine Tables for cycle efficiencies (comparing performances of steam engines and steam turbines), the construction strength of columns, and his Rankine absolute temperature scale. Many of his tables remain in use today.

Moving into other fields of practical applications, Rankine developed improved strength for iron rails by conducting fatigue testing of existing railroad ties. He was a pioneer in promoting a theory of open water sailing vessels that debunked a simple but common belief of his time—that metal ships could not be made to float. Through his successful research, the size, length, and safety of ships was significantly improved. Through his knowledge of thermodynamics, he played a key role in the advancements in refrigeration equipment. His study of soil mechanics (earth pressures) led to improved retaining walls.

Rankine's works placed him among the greatest of many famous scholars at the University of Glasgow—from Adam Smith in modern economic science, to James Watt with his invention of the double acting steam engine, to Rankine's peer in thermodynamics, William Thomson (Lord Kelvin), after whom the absolute scale of temperature was named. Before passing away on Christmas Eve, 1872, in Glasgow, Rankine served as president of the Institute of Engineers in Scotland and honorary member of the Literary and Philosophical Society of Manchester.

Making a statement on the importance of recording and teaching the history of technology, Rankine began his *A Manual of the Steam Engine and Other Prime Movers* as follows: "Nations are wrongly accused of having, in the most ancient of times, honored and remembered their conquerors and tyrants only, and of having neglected and forgotten their benefactors, the inventors of the useful arts. On the contrary, the want of authentic records of those benefactors of mankind has arisen from the blind excess of admiration, which led the heathen nations of remote antiquity to treat their memory with divine honors, so that their real history has been lost amongst the fables of mythology."

Dennis R. Diehl

See also: Steam Engines; Thermodynamics.

BIBLIOGRAPHY

Cardwell, D. S. L (1971). *From Watt to Clausius: The Rise of Thermodynamics in the Early Industrial Age*. Ithaca, NY: Cornell University Press.

Rankine, W. J. M. (1859). *Steam Engine and Other Prime Movers*. London and Glasgow: Richard Griffin & Company

Thurston, R. H. (1939). *A History of the Growth of the Steam Engine*. Ithaca, NY: Cornell University Press

RECYCLING

See: Materials

REFINERIES

The process of refining begins with the delivery of crude oil—a mixture of thousands of different hydrocarbon compounds that must be separated into fractions of varying boiling ranges in order to be used or converted into products such as gasolines, diesel fuels, and so on. Crude oils from different sources vary widely in terms of density (gravity), impurities (sulfur, nitrogen, oxygen, and trace metals), and kinds of hydrocarbon compounds (paraffins and aromatics). In general, the heavier the crude oils, the less transportation fuel (gasoline, diesel fuel and jet fuel) boiling-range components are contained in the crudes. The average crude oil charged to U.S. refineries in 1998 contained about 20 percent material in the gasoline boiling range and about 25 percent in the middle distillate (diesel and jet fuel) boiling range. The U.S. market requirements are about 50 percent gasoline and 35 percent middle distillate fuels per barrel of crude oil charged to U.S. refineries.

CRUDE OILS

Crude oils are classified by major hydrocarbon type (paraffinic, intermediate, or naphthenic), gravity (light or heavy), and major impurities—high sulfur (sour), low sulfur (sweet), high metals, etc. Everything else being equal, the heavier the oil the lower the value of the oil, because a smaller percentage of that particular crude is in the transportation fuel boiling range. The boiling ranges of the heavier fractions can be reduced, but the additional processing is costly, adding anywhere from fractions of a cent to several cents per gallon. The same is true of the sulfur, nitrogen, oxygen, and metal contents of crude oils. The impurities have to be removed by reaction with hydrogen, and hydrogen processing is very expensive.

Paraffinic crude oils are defined as those crude oils containing waxes, naphthenic crude oils as those containing asphalts, and intermediate crude oils those containing both waxes and asphalts. Usually the residues—material boiling above 1050°F (565°C)—of naphthenic crude oils make good asphalts, but the intermediate crude oils can contain enough wax so that the asphaltic part may not harden properly, and therefore they cannot be used to produce specification asphalts.

Refinery processes are available to reduce the boiling points of components of the crude oils that boil higher than the transportation fuels by cracking the large high-boiling-point molecules into smaller molecules that boil in the desired transportation fuel boiling range. Lighter materials can also be converted to transportation fuels (usually gasoline) by processes such as alkylation or polymerization. There are also processes to improve the quality of the transportation fuel products because the crude oil components with the proper boiling range do not have the other characteristics required. These include octane and cetane numbers for gasoline and diesel fuel, respectively, and low sulfur and nitrogen contents for gasoline, diesel, and jet fuels.

Over 50 percent of the crude oils used in the United States are imported, because U.S. crude oil production has passed its peak and has been declining in recent years. Also, other countries (Mexico, Saudi Arabia, and Venezuela) have been buying full or part ownership into U.S. refineries in order to ensure there will be a market for their heavy, high-sulfur crude oils. These efforts to ensure a market for their imported crude oils have resulted in more equipment being installed to crack the very heavy portions of the crude oils into as much transportation fuel boiling-range material as is economically feasible. As of January 1, 1999, U.S. refineries had the ability to crack more than 50 percent of the crude oil charged to the refineries. As the demands for heavier products decreases, additional cracking equipment will be added to some U.S. refineries to increase the cracking ability to 60 percent or higher. Also, as the demand for cleaner burning transportation fuels with lower concentrations of aromatics and olefins increases, additional process equipment will be added to produce larger volumes of alkylates.

Most of the U.S. refineries are concentrated along the Gulf Coast, because most imported crude oils come from Venezuela, Mexico and Saudi Arabia and a large amount of U.S. domestic crude oil is produced near the Gulf Coast in Texas, Louisiana, Mississippi and Oklahoma. California also has a concentration of refineries that process crude oil from deposits in California and Alaska, in order to avoid the costs of transferring crude oils and products over or around the Rocky Mountains while also satisfying the large demands for transportation fuels in the West. These logistics explain in large part why over 54 percent of U.S. refining capacity is in the states of Texas, Louisiana, and California. Texas contains the most with more than 26 percent, Louisiana is second with almost 16 percent, and California is third with more than 12 percent. Environmental regulations, resistance to "having a refinery in my neighborhood," and economics have stopped the construction of completely new refineries in the U.S. since the 1970s. It is improbable that any will be built in the foreseeable future. Since the mid-1970s, all new complete refineries have been built overseas.

Because transportation fuels vary in boiling range and volatility depending upon the type of engine for which they are suitable, refineries are designed to be flexible to meet changes in feedstocks and seasonal variations in transportation demands. Typically, spark ignition engines require gasolines and compression ignition engines (diesels) require diesel fuels. Jet fuel (similar to kerosene) is used for most commercial and military aircraft gas turbines. Jet fuel demand fluctuates widely depending on the economy and whether the United States is involved in military operations. Most of the refineries in the United States are fuel-type refineries; that is, more than 90 percent of the crude oil feedstocks leave the refineries as transportation or heating fuels. Although there are other profitable products, they are small in volume compared to the fuels and it is necessary for the refinery to make profits on its transportation fuels in order to stay in operation. In 1999, there were 161 refineries in the United States, owned by eighty-three companies, and having a capacity of over 16,000,000 barrels per day (672,000,000 gallons per day). Worldwide, there are over 700 refineries with a total crude oil feed rate of more than 76,000,000 barrels per day. In the United States, for refinery crude oil feedstock capacities and crude oil purchases throughout the world, the standard unit describing amount is the barrel. The term *barrel* is outmoded and refers to the 42 U.S. gallon capacity British

Starter Material	*Boiling Point or Range*		*Product*	*Uses*
	°F	*°C*		
Methane & Ethane	-259°F -132°F	161°C - 91°C	Fuel gas	Refinery fuel
Propane	- 44°F	- 42°C	LPG	Heating
i-Butane	+11°F	-12.2°C		Feed to alkylation unit
n-Butane	31°F	-0.5°C	LPG, lighter fluid, gasoline	Gasoline blending or to LPG
Propylene and butylenes				Feed to alkylation or polymerization units
Light naphthas	80-180°F	27-82°C	Light gasoline	Gasoline blending or isomerization unit
Heavy naphthas	180-380°F	82-193°C	Heavy gasoline	Catalytic unit feed
Kerosine	380-520°F	193-271°C	Kerosine	Jet fuel
Atmospheric gas oil	520-650°F	271-343°C	Light gas oil	Blending into diesel fuels and home heating oils
Vacuum gas oils	650-1050°F	343-566°C		Feeds to FCCU and hydrocracker
Vacuum resid	1050°F+	566°C+	Vacuum tower bottoms (VRC)	Heavy and bunker fuel oils, asphalts

Table 1.
Characteristics of Some Typical Refineries

wooden salted-fish barrel used to haul crude oil to the refineries in horsedrawn wagons in the later part of the nineteenth century. Most of the world uses either metric tons or cubic meters as the unit of measurement.

REFINING SYSTEMS

All refineries are similar and contain many of the same basic types of equipment, but there are very few that contain exactly the same configuration, due to differences in the crude oils used as feedstocks and the products and product distributions needed. In addition, most refinery processing is very expensive, and many of the smaller companies do not have the necessary capital to install the latest highly capital-intensive technology such as alkylation units and hydrocrackers. Instead they use less costly units such as fluid catalytic cracking units (FCC) and polymerization units. However, except for the least complicated refineries, topping and hydroskimming refineries, they all have processes to change boiling range and improve quality.

A small number of refineries produce lubricating oils, but the total volumes produced are only about 3 percent of crude oil charged to U.S. refineries. Processes to make lubricating oil blending stocks are expensive from both capital and operating costs viewpoints. Nevertheless, the business is still profitable because lubes demand high retail prices. Older processes use solvent extraction to improve the temperature-viscosity characteristics (viscosity index) and low-temperature refrigeration tech-nologies, such as dewaxing, to lower the pour points of the lube oil blending stocks. Newer plants use selective hydrocracking and high-pressure hydrotreating to produce high-quality lube oil blending stocks. The newer equipment is so expensive that many of the new facilities are joint projects of two or more companies in order to gain enough volume throughout to justify the high capital investments.

Some refineries also install processes to produce petrochemical feedstocks. However, from a volume viewpoint, it is not a large part of refinery income. Even though profit margins can be high, most refineries restrict their capital investments to those

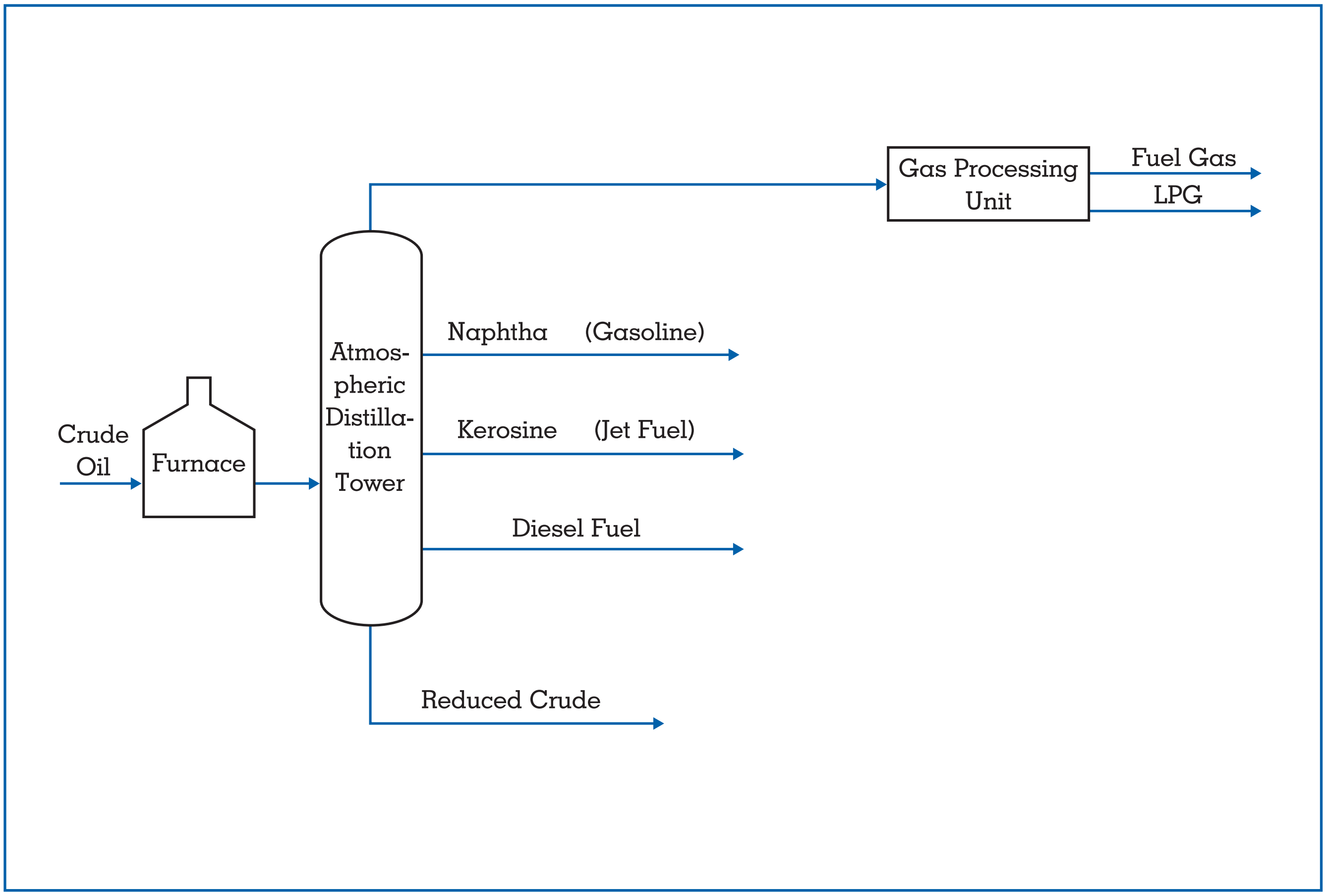


Figure 1.
Diagram of a Topping Refinery.

needed to produce transportation fuels. The most common petrochemical feedstocks produced are olefins (chiefly ethylene and propylene) and aromatics (benzene, methylbenzene [toluene], ethylbenzene, and xylenes [BTX compounds]). BTX compounds are produced in a catalytic reformer by operating at high reactor temperatures to produce 102 to 104 research octane products containing over 75 percent aromatics. Using solvent extraction, fractionation, and adsorption technologies the products are separated into individual aromatic compounds for sale as petrochemical feedstocks.

Topping and Hydroskimmimg Refineries

The *topping* refinery (see Figure 1) is the simplest type of refinery, and its only processing unit is an atmospheric distillation unit which *tops* the crude by removing the lower–boiling components in the gasoline, diesel fuel, and diesel fuel boiling ranges (those boiling up to 650°F [343°C]). The bottoms product from the atmospheric distillation unit (atmospheric resid, atmospheric reduced crude, or ARC) is sold as a heavy fuel oil or as a feedstock to more complex refineries containing conversion equipment that can convert the high–molecular–weight components into transportation fuel boiling–range products. Until recently the topped components could easily be converted into retail products by blending additives (such as lead compounds) into them. When lead compounds were banned from gasolines in the United States, it was necessary to add equipment to improve the octanes of the gasoline boiling–range components. Catalytic reforming units increased the octanes of the naphthas by converting paraffin molecules into aromatic compounds plus hydrogen. As a result, a topping refinery producing unleaded gasoline is a hydroskimming refinery.

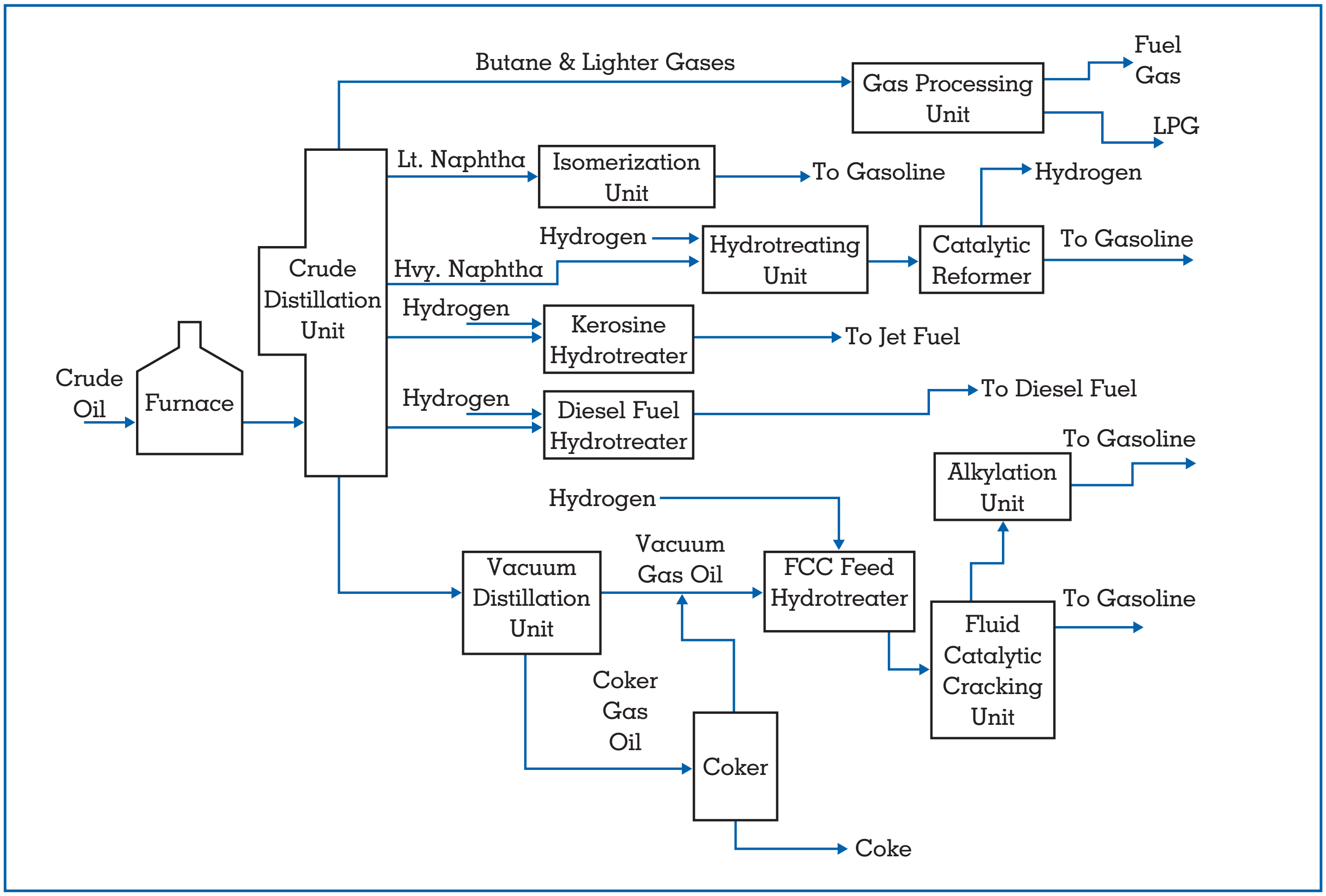


Figure 2.
Integrated Refinery.

Conversion and Integrated Fuels Refineries

Although a small amount of crude oils charged (less than 5 percent) to U.S. refineries is converted to specialty products, the remainder is either used in the refinery as a fuel or is processed into asphalt or fuel products including motor fuels and heating oils. The market for heavy fuel oils—those with boiling ranges above 650°F (343°C)—in the United States is less than 7 percent of the crude oil charged to U.S. refineries. It is therefore necessary to change the boiling ranges of the heavier portions of the crude to meet the product demands of the transportation fuels. Refineries with the necessary boiling-range conversion equipment are integrated or conversion refineries (see figure 2). The majority of the conversion equipment is to reduce boiling point—these are cracking units. When the molecules are cracked to reduce boiling range, secondary reactions produce some light hydrocarbons with boiling points too low for typical transportation fuels. Conversion equipment is also included to increase the boiling points of these components to permit blending into gasolines. As a result, about 70–80 percent of the crude oil charged to U.S. refineries is converted into gasolines, diesel and jet fuels, and home heating oils.

REFINERY PRODUCTS

Refineries produce more than 2,000 products, but most of these are very similar and differ in only a few specifications. The main products, with respect to volume and income, are liquefied petroleum gases (LPG), gasolines, diesel fuels, jet fuels, home heating oils (No. 1 and No. 2), and heavy heating oils (No. 4, No. 5, No. 6, and bunker fuel oil). Some refineries also produce asphalts and petroleum coke.

LPGs are gases that are liquefied at high pressures—up to 250 psig (17 bars)—and consist mainly

Generic name	*Characteristics*
Naphtha	Any liquid hydrocarbon material boiling in the gasoline boiling range
Middle distillate	Liquid hydrocarbons boiling in the jet fuel, diesel fuel, and home heating oil ranges
Gas oil	Hydrocarbons boiling at temperatures above 400°F (205°C) that have been vaporized and reliquefied
Coke	A solid residue from cracking operations, mostly carbon but with a small amount of high-molecular-weight hydrocarbons

Table 2.
Generic Names of Selected Refinery Products.

of mixtures of propane and butanes. These are sold in pressurized cylinders and are used for heating. Their selling price is related to the heating value.

Gasolines are the largest volume refinery products. In 1998, for instance, more than 330,000,000 gallons were consumed each day in the United States. This amounts to more than one million tons per day. The best known specifications for gasoline are octane numbers and Reid vapor pressure (RVP). The octanes are numbers determined by laboratory tests and can be related to performance of gasolines at normal road operating conditions. Octane numbers measure the resistance of the gasoline to self–ignition (knocking) and are determined by comparison with reference fuels. Self-ignition causes a decrease in engine efficiency and power, and when it occurs there are noticeable reductions in mileage and acceleration. Research octane number can be correlated with city driving performance (RON) and motor octane number (MON) with highway driving performance. Federal regulations require that the arithmetic average of these octanes ([RON + MON]/2) be posted on the pumps dispensing the gasolines. In 1998, about 70 percent of the gasolines sold in the United States were regular grade, and 30 percent premium grades. Regular gasoline posted octanes are 87 for lower elevations and 85 or 86 at higher elevations (above 3,000 feet). The octane requirement of an engine decreases with elevation. Refiners try to produce gasolines with the RON and MON as close together as possible, so their gasolines perform equally well for both city and highway driving but it is not practical. Refinery processing equipment has major effects on the sensitivity to driving conditions of the gasolines produced. Typically the RON and MON of gasolines will differ by six to twelve octane numbers (RON–MON). For gasolines the RON is always higher than the MON.

Reid vapor pressure determines ease of starting, engine warmup time, and tendency for vapor lock to occur. Vapor pressure is regulated by the addition of n–butane to the gasoline mixture. During summer months, EPA regulations limit the maximum vapor pressure of gasolines to as low as 7.2 psi in the southern U.S. and Rocky Mountain areas in order to reduce hydrocarbon vapor emissions into the atmosphere. During the cold months of winter, RVP is increased in northern areas to as high as 14 psi to make starting of cold engines possible and for the engines to warm up quickly. Normal butane not only increases RVP, but it also acts as an octane booster. During the summer months when less n–butane can be blended into gasolines, additional amounts of some other, more costly, octane booster must be added to compensate.

The cetane number determined in the laboratory by comparison with reference fuels is of value in that it is an indication of the ease of self–ignition. Diesel engines do not use a spark to ignite the fuel–air mixture in the cylinder but rely on the diesel fuel to self–ignite when injected with a high–pressure injection pump into the hot compressed air in the cylinder. The quicker the fuel self–ignites, the more efficient the engine. Many small refineries do not have cetane number determination equipment, so they substitute a cetane index equivalent to the cetane number and calculated from the fuel's gravity and average boiling point. The Environmental Protection Agency (EPA) requires that highway diesel fuels sold in the United States must have a cetane index (CI) of at least 40 and a sulfur content of less that 0.05% by weight (500 ppm). During 1998, almost 150 million gallons of diesel fuels and 60 million gallons of jet fuel were used in the United States per day.

REFINERY OPERATIONS

Refineries are unique businesses in that they operate on small profits per unit volume of product but produce very large volumes of product. This means that a regulatory adjustment (such as sulphur content) that creates a small change in operating costs per gallon of product (fractions of a cent) can mean the difference between profit and loss. The industry therefore strives to optimize refinery operations to lower costs and to improve efficiency.

One of the significant problems for refinery is the small amount of water emulsified in the crude oils coming into refineries. This water contains inorganic salts that decompose in the process units and form compounds that are corrosive to equipment and poison catalysts used in the processes. To prevent these costly problems, the first process in the refinery, desalting, reduces the water and salt content of the crude oils. Following desalting, the crude is sent to the distillation units (crude stills) to be separated into fractions by boiling range. The first of the crude stills, the atmospheric distillation unit, vaporizes and separates those components boiling below about 650°F (345°C) into four general categories; butane and lighter (methane, ethane, and propane) gases, light and heavy naphtha fractions, kerosene, and atmospheric gas oil. (Gas oil is a generic name for any hydrocarbon boiling above the end boiling point of gasoline that has been vaporized and condensed into a liquid.) The unvaporized material boiling above 650°F is removed from the bottom of the distillation column as vacuum resid or atmospheric reduced crude (ARC). The naphtha fractions are in the gasoline boiling range but, before blending into gasoline, their quality has to be improved by increasing their octanes and reducing the sulfur contents. The kerosene is hydrotreated (reacted with hydrogen) to reduce the sulfur content and blended into jet fuel. And the atmospheric gas oil is hydrotreated to reduce the sulfur content and is blended into diesel fuels and home heating oils (No. 1 and No. 2 heating oils).

The atmospheric reduced crude is the feedstock for the vacuum distillation unit. To prevent thermal decomposition (cracking) of the higher boiling point hydrocarbons in the crude oil, the pressure in the vacuum distillation fractionation column is reduced to about one-twentieth of an atmosphere absolute (one atmosphere pressure is 14.7 psia or 760 mm Hg). This effectively reduces the boiling points of the hydrocarbons several hundred degrees Fahrenheit. The components boiling below about 1050°F (565°C) are vaporized and removed as vacuum gas oils, the unvaporized material is removed from the bottom of the unit as vacuum reduced crude (vacuum resid or VRC). The vacuum gas oils are sent as feedstocks to either fluid catalytic cracking units or to hydrocrackers. The fluid catalytic cracking units crack the vacuum gas oils into gasoline blending stocks (up to 70 percent by volume on feed) and gas oil blending stocks for diesel fuels and home heating oils. The gasoline blending stocks have octane numbers in the high eighties and, if the sulfur contents are sufficiently low, can be blended directly into gasolines. The gas oil blending stocks have low cetane numbers and poor burning characteristics that limit the amounts that can be blended into diesel fuels and home heating oils.

If the crude oils from which the vacuum reduced crudes are made have the proper qualities (low wax and high asphaltenes), the vacuum-tower bottoms can be blended into asphalt products and sold to pave roads and make roofing materials and sealants. The vacuum reduced crudes also can be sold as heavy fuel oils or used as feedstocks to a severe thermal cracking unit (coker) to convert them into lower boiling hydrocarbons plus solid coke residues. The coker converts the vacuum tower bottoms into feedstocks for other refinery processing units to make gasolines, light heating oils, and, if used as hydrocracking unit feedstocks, produce gasoline, diesel fuel, and jet fuel blending stocks. Historically, the heavy fuel oils have sold, on a volume basis, for about 70 percent of the prices paid for the crude oils from which they were produced. When crude oil is selling for twenty dollars per barrel, heavy fuel oil will be selling for about fourteen dollars per barrel. This discounting is necessary because the amount of heavy fuel oil available exceeds demand. The alternate cost of processing the vacuum tower bottoms in the refinery is much higher than the cost of processing crude oil. The vacuum reduced crudes sell at a loss, or they are converted into salable products by processing the liquid products from a coker in other refinery processing units. The coke produced by the coker can be sold as a fuel, to make anodes for aluminum production, or to make electrodes for electric steel furnaces.

Fluid catalytic cracking units (FCC or FCCU) are the major processing units to reduce boiling ranges of those crude oil components that have boiling points higher than the final boiling points of the transportation fuels—typically above 650°F (343°C). These

Imperial Oil refinery in Edmonton, Alberta, Canada. (Jack Fields/Corbis)

units circulate finely divided (about the size of very fine sand particles) silica-alumina-based solid catalyst from the reactor to the regenerator and back. The reactor operates at temperatures between 950 and 1,100°F (510–595°C) to crack the large molecules in the gas oil feed selectively into smaller molecules boiling in the gasoline boiling range of 80–400°F (27–205°C). When the cracking takes place in the reactor, coke is also produced and is laid down on the surface of the catalyst. This coke layer covers the active cracking sites on the catalyst, and the cracking activity of the catalyst is reduced to a low level. Conveying the catalyst to the regenerator, where the coke is burned off with air, restores the catalyst activity. This cycle is very efficient: the heat generated by burning the coke off the spent catalyst in the regenerator supplies all of the heat needed to operate the unit. The regenerated catalyst leaving the regenerator has had its activity restored to its original value and at a temperature of about 1,300°F (705°C) is returned to the riser–reactor and mixed with fresh feed to continue the cracking reactions. By controlling the reactor temperature, the conversion of the feed to gasoline boiling range components can be maximized during the summer when gasoline demand is at its highest. In practice, up to 70 percent by volume of the product is in this boiling range. The product is rich in aromatic and olefinic compounds, which makes the high quality FCC gasoline very suitable for blending into the final refinery gasoline products. The remainder of the feed is converted into butane, lighter gases, and catalytic gas oils (typically called cycle gas oils) that are blended into diesel fuels or home heating oils, or used as feedstocks to hydrocrackers. During the winter, when gasoline demand is lower and home heating oil is needed, the severity of the operation is lowered by reducing the reactor temperature, and the yields of the cycle gas oils stocks used for blending into home heating oils and diesel fuels is more than doubled. The gasoline boiling-range products are reduced by one-quarter to one-third. The butane and lighter gases are separated into components in the FCC gas recovery plant. The propylenes, butylenes, and isobutane are used as feedstocks to alkylation units to make alkylate gasoline blending stocks. Alkylates have blending octanes in the nineties and have low sensitivities to driving conditions (they perform equally well for highway and city driving with RON and MON within two numbers of each other). The hydrocarbons in alkylates are also less environmentally damaging and show less health effects than olefins or aromatics. For those refineries without alkylation units, the propylenes can be used as feedstocks to polymerization units to produce polymer gasoline blending stocks. The polymer gasoline has a blending octane of approximately 90 but performs much more poorly under highway driving conditions than in the city (about sixteen numbers difference between the RON and MON).

Some refineries have hydrocracking units in addition to the FCC unit. Hydrocracking units operate at high hydrogen pressures and temperatures and are constructed of high-alloy steels. This makes them very expensive to build and operate. They are able to use feedstocks containing high concentrations of aromatics and olefins and produce jet fuel and diesel fuel products as well as naphthas for upgrading and blending into gasolines. The FCC units operate most efficiently with paraffinic feedstocks and produce high yields of high-octane gasoline blending stocks. Hydrocrackers give higher yields than FCC units, but the naphtha products have octanes in the seventies and the heavy naphtha fraction—180–380°F (82–195°C)—must be sent to the catalytic reformer to improve its octane. It is possible to make specification jet and diesel fuels as products from the hydrocracker.

In any cracking process, secondary reactions produce light hydrocarbons whose boiling points are too low to blend into gasolines. These include both olefins and paraffins. Olefins contain at least one double bond in each molecule and are very reactive whereas paraffin molecules contain only single bonds and are considered unlikely to undergo chemical reactions. Some of the light hydrocarbon olefins produced can be reacted with each other (polymerization) or with isobutane (alkylation) to produce high–octane hydrocarbons in the gasoline boiling. The olefins used are propylenes and butylenes; ethylene is also produced from cracking operations but is not used in refinery processing.

Crude oil components with boiling ranges in the gasoline blending range have low octanes, and the octanes must be increased in order to make a product suitable for modern automobiles. For light naphthas containing molecules with five and six carbons (pentanes and hexanes), octanes can be improved from thirteen to twenty numbers on the average by isomerization processes. These processes convert normal paraffins into isoparaffins that have more compact molecules with higher octane numbers. The heavy naphthas with hydrocarbons containing from seven to ten carbon atoms, have their octanes increased into the nineties by converting paraffin molecules into aromatic molecules by reforming the molecules into ring structures and stripping off hydrogen. This process is catalytic reforming. Catalytic reforming is the only refinery process in which the operator has any significant control over the octane of the product. The more severe the operation (the higher the temperature in the reactor), the greater the percentage of aromatics and the higher the octanes of the products—but the smaller the volume of products produced. Usually the units are operated to produce products (reformate) with research octane numbers (RON) in the range of 96 to 102 to be blended with other naphthas to give the desired quality (pool octanes) sold to consumers.

Other refinery processes used to improve product quality include hydrotreating (hydrodesulfurization, hydrodenitrogenation, and demetallization), sweetening, and visbreaking. Hydrotreating reduces impurity contents of all of the impurities except metals by reacting hydrogen with the impurity to produce low-boiling compounds that are gases and can be separated from the liquid material. These reactions are carried out at high temperatures and pressures in reactors filled with catalysts containing metals that promote the reaction of hydrogen with the impurity. Hydrogen reacts with sulfur to form hydrogen sulfide, with nitrogen to produce ammonia, and with the molecules containing metals to deposit the metals on the catalyst supports. The processes operate at temperatures between 600°F (315°C) and 750°F (400°C) and pressures from 400 psig (27 bars) to 1600 psig (110 bars). Because these processes operate with hydrogen at high temperatures, the equipment is made of very expensive high–alloy steels. Energy and hydrogen costs result in high operating costs, much higher per barrel of feed than the FCC unit.

Refiners use sweetening processes to remove mercaptans that give a very unpleasant odor to gasolines and middle distillates (the skunk uses mercaptans to protect itself). This is done by washing the hydrocarbon stream with a caustic solution followed by a wash with water to remove the caustic.

Visbreaking is a mild thermal cracking process that reduces the viscosity of heavy fuel oils and reduces the amount of low-viscosity blending stocks that must be added to the heavy residuals to meet viscosity specifications of the specific heavy fuel oil. The amount of heavy fuel oil production by a refinery is reduced by 20–30 percent if a visbreaker is used. The refinery profitability is improved with visbreaker operation, because heavy fuel oils are low value products.

Small refineries usually produce the products in greatest demand, such as gasolines, jet and diesel fuels, and heavy fuel oils. Larger refineries produce a much broader slate of products because, with much higher feed rates, they can make economical volumes of small volume products. In addition, environmental requirements have made it much more difficult for small refineries because of the large capital investments required for the equipment to reduce impurities. Published economic studies indicate that refineries of less than 80,000 barrels per day of capacity do not produce sufficient income to justify the expensive equipment necessary for reformulated fuel production. Having fewer large refineries increases the transportation costs of getting the products to the necessary markets, but the increased size gives economies of scale. There are always fixed costs of labor and management that are independent of equipment size, and the incremental costs of higher charge rates are much less per unit volume than the total of fixed and operating costs. Depreciation costs per unit volume of throughput also decrease with the size of the equipment. In 1998, U.S. refineries operated at above 95 percent of rated charge capacity. This is much higher than it has been in the past. Even so, they have not been able to keep up with domestic demand, and imports of finished products have been increasing.

Operating costs vary a great deal from one refinery to another. Factors include the types of processes and equipment, the amount of the crudes that have to be cracked and alkylated to achieve the products desired, the degree of hydrogenation needed to meet product and environmental specifications, and the complexities and locations of the refineries. Energy is the single largest component of operating cost, so the costs of crude oils are major factors. For a simple hydroskimming refinery, the energy needed to process the crude oil can be as low as 2 percent of that in the crude oil, while for a very complex refinery with a large number of hydrogenation processes, it can be 8 percent or more of that in the crude oil.

SAFETY AND ENVIRONMENTAL CONCERNS

The increasingly stringent environmental quality improvements for transportation fuels as specified by the EPA is putting constraints on present processing technologies. Process equipment needed for the newer technologies to meet future environmental restrictions is very costly and requires long lead times to design and build. Since the removal of low-cost lead compounds, used until the 1970s to increase the octanes of gasolines, refineries have been providing the replacement octane improvement needed by increasing the high-octane aromatic and olefin content of gasolines and also the amount of high-octane blending compounds containing oxygen (alcohols and ethers). The EPA is reducing the maximum aromatics and olefins content of reformulated gasolines for health and pollution reasons. There also is concern over the appearance of ethers in ground waters because of possible health effects, and it may be necessary to restrict their usage in gasolines. There is not enough ethanol available today to replace these components while providing the volumes of high-octane gasolines needed (over one million tons per day of gasolines used in the United States in 1998). Other ways must be used to provide the high octanes needed and to reduce degradation of the environment.

The United States is unique among the major countries in that supply and demand has determined price structures in the petroleum industry. Today, even though the products are much better than fifty years ago, the before-tax retail prices of gasolines, diesel fuels, and heating oils are much less on a constant-value dollar basis than they ever have been before. Even with the federal and state taxes included, the retail prices on a constant-value dollar basis are equivalent to those paid during the Depression years. In Europe, for example, taxes account for up to 80 percent of the costs of motor fuels that sell at retail for three to four times as much as they do in the United States.

U.S. production of crude oils has already peaked, and it is predicted that world production of crude oils will reach its maximum between 2010 and 2030. This does not mean that an adequate supply of crude oil will not be available, but supply and demand will create price structures that will make other sources of fuel more competitive. Global warming may also impose restrictions on the use of hydrocarbon fuels but, at the present time, there are no alternatives that do not require long lead times and very expensive and time consuming periods for building plants to produce the alternative fuels. If the alternatives cannot be marketed in existing petroleum facilities, it will also require that broad systems be built to market the fuels. It is difficult to conceive of the efforts that must be made to replace gasoline in the United States. Methanol has been mentioned as a replacement, but if the total annual production of methanol in the United States in 1998 was used in automobiles instead of gasoline, it would only be enough to operate them less than a week. During 1998, there were problems with electricity blackouts and restrictions of use in the United States without the added imposition of the power needed to charge batteries for replacement of gasolines in automobiles. conversion from petroleum fueled to electric powered vehicles will require the building of many more generating plants to supply the electricity necessary to meet transportation needs. When petroleum based fuels are replaced for transportation, several fuels probably will be used rather than one. Local deliveries, long distance deliveries, and travel will each use the most economical and environmentally friendly fuel.

James H. Gary

See also: Refining, History of.

BIBLIOGRAPHY

Chang, Thi. (1998). "Distillation Capacities Hold Steady, More Mergers Planned." *Oil & Gas Journal* 96(51):41–48.

Gary, J. H., and Handwerk, G. C. (1995). *Petroleum Refining, Technology and Economics*. New York: Marcel Dekker.

Radler, M. (1998). "1998 Worldwide Refining Survey." *Oil & Gas Journal* 96(51):49–92.

REFINING, HISTORY OF

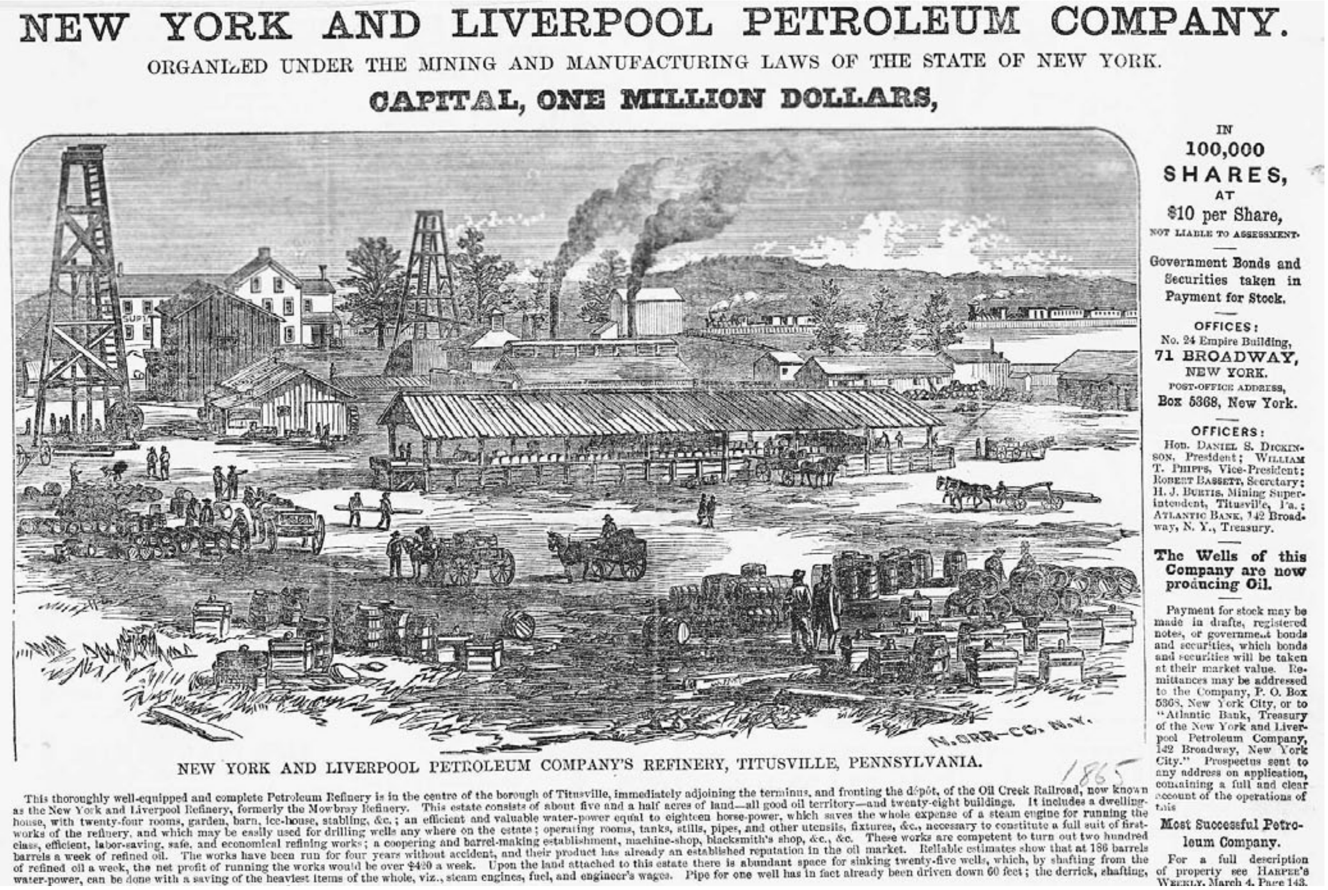
The U.S. petroleum refining industry has played a central role in the expansion of this country's energy capacity over the last century. Currently, petroleum refiners generate products which account for approximately 40 percent of the total energy consumed in the United States (with respect to Btu units). The industry is characterized by a small number of large, integrated companies with multiple high-capacity refining facilities.

Between 1982 and 1997, the total number of U.S. refineries had declined from 300 to 164 operating companies. This contraction was due for the most part to the closing down of the smaller refining operations, i.e., refineries with less than 50,000 barrels of crude oil per day (BPD) capacity. While the smaller refineries still generally account for up to half of all U.S. facilities, in aggregate they control barely 14 percent of total U.S. crude refining capacity.

PROCESS DESCRIPTION: OVERVIEW

In general, refining consists of two major phases of production. The first phase of production acts on the crude oil once as soon as it enters the plant. It involves distilling or separating of the crude oil into various fractional components. Distillation involves the following procedures: heating, vaporization, fractionation, condensation, and cooling of feedstock.

Distillation essentially is a physical operation, i.e., the basic composition of the oil remains unchanged. The second phase of production, which follows distillation, is chemical in nature in that it fundamentally alters the molecular composition of the various oil fractions. These processes depend on heat and pressure and, virtually in all cases, the use of catalysts. These processes are designed to improve the octane number of fuel products. For example, in the process of isomerization, rearrangement of the molecules occurs, resulting in different chemical configurations (e.g., branched vis-à-vis linear structures) but with the atomic composition remaining constant. Polymerization involves the formation of larger molecules (polymers) from smaller ones. Reforming transforms hydrocarbons into other hydrocarbons, often to make aromatics from nonaromatic petroleum constituents (e.g., paraffins and naphthenes).



NEW YORK AND LIVERPOOL PETROLEUM COMPANY.

ORGANIZED UNDER THE MINING AND MANUFACTURING LAWS OF THE STATE OF NEW YORK.

CAPITAL, ONE MILLION DOLLARS,

IN
100,000
SHARES,
AT
$10 per Share,
NOT LIABLE TO ASSESSMENT.

Government Bonds and Securities taken in Payment for Stock.

OFFICES:
No. 24 Empire Building,
71 BROADWAY,
NEW YORK.
POST-OFFICE ADDRESS,
Box 5368, New York.

OFFICERS:
Hon. DANIEL S. DICKINSON, President; WILLIAM T. PHIPPS, Vice-President; ROBERT BASSETT, Secretary; H. J. BURTIS, Mining Superintendent, Titusville, Pa.; ATLANTIC BANK, 142 Broadway, N. Y., Treasury.

The Wells of this Company are now producing Oil.

Payment for stock may be made in drafts, registered notes, or government bonds and securities, which bonds and securities will be taken at their market value. Remittances may be addressed to the Company, P. O. Box 5368, New York City, or to "Atlantic Bank, Treasury of the New York and Liverpool Petroleum Company, 142 Broadway, New York City." Prospectus sent to any address on application, containing a full and clear account of the operations of this

Most Successful Petroleum Company.

For a full description of property see HARPER'S WEEKLY, March 4, Page 143.

NEW YORK AND LIVERPOOL PETROLEUM COMPANY'S REFINERY, TITUSVILLE, PENNSYLVANIA.

1865

This thoroughly well-equipped and complete Petroleum Refinery is in the centre of the borough of Titusville, immediately adjoining the terminus, and fronting the dépôt, of the Oil Creek Railroad, now known as the New York and Liverpool Refinery, formerly the Mowbray Refinery. This estate consists of about five and a half acres of land—all good oil territory—and twenty-eight buildings. It includes a dwelling-house, with twenty-four rooms, garden, barn, ice-house, stabling, &c.; an efficient and valuable water-power equal to eighteen horse-power, which saves the whole expense of a steam engine for running the works of the refinery, and which may be easily used for drilling wells any where on the estate; operating rooms, tanks, stills, pipes, and other utensils, fixtures, &c., necessary to constitute a full suit of first-class, efficient, labor-saving, safe, and economical refining works; a coopering and barrel-making establishment, machine-shop, blacksmith's shop, &c., &c. These works are competent to turn out two hundred barrels a week of refined oil. The works have been run for four years without accident, and their product has already an established reputation in the oil market. Reliable estimates show that at 186 barrels of refined oil a week, the net profit of running the works would be over $420 a week. Upon the land attached to this estate there is abundant space for sinking twenty-five wells, which, by shafting from the water-power, can be done with a saving of the heaviest items of the whole, viz., steam engines, fuel, and engineer's wages. Pipe for one well has in fact already been driven down 60 feet; the derrick, shafting,

New York and Liverpool Petroleum Company refinery in Titusville, Pennsylvania, 1865. (Corbis Corporation)

The most important of the second phase refinery operations is cracking. Cracking breaks down large hydrocarbon molecules into smaller and lighter molecular units. At first performed using high temperatures and pressures only, by the 1930s, catalytic cracking was begun. Catalytic cracking went further than earlier cracking technology to reduce the production of less valuable products such as heavy fuel oil and cutter stock.

Cracking represents a fundamental advance in refining process technology. Innovation here was mostly responsible for increasing the throughput capability of the industry and heightening octane numbers for both motor and aviation fuel. Cracking technology was central to the success of the U.S. petroleum industry in delivering needed fuel for both the marketplace and the strategic requirements of World War II.

Cracking process technology has been a distinctly American achievement. Although Europeans helped transfer some important cracking innovations to the United States, it was the U.S.-based companies and their engineers who converted these yet unrealized prototypes into commercially viable, full-scale plants.

Cracking technology evolved out of the need to solve a series of increasingly complex technical problems. Ultimate success in handling such difficult problems as carbon buildup on the catalyst surface, non-uniform distribution of thermal energy, and catalyst breakage and equipment failure came with development of the fluid catalytic cracking process in the early 1940s.

EARLY REFINING TECHNOLOGY

In the late nineteenth and early part of the twentieth centuries, refining operations were in essence distillation procedures. The refinery did little more than separate the petroleum into various fractions for commercial use. Prior to World War I, gasoline was not a dominant product of the refinery. The most commercially useful products resided in the lower (e.g., denser) range of the petroleum fractions. In order of importance, these products consisted of kerosene (for heating and light), and a variety of oils, waxes, and lubricants for industrial and home use.

The machinery employed in the early refineries was rather small in scale and operated inefficiently. Generally, equipment consisted of a series of shell tubes or stills. These were placed in the horizontal position and were connected one to another from the top through the use of vapor pipes. These pipes directed the vapors from the stills into condensers which cooled the gases and so caused the products to separate out. These products were then collected in sequence, often at one point in the plant, as liquids of varying densities and properties.

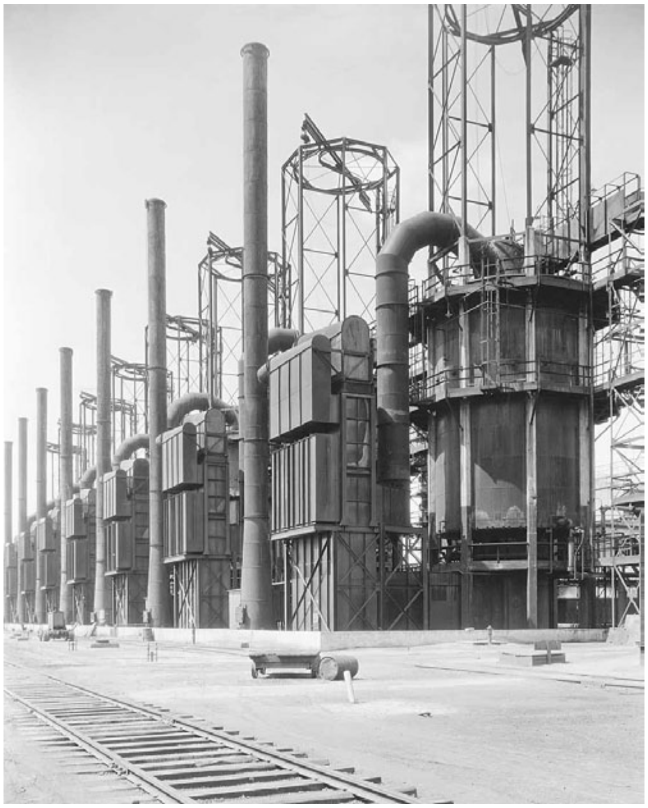
Despite various mechanical attempts to increase throughput, operations were at first conducted in batches, which required the plant to be shut down for the still and ancillary equipment to be cleaned out. Until cracking technology entered the picture, refinery operations were inefficient as they captured for use barely 50 percent of the available petroleum.

THERMAL CRACKING TECHNOLOGY

The initial impetus for development of a cracking process came in the years prior to World War I with the greater demand for gasoline products for the emerging automotive industry. Although early experimental work on the high-pressure cracking of petroleum was conducted in Europe in the late nineteenth century, the commercial breakthrough came in the years leading up to World War I by Standard Oil (Indiana). In 1913, William Burton, a petroleum chemist at Standard Indiana, conceived of commercially breaking down the molecular components comprising petroleum into smaller molecular units by the force of high pressure.

While his experiments yielded acceptable gasolines, serious problems arose as Burton attempted to scale up his process. The short cycles forced on producers by extensive coking worked against continuous operations. The carbon buildup interfered with heat transfer from the furnace to the petroleum and resulted in the formation of "hot spots" and in turn damage to the vessel. Pressure control was also a problem.

An assistant of Burton at Standard Indiana, E. M. Clark modified Burton's still in a fundamental way. Clark perceived that Burton's problems arose because heat and pressure were being applied to a static mass of petroleum within the autoclave. This situation was conducive to the formation of carbon at the bottom of the vessel and made it difficult to control pressure and obtain a uniform cracking throughout the charge. Clark retained Burton's use of high pressures but within the context of the tube still. In this design, oil flowed through, and was cracked within, banks of tubes. Each tube measured five feet in length. The hydrostatic pressures within the tubes produced uniform flow conditions. These banks of tubes were suspended in a furnace. Partially cracked charge was then directed upward to an overhead "soaking" tank. Here, cracking continued to completion under pressure but without the application of heat.



View of cracking stills at a Sinclair oil refinery. (Corbis Corporation)

An important aspect of Clark's technology was that the oil being cracked, which flowed through the tubes and in a near vapor state, was maintained in a continually dynamic or turbulent condition. This meant that the coke particles which formed were prevented from adhering to the sides of the tubes. Further, as a result of the high surface/volume ratio of the still, heat transfer through the tubes was facilitated and higher pressures could be applied. Thus, whereas the Burton process operated at a maximum of 100 psi, Clark's tube still handled pressures of up to 1,000 psi.

The first commercial Tube-and-Tank cracking plant came on line in 1922. Overall, compared to the Burton Process, the Tube-and-Tank Process allowed larger volumes of petroleum to be processed under conditions of intense cracking and longer production cycles.

In the 1920s, a number of new thermal cracking technologies emerged that, in essence, were variations of the Burton-Clark designs. These included most notably the Cross and Holmes-Manley processes. These processes provided innovations in such parameters as operating conditions, methods of removing coke, and the number and configuration of the overhead soaking chambers.

These processes remained batch-type operations. As such they reached capacities in the range of 1,000–3,000 barrels per day (BPD). A truly novel approach to thermal cracking emerged from the facilities of the Universal Oil Products company in the late 1920s. To a greater extent than the other thermal cracking technologies, the so-called Dubbs Process approached continuous operations. Important elements of this technology were the continuous recycling of heavier byproducts back into the cracking section for further processing and the so-called bottom draw-off technique that continuously removed heavy liquid from the bottom portion of the soaking drum. This procedure reduced the rate of carbon buildup in the system and so increased the time over which the still could operate before being shut down for cleaning.

CATALYTIC CRACKING TECHNOLOGY

By the early 1930s, thermal cracking had achieved a fairly high level of operation. Both the Dubbs (UOP) and Tube-and-Tank (Jersey Standard) Processes represented the state of the art in the field. Between the end of World War I, when the Burton Process was still revolutionary, and the early 1930s, octane ratings of gasoline increased 36 percent. This improvement resulted from the existence of more advanced thermal plants and the increasing use of additives, especially tetraethyl lead. Moreover, the quantity of gasoline produced increased. Not only were the Dubbs and Tube-and-Tank technologies inherently more efficient than earlier thermal designs, but they were also readily scaled up in capacity. Through the 1930s, the size of thermal plants increased. The dimensions and amount of equipment used within a facility grew.

By the mid-1930s, thermal cracking reached its zenith in scale and sophistication. At this time the world's largest thermal cracking unit, built by Indiana Standard in Texas city, Texas, had a capacity of 24,000 BPD. This represents a capacity over 270 times larger than the first Burton stills and over forty-two times larger than the early Tube-and-Tank units.

Fixed Bed Processes: Houdry Catalytic Cracking

At this time, forward-looking companies understood that the ability of a refiner to control the highest octane fuels could corner critical and specialized niche automotive and aviation fuel markets. Refiners looked on catalytic cracking as a way to more finely tune their products for these market segments.

By the mid-1930s, catalytic technology entered into petroleum refining. To a greater extent than thermal cracking, catalysis permitted the close control of the rate and direction of reaction. It minimized the formation of unwanted side reactions, such as carbon formation, and overall improved the yield and quality of fuel output.

Within the United States, catalytic cracking was carried out on petroleum vapors. In contrast to thermal cracking, catalytic cracking did not require the application of ultra-high pressures.

The first attempt at a commercial catalytic process was the Macafee Process. In this design, cracking took place within a circular vessel packed with aluminum chloride catalyst. Cracked gasoline vapor products exited the reactor through a series of chimney-like structures. The Gulf Refining Company built a working plant at Port Arthur, Texas, just after World War I. However, the rapid accumulation of carbon deposits on the inside of the reactor and the high cost of the catalyst precluded further development in this direction.

Further catalytic cracking efforts did not take place in earnest in this country until the early 1930s. Sun Oil undertook the first significant commercial effort in catalytic cracking beginning in 1932. At this time, Sun Oil, based in Marcus Hook, Pennsylvania, was a relatively small company. It operated a number of oil tankers and had at its disposal a growing pipeline network in the Northeast and increasingly the South and West. In addition to a refinery at Marcus Hook, it operated a large refinery at Toledo, Ohio. Sun established its reputation as a producer of a variety of high-quality fuels and related oil products. Having taken thermal cracking to its limit through the use of high pressures, Sun embraced a promising new method developed by Eugene Houdry, a French mechanical and automotive engineer.

Through the 1920s Houdry had experimented in France on a number of possible catalytic routes to higher octane fuel. Finding little success in France, he came to the United States to further develop his process. After initial attempts at commercialization under the sponsorship of Sacony-Vacuum Company (currently Mobil) in Paulsboro, New Jersey, failed, Houdry and his development company, Houdry Process Corporation, moved to Sun.

Over the next four years, Houdry, working closely with Sun's engineering team headed by Clarence Thayer, worked to build a commercial plant. The limitations imposed by a static catalyst bed design imposed a major obstacle, particularly in the formation of carbon deposits that fouled the catalyst mass and impeded a continuous system of production.

The commercialization of the Houdry Process involved borrowing and integrating mechanical designs from the automotive, electrical, and metallurgical industries. The Houdry Process consisted of a series of catalyst cases containing the cracking catalyst. Each case went through a succession of operations: cracking (which over time resulted in carbon deposits on the catalyst), preparation of the catalyst for regeneration (via the purging of oil vapors), regeneration of the spent catalyst (burning off of carbon deposits from the catalyst surface), preparations for the next cracking operation (i.e., purging the newly regenerated catalyst of combustion or flue gases), cracking, and so on. To approximate continuous operation, the process placed the catalyst cases on different phases of the cycle (via a staggered arrangement) at any one time. In this way, cracking, purging, and regenerations were carried on simultaneously in different cases.

Internal tube components were an important part of the Houdry Process. Tubes were used to distribute oil vapors over the catalyst during cracking and to receive and direct cracked vapors and combustion gases (after regeneration) to various parts of the system. Tubes also played a central role in heat control. Fluid-filled, heat control tubes were placed deep within the catalyst bed so that they tapped and transferred heat which built up throughout the mass.

The hot fluid moved through the tubes to a boiler into which heat energy was transferred. Steam generated here was transported back through the tubes to heat the catalyst for the cracking operation. Initially, the heat transfer fluid used was water or superheated steam. Because these fluids tended to cause tubes to crack, a more advanced design was developed that utilized molten salts, chemically and physically more stable under the rigorous operating conditions.

By the early 1940s, a Houdry plant capacity ranged from 7,000 to 15,000 barrels per day. Houdry's contribution to the U.S. energy industries extends beyond fuels. The Houdry Process was the first industrial technology in the United States to employ on a large scale the gas turbine system, adapted from European designs, and the heart of the energy recycling process. Capturing and turning into useful mechanical energy the heat generated during combustion in the regeneration cycle was especially critical. The Houdry Process paved the way for energy recycling technology, including the process commonly called cogeneration, in such industries as power generation and iron and steel production.

But there were many problems in the operation of the Houdry Process that could not be resolved. Even with its sophisticated regeneration system, carbon continued to form on the catalyst over a period of time. The heat transfer system was not as efficient as it needed to be. The process never achieved fully continuous operations, limiting the quantity of oil that could be processed and the quality of the gasoline produced. At its height (late 1930s, early 1940s) the process did not produce gasoline with an octane rating above 87. It never captured more than 10 percent of the total U.S. cracking capacity, although in the early 1940s it controlled over 90 percent of the total catalytic cracking market.

Moving-Bed Processes: The Thermofor and Air-Lift Systems

Sun attempted to resolve these problems by improving upon the fixed-bed process. The Sacony-Vacuum Company (formerly the Standard Oil company of New York and the future Mobil Oil) was the first to develop a different type of catalytic cracking process. While involved with Sun Oil in development of the fixed-bed process, Sacony early on understood the limitations of that technology. Beginning in 1932, Sacony began conceiving of the moving-bed concept.

In its development, it adapted two existing technologies. In the agricultural sector, the mechanics of grain elevators provided a model for how to move solids vertical distances and in closed-loop flow arrangements. Sacony engineers modified the elevator bucket systems traditionally used by the grain industry to carry hot catalyst from the bottom to top of vessels and between vessels.

Then too, Sacony modeled its regenerator vessel after a certain type of metallurgical furnace (known as the Thermofor kiln). The vessel consisted of a series of semi-independent burning zones. Distribution channels delivered compressed oxygen to each zone to fuel combustion. A series of baffles and ducts within each combustion compartment produced a uniform distribution of air through the catalyst. Flue gases emitted by the regeneration process collected in common headers located between the burning zones.

The so-called Thermofor moving-bed process borrowed two important techniques from Houdry's technology: a molten salt cooling system for the kiln section and gas turbine technology that generated power to pressurize more air for the regenerator. The catalyst was stored in an overhead hopper, from which it was fed into the catalyst-filled reactor. The catalyst then traveled down the vessel under the influence of gravity. In the commercial plants, oil, injected as a liquid spray near the top of the reactor, moved down along with the catalyst. During this time, the latter transferred its heat, obtained and stored during the previous regeneration cycle, to the oil. The system conserved on energy by thus recycling heat units through the catalyst which simultaneously vaporized and cracked the oil particles that descended with it.

Following cracking, the spent catalyst and oil descended to a disengager that separated the gasoline from the catalyst. The catalyst, with oil residue entrained on its surface, then moved through a purging section where superheated steam thermally removed oil remnants. The oil-free catalyst, still laden with carbon deposits, was then lifted by elevator from the bottom of the reactor to the top of the regenerator.

Regeneration was carried out as the catalyst fell through various burning zones. The flue gases were recycled by being directed to the turbo-compressors that the pressurized air used as fuel in the combustion process.

The first Thermofor cracking unit came on line in late 1942. By March of 1943, twenty Thermofor units had been completed or were under construction. The larger Thermofor plants could circulate 100-150 tons of catalyst per hour and could process up to 20,000 barrels of petroleum per day.

But problems persisted. The catalyst, moving at rapid rates, tended to disintegrate as it impacted the inside surface of equipment . Dust particles formed, clogging pipes and transfer lines and disrupting the smooth flow of operation. This difficulty in turn prevented uniform heat distribution through the system and affected the rate and extent of both cracking and regeneration. Both the volume and quality of fuel produced suffered.

An improved design undertaken by Sacony used high-velocity gases to replace the mechanical elevator systems as catalyst carriers. These so-called "air-lift" units improved upon the Thermofor process both in terms of economies and octane numbers. It was, however, only with the fluid cracking process that catalytic technology realized fully continuous production.

Fluid Catalytic Cracking

Fluid catalytic technology addressed two major shortcomings of previous moving-bed systems: the slowness with which catalyst traveled through the vessels and the inability to sustain the cracking process continuously over an indefinite period of time. The central problem facing the industry was how to move the catalyst around the production circuit at rapid rates and at the same time fully control the intensity and duration of cracking (and regeneration). Jersey Standard (Exxon), up to this point not a major player in advanced catalytic cracking, addressed these issues as it continued to improve upon more traditional cracking processes.

By the time Jersey Standard moved into fluid research in the late 1930s, it possessed both cracking (thermal) and catalytic expertise. As noted above, Jersey developed its most important cracking technology up to that point with its Tube-and-Tank process, a thermal approach. And beginning in 1927, the company formed a patent-sharing agreement with the German giant, I. G. Farben, for development of a high-pressure catalytic hydrogenation process. Jersey's early catalytic work resulted in a number of reforming practices, such as hydroforming (1938) and steam reforming (1942) which supplemented cracking processes.

By the late 1930s, Jersey, associated with a group of companies looking for improved catalytic processes (the Catalytic Research Associates), combined its past cracking and catalytic experience in developing a fully continuous catalytic cracking system.

Prior to coming upon the notion of fluidization as a basic principle of continuous catalytic cracking, Jersey experimented with a series of fixed-and moving-bed systems. A unique approach emerging out of these efforts involved the design of "slurry" systems, by which a pumping mechanism propelled catalyst particles and oil fluid together in concurrent fashion along a horizontal reactor.

While problems with the circulatory (i.e., pumping) system limited the usefulness of this approach, it led to the next major leap for fluid cracking. Warren K. Lewis, head of the Department of Chemical Engineering at MIT and consultant to Jersey, was the leading creative force behind fluid cracking. He led the initial experiments (at MIT) identifying the commercial viability of, and establishing the preconditions for, fluidization phenomena, including the existing of the so-called turbulent bed. He also directed the design of the first semi-commercial units.

Fluid cracking retained the moving bed concept of the catalyst transported regularly between the reactor and regenerator. And as with the air-lift systems, the fluid plant rejected mechanical carrying devices (elevators) in favor of standpipes through which the catalyst fluid traveled.

Fluid technology exceeded gas-lift technique by incorporating two innovative concepts, that is, under certain conditions of velocity flow and solid/vapor concentration: (1) a catalyst-gas mixture traveling around the plant behaves in just the same way as a circulating liquid; and (2) within vessels, rather than the catalyst and oil falling under gravity, they closely intermingle indefinitely in a dense turbulent but stable, well-defined and carefully (and indefinitely) controlled cracking "bed."

The fact that a flowing catalyst-vapor mixture acted just as a traditional liquid was a crucial point. It meant that the cracking plant was in essence a hydrodynamic system readily controlled over a range of velocities and throughput by simple adjustment of pressure gradients and such variables as the height of the catalyst (i.e., the head of pressure) in the standpipe and the velocity of upward moving gas which carried the catalyst from the bottom of the standpipe to processing vessels. An important control element also was the "aeration" of the moving catalyst at strategic points along the standpipe (and transfer and carrier lines) by use of jets of air to adjust pressure gradients as required.

At the heart of the process was the turbulent bed. As catalyst rose with the carrier gas up the standpipe and into the reactor, it tended to "slip back" and concentrate itself into a boiling but well-delineated mass. The boiling action served to keep the catalyst particles tumbling about. Rather than falling via gravity, the boiling action in effect served to counteract the force of gravity and so maintain the catalyst in an indefinitely suspended state. Particles simply moved from one part of the bed—the turbulent mass—to another, all the while undergoing cracking. This design applied as well during combustion in the regenerator.

The plant was able to operate continuously. The continual and controlled state of turbulence in the bed assured close intermixing between solids and vapors and an even distribution of thermal energy throughout the bed, and the "liquid" catalyst flowed smoothly and rapidly from one vessel to the next.

An early fluid cracking unit removed spent catalyst from reactors (to be directed to the regenerator) using an overhead cyclone system. A more efficient technique, the so-called down-flow design, followed. It altered operating and flow conditions so that spent catalyst concentrated in the bottom part of vessels, where they could be removed, resulting in greater ease of catalyst recovery, simplification of plant layout, and improvement in operating flexibility.

Fluid catalytic cracking rapidly overtook its competitors as both a source of fuel and of critical organic intermediates. Prior to 1942, the Houdry Process controlled 90 percent of the catalytic fuel market. But only three years later, in 1945, fluid cracking led all other catalytic cracking processes in market share (40 percent). At this time Thermofor technology stood at 31 percent, and Houdry at less than 30 percent.

After 1942, fluid cracking technology increasingly dominated U.S. petrochemical production. It manufactured high-tonnage fuels for both motor vehicles

A fluid catalytic cracking unit in Joliet, Illinois, converts heavy components of crude oils into high octane gasoline and distillates. (Corbis Corporation)

and aircraft for the wartime effort. The quality of the gasoline was unprecedented. Octane ratings of fluid-produced fuels exceeded 95, an unheard-of figure only a few years before, and critical for aviation gasoline. Fluid cracking technology played a central role the U.S. synthetic rubber program: its byproduct gases supplied the butylenes which were essential in the making of butadiene, the essential rubber intermediate.

RECENT DEVELOPMENTS IN PETROLEUM REFINING

Since 1945, the fluid catalytic cracking process has rapidly overtaken fuel production and has become the central technology in the U.S. petrochemicals industry. With fluid cracking, the scale of petrochemical operations grew enormously. For the first time, refiners could process virtually any volume of oil rapidly and efficiently.

Accordingly, technological change in refining technology has centered on alterations and improvements made to the fluid cracking process. These modifications have included improvements made to catalysts, materials of construction, interior linings, and a variety of mechanical details. Since 1945, the process has operated on progressively higher temperatures for both cracking and regeneration. It has also incorporate newer generations of catalysts.

By the early 1960s, fluid cracking had become the workhorse of the refining industry. It was the central process in the production of over 70 percent of all high-octane fuel. From early 1940s to mid-1960s, capacities of fluid units have grown from less than 20,000 BPD to between 100,000–200,000 BPD.

The flexibility of the process was manifested in its ability to economically process smaller volumes of feedstock. Between the 1950s and 1980s, refiners and engineering firms developed smaller, lower investment units which could be readily scaled up as required. Recently, such fluid crackers have been simplified to the point that cracking and regeneration take place within a single, vertical standpipe or riser tube. The capacities of such units often do not exceed 1,000 BPD.

By the mid-1990s, the technology was virtually the only catalytic cracking process in operation in the major refineries. The technology accounted for over 95 percent of all high-octane fuel within the United States. By 1997, there were approximately 350 fluid cracking facilities in operation worldwide, most located within the United States. Between 8 and 9 percent of the fluid units existing worldwide (25-30 units) are owned and operated by Exxon, the original innovator of fluid cracking.

In the late 1990s, the growth of fluid cracking as a major petroleum refining process was about 2 percent per year. Total world fresh feed capacity for the fluid process now stands at more than 11 million barrels/day. Worldwide, FCC makes 80 billion gallons of high-grade gasoline annually. This represents nearly half of total world gasoline production.

A major recent development for the technology has been its closer integration with the large petrochemical complex. Essential petrochemical activity has been relying more on fluid technology and less on thermal units for their intermediate feedstock. The big development for the 1990s is increasing integration of FCC with the large petrochemical plant.

In the postwar period, fluidization influenced U.S. manufacture in other areas as well. Most critically, it has been applied in combustion processes for the making of metallurgical materials and by the utilities for generating heat and electricity for industrial and residential use. Moreover, the technology is a "cleaner" method for producing energy and so is an important means by power-based companies to comply with stricter environmental regulations.

Sanford L. Moskowitz

See also: Refineries.

BIBLIOGRAPHY

Enos, J. (1962). *Petroleum, Progress and Profits: A History of Process Innovation.* Cambridge, MA: MIT Press.

Frankenburg, W. G. et al., eds. (1954). *Advances in Catalysis and Related Subjects, Vol. 6.* New York: Academic Press.

Grace, J. R., and Matsen, J. M., eds. (1980). *Fluidization.* New York: Plenum Publishing.

Giebelhaus, A. W. (1980). *Business and Government in the Oil Industry: A Case Study of Sun Oil, 1876–1945.* Greenwich, CT: JAI Press.

Larson, H. M. et al. (1971). *History of Standard Oil Company (New Jersey): New Horizons, 1927–1950.* New York: Harper and Row.

Moskowitz, S. L. (1999). "Science, Engineering and the American Technological Climate: The Extent and Limits of Technological Momentum in the Development of the U.S. Vapor-Phase Catalytic Reactor, 1916–1950." Diss. Columbia University.

Murphree, E. V. et al. (1945). "Improved Fluid Process for Catalytic Cracking." *Transactions of the American Institute of Chemical Engineers.* 41:19–20.

Nelson, W. L. (1949). *Petroleum Refining Engineering.* New York: McGraw-Hill.

Othmer, D. F. (1956). *Fluidization.* New York: Reinhold Publishing Corp.

Popple, C. S. (1952). *Standard Oil Company (New Jersey) in World War II.* New York: Standard Oil Company (New Jersey).

Russell, R. P. (1944). "The Genesis of a Giant." *Petroleum Refiner* 23:92–93.

Spitz, P. (1988). *Petrochemicals: The Rise of an Industry.* New York: John Wiley and Sons.

Squires, A. M. (1986). "The Story of Fluid Catalytic Cracking: The First Circulating Fluid Bed." *Proceedings of the First International Conference on Circulating Fluid Beds, November 18-20, 1985.* New York: Pergamon Press.

Williamson, H. F., et al. (1963). *The American Petroleum Industry, Vol. II: The Age of Energy, 1899–1959.* Evanston, IL: Northwestern University Press.

REFRIGERATORS AND FREEZERS

With 120 million household refrigerators and freezers in operation in the United States, their consumption of electrical energy is of major concern not only to the consumer, but also to the power generating utilities that have to provide the power. The government, charged by Congress with guarding against air pollution, protecting the Earth's ozone layer, and fighting global warming, has a keen interest in refrigerator energy consumption.

Vapor compression, absorption refrigeration (instead of electric power, uses heat as the source of energy), and thermoelectric refrigeration (the direct conversion of electrical energy to cooling effect), are the principal means of refrigeration. Of the three methods, vapor compression, often referred to as mechanical refrigeration, is the most energy efficient, approximately two times more efficient than absorption refrigeration, and four times more efficient than thermoelectric refrigeration. Vapor compression is by far the most popular means for refrigerating household refrigerators and freezers, although the other two technologies have unique advantages in some specific applications.

VAPOR COMPRESSION BASICS

Figure 1 shows the four basic elements of the vapor compression system: (1) the evaporator, where the refrigerant vaporizes, and thus absorbs heat from the surroundings; (2) the compressor, where the refrigerant vapor is compressed (typically in the ratio of ten to one); (3) the condenser, where the refrigerant vapor of high pressure and high temperature is condensed by rejecting the heat absorbed by the evaporator, together with the heat of compression, to the atmosphere; and (4) the expansion device, be it an expansion valve or a capillary tube, that allows the liquid refrigerant arriving from the condenser at high pressure and at room temperature, to enter the evaporator, and repeat the refrigeration cycle.

Figure 2, the pressure-enthalpy plot of the standard vapor compression cycle, traces the state of the refrigerant through the refrigeration system. (Enthalpy represents the energy of the refrigerant as

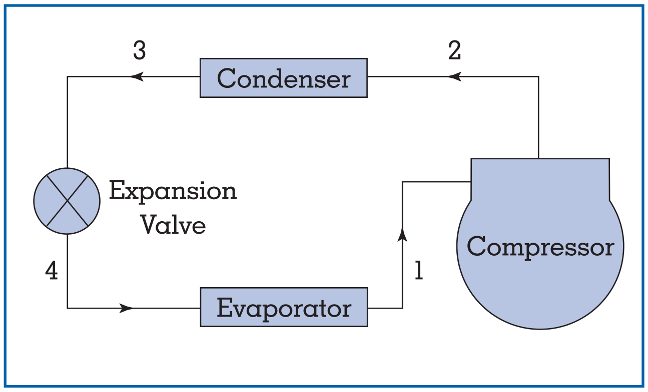


Figure 1.
Flow diagram of the standard vapor compression system.
SOURCE: W. F. Stoecker and J. W. Jones. (1982). *Refrigeration and Air Conditioning,* p. 201. New York: McGraw-Hill Book Company. Modified by author.

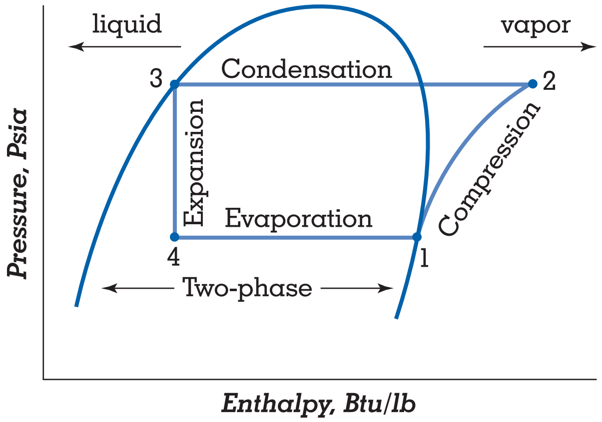


Figure 2.
The standard vapor compression cycle on the pressure-energy diagram.
SOURCE: W. F. Stoecker and J. W. Jones. (1982). *Refrigeration and Air Conditioning,* p. 201. New York: McGraw-Hill Book Company. Modified by author.

it circulates through the various components.) Note that the evaporator absorbs heat at low pressure, and the condenser rejects the heat absorbed by the evaporator and the work of compression, at high pressure. The curvature is the saturation curve of the refrigerant. It delineates the various *states* that the refrigerant passes through during the cycle, liquid to the left of the curve, two-phase mixture within the curve, *vapor* to the right of the curve.

The efficiency of any device is the ratio between what we are after to what we have to pay for it. In the case of a refrigerated appliance, we are after the refrigeration effect, (i.e., heat absorbed by the evaporator) and we pay for it with electric power that runs the compressor. In Figure 2, the evaporator absorbs more energy in the form of heat than the energy supplied to the *compressor* in the form of electricity. The ratio of the two is called the coefficient of performance (COP), a dimensionless number. Compressor efficiency is commonly expressed as the energy efficiency ratio (EER), the ratio of the refrigeration effect in Btu per lb of refrigerant, to the *energy input* to the compressor, in watt-hour per lb of refrigerant. The EER of a typical refrigerator compressor in the 1990s was about 5.5 Btu per watt-hour.

THE HISTORY OF VAPOR COMPRESSION

In 1748, William Cullen of the University of Glasgow in Scotland made the earliest demonstration of man-made cold when he evaporated ether in a partial vacuum. In 1824, Michael Faraday of the Royal Institute in London, found that ammonia vapor, when condensed by compression, would boil violently and become very cold when the pressure is removed. Only ten years later, in 1834, Jacob Perkins, an American living in London, patented the first closed vapor compression system that included a compressor, condenser, expansion device, and evaporator. This is the same basic system as in today's domestic refrigerator. He used ethyl ether as the refrigerant, and received British Patent 6662, dated 1834.

In 1844, John Gorrie, Florida, described in the Apalachicola Commercial Advertiser his new machine for making ice. The delays in ice delivery from the Boston lakes forced Gorrie to build his ice machine so that his hospital fever patients could be assured that ice would always be available.

In the early days ethyl ether and methyl ether were the refrigerants of choice. In 1869 the first ammonia system was introduced. Today ammonia is still a very common refrigerant not only in commercial vapor compression systems, but also in absorption refrigerators.

In 1915 Kelvinator marketed the first mechanical domestic refrigerator, followed shortly by Frigidaire. These models used belted compressors underneath wooden ice boxes that required frequent maintenance because of leaky shaft seals. In 1910 General Electric took out a license to build a sealed compressor that eliminated the need for shaft seals. The machine was originally invented in 1894 by Abby Audiffren, a French monk, who developed it for the purpose of cooling the monastery wine. After a major redesign it was introduced as the Monitor Top refrigerator in 1927, and was an instant success. Up to this time refrigerator cabinets were made principally of wood, with a great deal of carpentry, and not suitable for mass production. The Monitor Top had a steel cabinet, fabricated by punch presses and welding. The steel inner liner was attached to the outer shell by phenolic plastic breakerstrips. A more current production is the 27 cu ft refrigerator with through-the-door water and ice dispenser.

The *Monitor Top* used a toxic refrigerant, sulfur dioxide. In the late 1920s, Frigidaire Corporation, then a leading manufacturer of household refrigerators, asked the General Motors research laboratory to develop a refrigerant that is non-toxic and non-flammable. The result was a chlorofluorocarbon (CFC), namely dichlorodifluoromethane, commonly known as Refrigerant 12 of the Freon family. By the time of World War II, it completely replaced sulfur dioxide. In the 1980s, it was discovered that CFCs deplete the ozone layer surrounding the Earth, thus increasing the likelihood of skin cancer. In 1987 the United States joined other industrial nations in signing the

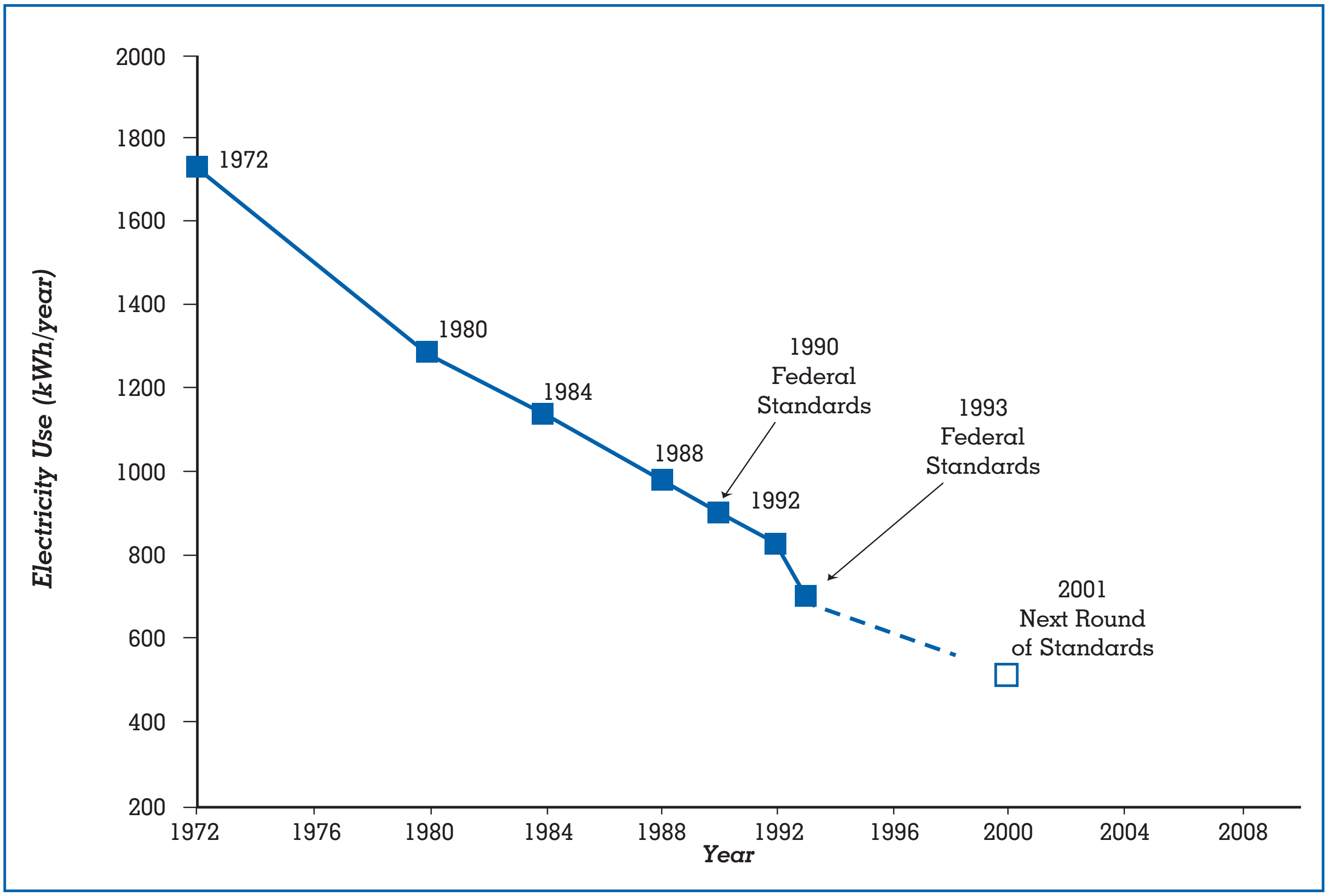


Figure 3.

Actual and projected refrigerator-freezer energy improvements from 1972 to 2001.

source: E. A. Vineyard, J. B. Horvay, and E. R. Schulak. "Experimental and Analytical Evaluation of a Ground-Coupled Refrigerator-Freezer." ASHRAE paper, p.2, presented in Toronto, Canada, on June 23, 1998.

"Montreal Protocol on Substances that Deplete the Ozone Layer." The phase-out of CFCs began on July 1, 1989, and by 1997, a hydrofluorcarbon, HFC134a, with zero ozone depletion potential, became the dominant refrigerant in the United States. The phase-out of CFCs in developing countries is on a slower schedule.

In the years after World War II many new customer-oriented features were introduced : plastic food liners, foam insulation, combination refrigerator-freezer, automatic defrosting, ice makers, and child-safe door closures. In the 1940s, the typical refrigerator was a single door cabinet, requiring periodic defrosting, with a storage volume of about 8 cu ft. By the 1990s, it developed into a 20 cu ft automatic defrosting two door combination refrigerator-freezer, with an ice maker.

ENERGY CONSIDERATIONS

The energy consumption of refrigerators and freezers is regulated by the Congress of the United States. The "National Appliance Energy Conservation Act of 1987" (NAECA) established an energy conservation standard "which prescribes a minimum level of energy efficiency or a maximum quantity of energy use". The Energy Standards were revised effective January 1, 1993, and again, effective July 2001. The first revision resulted in a cumulative 40 percent reduction in energy consumption when added to the initial standards. The result of the second revision in 2001 is an additional 30 percent reduction. A historical chart, Figure 3, shows actual and projected improvements in the use of electrical energy for refrigerator-freezers. In 1978 one manufacturer offered a 20 cu ft refrigerator-freezer that consumed 1,548 kWh per year. In 1997, the same manufacturer had a 22 cu ft refrigerator-freezer on the market that consumed 767 kWh per year. The 2001 target for the same product is 535 kWh per year. The latest mandatory energy reduction will add about $80 to the cost of the product, and result in an annual saving of $20 for the customer. Based on an annual production of 8.5 million refrigerators and freezers in the United States, this significant reduction of energy use will eliminate the need for building eight new power plants.

The energy consumption of a refrigerator is a function of two distinct elements: the heat that the evaporator needs to absorb to maintain the specified storage temperatures and the efficiency of the refrigeration system to reject that heat, and also the heat of compression, by the condenser. The heat load is determined by the storage volume of the appliance, the interior temperatures to be maintained, and the effectiveness of the insulation surrounding the storage space. The engineering challenge resides in reducing the heat load, and at the same time, improving the efficiency of the refrigeration system.

THE HEAT LOAD

Figure 4 shows the cabinet cross section of a typical automatic defrosting refrigerator. To defrost a refrigerator, forced convection heat transfer is needed between the evaporator and the load, thus the electric fan. The role of external heaters is to prevent the condensation of water vapor (sweating) on external cabinet surfaces. As the figure indicates, the heat flow through the walls of the cabinet represents 52 percent of the total heat load. Consequently, the thickness and thermal conductivity of the insulation has a major impact on the energy consumption of the appliance. Until the early 1960s the insulating material was glass fiber. Then the development of polyurethane insulation revolutionized the refrigeration industry. With an insulating value twice as good as glass fiber, wall thickness could be reduced by one-half, resulting in additional cubic feet of useful storage. In contrast to applying the glass fiber by hand, the assembly line for polyurethane foam is completely automated. The foam adds structural strength to the cabinet, allowing significant reduction in the thickness of the materials used on the inner and outer surfaces.

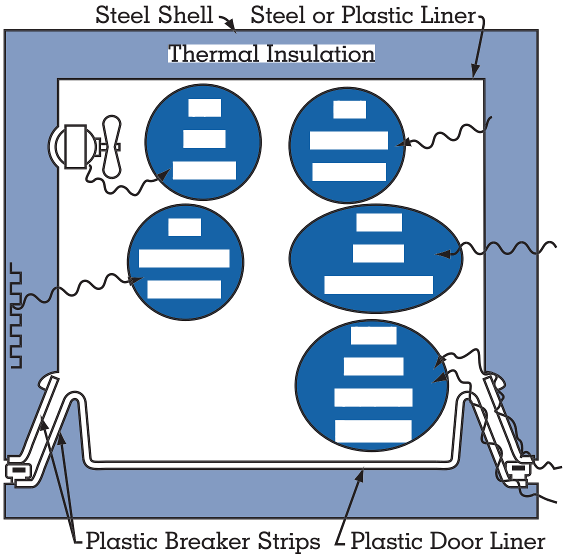


Figure 4.
Cabinet cross section showing typical contributions to the total basic heat load.
SOURCE: ASHRAE Handbook. (1998). *Refrigeration.* Atlanta, GA: ASHRAE.

In the 1980s, with the discovery of the ozone depletion potential of CFC refrigerants, the R11 blowing agent used in polyurethane foam was destined for phase-out in the United States by 1996, just like refrigerant R12 of the refrigeration system. A frantic search began for environmentally acceptable substitutes and Table 1 lists some of the more promising candidates. Note that in addition to ozone depletion potential (ODP), a relatively recent environmental concern, global warming potential (GWP) (sometimes referred to as the “greenhouse effect”) is also listed. The numbers are all relative to the properties of CFC 11, the worst offender on the list. In their search for an alternative to R11, the United States and Western Europe went their separate ways. The American choice was hydrochlorofluorocarbon (HCFC) refrigerant R141b. Europe, with an annual production of some 17 million refrigerators and freezers, and under pressure from environmentalists,

Blowing Agent	*ODP*[1]	*GWP*[2]
CFC 11	1	1
HCFC 22	0,05	0,36
HCFC 123	0,02	0,01
HFC 134a	0	0,27
HCFC 141b	0,11	0,1
HCFC 142b	0,06	0,36
Pentane	**0**	**0,001**
Air	0	0
CO_2	0	0,00025

[1] Ozone Depletion Potential (ODP)
[2] Global Warming Potential (GWP)

Table 1.
Ozone depletion potential (ODP) and global warming potential (GWP) of various foam blowing agents.
SOURCE: Machinenfabrik Hennecke GmbH, Sankt Augustin, Germany, Brochure #302. Modified by author.

went to pentane.. While both substances have zero ODP, it is the difference in GWPs that tilted the Europeans toward pentane. As a matter of fact, R141b is scheduled to be phased out in the United States by January 2003, due to its residual ODP. Pentane is a hydrocarbon, like gasoline, and highly flammable. Due to the mandated safety measures, the conversion from R11 to pentane at the manufacturing sites is very extensive and expensive. Both foams, whether blown with R141b or with cyclopentane, have a higher thermal conductivity (poorer insulating value) than R11 blown foam. Consequently, these new formulations, require considerable redesign. For the energy consumption to remain the same, either the cabinet wall thickness needs to be increased or the efficiency of the refrigeration system needs to be improved.

To meet the 2001 U.S. energy standards and the 2003 phase-out of HCFCs, there is a great incentive to develop a significantly better thermal insulation. The most dramatic approach would use vacuum panels for insulating the cabinet. A number of U.S. and Japanese manufacturers have developed such panels and placed these kinds of refrigerators in homes. The panels consist of multilayer plastic envelopes filled with precipitated (fumed) silica. The claimed thermal conductivity is one-fourth that of polyurethane foam. The two major obstacles are cost and the maintenance of vacuum for twenty years.

THE REFRIGERATION SYSTEM

To meet the 1993 Energy Standards, the industry undertook , at considerable cost, the optimization of the various refrigeration system components. The most significant improvement was the increase in compressor efficiency, from an EER of about 4 to about 5.5. Other system improvements included more efficient fan motors, more effective heat transfer by the evaporator and the condenser, and less defrost energy. In the early 1980s, both the Whirlpool Corporation and White Consolidate Industries introduced electronic defrost controls. Heretofore, an electric timer initiated the defrost cycle, typically every twelve hours, whether the evaporator needed it or not. With the electronic control the defrost interval is more a function of frost accumulation than of time, and thus referred to as a “variable defrost control” or as “adaptive defrost.”It saves energy by being activated only when needed.

Further improvements in system efficiency will be difficult to achieve with evolutionary changes. Following are some of the more promising areas of development:

Rotary Compressor. Inherently a rotary compressor is more efficient than the current reciprocating compressor. (room air-conditioners have been using rotary compressors for decades.) Several manufacturers in the U.S. and Japan have produced refrigerators with rotary compressors, but experienced long-term quality problems.

Dual Evaporators. A typical refrigerator-freezer, with automatic defrost, has a single evaporator for refrigerating both compartments, at a pressure that is determined by the temperature of the freezer. With separate evaporators in each compartment, the compressor would alternate between the two evaporators in providing refrigeration. The fresh food compartment has a significantly higher temperature than the freezer compartment, so it would requier less pressure to refrigerate, resulting in energy savings.

Variable Speed Compressor. Every time the compressor cycles on during the first few minutes of running, the compressor works on building up the pressure difference between the evaporator and the condenser, and is extremely inefficient. By letting the compressor run all the time, and modulating its speed according to the refrigeration needs, these periods of inefficiency can be eliminated.

Sharp Corporation promotes its new refrigerator, which includes the first built-in LCD. (Corbis Corporation)

Sonic Compression. Sonic compression is appealing as a potential low-cost, high efficiency oil-free technology. It should work with a wide range of refrigerants, including hydrocarbons, fluorocarbons, and ammonia. Theoretical COP is comparable with current vapor compression refrigeration cycles. The initial prototypes have a target cooling capacity of 110 to 250 W.

ABSORPTION REFRIGERATION.

Absorption refrigeration is based on the great affinity between certain liquids and gases. For example, 1 cu ft of water is capable of absorbing 800 cu ft of ammonia gas. One might look at this process as compression of a sort. Starting out with a small volume of absorbent (water) and a large volume of refrigerant gas (ammonia), the process ends up with a small volume of liquid solution of the refrigerant in the absorbent. In comparison to the vapor compression system (Figure 1), the compressor is replaced by two components, the absorber and the generator (Figure 5). The condenser and the evaporator function just as in the vapor compression cycle. Upon entering the generator, the aqua-ammonia solution is heated, and the ammonia in the solution vaporizes first, due to it's lower boiling point. The ammonia vapor then enters the condenser as the water drains back to the absorber. The condensed liquid ammonia is allowed to expand into the evaporator. As it vaporizes it absorbs heat from the load, and then it is reabsorbed by the water in the absorber. The cycle is ready to repeat itself.

Vapor compression uses the highest form of energy, namely electrical energy. In absorption refrigeration, the energy input is any source of heat (e.g., electrical energy, bottled gas, kerosene, or solar energy).

The first practical absorption refrigerator was developed in 1850 by Edmund Carre of France, who

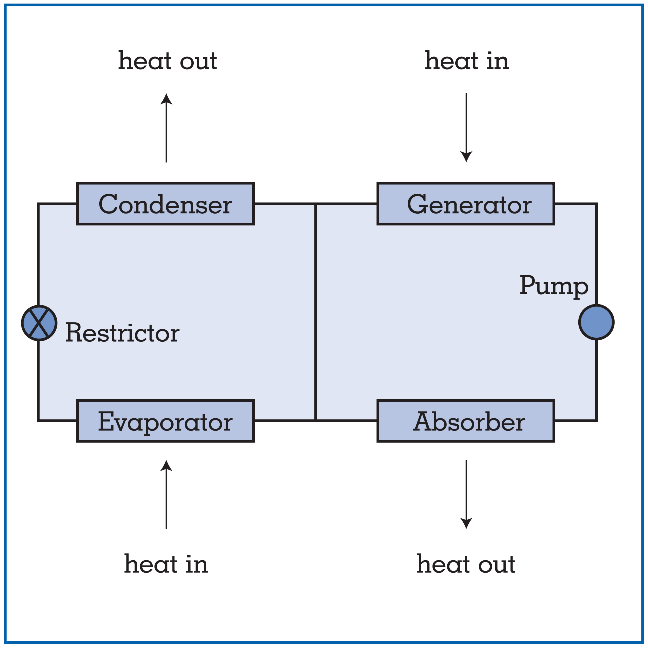


Figure 5.
Heat flow diagram of the absorption system.

was granted a U.S. patent in 1860. During the Civil War, the supply of ice from the North was cut off and Edmund's brother, Ferdinand Carre, shipped a 500 lb per day ice machine through the federal blockade to Augusta, Georgia, to be used in the convalescent hospital of the Confederate Army. Absorption refrigeration enjoyed it's heyday in the United States in the 1920s, with several manufacturers, like Servel and Norge, providing products. As larger refrigerators were demanded by the public, absorption refrigerators were unable to compete due to their inefficiency. The only remaining role for absorption refrigeration in the United States is in remote areas that are without electricity, and in boats and recreational vehicles that have bottled gas. There are many under-developed countries where absorption refrigeration is still the principal means of food preservation, due to lack of electricity.

THERMOELECTRIC REFRIGERATION

In 1821, Thomas Seebeck, an Estonian physician, discovered the existence of an electric current in a closed circuit consisting of unlike conductors, when the junctions between the conductors were at different temperatures. This discovery is the basis for thermocouples, used for all kinds of temperature measurements. In 1834, Jean Peltier, a French watchmaker, discovered the reverse of the "Seebeck Effect"; namely, if a direct current passes through a junction of two dissimilar metals in the appropriate direction, the junction will get cold. When thermocouples are put in series, one face of the assembly will get cold, the other hot.

The "Peltier Effect" was a laboratory curiosity until 1954, when Goldsmid and Douglass of England achieved a temperature difference of 47°F between the two junctions using a semiconductor, bismuth telluride, types negative and positive. It was predicted at that time that by the 1970s thermoelectric refrigeration would replace the vapor compression system in refrigerators and freezers. The 1966 Sears catalog offered a "Coldspot Thermo-electric Buffet Bar" that "chills foods and drinks to 40 degrees...." The buffet bar took advantage of an exciting feature of thermoelectric devices; namely, by reversing the electric current flow, the cold junctions will get hot, and the hot junctions cold. By the flick of a switch, the 2 cu ft compartment would change from a refrigerator to an oven. Due to its inefficiency, the thermoelectric refrigerator did not replace vapor compression.

The key to future advances in thermoelectric refrigeration is the discovery of a semiconductor more efficient than bismuth telluride. Until then, the only marketable applications are small portable refrigerators for cars and boats. They take advantage of the readily available direct current provided by the battery. Just plug it into the cigarette lighter and it begins to refrigerate.

J. Benjamin Horvay

See also: Air Conditioning; Appliances.

BIBLIOGRAPHY

American Home Appliance Manufacturers. (1988). *AHAM Standard for Household Refrigerators and Household Freezers.* Chicago: Author.

Appliance. (June 1992). "Tomorrow's Insulation Solution?" Oak Brook, IL: Dana Chase Publication.

ASHRAE Handbook (1981). *Fundamentals.* Atlanta, GA: American Society of Heating, Refrigerating and Air-Conditioning Engineers.

ASHRAE Handbook (1998). *Refrigeration.* Atlanta, GA: American Society of Heating, Refrigerating and Air-Conditioning Engineers.

ASHRAE Journal. (January 1999). "Washington Report—Montreal and Kyoto Protocols' Relevance to HFC's."

Bansal, P. K., and Kruger, R. (1995), "Test Standards for Household Refrigerators and Freezers I: Preliminary Comparison." *International Journal of Refrigeration* 18(1):4–20

Environmental Protection Agency. (1993). *Multiple Pathways to Super-Efficient Refrigerators*. EPA-430-R-93-008. Washington, DC: Author.

Holladay, W. L. (1994), "The General Electric Monitor Top Refrigerator." *ASHRAE Journal* September. 49–55.

National Appliance Energy Conservation Act. (1987). *Public Law 100-12*. March 17 (as amended by National Conservation Policy Act, Public Law 95-619), Part B—Energy conservation for consumer products other than automobiles. Washington, DC: U.S. Department of Energy.

United Nations Environmental Program. (1995). *Montreal Protocol on Substances that Deplete the Ozone Layer—Report on the Refrigeration, Air Conditioning and Heat Pumps Technical Options Committee*. New York: Author.

Stoecker, W. F., and Jones, J. W. (1982). *Refrigeration and Air Conditioning*. New York: McGraw-Hill Book Company.

Woolrich, W. R. (1969). "The History of Refrigeration: 220 Years of Mechanical and Chemical Cold: 1748–1968." *ASHRAE Journal* July:31–39.

REGULATION AND RATES FOR ELECTRICITY

Electric utilities have historically been franchise monopolies, vertically integrated from power production through transmission, distribution, and customer service with no competition from other electric utilities. However, in many parts of the country, electric and gas utilities do compete. Rates charged by these utilities were determined in a regulatory proceeding: Electric utilities proposed rates that compensated them for their expenses and allowed them to earn a reasonable return; state regulatory commissions reviewed and approved the proposals.

This historical relationship between electric utilities and their regulators is undergoing a dramatic change. Policymakers are restructuring and deregulating portions of the electric industry. Restructuring of the electric industry is consistent with the deregulation of other U.S. industries since the 1980s. The objective in "restructuring " is to increase efficiency, lower costs, increase customer choices, and lower the prices paid by consumers in restructured industries.

As in other industries, restructuring of the electric industry is a response to underlying conditions in the industry. Policymakers are responding to two phenomena. First, there is a disparity in retail electric prices among states and regions. For example, the average price of electricity in New England states such as New Hampshire (nearly 12 c/kWh) is almost three times as large as the price in low-cost states such as Idaho and Washington (approximately 4c/kWh). Second, advances in information technology will make it possible to perform complex real-time functions.

In restructured electric markets, the vertical electric monopoly will no longer be the sole provider of electricity. The generation, transmission, distribution, and customer service functions will be separated. The upstream generation function will be competitive, allowing new, any power producer to produce and sell electricityin any service territory. The transmission and distribution functions will continue to be regulated, but will be required to allow access to power suppliers and marketers. This separation or "unbundling" of the industry is necessary to provide nondiscriminatory access for all suppliers of electricity. Customers will have their choice of electric suppliers.

In these restructured electric markets, prices will be determined more by market forces and less by regulatory proceedings. To some, introducing competition will promote more efficient markets, providing the proper financial incentives for firms to enter or leave the industry. In this way, consumers will benefit from lower production costs and, hence, reduced electricity prices. To others, restructuring will increase electricity prices for some customers, sacrifice the current environmental and social benefits, and jeopardize system reliability of the status quo.

ADVANTAGES OF RESTRUCTURING

Consider a typical rate-making proceeding for a regulated utility. Electric utilities can recover all prudently incurred operating and maintenance costs plus an opportunity to earn a fair return on their investment. This process involves three steps: (1) determining the total amount of revenues (i.e., required revenues) that an electric utility needs; (2) allocating the total to individual customer groups (e.g., residential, commercial, and industrial customers); and (3) designing a rate structure for each customer group that allows the utility to recoup costs.

Required revenues are the sum of operation and maintenance expenses, depreciation, taxes, and a return on rate base. The rate base is the total amount of fixed capital used by the utility in producing, transmitting, and delivering electricity. The return on rate base is the weighted average cost of capital, including debt and equity sources.

Allocating required revenues to customer groups involves four steps: (1) categorizing customers into groups with similar characteristics (e.g., low-voltage customers); (2) functionalizing costs into those pertaining to production, transmission, distribution, and administration; (3) classifying functionally assigned costs into those attributable to the customer (e.g., metering costs), energy (e.g., amount of consumption), and capacity (e.g., instantaneous demand); and (4) allocating costs to customer groups. The result of this process is the allocation of all of a utility's revenue needs into customer groups. The number of groups depends on the characteristics of a utility's service territory.

Rates are then designed to recoup the revenues for each of the customer groups. Rates can be based on the amount of consumption or the type of service. Consumption-based rates are either flat, or increasing or decreasing in steps. Service-based rates depend on the type of service a utility offers its customer classes: firm rates, lifeline rates, interruptible rates, stanby rates, various incentive rates, or time-of-use rates.

The problems with this cost-of-service approach are the incentives and opportunities given to electric utilities. They are not the type of incentives that characterize an efficient market and that balance the additional risk of operating in an efficient market.

Industry restructuring advocates believe that a competitive market will create incentives to operate more efficiently and make better economic investments in utility plant and equipment. In contrast to a competitive market, cost-based ratemaking does not generally reward utilities with a higher return for making an especially good investment in plant and equipment, or penalize them for making an especially bad investment. The return that a utility earns on a highly successful investment is generally the same as its return on a less successful one. The opportunity to prosper—or, alternatively, to go bankrupt—is generally not part of this regulatory process.

In many cases, the regulators themselves distort the ratemaking process. Historically they have allowed or even encouraged utility investments that would not otherwise have been made in competitive markets.

Makers of public policy in the U.S. Department of Energy (DOE) and the Federal Energy Regulatory Commission (FERC) believe that industry restructuring will change these incentives and introduce more efficient practices and technologies in the industry. Proper pricing of electricity is one such change. Time-of-use (TOU) rates is one example. TOU rates vary over the course of a year: hour by hour, day by day, or season by season. They are theoretically appealing: Consumers who cause daily peaks bear the burden of paying higher costs during those periods. Likewise, consumers contributing to seasonal peaks—such as using air conditioners—bear the cost of building the capacity needed to meet the peaks. In practice, TOU pricing has proved effective in shaping load for electric utilities, reducing peak demand, and lowering production costs.

When initially adopted, TOU rates were based on projections of future costs by season, month, day, or hour. However, advances in metering and communications technology now afford utilities the ability to transmit prices to customers based on actual operating costs and to read meters in real time. This "real-time TOU pricing" is one of the most important aspects of many of the restructuring efforts to date. They can provide customers with direct access to the prices arising in competitive electric markets.

Under restructuring, customer exposure to market-based electric rates will broaden in other ways. Because deregulation allows customers to choose their own electric suppliers, potential suppliers are becoming more innovative in attracting customers. Although the Internet is in its infancy now, entrepreneurs are in the process of harnessing it for electricity sales. Some Internet companies are purchasing wholesale power from generating plants and reselling it to customers on the Net. Other companies are aggregators. They enlist customers on the Net and create buying pools from which they can extract lower prices from suppliers.

Restructuring will promote the introduction of other advanced technologies and practices as well. For example, the use of combined-cycle, gas turbine power plants are expected to proliferate under restructuring. These plants are generally more efficient and more environmentally benign than many fossil fuel plants currently in use.

DRAWBACKS OF RESTRUCTURING

Competition can promote efficiency and lower average prices for electricity. However, there is no guarantee that all customers will benefit equally from lower prices. Larger commercial and industrial customers generally have the wherewithal to obtain better rates in competitive markets than do smaller residential customers. Because of fewer options, low-income households are especially vulnerable to competitive markets for electricity.

Environmental and social benefits could also be jeopardized under restructuring. The elimination of integrated resource planning (IRP) is particularly of concern. In an IRP process, energy-efficiency programs and renewable-energy technologies "compete" with conventional generating plants for the resource investment expenditures of electric and gas utilities. The competition takes into account the environmental consequences of producing electricity from fossil fuels. At IRP's peak in the early 1990s, more than thirty-three states mandated the use of IRP processes. These mandates came from state legislation or from state regulatory commission orders.

As the electric industry undergoes reorganization, the retail price of electricity will be determined more and more in markets with the participation of multiple parties who do not themselves generate electricity. The "wires" portion of utilities that historically ran energy-efficiency programs argue that they are unfairly burdened by running these programs if their competitors are not also obligated to do so. Therefore, IRP processes are jeopardized in restructured electric markets. Without IRP, the environmental consequences of relying more on fossil fuels—and less on energy efficiency and renewables—are obvious.

Finally, a number of industry engineering experts and industry engineering organizations voiced concern that the electric grid may become less reliable after restructuring. The operation and maintenance of the North American electric grid depends on the coordinated interaction of more than a hundred control areas. The incentives to buy and sell power, retain adequate surplus capacity, and maintain the grid will change in a restructured electric industry. The reliability of the electric grid may suffer as a consequence.

Lawrence J. Hill

See also: Capital Investment Decisions; Eco-nomically Efficient Energy Choices; Economic Growth and Energy Consumption; Electric Power, System Protection, Control, and Monitoring of; Energy Management Control Systems; Government Intervention in Energy Markets; Subsidies and Energy Costs; Supply and Demand and Energy Prices; Utility Planning.

BIBLIOGRAPHY

Fox-Penner, P. (1997). *Electric Utility Restructuring: A Guide to the Competitive Era.* Vienna, VA: Public Utilities Reports.

Phillips, C. F., Jr. (1993). *The Regulation of Public Utilities.* Arlington, VA: Public Utilities Reports.

RENEWABLE ENERGY

Renewable energy is energy that can be replenished on a time scale appropriate to human use. Solar energy for growing plants for food is renewable because light flows continuously from the sun, and plants can be reproduced on a time scale suitable for human needs. Coal is produced continually in some geologic formations, but the time scale is on the order of hundreds of thousands of years. Accordingly, coal is considered a nonrenewable energy.

The Department of Energy categorizes sources of nonrenewable energy production as fossil fuels, nuclear electric, and pumped hydroelectric. Sources of renewable energy are delineated as conventional hydroelectric, geothermal, biofuel, solar, and wind. Conventional hydroelectricity and energy from biofuels are considered mature sources of renewable energy; the remaining types are thought of as emerging. The chart in Figure 1 depicts energy production for 1997.

More electric energy is invested in pumping water for pumped storage hydroelectricity than is produced by the electric generators. Therefore, the net production of electricity from pumped hydro sources is slightly negative and does not appear on the chart in

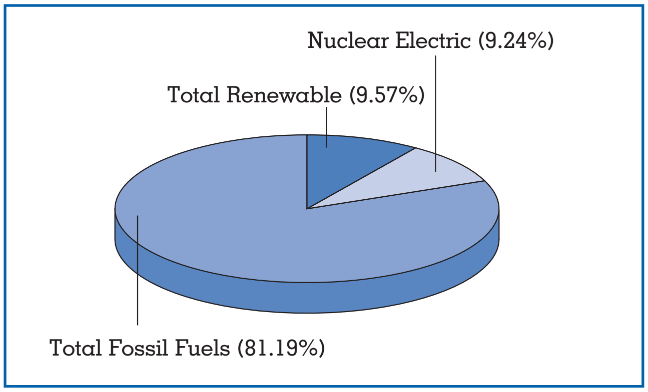


Figure 1.
Percentage of total energy production for 1997.
SOURCE: U.S. Energy Information Agency

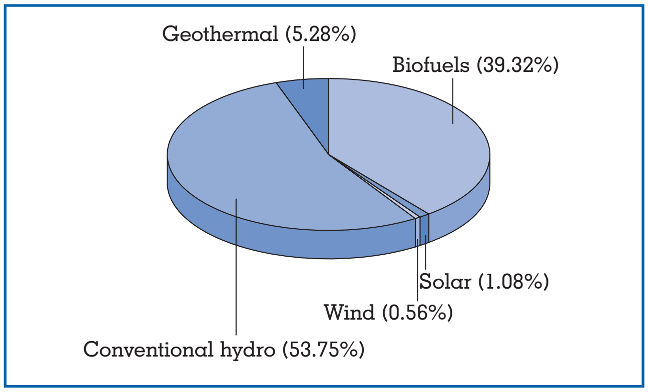


Figure 2.
Breakdown of energy production from renewable sources.
SOURCE: U.S. Energy Information Agency

Figure 1. A breakdown of the 9.57 percent contribution from renewable sources is shown in Figure 2. The great bulk of the energy produced from renewable sources is from biofuels and conventional hydroelectric systems, both of which are considered mature.

Nonrenewable energy such as coal, petroleum, and natural gas are vast but not inexhaustible. A nonrenewable resource can be preserved if it can be replaced with a renewable one. This is the prime attraction of renewable energy. Sometimes a renewable resource offers an environmental advantage. If a house heated by burning fuel oil could instead use a solar heating system, then products from burning that might contribute to environmental problems would not be released into the atmosphere. On the other hand, damming a river to make a reservoir for a hydroelectric power plant might inflict damage to the environs by destroying plant and animal habitat.

Most use of renewable energy involves the sun. Solar energy is converted to thermal energy for space heating, heating hot water for domestic use and, to some extent, for generating electricity. A photovoltaic cell, which functions like a battery, converts solar energy directly to electric energy. Biomass energy has its origin in plants grown with the help of solar radiation. Wind energy is due to unequal heating of Earth's surface by the sun. A hydroelectric plant converts the gravitational energy of water into electricity, but the mechanism that replenishes the water is powered by the sun. Roughly, 95 percent of all renewable energy is solar in origin. The remainder is from geothermal energy. Tidal energy, having its origin mainly in the gravitational force between Earth and Earth's moon, is used in some parts of the world (France, for example), but is yet to be exploited in the United States.

Figure 2 shows that biomass and hydroelectricity account for about 93 percent of renewable energy. Biomass includes wood, ethanol, and municipal solid waste. Wood pellets, manufactured from finely ground wood fiber, represent a growing market for biomass fuels for specially designed stoves and furnaces. Paper products, a biomass fuel, are the major component of municipal solid wastes. Disposal of these wastes is a major societal problem. Burning the wastes for their energy content was a popular option during the 1980s because of federal, state, and local policies promoting the construction of waste-to-energy facilities. Since then, environmental policies encouraging recycling and requiring costly pollution control have made the economics of operating a waste-to-energy facility far less favorable. Accordingly, the use of waste-to-energy facilities has declined. The technology for converting corn to ethanol is very well developed, but the degree to which ethanol becomes a substitute for gasoline will depend strongly on economics.

Solar and wind energy account for less than 2 percent of the use of renewable energy (see Figure 2). Nevertheless, there are some important developments. The greatest increase in wind energy is outside the United States; however, in 1997 the United States led the world for new wind generating systems by adding 1,620 megawatts of capacity. A power output of 1,620 megawatts is the equivalent of the power output of two large electric power plants. However, wind systems do not operate as continuously as large electric power plants, so in the long run they would not deliver nearly as much energy as two large electric power plants. Most wind energy projects are in California, but significant projects are in Texas and Minnesota.

When assessing ways to take advantage of solar energy, one should not necessarily think of sophisticated technological schemes such as roof-top collectors, photovoltaic cells, and wind-powered electric generators. Designing a house or a building to maximize the input of solar radiation during the heating season can produce significant savings in energy. If in the southern United States all the worn-out, dark-colored roofs were replaced with white ones, ten times more energy would be saved in air conditioning than is produced by all wind generators in the United States (Rosenfeld 1997).

The major facility for generating electricity from geothermal sources is at The Geysers in Northern California. Generation at The Geysers is declining both for economic reasons and because of reduced steam pressure. However, other facilities continue to produce steady quantities of electricity.

Joseph Priest

See also: Solar Energy; Hydroelectric Energy; Turbines, Wind.

BIBLIOGRAPHY

Energy Efficiency and Renewable Energy Network (EREN) Home Page. Energy Efficiency and Renewable Energy Network. August 1, 2000 <http://www.eren.doe.gov/>.

National Renewable Energy Laboratory (NREL) Home Page. National Renewable Energy Laboratory. August 1, 2000 <http://www.nrel.gov/>.

Oak Ridge National Laboratory Energy Efficiency and Renewable Energy Program Web Site. Oak Ridge National Laboratory. August 1, 2000 <http://www.ornl.gov/ORNL/Energy_Eff/Energy_Eff.html>.

Renewable Energy Technologies. Sandia National Laboratories, Renewable Energy Technologies Division. August 1, 2000 <http://www.sandia.gov/Renewable_Energy/renewable.html>.

Rosenfeld, A. H. et al. (1997). "Painting the Town White—And Green." *Technology Review* 100:52–59.

RESERVES AND RESOURCES

We need to know about the quantity and quality of oil and gas reserves because the prosperity of the world is dependent upon petroleum-based fuels. The debate about forthcoming oil shortages—as soon as 2004 or perhaps later in the twenty-first century—hinges on our understanding of petroleum reserves and future resources. Global catastrophic changes are predicted by some if a shortage occurs early in the twenty-first century; however, others are less concerned, because of new estimates of reserves and potential petroleum resources.

Our concepts of petroleum reserves and resources and their measurements are changing to reflect the uncertainty associated with these terms. Petroleum reserves have been largely calculated deterministically (i.e. single point estimates with the assumption of certainty). In the past decade, reserve and resource calculations have incorporated uncertainty into their estimates using probabilistic methodologies. One of the questions now being addressed are such as "how certain are you that the reserves you estimate are the actual reserves and what is the range of uncertainty associated with that estimate?" New techniques are required to address the critical question of how much petroleum we have and under what conditions it can be developed.

The goal of most industry and financial groups is to forecast future production rates, cash flow, net present values (NPV) and other measures designed to prudently manage financial and production aspects of the petroleum reserve. The determination of the appropriate category to which a reserve is assigned is a critically important decision. These groups emphasize the least risky category of reserves—proved (measured) reserves.—because industry and financial groups can determine the present value of a reserve from existing infrastructure, whereas other resources require additional investment and increased uncertainty regarding the ultimate asset value.

Revisions of original reserve estimates, that generally increase the earlier estimates, have caused several groups to undertake studies of these changes. Many such studies focus on the original underestimation of reserves and the supporting analysis of reserve growth. However, there are instances of original overestimation of reserves. U.S. Geological Survey (USGS) studies on reserve growth of the world's oil fields demonstrate the significant potential of reserve growth, a potential that may be greater than that of undiscovered resources. Close estimation of the true petroleum reserves is critically important to both producers and consumers of hydrocarbons, and opinions about estimates are numerous, conflicting, and often contentious. To understand the current flux of reserves and resources, definitions are provided followed by a discussion. Definitions given are those used in western literature. Countries of the former Soviet Union (FSU) use different reserve and resource definitions. Our analysis of databases in the Volga-Ural and West Siberian Basins shows that when FSU data are converted to our western definitions, similar uncertainties exist.

DEFINITIONS

There are many definitions of reserves and resources, and we need some historical perspective to understand the current debates.

U.S. Geological Survey/U.S. Bureau of Mines

In 1976 the USGS/U.S. Bureau of Mines (USBM) defined resources and reserves (Figure 1) as follows.

Resources: A concentration of naturally occurring solid, liquid, or gaseous materials in or on the Earth's crust in such form that economic extraction of a commodity is currently or potentially feasible.

Reserves: That portion of the identified resource from which a usable mineral and energy commodity can be economically and legally extracted at the time of determination. According to the American Geological Institute (AGI), by using this resource/reserve scheme, we can update the definition: an estimate within specified accuracy limits of the valuable metal or mineral content of known deposits that may be produced under current economic conditions and with present technology. The USGS and USBM do not distinguish between extractable and recoverable reserves and include only recoverable materials.

Securities and Exchange Commission

The critical linkage between financial and accounting considerations and reserve definitions requires an understanding of the definitions given in Securities and Exchange Commission Regulation (SEC) Regulation 4-10. The SEC was empowered to develop energy accounting standards for the United States by the Energy Policy and Conservation Act of 1975 (EPCA), which was passed largely in response to the 1973 oil embargo. This act required the establishment of a national energy database including financial information. The SEC's definitions may be inherently conservative or restrictive thereby causing initial conservative estimates of reserves. The SEC definitions largely focus on proved oil and gas reserves, but are divided into two subcategories: (1) proved developed oil and gas reserves and (2) proved undeveloped reserves.

From SEC Regulation 4-10, Section 210.4-10 and as used by the industry, proved oil and gas reserves are the estimated quantities of crude oil, natural gas, and natural gas liquids that geological and engineering data demonstrate with reasonable certainty to be recoverable in future years from known reservoirs under existing economic and operating conditions (i.e., prices and costs as of the date the estimate is made). It goes on to define proved reserves if economic producibility is supported by either actual production or conclusive formation test and include that portion delineated by drilling and defined by gas-oil and/or oil-water contacts, and immediately adjoining portions not yet drilled, but which can be reasonably judged as economically productive on the basis of available geological and engineering data. In the absence of information on fluid contacts, the lowest known structural occurrence of hydrocarbons controls the lower proved limit of the reservoir. Proved reserves do not include: oil that may become available from known reservoirs but is classified separately as "indicated additional reserves"; crude oil, natural gas, and natural gas liquids, the recovery of which is subject to reasonable doubt because of uncertainty as to geology, reservoir characteristics or economic factors; crude oil, natural gas, and natural gas liquids that may occur in undrilled prospects; and crude oil, natural gas, and natural gas liquids, that may be recovered from oil shales, coal, gilsonite, and other such sources.

Proved developed oil and gas reserves are reserves that can be expected to be recovered through existing wells with existing equipment and operating methods. Additional oil and gas expected to be obtained through the application of fluid injection or other improved recovery techniques for supplementing the natural forces and mechanisms of primary recovery

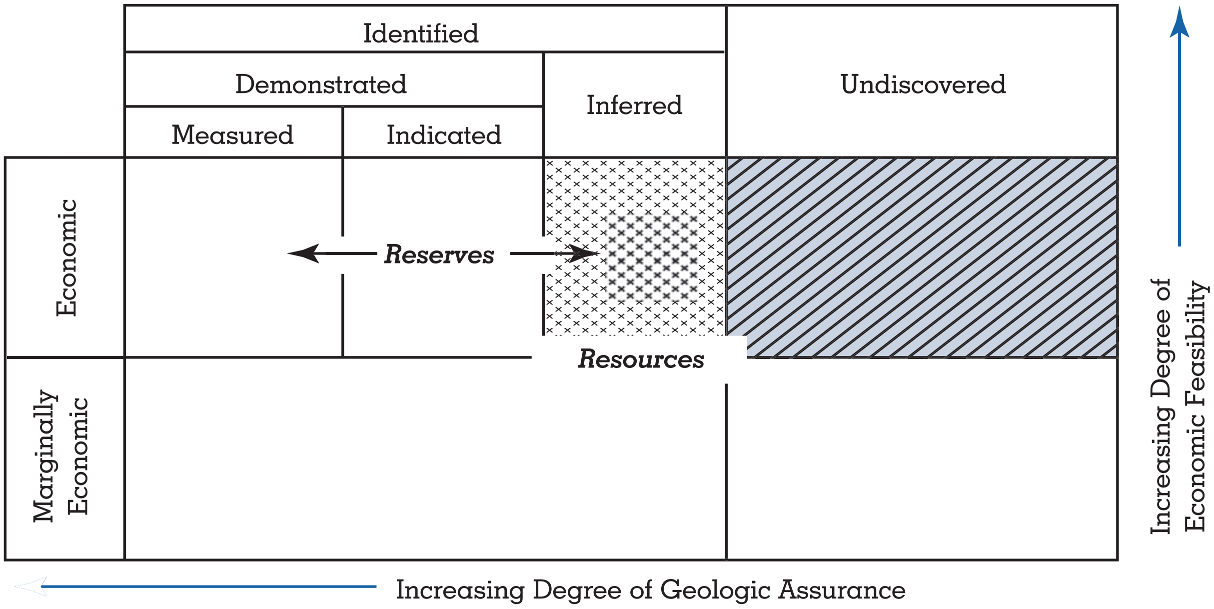


Figure 1.
Reserves vs. Resources Chart.
SOURCE: USGS/U.S. Bureau of Mines, 1976

should be included as "proved developed reserves" only after testing by a pilot project or after the operation of an installed program has confirmed through production response that increased recovery will be achieved.

Proved undeveloped oil and gas reserves are reserves that are expected to be recovered from new wells on undrilled acreage, or from existing wells where a relatively major expenditure is required for recompletion. Reserves on undrilled acreage shall be limited to those drilling units offsetting productive units that are reasonably certain of production when drilled. Proved reserves for other undrilled units can be claimed only where it can be demonstrated with certainty that there is continuity of production from the existing productive formation. Under no circumstances should estimates for proved undeveloped reserves be attributable to any acreage for which an application of fluid injection or other improved recovery technique is contemplated, unless such techniques have been proved effective by actual tests in the area and in the same reservoir.

The World Petroleum Congress/Society of Petroleum Engineers

In 1997, the World Petroleum Congress (WPC) and the Society of Petroleum Engineers (SPE) jointly published petroleum reserve definitions that added the element of probability to the deterministic definitions in common use. The WPC/SPE definitions build on the SEC definitions by including probabilistic estimates.

Reserves are those quantities of petroleum which are anticipated to be commercially recovered from known accumulations from a given date forward.

Proved reserves are those quantities of petroleum which, by analysis of geological and engineering data, can be estimated with reasonable certainty to be commercially recoverable, from a given date forward, from known reservoirs and under current economic conditions, operating methods, and government regulations. Proved reserves can be categorized as developed or undeveloped. If deterministic methods are used, the term reasonable certainty is intended to express a high degree of confidence that the quantities will be recovered. If probabilistic methods are used, there should be at least a 90 percent probability that the quantities actually recovered will equal or exceed the estimate.

Unproved reserves are based on geologic and/or engineering data similar to that used in estimates of proved reserves; but technical, contractual, economic, or regulatory uncertainties preclude such reserves being classified as proved. Unproved reserves may be further divided into two subcategories: probable reserves and possible reserves.

Probable reserves are those unproved reserves which analysis of geological and engineering data suggest are more likely than not to be recoverable. In this context, when probabilistic methods are used, there should be at least a 50 percent probability that the quantities actually recovered will equal or exceed the sum of estimated proved plus probable reserves.

Possible reserves are those unproved reserves which analysis of geological and engineering data suggests are less likely to be recoverable than probable reserves. In this context, when probabilistic methods are used, there should be at least a 10 percent probability that the quantities actually recovered will equal or exceed the sum of estimated proved plus probable plus possible reserves.

The reserves identified in the upper left part of Figure 1, measured reserves, have greater economic feasibility and greater geologic certainty than indicated or inferred reserves. Undiscovered resources have even greater uncertainty than the previous categories. Many industry groups use the terms proved, probable, possible and speculative that roughly correlate to USGS categories of measured, indicated, inferred, and undiscovered. Proved, probable and possible reserves are commonly called P1, P2, and P3 respectively (Figure 2) and considerable analysis has been applied to relationships among these reserve categories to estimate ultimate recoverable resources.

Technological advances, particularly 3-D seismic techniques and deepwater drilling technology, have revolutionized our ability to discover and develop petroleum reserves. Technology has accelerated the rate of discovery of new global petroleum resources throughout the 1990s. Deepwater offshore drilling technology (up to 3,000 m water depths) permit exploration of many areas that were previously inaccessible, such as offshore west Africa and offshore South America. This geographic expansion of petroleum resource development into offshore areas provides opportunities for many non-OPEC nations. The discovery of significant oil resources in the Western Hemisphere, particularly in deepwater areas of the south Atlantic Ocean and Gulf of Mexico, are important to the United States, which seeks oil from more widely distributed potential supply sources. Any regional optimism springing from this discovery is mitigated by the realization that future contributions from field growth, which according to the USGS are nearly as large as undiscovered petroleum resource potential, will come from the areas where fields are already discovered.

Natural gas is an underutilized resource worldwide relative to oil but will play an increased role in the twenty-first century. New technologies such as gas-to-liquids and large-scale liquefied natural gas projects will also dramatically alter the fossil fuel landscape.

DISCUSSION

The definitions above are an abbreviated version of those used in a very complex and financially significant exercise with the ultimate goal of estimating reserves and generating production forecasts in the petroleum industry. Deterministic estimates are derived largely from pore volume calculations to determine volumes of either oil or gas in-place (OIP, GIP). This volume when multiplied by a recovery factor gives a recoverable quantity of oil or natural gas liquids—commonly oil in standard barrels or natural gas in standard cubic feet at surface conditions. Many prefer to use barrels of oil equivalency (BOE) or total hydrocarbons for the sum of natural gas, natural gas liquids (NGL), and oil. For comparison purposes 6,000 cubic feet of gas is considered to be equivalent to one standard barrel on a British thermal unit (Btu) basis (42 U.S. gallons).

Accommodation of risk (or uncertainty) in reserve estimations is incorporated in economic decisions in several ways. One common method

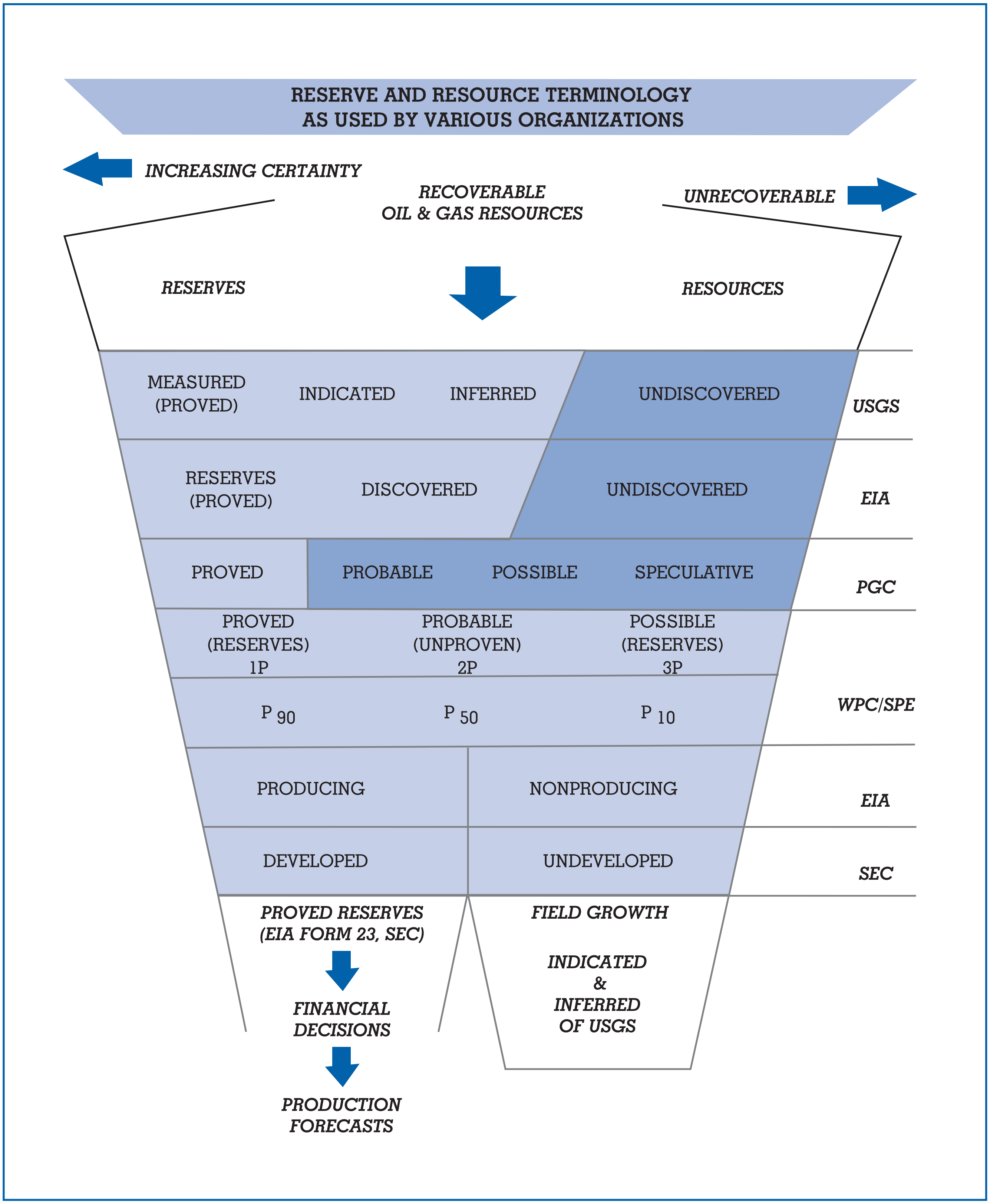


Figure 2.
Reserve and Resource Terminology.

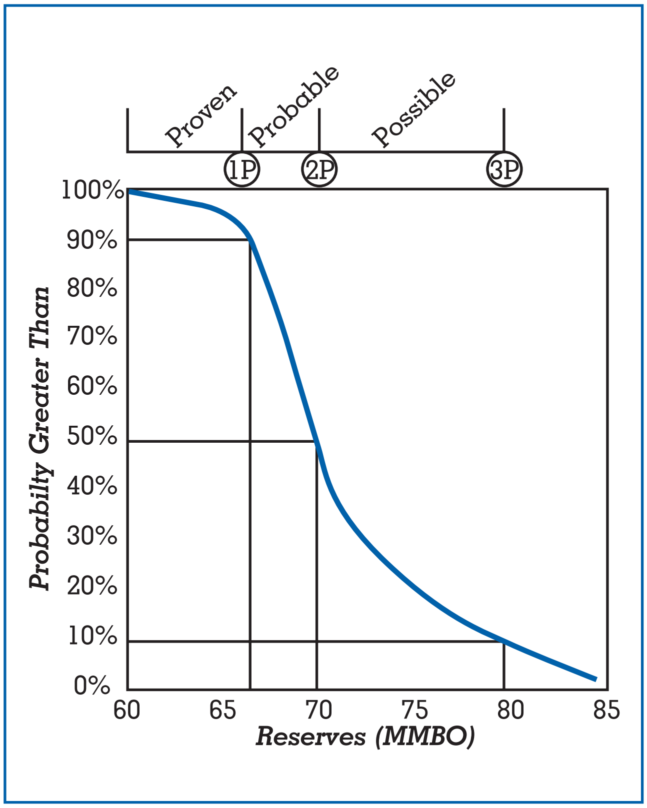


Figure 3.
Proven, Probable, and Possible Reserves and Associated Probabilities

As defined by the World Petroleum Congress/Society of Petroleum Engineers (modified from Grace, J. D. et al. (1993). "Comparative Reserves Definitions: U.S.A., Europe, and the Former Soviet Union." *Journal of Petroleum Technology* 45(9):866-872.

incorporates the expected value (EV) which is the probability-weighted average of all possible outcome values. This process is called "decision analysis"; however, the estimation or production forecast process has become so important that even more sophisticated analysis is commonly undertaken. In regions such as the North Sea, operating costs are high and critical infrastructure decisions are needed early in order to maximize the value of the reserve. Oil price, transportation costs, capital expense, operating expenditures, production rates, and closely estimated reserve volumes are the principal parameters used in developing an evaluation strategy. Reserve estimates can change quite dramatically. The United States is by far the most intensely developed petroleum region in the world with more than 3.5 million wells drilled and a long history of reserve estimation that shows substantial increases in ultimate recoverable reserve estimates. However, negative revisions also occur; either overestimation or underestimation of reserves can be costly. Petroleum reserve estimates are one of the parameters used in determining the EV of a field's reserves. In order to accommodate uncertainty in the various reserve categories, the reserve estimates are probability-weighted and used to determine the EV as shown in the following example and in Figure 3. Although a company may believe they have 3,525,000 barrels of recoverable oil, their probability-weighted reserve is only 31 percent of that figure or 1,093,750 barrels of recoverable oil. This procedure, repeated over and over in many fields throughout the world, has led to underestimation of ultimate recoverable reserve estimates. The subsequent upward revision of these initial conservative estimates is known as reserve growth or field growth. In the United States field growth accounts for twice the volume of future potential reserves of oil (i.e., 60 billion barrels) as undiscovered oil resources (i.e., 30 billion barrels) according to USGS 1995 estimates. In the world as a whole, field growth, based on analysis of the PetroConsultants datafile for a sixteen year period (1981–1996), accounts for nearly 500 billion barrels of oil for just those fields discovered since 1981.

Although, initially, deterministic or point value estimates of reserves were largely believed to be adequate, it is now clear that many factors complicate reserve estimation for recoverable oil or gas. Some complicating geologic and reservoir factors include: variations in reservoir architecture and quality (low complexity or high complexity), microscopic displacement efficiency, volumetric sweep efficiency (the effectiveness of hydrocarbon recovery from secondary and tertiary recovery processes), data quality and analysis. In the North Sea study, these variables are quantified and scored to improve field reserve estimates. In another study of North Sea fields, 65 percent of the responding countries used deterministic ultimate recovery estimates, whereas 53 percent used probablistic estimation methods. Inaccuracies of ultimate recovery and production forecasts were very expensive to those North Sea operators who either underestimated or overestimated facilities requirements. The conclusion that more unified and refined reserve definitions are needed is almost uni-

Reserve Category	*Total Company (boe)*	*Probability*	*Probability Weighted Reserve (boe)*
Proved producing	600,000	1.00	600,000
Proved developed non-producing (waiting on pipeline connection)	50,000	0.90	45,000
Proved behind pipe	125,000	0.75	93,750
Proved undeveloped (within 1 kilometer of producing well)			
Basin A	200,000	0.50	100,000
Basin B	50,000	0.40	20,000
Probable	500,000	0.35	175,000
Possible (2,000 hectares of exploratory leases)	2,000,000	0.03	60,000
	3,525,000		**1,093,750**

Table 1.
Probability Weighted Reserve Estimates
Modified from Seba, R. D. (1998). *Economics of Worldwide Petroleum Production*. Tulsa, OK: Oil and Gas Consultants International Inc.

versally supported. Appraisals at reservoir level are on a smaller scale than at field level. There are three basic ways to describe uncertainty in relation to reservoir and estimation models: fuzziness, incompleteness, and randomness. Fluctuations in the price of oil or natural gas affect the recoverability of the resource base and as prices increase, reserves increase as they become more economically viable. Advances in exploration and production technology also make recoverability more viable and enhance the amount of reserves that can be extracted from any given reservoir.

IMPLICATIONS

It is well known that reserve estimates made early in the development of a field are often wrong and that there's a 50 percent change in estimated ultimate recovery in many fields during the first ten years. In addition, the average field lifetime has been a decade longer than initially expected in the North Sea. If you believe in reserve growth, you would conclude that there will not be a petroleum crisis anytime soon. If, on the other hand, you believe that reserves have been overstated and you make negative revisions to reserve estimates, there will be a crisis soon. Current known petroleum reserves (proved + P50 probable) in the world are 1.44 trillion barrels of oil, 5,845 trillion cubic feet of gas, and 80 billion barrels of natural gas liquids based on 1996 PetroConsultants data. Depending upon the volume of undiscovered and field growth reserves added to these known reserves, you can develop either pessimisitic or optimistic scenarios.

In 1997, net U.S. petroleum imports of 8.9 million barrels of oil per day were worth $67 billion and exceeded U.S. petroleum production of 8.3 million barrels of oil per day. Concerns about oil shortages caused the United States to build a strategic petroleum reserve in 1977 that currently holds 563 million barrels of oil or about two months of net imported petroleum. Although domestic oil and gas production declined from 1970 to 1993, natural gas production has increased since the mid-1980s and energy equivalent production from natural gas exceeded domestic oil production in the late 1980s. Estimates in 1999 of per well oil production for the United States have fallen to 11.4 barrels a day, the lowest value in the last forty-five years. Considering this decline, the world's reserves and resources are critical to our future because the United States is the largest consumer of petroleum resources (~22% of world consumption). Increasingly, the United States relies on foreign oil supplies. Determining how much oil the world has and where it is located will continue to focus our attention on reserve estimation.

Thomas S. Ahlbrandt

See also: Energy Economics; Oil and Gas, Drilling for; Oil and Gas, Exploration for; Oil and Gas, Production of; Supply and Demand and Energy Prices.

BIBLIOGRAPHY

Ahlbrandt, T. S. (1999). "USGS New Millennium World Oil and Gas Assessment." *Energy Mix of the Future Session, Abstracts with Program*, vol. 31, no.7 (October). Denver, CO: Geological Society of America Annual Meeting.

American Geological Institute. (1997). *Glossary of Geology*, 4th ed., edited by J. A. Jackson. Alexandria, VA: Author.

Attanasi, E. D. and Root, D. H. (1994). "The Enigma of Oil and Gas Field Growth." *American Association of Petroleum Geologists* 78:321–332.

Brock, H. R.; Jennings, D. R.; and Feiten, J. B. (1996). *Petroleum Accounting: Principles, Procedures and Issues*, 4th ed. Denton, TX: Price Waterhouse Coopers, Professional Development Institute.

Campbell, C. J. (1997). *The Coming Oil Crisis*. Brentwood, United Kingdom: Multi-Science Publishing Company and Petroconsultants

Curtis, J. B. (1999). "Comparison of Estimates of Recoverable Natural Gas Resources in the United States: A Report of the Potential Gas Committee." *Gas Resource Studies Number 7*. Golden, CO: Colorado School of Mines.

Dromgoole, P., and Speers, R. (1997). "Geoscore: A Method for Quantifying Uncertainty in Field Reserve Estimates." *Petroleum Geoscience* 3:1–12.

Edwards, J. D. (1997). "Crude Oil and Alternate Energy Production Forecasts for the Twenty-first Century: The End of the Hydrocarbon Era." *American Association of Petroleum Geologists Bulletin* 81:1292–1305.

Energy Information Administration. (1998). *International Energy Outlook 1998*. Washington, DC: Department of Energy/Energy Information Administration.

Energy Information Administration. (1999). *Annual Energy Review 1998*. Washington, DC: Department of Energy/Energy Information Administration.

Foley, L.; Ball, L.; Hurst, A.; Davis, J.; and Blockley, D. (1997). "Fussiness, Incompleteness and Randomness: Classification of Uncertainty in Reservoir Appraisal." *Petroleum Geoscience* 3:203–209.

Grace, J. D.; Caldwell, R. H.; and Heather, D. I. (1993). "Comparative Reserves Definitions: U.S.A., Europe, and the Former Soviet Union." *Journal of Petroleum Technology* 45(9):866–872.

International Energy Agency. (1998). *World Energy Outlook*. Paris, France: International Energy Agency/OECD.

Klett, T. R.; Ahlbrandt, T. S.; Schmoker, J. W.; and Dolton, G. L. (1997). *Ranking of the World's Oil and Gas Provinces by Known Petroleum Volumes*. Denver, CO: U.S. Geological Survey, Open File Report 97–463, CD-ROM.

Masters, C. D.; Attanasi, E. D; and Root, D. H. (1994). *World Petroleum Assessment and Analysis: Proceedings of the 14th World Petroleum Congress*. London, England: John Wiley and Sons.

Petroconsultants, Inc. (1996). *Petroleum Exploration and Production Database*. Houston, TX: Author.

Potential Gas Committee. (1999). "Potential Supply of Natural Gas in the United States. A Report of the Potential Gas Committee." *Potential Supply of Natural Gas, 1998*. Golden, CO: Colorado School of Mines.

Schmoker, J. W., and Klett, T. R. (1999). "U.S. Geological Survey Assessment Model for Undiscovered Conventional Oil, Gas, and NGL Resources—The Seventh Approximation." *U.S. Geological Survey Electronic Bulletin 2165*. Golden, CO: U.S. Geological Survey.

Schuyler, J. (1999). "Probabilistic Reserves Definitions, Practices Need Further Refinement." *Oil and Gas Journal*, May 31.

Seba, R. D. (1998). *Economics of Worldwide Petroleum Production*. Tulsa, OK: Oil and Gas Consultants International Inc.

U.S. Geological Survey. (2000). *U.S. Geological Survey World Petroleum Assessment 2000—Description and Results*. Washington, DC: U.S. Government Printing Office.

U.S. Geological Survey. (1995). "National Assessment of United States Oil and Gas Resources." *U.S. Geological Survey Circular 1118*. Washington, DC: U.S. Government Printing Office.

U.S. Geological Survey and U.S. Bureau of Mines. (1976). "Principles of the Mineral Resource Classification System of the U.S. Bureau of Mines and U.S. Geological Survey." *U.S. Geological Survey Bulletin 1450-A*. Washington, DC: U.S. Government Printing Office.

RESIDUAL FUELS

A residual fuel oil is any petroleum-based fuel which contains the undistilled residue from atmospheric or vacuum distillation of crude oil and may be called Bunker Fuel Oil, No. 6 Fuel Oil, or heavy Fuel Oil. When diluted with distillate to reduce viscosity, it may be called Marine Diesel Intermediate Fuel, No. 5 Fuel Oil, No. 4 Fuel Oil, or Light Fuel Oil) (Table 1). It is higher in viscosity than a distillate fuel, so it normally requires preheating for pumping and atomization. Asphaltenes, which are high molecular weight condensed aromatics in residual fuel, are difficult to burn completely, so combustion tends to give particulate emissions. Residual fuel usually contains organometallic compounds of vanadium, which leave an ash residue that may cause deposits and corrosion in boilers and engines because residual fuels are higher in sulfur and nitrogen than distillates. The emission of sulfur and nitrogen oxides during combustion is much greater. Carbon content is also higher than for distillate fuels or gas, so it emits more CO_2 than they do, but less than coal. Basically residual fuels are more difficult to handle and tend to pollute more than distillate or gaseous fuels, but less than solid fuels such as coal.The major advantage of residual fuel is that it is the cheapest liquid fuel available, in large part because it is a byproduct of the refining of gasoline and diesel fuel. As such, it sells for less per gallon than crude oil, the raw material from which it is made. Since heavy fuel prices are usually 50 percent to 75 percent of the price of crude, refiners try to minimize production of heavy fuel oil and maximize the production of higher-value gasoline and distillate fuels.

In order to use this low-priced fuel, consumers must make a significant investment in handling equipment. Heavier grades need preheat equipment to maintain 50°C (122°F) for pumping in the distribution system, and 135°C (275°F) for atomization in boilers or diesel engines. A relatively long residence time is needed in the boiler or engine to give complete burnout of the fuel asphaltenes, which tend to produce particulates. Thus, residual fuels are normally used only in relatively large installations.

No. 6 Fuel Oil, Bunker Fuel Oil	
Viscosity, Cst @ 50C	500
Density, Kg/Cu.Meter	985
Flash Point, Deg C	60
Energy Content, KJoule/Kg.	43,000
Chemical Composition	
Carbon, Wt.%	86
Hydrogen, Wt.%	11.5
Sulfur, Wt.%	2.5
Ash Content, Wt.%	0.08
Vanadium, PPM	200
Lower Viscosity Blends	
Marine Diesel Fuel	
Intermediate Fuel, IF 180 Cst @ 50 C	180
Intermediate Fuel, IF 380 Cst @ 50 C	380
No. 5 Fuel Oil, Cst @ 50 C	40
No. 4 Fuel Oil, Cst @ 40 C	20

Table 1.
Typical Residual Fuel Properties
SOURCE: ASTM T.P.T. Course Notes "Marine Fuels: Specifications, Testing, Purchase, and Use."

In the earliest days of refining, kerosene was the principal refinery product. All the crude heavier than kerosene was sold as residual fuel and used initially for industrial and commercial steam generation. Residual fuel soon began to displace coal in steamships and railroad locomotives because of its portability and relatively low ash content. In the 1920s, residual fuel underwent a drastic change in composition because of the installation of thermal cracking (a refinery process in which a heavy distillate or residual fraction of petroleum is subjected to high temperature, which causes large, high boiling molecules to crack into smaller ones which boil in the gasoline or heating oil range) to meet gasoline demand. Most heavy fuels contained large amounts of thermal tars and visbreaker products (products of a thermal racking process). Great care had to be taken when blending these fuels to avoid sediment formation due to incompatibility.

With the development of catalytic cracking in the 1930s and 1940s, residual fuel composition changed again. Vacuum distillation came into use to provide additional clean feed for catalytic cracking. Vacuum distillates, which had previously been part of residual fuel, were now converted to gasoline and heating oil in the catalytic cracker.

The residuum from vacuum distillation became, and still is, the basic component of residual fuel oil. It contains the heaviest fraction of the crude, including all the ash and asphaltenes. It is extremely high in viscosity and must be diluted with light distillate flux (a low viscosity distillate or residual fraction which is blended with a high viscosity residual fraction to yield a fuel in the desired viscosity range) to reach residual fuel viscosity. The lowest value distillates, usually cracked stocks, are used as flux. In some cases the vacuum residuum is visbroken to reduce its viscosity so that it requires less distillate flux.

By the end of World War II the use of residual fuel oil in the United States had reached about 1.2 million barrels per day. The bulk of this use was in industrial/commercial boilers, railroad locomotives, and steamships. Shortly thereafter, railroad use declined rapidly as diesel engines, which used distillate fuel, replaced steam locomotives. In the 1950s and 1960s residual fuel oil use for marine and industrial applications, as well as for electric power generation, con-

	United States		Worldwide	
	Electric Power Generation	Industrial & Commercial	Diesel Ships	Steam Ships
1973	1406	1421	2591	1085
1978	1612	1411	2101	700
1983	652	769	1418	245
1988	646	732	1803	240
1993	424	656	2171	227
1997	301	496	2451	210

Table 2
Residual Fuel Oil Consumption
SOURCE: Source: U.S. Energy Information Administration. (1998). *Monthly. Energy Review* (October).

tinued to grow. During this period, the use of fuel oil for marine propulsion gradually shifted from steamships to low and medium speed diesel engines, which were gradually displacing steam turbine propulsion because they were more efficient.

By 1973 about 1.4 million barrels per day of residual fuel oil were used for electric power generation in the United States. This accounted for 16.8 percent of U.S. electricity generation, mostly in areas where cheap, foreign heavy fuel could be delivered by tanker. That same year, another 1.4 million barrels per day of heavy fuel oil were used in the United States for industrial and commercial applications. Worldwide during 1973 about 2.6 million barrels per day of residual fuel oil were used in marine diesel engines, and another 1.1 million barrels per day were used for steamship propulsion.

After the energy crisis of the 1970s, the price of heavy fuel oil rose dramatically and demand dropped drastically. By 1983, heavy fuel oil for electric power generation in the United States fell to 650 thousand barrels per day, providing just 6.2 percent of the total generated. Industrial/commercial demand fell almost as far, to 770 thousand barrels per day. Worldwide demand for marine diesel fuel had dropped to 1.4 million barrels per day and steamship demand to 245 thousand barrels per day. Since 1983, the demand in each of these sectors except marine diesel fuel has continued to decrease. Diesel engines, which are more efficient than steam turbine systems, have become the dominant means of ship propulsion. By 1997, the demand for marine diesel fuel was back to about 2.4 million barrels per day. It is forecast to stay at about that level or slightly higher.

Refiners have coped with this decrease in demand for residual fuel oil by making as little as possible and by shifting their production to heavier grades. They no longer make lighter grades, such as No. 4 or No. 5 fuels, except by special arrangement. Consumers who previously used these grades now use No. 6 fuel oil or gas. Similarly, most marine diesel fuel oil is now supplied as IF 180 (180 Cst @ 50°C) or IF 380 (380 Cst @ 50°C). Prior to 1973, many marine diesel engines were operated on lighter grades. Switching from IF 60, which is roughly 29 percent distillate and 71 percent bunker fuel (the residual fuel burned in boilers on steamships), to IF 380, which is 2 percent distillate and 98 percent bunker fuel, would increase bunker fuel consumption by 38 percent and reduce distillate consumption by 93 percent. Such switches have been beneficial to both refiners and consumers since the energy crisis. Refiners supply more residual and less distillate to a given consumer, and the consumer pays a lower price for the heavier grade of fuel. In most cases, the diesel engine operates properly on the heavier fuel.

While some refiners have reduced residual fuel production by supplying heavier grades, others have eliminated residual fuel completely by installing cokers or hydrocrackers. These are process units that convert residua to gasoline or distillate. They are very expensive to install and operate, but can be justified when there is an oversupply of residual fuel.

Charles W. Siegmund

BIBLIOGRAPHY

Guthrie, V. B. (1960). *Petroleum Products Handbook*. New York: McGraw-Hill.

Energy Information Administration. (1998). *Monthly Energy Review* (October). Washington, DC: U.S. Depart-ment of Energy.

Siegmund, C. W. (1997). *Marine Fuels: Specifications, Testing, Purchase, and Use*. ASTM Technical & Professional Training Course Notes. West Conshocken, PA: ASTM.

REST ENERGY

See: Einstein, Albert; Matter and Energy

RIDESHARING

See: Traffic Flow Management

RISK ASSESSMENT AND MANAGEMENT

By their very nature, operations in the energy industry are characterized by high risk. Operating an oil well, a power plant, or petrochemical plant is considerably more complex and costly than running most other operations. Losses may be infrequent, yet when they occur they could be substantial. The risks involved in these kinds of operations can be classified under two categories—technical (engineering) risk and financial (price) risk. This article focuses on Financial Risk.

PRICE RISK

The price risk can be defined and understood in alternative ways. One can view the risk as the probable fluctuation of the price around its expected level (i.e., the mean). The larger the deviation around the mean the larger is the perceived price risk. The volatility around the mean can be measured by standard deviation and be used as a quantitative measure for price risk. At the same time, in the industry it is common to define "risk" referring only to a price movement that would have an adverse effect on the profitability. Thus, one would talk about an "upward potential and downside risk."

ENERGY PRICE DYNAMICS

Unpredictable movements in the price level are uncommon in energy markets. The magnitudes of these price shocks can be substantial: 1973 and 1979 oil price shocks, electricity price swings in June 1998, and 2000 oil price increases exemplify the potential magnitudes of these price fluctuations. Energy price dynamics usually consist of three components: deterministic part, seasonal and cyclical influences, and noise. In a market situation with no demand or supply shocks, the only uncertainty observed in the market will be due to noise. A number of variables, such as import policies, policy changes of industry organizations (e.g., OPEC); unexpected weather and atmospheric conditions, tax and regulatory policy changes, legal, environmental, and economic problems, and political and currency crises can affect the demand/supply for energy products and thus lead to large price swings.

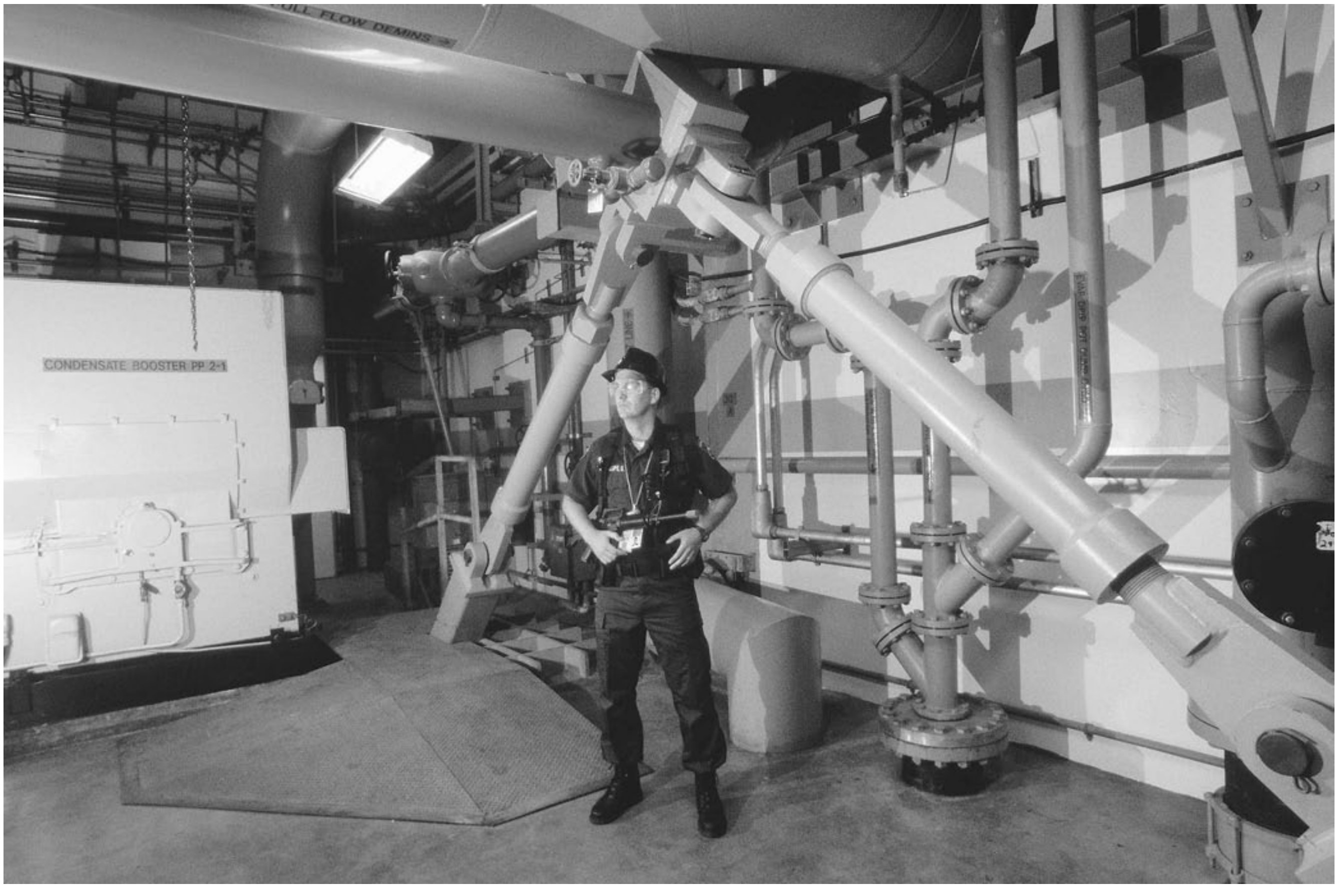
DIVERSIFIABLE AND NONDIVERSIFIABLE RISK

Market-wide (governmental, political, legal, environmental, or economic) events may affect individual firms in the energy industry differently. The magnitude of the individual impact will be determined by the co-movement of the market and individual firms (systematic risk). At the micro level, companies are not only exposed to price and systematic risk, but also to firm-level, nonsystematic risk factors. At the market level, these individual firm-level risks can be reduced or diversified, making it possible to focus on price management only. But at the individual firm level, risk managers also have to address a number of engineering risk factors, such as windstorms that may damage offshore platforms, explosions or fire in the well, and vapor cloud explosions.

MANAGING RISK

Beause energy firms face a variety of potential risk factors that may lead to substantial price fluctuations, risk managers in energy industries combine traditional insurance strategies with financial instruments and hedging policies to manage the risk. The standard financial hedging instruments that are employed are futures and options contracts traded on several energy products, such as crude oil, heating oil, gasoline, natural gas, and electricity.

Energy companies' property/casualty insurance programs typically address a wide spectrum of exposures, such as political risk, earthquake, and workers' compensation. Traditional insurance contracts have been used in energy markets for years against property damage and casualty, for reducing coastal hurricane liability and similar cost reduction purposes. Due to the many great risks that energy companies face, they commonly have to reduce some of the total



The seismic braces seen here in the Diablo Canyon Nuclear Power Plant reduce the risks associated with an area prone to earthquakes. (Corbis Corporation)

risk so that they can afford the insurance to undertake a project.

A recent financial concept that is becoming popular is reinsurance, a transaction whereby one insurance company agrees to indemnify another insurance company against all or part of the loss that the latter sustains under a policy or policies that it has issued. For this service, the ceding company pays the reinsurer a premium. The purpose of reinsurance is the same as that of insurance (i.e., to spread risk). With financial reinsurance in place, energy companies can use traditional excess insurance to address their higher levels of coverage. The combined risk management strategy uses a financial instrument in combination with an insurance policy. This improves the ability of energy companies to fund their own losses during periods of high prices and profits, yet in unfavorable market conditions, the companies can tap an arranged insurance capacity to decrease their financial vulnerability. This strategy is attractive for energy companies because it limits the probability of having to pay for losses to cases where the insurance is actually needed.

There are several other financial risk management tools used in energy markets. For example, option-trading strategies are employed to avoid the adverse effects of price movements. Companies can also protect themselves from volatile (or stable) prices by trading various option strategies. They can trade on the cost of processing products by buying and selling contracts on the raw and processed goods. Recent contributions to risk management in energy industries comprise instruments and strategies such as bidding hedges and weather swaps. A bidding hedge is used to protect against the financial exposures in a competitive bid to construct an energy facility (e.g., currency risks, interest rate, and price risks). Weather swaps allow companies to exchange a series of variable cash flows for a series of fixed cash flows dependent on an index based on weather statistics. One common application of this class of instruments is the temperature swap that attracts many energy companies' interest whose products are sensitive to weather conditions and temperature (e.g., heating oil).

Recent trends in energy finance can be outlined as follows: The focus of main energy sources is turning away from coal and other high CO_x emitting sources to natural gas. Investment in crude oil is slowing. Nuclear power plants remain unpopular due to high operational risk and high waste clean-up costs. On the other hand, investments in the electricity sector are growing at a rapid rate; they have increased from $500 million in 1988 to over $10 billion in 1997.

Ahmet E. Kocagil

See also: Efficiency of Energy Use; Energy Economics; Futures; Supply and Demand and Energy Prices.

BIBLIOGRAPHY

Buljevich, E. C., and Park, Y. S. (1999). *Project Financing and the International Financial Markets*. Boston: Kluwer Academic Press.

Fusaro, P. (1998). *Energy Risk Management*. New York: McGraw-Hill.

Helbling, C. P.; Fallagger, G.; Hill, D. (1996). *Rethinking Risk Financing*. Zurich. Switzerland: Swiss Reinsurance Company.

McKechnie, G. (1983). *Energy Finance*. London: Euromoney Publications.

The Petroleum Economist. (1991). *Energy Finance Handbook*, London: Author.

Risk Books. (1999). *Energy Modelling and the Management of Uncertainty*. New York: Author.

ROCKET PROPELLANTS

Rocket propellant is a mixture of combustible substances that is burned inside the combustion chamber of a rocket engine. Burning is the chemical process of decomposition and oxidation of the propellant. The resulting highly heated and compressed gas (propulsive mass) is ejected from a combustion chamber and facilitates propulsion—movement of the aggregate attached to the rocket engine. In physical terms, combustion converts chemical energy into kinetic energy.

Rocket propellants possess unique properties, such as a capability to self-sustain the burning process, generate thermal energy, and simultaneously produce propulsive mass. Some types of propellants are even able to self-ignite (initiate burning without outside power input). Unlike most other combustible chemicals, rocket propellants can burn in vacuum. This is because a propellant consists of two integral components: a fuel that burns and produces propulsive mass and an oxidizer that facilitates and sustains oxidation. In this respect, rocket propellants are more like explosives than like automobile and aviation fuels used that require atmospheric air for oxidation. The major difference between explosives and rocket propellants is the gas expansion rate which is much slower in propellants and makes it possible to contain and control the process of burning inside the rocket engine. Most modern rockets use solid and liquid propellants.

SOLID PROPELLANTS

A solid propellant is a mechanical (heterogeneous) or a chemical (homogeneous, or colloidal) mixture of solid-state fuel and oxidizer-rich chemicals. Specially-formed charges of solid propellant (grains) are placed in the combustion chamber of the solid rocket motor (SRM) at a production facility. Once assembled, the engine does not require additional maintenance, making it simple, reliable and easy to use.

The earliest known rocket engines used gunpowder as a solid propellant. Composed of 75 percent potassium nitrate, 10 percent sulfur, and 15 percent charcoal, gunpowder was probably invented in China in the tenth century. One of the first references to rockets can be found in Chinese chronicles from 1232. Rockets began to appear in Europe in thirteenth century. A century later, they were used for various purposes, mostly for fireworks, in France and Italy. The term "rocket," which became common in fifteenth or sixteenth century, originated from the Italian word *rocchetta* (spindle), a reference to the shape of early rockets. In seventeenth and eighteenth centuries, rockets spread all around Europe. In the nineteenth century they were widely used in Great Britain, Russia, France, Austria,

The space shuttle *Challenger* lifting off. (NASA)

Germany and other European countries as a weapon: an auxiliary light-weight artillery. During the same period, British rockets were brought to the American continent. Solid rocket motors evolved with the introduction of smokeless gunpowder, a far more energy-efficient propellant. Unlike the black gunpowder that was a mechanical mixture, smokeless gunpowder was a complex compound developed chemically from cellulose. Rockets of this type first appeared in the early twentieth century but were not widespread until the 1940s. Unguided short-range barrage missiles with smokeless gunpowder propellant were used extensively in World War II by the Soviet Union, United States and Germany. Germans built the first long-range solid propellant missile, called *Rheinbote*. A composite solid propellant developed in the United States in 1941 facilitated a real revolution in solid rocket technology in the 1950s and1960s. It allowed the production of very large and powerful solid propellant engines that became suitable for long-range ballistic missiles and space launch systems.

Modern composite solid propellant is a mechanical mixture of the powder-like chemicals and a binding resin. The propellant used for the Space Shuttle solid rocket boosters (SRBs) is a typical example of such mixture:

- Ammonium perchlorate (oxidizer), 69.93%
- Aluminum powder (fuel), 16.00%
- Rubber-based polymer, called PBAN (binder), 12.04%
- Epoxy-based curing agent, 1.96%
- Iron oxide powder (burning rate catalyst), 0.07%

After these components are mixed, the propellant looks like a thick syrup. It is poured into the engine's casing, where it cures and solidifies. During solidification process, the propellant glues itself to the walls of the casing and special tools are used to produce a cylindrical or star-shaped channel in the center of the grain. Due to changing area of the burning surface as propellant is consumed, the channel's diameter and shape determine the initial thrust and how it changes during the burn time. Since the burning starts inside the channel, the propellant itself shields the engine's casing from the hot gases until the very last seconds of the burn time. This is the most common configuration, but other shapes of the propellant grains can be used, if necessary.

In its final state, the composite solid propellant looks like a gray rubbery material and must be flexible enough to resist stress and vibration. Even small imperfections of the propellant surface, such as air bubbles, caverns and cracks, are very dangerous. Imperfections may cause a sudden radical increase of internal pressure (due to increase of the burning surface area) and rupture the engine's casing, leading to a catastrophic explosion. Integrity of the engine's body is especially important for very large solid rocket motors, which are assembled from separate sections. In 1986, loss of the engine's casing integrity caused a national tragedy: hot gases leaked through a faulty joint between two sections of the SRB and destroyed the Space Shuttle *Challenger*, killing seven astronauts. Production of the large solid propellant engines require a substantial amount of time, high precision equipment, and a tightly controlled environment. As a result, the modern powerful solid rocket motors are as expensive as the much more complex liquid propellant engines.

Robert H. Goddard standing to the left of the first flight of a liquid-propellant rocket. (Public Domain)

Ammonium perchlorate (oxidizer)	69.93%
Aluminum powder (fuel)	16%
Rubber-based polymer, called PBAN (binder)	12.04%
Epoxy-based curing agent	1.96%
Iron oxide powder (burning rate catalyst)	0.07%

Table 1.
Energy Efficiency of Modern Rocket Propellants

LIQUID PROPELLANTS

A liquid propellant consists of two liquid chemicals, fuel and oxidizer, which are delivered from separate tanks into the combustion chamber of a liquid propellant rocket engine (LPRE).

The use of a liquid propellant, namely liquid oxygen and liquid hydrogen, was proposed by the Russian space pioneer Konstantin Tsiolkovsky in 1903. He hypothesized that in addition to supplying more energy than gunpowder, such propellant would also supply water and oxygen to the crew of the future spacecraft. Space proponents from other countries independently came to the same conclusion. One of them was the American physics professor Robert Goddard, who built and launched the world's first liquid propellant rocket in 1926. LPRE technology received a major boost in the 1940s, when German engineers under Wernher von Braun developed the first operational long-range ballistic missile, *A-4* (better known as the *V-2*).

Despite the variety of liquid propellants, all of them can be divided into two major categories: storable and cryogenic. Physical properties of the storable propellants are very attractive from a technological standpoint. They have high density, do not boil under normal temperature, and degrade slowly. That makes it possible to keep the propellant inside a rocket for long periods, which is especially important for military applications and lengthy space missions. Among numerous possible combinations, the derivatives of nitric acid and hydrazine are the most widely used storable propellants today. Such propellants are self-igniting (hypergolic): the rocket engine fires as soon as the propellant components are mixed in a combustion chamber. An added advantage is that no electrical ignition mechanism is required. A typical example of a hypergolic propellant is a propellant for the Russian space launch vehicle Proton. It contains 73 percent of nitrogen tetroxide (NT) as an oxidizer, and 27 percent of unsymmetrical dimethyl hydrazine (UDMH) as a fuel.

Convenient physical properties cannot overshadow the extremely dangerous nature of storable propellants. Practically all of them are highly toxic, corrosive toward many materials, and cause severe burns on human skin. The history of the world rocketry reveals numerous cases of injures and deaths among rocket service personnel due to mishandling of these deadly chemicals. The most tragic of such accidents happened in the Soviet Union on October 24, 1960, when a ballistic missile filled with storable propellant exploded while being serviced by a ground crew. About a hundred people were fatally burned, or suffocated in toxic fumes. The founder of the Soviet space program, Dr. Sergei Korolev, called storable propellants "the devil's venom."

PROPELLANT	***SPECIFIC IMPULSE AT SEA LEVEL***	***COMMENTS***
Composite solid propellant	270 seconds	Used for modern solid propellant engines in military missiles, space launch vehicles and spacecraft. Pollutes environment by toxic exhaust.
Nitrogen Tetroxide + UDMH (liquid storable)	296 seconds	The best modern storable propellant. Widely used in military missiles, space launch vehicles and spacecraft. Both components are toxic before burning and pollute environment by toxic exhaust.
LOX + Kerosene (cryogenic, non-storable)	311 seconds	Widely used in space launch vehicles. Non-toxic, environmentally clean.
LOX + Liquid Hydrogen (cryogenic, non-storable)	388 seconds	The best modern cryogenic propellant. Used in the most advanced space launch vehicles. Non-toxic, environmentally clean.
Liquid Fluorite + Liquid Hydrogen (cryogenic, non-storable)	411 seconds	Not Used. Oxidizer is very toxic before burning and pollutes environment by toxic exhaust.

Table 2
Specific impulses of Rocket Propellants.

At first glance, the physical properties of the cryogenic propellants seem far less suitable for rocket technology since they consist of volatile liquified gases, such as liquid oxygen (LOX). Due to extremely low temperature (-183° Celsius for LOX), cryogenic propellant components require special care and complex equipment. These components are constantly boiling and evaporating under normal conditions and cannot be stored for a long time. After the cryogenic rocket is fueled and prepared for launch, it must constantly be fed by new portions of propellant to compensate for evaporation losses. This feeding can only be stopped just seconds before the takeoff. In spite of these disadvantages, cryogenic propellants provide superior energy characteristics. Liquified gases represent pure chemical elements that fully participate in a process of oxidation. For example, nitrogen tetroxide contains only 70 percent of oxygen compared to 100 percent for LOX. Thus, the energy output of complete oxidation is greater than in the case of the complex chemical compounds. Cryogenic components cost much less to produce than solid composite and liquid storable propellants. Natural resources of oxygen is also unlimited: the Earth's atmosphere. Yet the complexity of the cryogenic rocket hardware makes this technology very expensive.

Liquid oxygen is the most common cryogenic material today and the oldest liquid oxidizer in the rocket technology: it was used for the first liquid propellant rocket of Robert Goddard. It remains unsurpassed in its combination of effectiveness, low cost, and environmental "friendliness." Although liquid oxygen can be used with a variety of storable fuels (alcohol, kerosene, hydrazine, etc.), the best result is achieved with another cryogenic component, liquid hydrogen (LH_2). This combination is the most energy-efficient propellant of today's rocket technology. Yet it is probably one of the most difficult propellants to use. The major problem with liquid hydrogen is its low density (only 7% of water density). Thus, LH_2 occupies large volume and requires fuel tanks to be so spacious and heavy that it sometimes makes the use of LH_2 prohibitive. Rocket engineers are trying to resolve this problem by various methods. For example, the introduction of the new lightweight alloys decreased the mass of the Space Shuttle External

Tank (containing LOX/LH_2 propellant) by 13 percent, and NASA is willing to make it 26 percent lighter. Another solution is the "hydrogen slush," which is liquid hydrogen hyper cooled almost to the point of solidification (–259° Celsius compared to –252° Celsius for "normal" LH_2). Due to 15 percent greater density, hydrogen slush occupies lesser volume and has lower evaporation losses.

Although modern chemistry allows development of even more effective rocket propellants, energy efficiency is not the only consideration factor. For example, fluorine and its derivatives are better oxidizers than oxygen, but their extreme toxicity make them environmentally dangerous. The same concerns prevent the use of beryllium hydride—an excellent fuel that combines high density with the energy efficiency comparable to liquid hydrogen.

An increasing number of space launches worldwide makes environmental and economic problems more acute even with the existing technology. Most modern space launch vehicles use solid and storable propellants that pollute air, water and soil by toxic exhausts and by the remains of unburned propellant inside the ejected stages. One possible solution is replacement of the toxic propellants with the less expensive non-toxic components on the existing launch vehicles. Russian engineers, for example, proposed the use of liquid methane (or another natural gas) in combination with liquid oxygen. This new propellant would completely eliminate pollution and allow a lower cost per launch. At the same time its energy-efficiency would be second only to the liquid hydrogen/liquid oxygen propellant.

Apart from its chemical content and physical properties, the performance of a rocket propellant depends on engine design. Assuming that such design is optimal for a particular propellant, it is possible to compare various propellants using the energy-efficiency criteria of rocket engines, the specific impulse. Greater specific impulse indicates better propellant. Table 2 shows specific impulses of several typical propellants achievable by the best modern rocket engines.

Peter A. Gorin

See also: Propulsion; Spacecraft Energy Systems.

BIBLIOGRAPHY

Gatland, K., ed. (1984). *The Illustrated Encyclopedia of Space Technology*. 4th ed. London: Salamander Books.

Rycroft, M., ed. (1990). *The Cambridge Encyclopedia of Space*. Cambridge, Great Britain: Cambridge University Press.

ROCKETS

See: Spacecraft Energy Systems

ROTARY ENGINES

See: Engines

S

SAKHAROV, ANDREI DMITRIEVICH (1921–1989)

Andrei Sakharov was a Soviet physicist who became, in the words of the Nobel Peace Prize Committee, "a spokesman for the conscience of mankind." He made many important contributions to our understanding of plasma physics, particle physics, and cosmology. He also designed nuclear weapons for two decades, becoming "the father of the Soviet hydrogen bomb" in the 1950s. After recognizing the dangers of nuclear weapons tests, he championed the 1963 U.S.-Soviet test ban treaty and other antinuclear initiatives.

From the 1960s onward, at great personal risk, Sakharov severely criticized the Soviet regime and ardently defended human rights against it. He won the Nobel Peace Prize in 1975.

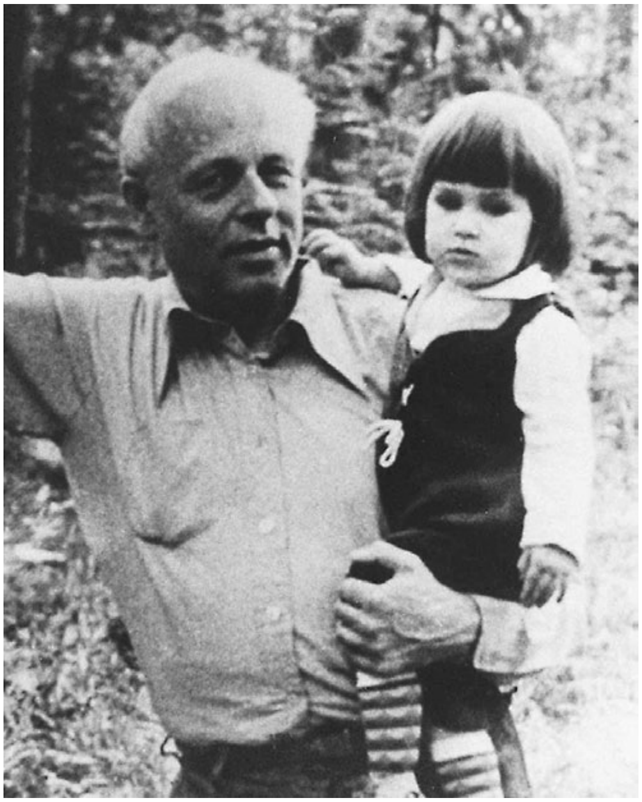
Andrei Sakharov was born in Moscow, Russia, to a family of the intelligentsia on May 21, 1921. His father, Dmitri, taught college physics and wrote textbooks and popular science books. Sakharov studied at home until the seventh grade. Dmitri, a man of warmth and culture, was his first physics teacher.

In 1938 Sakharov entered the Physics Department of Moscow State University. When World War II began, his academic prowess exempted him from military service. He and the remaining students and teachers were evacuated to Ashkhabad in Soviet Central Asia. Sakharov graduated with honors in 1942, with the war at its height, but joined a factory rather than continue school. While doing routine laboratory work at a munitions factory in Ulyanovsk on the Volga River, his engineering talent showed through a number of inventions. He also met Klavdia Vikhireva, a laboratory technician. They married in 1943.

Sakharov returned to Moscow in early 1945, as a graduate student at FIAN, the Physical Institute of the USSR Academy of Sciences. Igor Tamm, head of FIAN's Theoretical Physics Department, influenced him greatly. In 1947, Sakharov received his Ph.D. for work on particle physics.

In June 1948 Tamm was commissioned to help Yakov Zeldovich and his research team, which for two years had studied the feasibility of a thermonuclear, or hydrogen, bomb. Tamm and several of his students, including Sakharov, formed a special auxiliary group at FIAN to work on an H-bomb proposal. Sakharov suggested a radically new scheme for the bomb, and Vitaly Ginzburg added a key idea concerning thermonuclear explosive material. The U.S.-Soviet arms race took off. In the spring of 1950, Tamm and Sakharov moved to the Installation, a secret city in the central Volga region of the USSR. They worked on Sakharov's scheme and successfully tested the first prototype Soviet H-bomb on August 12, 1953.

Also in 1950 Sakharov and Tamm proposed an idea for a controlled thermonuclear fusion reactor, the TOKAMAK (acronym for the Russian phrase for "toroidal chamber with magnetic coil"), which achieved the highest ratio of output power to input power of any fusion device of the twentieth century. This reactor grew out of interest in a controlled nuclear fusion reaction, since 1950. Sakharov first considered electrostatic confinement, but soon came to the idea of magnetic confinement. Tamm joined the effort with his work on particle motion in a magnetic field, including cyclotron motion, drifts, and magnetic surfaces. Sakharov and Tamm realized that



Andrei Sakharov with his granddaughter Anya, in the 1970s. (Library of Congress)

destructive drifts could be avoided either with current-carrying rings in the plasma or with an induction current directly in the plasma. The latter is essentially the TOKAMAK concept.

At the Installation, Sakharov worked with many colleagues, in particular Yakov Zeldovich and David Frank-Kamenetskii. Sakharov made key contributions to the Soviets' first full-fledged H-bomb, tested in 1955. He also made many contributions to basic physics, perhaps the most important being his thesis that the universe is composed of matter (rather than all matter having been annihilated against antimatter) is likely to be related to charge-parity (CP) noninvariance.

Sakharov received many honors. He was elected as a full member of the Soviet Academy of Sciences in 1953 (at only age thirty-two); he was awarded three Hero of Socialist Labor Medals; he received a Stalin Prize; and he was given a country cottage. Sakharov's anti-Soviet activism cost him these rewards.

In May 1968 he completed the essay "Reflections on Progress, Peaceful Coexistence, and Intellectual Freedom." It proposed Soviet cooperation with the West, which the Soviets flatly rejected. The manuscript circulated in several typewritten copies known as *samizdat* ("self-print" in Russian) and was widely read outside the USSR. Sakharov was summarily banned from military-related research.

Also in May 1968, Sakharov accepted an offer to return to FIAN to work on academic topics. He combined work on fundamental theoretical physics with increased political activism, developing contacts to the emerging human rights movement. His wife, Klavdia, died of cancer in March 1969. In 1970 Sakharov and Soviet dissidents Valery Chalidze and Andrei Tverdokhlebov founded the Moscow Human Rights Committee. In the movement he met Elena Bonner, who became his companion-in-arms. They married in 1972.

Although Sakharov won the 1975 Nobel Peace Prize, and was the only Soviet ever to win it, he was barred from leaving Russia to receive it. The Nobel Committee's official citation praised Sakharov for his "fearless personal commitment in upholding the fundamental principles for peace…. Uncompromisingly and with unflagging strength Sakharov has fought against the abuse of power and all forms of violation of human dignity, and he has fought no less courageously for the idea of government based on the rule of law."

The Soviet regime persecuted Sakharov for his activism on behalf of dissidents and those seeking to emigrate. After he spoke out against the Soviet invasion of Afghanistan in 1979, he was picked up by the KGB and exiled to Gorky, under house arrest. There was no trial. In 1986 he was released by Premier Mikhail Gorbachev and returned to Moscow.

Sakharov made his first trip outside the Soviet Union in late 1988. In 1989 he was elected to the Congress of People's Deputies, the supreme legislative body of the Soviet Union. He died on December 14, 1989.

Andrew M. Sessler

See also: Ethical and Moral Aspects of Energy Use; Military Energy Use, Historical Aspects of; Nuclear Energy, Historical Evolution of the Use of; Nuclear Fusion.

BIBLIOGRAPHY

Anderson, R. H. (1989). "Andrei D. Sakharov, 68, Nobel Laureate and Wellspring of the Soviet Conscience." *New York Times*, December 16.

Andrei Sakharov: Soviet Physics, Nuclear Weapons, and Human Rights. American Institute of Physics. Niels Bohr Library. August 1, 2000 <http://www.aip.org/history/sakharov/>.

Kapitza, S. P., and Drell, S., eds. (1991). *Sakharov Remembered*, New York: American Institute of Physics.

Sakharov, A. (1990). *Memoirs*. New York: Alfred A. Knopf.

SANDIA NATIONAL LABORATORY

See: National Energy Laboratories

SAVERY, THOMAS (1650–1715)

In the study of seventeenth-century life, mystery often pervades the lives and deaths of even the most famous people and Thomas Savery is no exception. There is no record of his birth in local registers, but it is believed that he was born near Plymouth in Shilston, England, around 1650. There is no known portrait of him available today, although there was purportedly a drawing of him in a reprint of *A Miner's Friend*, published in 1827. After 1700 he became known as Captain Savery, with little evidence as to why, although he had done service as a military engineer, working as a Trench Master in 1696. According to folklore, many who were placed in charge during that era were likely to be called "Captain."

What *is* known is that Savery is sometimes left out of listings of the great mechanical inventors, even though he played a decidedly major role in the invention of the first steam engine. In fact, he was called "the most prolific inventor of his day" by the very book that set out to give proper due to the oft-neglected Thomas Newcomen as the steam engine's true inventor—*The Steam Engine of Thomas Newcomen* (Rolt and Allen, 1997).

Savery's career was hardly preordained, although he came from a family of prosperous merchants in Totnes, who acquired the manors of Shilston and Spriddlescombe in the parish of Modbury in the early seventeenth century. Savery's family was well known in the West Country and he may have been a merchant in Exeter for a period of time. At the age of twenty-three, he was protected by a writ from Charles II forbidding the local politicians to molest him. The writ in 1673 cited Savery's "many losses, particularly in the last Dutch wars," while serving as a freeman for the Merchant Adventurers Company.

Continuing a career that began in military engineering, his first patent was granted on July 25, 1698, for his historic work entitled "Raising water by the impellent force of fire." He called it "The Miner's Friend." Captain Savery's most important contribution was to evaluate the advances and frustrations of more than a century of experiments with steam power and create important innovation by combining steam with the effects of atmospheric pressure. His inventiveness was stimulated by his keen knowledge of copper and tin mining operations.

The principle of his pump was that steam was passed from a boiler into a closed receiver filled with water where its pressure forced the water through a nonreturn valve and up an ascending delivery pipe. When all the water was expelled, the steam supply was shut off and a new supply of cold water was poured over the outside walls of the receiver. This cooled the receiver and condensed the steam within. A vacuum was therefore created in the receiver, forcing water up a suction pipe, though a second nonreturn valve by atmospheric pressure. When the receiver was refilled the cooling water was shut off, steam turned on, and the cycle was repeated.

In subsequent years, Savery made important improvements that benefited future steam inventions. In June 1699 he demonstrated to the Royal Society a pump with two receivers, each with a separate, hand-controlled steam supply. This ensured improved continuity of operation, allowing one receiver to operate in its vacuum stage and the other under steam pressure. In 1701, he added two more critical steps: a second boiler, avoiding the need to shut down the fire and pump, between stages; and he replaced the two interconnected steam cocks with a single valve, run with a manually operated long lever. This may have been the inspiration for the modern slide valve and his inventiveness created, in effect, the world's first feed-water heater.

Savery appeared to be the first to take the huge step out of the lab and into the practical workshop. His equipment was made of brass and beaten copper, using firebrick furnaces. It was said that Salisbury Court (extending from Fleet Street to the river Thames) was the site of the world's first steam pump factory, although there is evidence that Savery abandoned his project in 1705. The limitations of his progress became known, literally under fire.

When tested, his engines generated too much heat and steam pressure for the technology of the times. Soldering would melt or machine joints would split. In order to solve the need for pumping water from mines, the necessary system of pumps would be far too costly and dangerous. In light of the dangers, both historians and scientists might find it intriguing that Savery never saw fit to add a safety valve, invented around that time.

In one of the more amusing assessments of this danger, Steven Switzer commented in *Hydrostatics and Hydraulics* (1729) about the steam power, "How useful it is, in gardens and fountain works ... in the garden of that noble peer, the Duke of Chandos, where the engine was placed under a delightful banqueting house, and the water was being forced up into a cistern on top, used to play a fountain in a delightful manner." Rival steam inventors derisively chided that the guests might forego the delight of the water fountain, had they known exactly what pressures were building under their feet.

Savery's work at the turn of the century preceded Newcomen's radically different and more successful engine in 1712. It also followed on the heels of noteworthy early experimenters, such as Giambattista della Porta in Italy in the early 1600s, Salomon de Caus in England in the 1620s, David Ramsay of Scotland in 1631, Edward Somerset of England in 1663, Sir Samuel Morland in 1667, Otto von Guericke of Magdeburg in 1672, and Denis Papin in the 1690s.

Savery's name and work remained in the public eye, followed by many scientists and inventors during the eighteenth century. The most important was Englishman Thomas Newcomen, who is believed to have known Savery and also to have seen his engines at work. Although Newcomen created a very different engine to solve the needs of the mines, the similarities were sufficient and Savery had earned his patent first. As a result, Newcomen and Savery entered into a joint patent agreement, whereby each shared the benefits of the other's work, while saving themselves the expense of a costly battle over separate patents.

The practical, if limited, use of Savery's incrementally successful efforts were permanently inscribed in history, as highly important advancements of his time.

Dennis R. Diehl

See also: Steam Engines; Turbines, Steam.

BIBLIOGRAPHY

Cannon, J. (1997). *The Oxford Companion to British History*. Frome, Somerset: Butler & Tanner, Ltd.

Rolt, L. T. C. and Allen, J. S. (1997). *The Steam Engine of Thomas Newcomen*. Ashbourne, UK: Landmark Publishing.

SCIENTIFIC AND TECHNICAL UNDERSTANDING OF ENERGY

The word "energy" entered English and other European languages in the sixteenth century from Aristotle's writings, and was restricted to meanings close to his until the nineteenth century. The total entry for "Energy" in the first edition (1771) of the *Encyclopaedia Britannica* is as follows: "A term of Greek origin signifying the powers, virtues or efficacy of a thing. It is also used figuratively, to denote emphasis of speech." These meanings survive; Swift's energetic pronouncement in 1720 that "Many words deserve to be thrown out of our language, and not a few antiquated to be restored, on account of their *energy* and sound" still rings a bell. But they are no longer dominant. Stronger by far is the scientific meaning, fixed between 1849 and 1855 by two men, William Thomson (the future Lord Kelvin) and W. J. Macquorn Rankine, professors, respectively, of natural philosophy and engineering at the University of Glasgow in Scotland.

Though Rankine and Thomson gave the term "energy" its modern scientific meaning, they did not originate the concept. The idea that through every change in nature some entity stays fixed arose in a complex engagement of many people, culminating between 1842 and 1847 in the writings of four men: Robert Mayer, James Prescott Joule, Hermann Helmholtz, and Ludvig Colding. The later choice of a word, easy as it may seem, is no mere detail. In the 1840s the lack of a word for the concept struggling to be born was a heavy impediment. The two nearest "Kraft" in German, and "force" in English—were riddled with ambiguity. With "energy" everything fell into place.

Three issues are involved in the coming-into-being of both concept and term: (1) the eighteenth-century debate on Leibniz's notion of *vis viva* ("living force"), (2) the unfolding eighteenth- and nineteenth-century understanding of steam engines, and (3) the search after 1830 for correlations among "physical forces." Concept must come first, with (3) treated before (2). On force as a term, compare (1) and (3) with Newton's definition. For the Newtonian, a force is a push or a pull, which, unless balanced by another, makes objects to which it is applied accelerate. "Living force" was not that at all: It corresponded to the quantity that we, under Thomson's tutelage, know as kinetic energy. As for forces being correlated, force in that context bore a sense not unlike the modern physicist's four "forces of nature," gravity, electromagnetism, and the two nuclear ones. It referred to the active principle in some class of phenomena, vital force, chemical force, the forces of electricity, magnetism, heat, and so on. To correlate these—to unify them under some overarching scheme—was as much the longed-for grail in 1830 as Grand Unification would be for physicists after 1975.

CORRELATION OF FORCES AND CONSERVATION OF ENERGY

Sometime around 1800 there came over the European mind a shadow of unease about Newtonian science. Beauty, life, and mystery were all expiring in a desert of atoms and forces and soulless mechanisms—so the poets held, but not only they. Science also, many people felt, must seek higher themes, broader principles, deeper foundations—symmetries, connections, structures, polarities, beyond, across, and above particular findings. Elusive, obscure, sometimes merely obscurantist, remarkably this inchoate longing actualized itself in a few decades in a succession of luminous discoveries, and flowed forward into the idea of energy.

In Germany it took shape in a much-praised, much-derided movement, interminable in discourse about Urprinciples and life forces, spearheaded by Lorenz Oken and Friedrich Schelling under the name *Naturphilosophie*, turgid indeed, but source of a powerful new credo. Nature is one, and so must all the sciences be; the true philosopher is one who connects. Thus in 1820 Hans Christian Oersted joined magnetism to electricity with his amazing discovery that an electric current deflects a magnet. Readers of Oersted baffled by strange talk of "polarities" and the "electric conflict" should know that it stems from Schelling: Oersted was a Naturphilosoph. But there were other voices. Similar hopes expressed in a more down-to-earth English tone mark John Herschel's *Preliminary Discourse on the Study of Natural Philosophy* (1830), Mary Somerville's *Connexion of the Physical Sciences* (1839), and W. R. Grove's *On the Correlation of Physical Forces* (1843). Fluent in German and Hanoverian in descent, Herschel has special interest because his "natural philosophy" had so resolutely English a cast, remote from Naturphilosophie. His master was Bacon, yet Herschel sought connection. So above all did the man whose first published paper (1816) was on the composition of caustic lime from Tuscany, and his last (1859) on effects of pressure on the melting point of ice, the profoundly English Michael Faraday.

From worldviews about the connectedness of everything to the correlation of specific "forces" to enunciating the law of energy was a long journey requiring two almost antithetical principles: great boldness in speculation and great exactness in measurement. Two eras may be identified: an era of correlation from the discovery of electrochemistry in 1800 through much of Faraday's career, and the era of energy beginning in the late 1840s. The two are connected: Without correlation there might have been no energy, and with energy new light was thrown on correlations, but they are not the same. In the great correlations of Oersted and Faraday, the discovery, when finally made, often embodied a surprise. An electric current exerts a magnetic force, as

Oersted had expected, but it is a transverse force. A magnet generates an electric current, as Faraday had hoped, but it must be a moving magnet, and the current is transient. A magnetic field affects light traversing glass, as Faraday had surmised, but its action is a twisting one far from his first guess. Not number but vision was the key, an openness to the unexpected. Energy is different. To say that energy of one kind has been transformed into energy of another kind makes sense only if the two are numerically equal.

Between Faraday and Joule, the man after whom very properly the unit of energy, the joule, is named, lay an instructive contrast. Both master experimenters with a feeling for theory, they stood at opposite conceptual poles: geometry versus number. For Faraday, discovery was relational. His ideas about fields of force reveal him, though untrained, as a geometer of high order. His experiments disclose the same spatial sense; number he left mainly to others. "Permeability" and "susceptibility," the terms quantifying his work on magnetism are from Thomson; the quantity characterizing his magnetooptical effect is (after the French physicist Marcel Verdet) Verdet's constant. For Joule, number was supreme. His whole thinking on energy originated in the hope that electric motors run from batteries would outperform steam engines. His best experiments yielded numbers, the electrical and mechanical equivalents of heat. His neatest theoretical idea was to compute one number, the extraordinarily high velocity gas molecules must have if pressure is to be attributed to their impacts on the walls of the container.

The doctrine of energy came in the 1840s to the most diverse minds: to Mayer, physician on a Dutch ship, off Java in 1842; to Joule, brewer turned physicist, at Manchester in 1843; to Colding, engineer and student of Oersted's, at Copenhagen in 1843; to Helmholtz, surgeon in the Prussian Army, at Potsdam in 1847. In a famous text, "Energy Conservation as an Example of Simultaneous Discovery" (1959), Thomas Kuhn listed eight others, including Faraday and Grove, directly concerned with interconvertibility of forces, and several more peripherally so, arguing that all in some sense had a common aim. Against this it is necessary to reemphasize the distinction between correlation and quantification (Faraday, the geometrist, found more "correlations" than any energist) and the desperate confusion about force. In exchanges in 1856 with the young James Clerk Maxwell, Faraday discussed "conservation of force"—meaning what? Not conservation of energy but his geometric intuition of a quite different law, conservation of *flux*, the mathematical result discovered in 1828 by M. V. Ostrogradsky, afterward Teutonized as Gauss's theorem. In 1856, energy—so natural to the twenty-five-year-old Maxwell—was utterly foreign to the sixty-four-year-old master from another age.

Physicists regard energy, and the laws governing it, as among the most basic principles of their science. It may seem odd that of the four founders two were doctors, one an engineer, and one only, Joule (and he only inexactly), could have been called then a physicist. The surprise lessens when we recognize that much of the impetus of discovery came from the steam engine, to which one key lay in that potent term coined and quantified in 1782 by James Watt, "horsepower." Seen in one light, horses and human beings are engines.

WORK, DUTY, POWER, AND THE STEAM ENGINE

The steam engine was invented early in the eighteenth century for pumping water out of mines. Later Watt applied it to machinery, and others to trains and ships, but the first use was in mines, especially the deep tin and copper mines of Cornwall, in southwestern England. Crucial always, to use the eighteenth-century term, was the *duty* of the engine: the amount of water raised through consuming a given quantity of fuel: so many foot-pounds of water per bushel of coal. To the mineowner it was simple. Formerly, pumping had been done by horses harnessed to a rotary mill. Now (making allowance for amortization of investment), which was cheaper: oats for the horses or coal for the engine? The answer was often nicely balanced, depending on how close the metals mine was to a coal mine.

Improving the duty meant using fuel more efficiently, but to rationalize that easy truth was the work of more than a century. Theoretically, it depended, among other things, on the recognition in 1758 by Adam Black—Watt's mentor at Glasgow—of a distinction between heat and temperature, and the recognition by nineteenth-century chemists of an absolute zero of temperature. On the practical side lay a series of brilliant inventions, beginning in 1710 with Thomas Newcomen's first "atmospheric" engine, preceded (and impeded) by a patent of 1698 by Thomas Savery. Eventually two kinds of efficiency were distinguished, thermal and thermodynamic. Thermal efficiency means avoiding unnecessary loss, heat going up the chimney, or heat dissipated in the walls of the cylinder. Thermodynamic efficiency, more subtly, engages a strange fundamental limit. Even in an ideal engine with no losses, not all the heat can be converted into work.

Both Savery's and Newcomen's engines worked by repeatedly filling a space with steam and condensing the steam to create a vacuum. Savery's sucked water straight up into the evacuated vessel; Newcomen's was a beam engine with a pump on one side and a driving cylinder on the other, running typically at ten to fifteen strokes a minute. The horsepower of the earliest of which anything is known was about 5.5. Comprehensively ingenious as this engine was, the vital difference between it and Savery's lay in one point: how the vacuum was created. Savery poured cold water over the outside, which meant that next time hot steam entered, much of it was spent reheating the cylinder. Newcomen sprayed water directly into the steam. The higher thermal efficiency and increased speed of operation made the difference between a useless and a profoundly useful machine.

Newcomen engines continued to operate, especially in collieries, almost to the end of the eighteenth century. Yet they, too, were wasteful. The way was open for a great advance, James Watt's invention in 1768 of the separate "condenser." Next to the driving cylinder stood another cylinder, exhausted and permanently cold. When the moment for the downstroke came, a valve opened, the steam rushed over and condensed; the valve then closed and the cycle was repeated. The driving cylinder stayed permanently hot.

Watt and Matthew Boulton, his business partner, were obsessed with numbers. Testing Clydesdale dray horses, the strongest in Britain, Watt fixed their lifting power (times 1.33) at 33,000 ft-lbs per minute. Thus was horsepower defined. Comparing his first engines with the best Newcomen ones, he found his to have four times higher duty. His next leap was to make the engine double acting and subject to "expansive working." Double action meant admitting steam alternately above and below the piston in a cylinder closed at both ends; for an engine of given size it more than doubled the power. Expansive working was one of those casual-seeming ideas that totally reorder a problem. Steam expands. This being so, it dawned on Watt that the duty might be raised by cutting off the flow into the cylinder early—say, a third of the way through the stroke. Like something for nothing, the steam continued to push without using fuel. Where was the effort coming from? The answer, not obvious at the time, is that as steam expands, it cools. Not pressure alone, but pressure and temperature together enter the lore of engines.

Behind the final clarifying insight provided in 1824 by Sadi Carnot, son of the French engineer-mathematician and revolutionary politician Lazare Carnot, lay three decades of invention by working engineers. One, too little remembered, was a young employee of Boulton and Watt, James Southern. Sometime in the 1790s he devised the indicator diagram to optimize engine performance under expansive working. In this wonderfully clever instrument a pressure gauge (the "indicator") was mounted on the cylinder, and a pen attached by one arm to it and another to the piston rod drew a continuous closed curve of pressure versus volume for engines in actual operation. The area of the curve gave the work (force times distance) done by the engine over each cycle; optimization lay in adjusting the steam cutoff and other factors to maximize this against fuel consumption. Known to most purchasers only by the presence of a mysterious sealed-off port on top of the cylinder—and exorbitant patent royalties for improved engine performance—the indicator was a jealously guarded industrial secret. The engineer John Farey, historian of the steam engine, had heard whispers of it for years before first seeing one in operation in Russia in 1819. To it as background to Carnot must be added three further points: (1) the invention in 1781 by one of Watt's Cornish rivals, J. C. Hornblower, of compounding (two stages of expansion successively in small and large cylinders); (2) the use of high-pressure steam; and (3) the recognition, especially by Robert Stirling, inventor in 1816 of the "Stirling cycle" hot air engine, that the working substance in engines need not be steam. Of these, high pressure had the greatest immediate impact, especially in steam traction.

Because Watt, who had an innate distaste for explosions, insisted on low pressures, his engines had huge cylinders and were quite unsuitable to vehicles. No such craven fears afflicted Richard Trevithick in Cornwall or Oliver Evans in Pennsylvania, each of whom in around 1800 began building high-pressure engines, locomotive and stationary, as a little later did another Cornish engineer, Arthur Woolf. Soon a curious fact emerged. Not only were high-pressure engines more compact, they were often more efficient. Woolf's stationary engines in particular, combining high pressure with compounding, performed outstandingly and were widely exported to France. Their success set Carnot thinking. The higher duty comes not from pressure itself but because high-pressure steam is hotter. There is an analogy with hydraulics. The output of a waterwheel depends on how far the water falls in distance: the output of a heat engine depends on how far the working substance falls in temperature. The ideal would be to use the total drop from the temperature of burning coal—about 1,500°C (2,700°F)—to room temperature. Judged by this standard, even high-pressure steam is woefully short; water at 6 atmospheres (a high pressure then) boils at 160°C (321°F).

Carnot had in effect distinguished thermal and thermodynamic inefficiencies, the latter being a failure to make use of the full available drop in temperature. He clarified this by a thought experiment. In real engines, worked expansively, the entering steam was hotter than the cylinder, and the expanded steam cooler, with a constant wasteful shuttling of heat between the two. Ideally, transfer should be isothermal—that is, with only infinitesimal temperature differences. The thought experiment—put ten years later (1834) into the language of indicator diagrams by the French mining engineer Émil Clapeyron—comprised four successive operations on a fixed substance. In two it received or gave up heat isothermally; in the other two it did or received work with no heat transfer. Neither Carnot nor Clapeyron could quantify thermodynamic efficiency; Thomson in 1848 could. He saw that the Carnot cycle would define a temperature scale (now known as the Kelvin scale) identical with the one chemists had deduced from gas laws. For an ideal engine working between temperatures T_1 (cold) and T_2 (hot) measured from absolute zero, the thermodynamic efficiency is $(T_2 - T_1)/T_2$. No engine can surpass that.

Nothing in nature is lost; that was the message of the 1840s. Chemical, mechanical, and electrical processes are quantifiable, exact, obey a strict law of energy. Equally exact are the amounts of heat these processes generate. But the reverse does not hold. The history of steam was one long struggle against loss; not even Carnot's engine can extract all the energy the fuel has to give. This was the conceptual paradox three men, Thomson and Rankine in Scotland, and Rudolf Clausius in Germany, faced in 1849–1850 in creating the new science of thermodynamics. Out of it has arisen a verbal paradox. For physicists conservation of energy is a fact, the first law of thermodynamics. For ordinary people (including off-duty physicists), to conserve energy is a moral imperative.

Heat as energy is somehow an energy unlike all the rest. Defining how takes a second law, a law so special that Clausius and Thomson found two opposite ways of framing it. To Clausius the issue was the age-old one of perpetual motion. With two Carnot engines coupled in opposition, the rule forbidding perpetual motion translated into a rule about heat. His second law of thermodynamics was that heat cannot flow without work uphill from a cold body to a hot body. Refrigerators need power to stay cold. Thomson's guide was Joseph Fourier's great treatise on heat *Théorie analytique de la chaleur* (1827), and the fact that unequally heated bodies, left to themselves, come always to equilibrium. His version was that work has to be expended to maintain bodies above the temperature of their surroundings. Animals need food to stay warm.

In a paper with the doom-laden title "On a Universal Tendency in Nature to a Dissipation of Mechanical Energy," Thomson in 1852 laid out the new law's dire consequences. How oddly did the 1850s bring two clashing scientific faiths: Darwin's, that evolution makes everything better and better; and Thomson's, that thermodynamics makes everything worse and worse. Thirteen years later Clausius rephrased it in the aphorism "The energy of the world is constant, the entropy of the world strives toward a maximum." After Maxwell in 1870 had invented his delightful (but imaginary) demon to defeat entropy, and Ludwig Boltzmann in 1877 had related it formally to molecular disorder through the equation which we now, following Max Planck, write $S = k \log W$, the gloom only increased. With it came a deep physics problem, first studied by Maxwell in his earliest account of the demon but first published by Thomson in 1874, the "reversibility paradox." The second law is a consequence of molecular collisions, which obey the laws of ordinary dynamics—laws reversible in time. Why then is entropy not also reversible? How can reversible laws have irreversible effects? The argument is sometimes inverted in hope of explaining another, even deeper mystery, the arrow of time. Why can one move forward and backward in space but only forward in time? The case is not proven. Most of the supposed explanations conceal within their assumptions the very arrow they seek to explain.

In the history of heat engines from Newcomen's on, engineering intuition is gradually advancing scientific knowledge. The first engines strictly designed from thermodynamic principles were Rudolf Diesel's, patented in 1892. With a compression ratio of fifty to one, and a temperature drop on reexpansion of 1,200°C (~2,160°F), his were among the most efficient ever built, converting 34 percent of the chemical energy of the fuel into mechanical work.

SOLIDIFICATION OF THE VOCABULARY OF ENERGY

Two concepts entangling mass and velocity made chaos in eighteenth-century dynamics, Newton's *vis motrix* mv and Leibniz's *vis viva* mv^2 (now written ½ mv^2, as first proposed by Gaspard Coriolis in 1829). It was a battle of concepts but also a battle of terms. Then casually in 1807 Thomas Young in his *Course of Lectures on Natural Philosophy and the Mechanical Arts*, dropped this remark: "The term energy may be applied with great propriety to the product of the mass or weight of a body times the number expressing the square of its velocity." Proper or not, *energy* might have fallen still born in science had not John Farey also, with due praise of Young, twice used it in 1827 in his *History of the Steam Engine* for *vis viva*. There, dormant, the word lay, four uses in two books, until 1849, when Thomson in last-minute doubt added this footnote to his second paper on Carnot's theorem, "An Account of Carnot's Theory on the Motive Power of Heat": "When thermal agency is spent in conducting heat through a solid what becomes of the mechanical effect it might produce? Nothing can be lost in the operation of nature—no energy can be destroyed. What then is produced in place of the mechanical effect that is lost?"

In that superb instance of Thomson's genius for producing "science-forming" ideas (the phrase is Maxwell's), thermodynamics is crystallized and energy is given almost its modern meaning. His next paper, "On the Quantities of Mechanical Energy Contained in a Fluid in Different States as to Temperature and Density" (1851), elevated energy to the title and contained the first attempt to define it within the new framework of thought. The next after was the sad one on dissipation. The torch then passed to Rankine, who in two exceptionally able papers, "On the General Law of the Transformation of Energy" (1853) and "Outlines of the Science of Energetics" (1855), completed the main work. The first introduced the term (not the concept) "conservation of energy," together with "actual or sensible energy" for *vis viva*, and "potential or latent energy" for energy of configuration. "Potential energy" has stuck; "actual energy" was renamed "kinetic energy" by Thomson and P. G. Tait in their *Treatise on Natural Philosophy* (1867); their definition (vol. 1, sec. 213) nicely explains the Coriolis factor: "The *Vis Viva or Kinetic Energy* of a moving body is proportional to the mass and the square of the velocity, conjointly. If we adopt the same units for mass and velocity as before there is particular advantage in defining kinetic energy as half the product of the mass and the square of the velocity."

A paper in French by Thomson, "Mémoire sur l'energie mécanique du système solaire" (1854), was the first use in a language other than English; but it was Clausius who by adapting the German *Energie* to the new scientific sense assured its future. He met much opposition from Mayer and some from Helmholtz, both of whom favored Kraft. Mayer indeed held fiercely that Kraft or force in the sense "forces of nature" had conceptual priority, so that Newton's use of "force" should be disowned. More was at stake here than words. The dispute goes back to the three meanings of "force." For Mayer, conservation and correlation were two sides of the same coin, and the single word "Kraft" preserved the connection. For advocates of "energy," "Kraft/force" in this general sense was a vague term with no quantitative significance. "Energy" was so precise, explicit, and quantitative a concept that it required a term to itself.

Beyond the expressions "conservation of energy" and "potential energy," and the notion and name of a science of energetics (revived in the 1890s by Wilhelm

Ostwald), Rankine gave far more in concept and term than he receives credit for. He supplied the first theory of the steam engine, the important term "adiabatic" to denote processes in which work is done with no transfer of heat, and—by a curious accident—the name "thermodynamics" for the science. In 1849 Thomson, seeking a generic name for machines that convert heat into work, had used "thermo-dynamic engine." Rankine followed suit, introducing in 1853 the vital concept of a thermo-dynamic function. Then in 1854 he submitted to the Philosophical Transactions of the Royal Society a long paper, "On the Geometrical Representation of the Expansive Action of Heat and the Theory of Thermo-Dynamic Engines." Papers there carried a running head, author's name to the left, title to the right. Faced with Rankine's lengthy title, the compositor set "Mr. Macquorn Rankine" on the left and "On Thermo-Dynamics" on the right. So apt did this seem that "thermodynamics" (without the hyphen) quickly became the secure name of the science. Later critics, unaware of its roots, have accused the inventors of thermodynamics of muddled thinking about dynamics; in fact, this and the history of Rankine's function provide one more instance of the power and subtlety of words. The "thermo-dynamic function" is what we, after Clausius who rediscovered it and gave it in 1865 its modern name, call entropy. In its original setting as the mathematical function proper to the efficient functioning of engines, Rankine's name was a good one. In the long view, however, Clausius's name for it, formed by analogy with energy, was better, even though the etymological reading he based it on (energy = work content, entropy = transformation content) was false. It was shorter and catchier but, more important, stressed both the parallel and the difference between the two laws of thermo-dynamics.

In 1861, at the Manchester meeting of the British Association for the Advancement of Science, two engineers, Charles Bright and Latimer Clark, read a paper "On the Formation of Standards of Electrical Quantity and Resistance," which had an influence on science and everyday life out of all proportion to the thirteen lines it takes in the Association Report. Its aim was simple. As practical telegraphers, Bright and Clark wished to attain internationally agreed definitions, values, and names for working quantities in their field. Their proposal led to the setting up under Thomson of a famous British Association committee to determine—from energy principles—an absolute unit of electrical resistance. Another, and fascinating, topic was Clark's suggestion made verbally at Manchester that units should be named after eminent scientists. He proposed "volt," "ohm," and "ampere" for units of potential, resistance, and current, though on ampere there was dispute, for Thomson's committee preferred to name the unit of current for the German electrician Weber.

International blessing and the setting of many names came at the Paris Congress of Electricians in 1881, but with no choice yet for units of energy and power. A year later C. William Siemens, the Anglicized younger brother of Ernst Werner von Siemens, was president of the British Association. In his address reported in the British Association report in 1882 he proposed "watt" and "joule"—the one "in honour of that master mind in mechanical science, James Watt … who first had a clear physical conception of power and gave a rational method of measuring it," the other "after the man who has done so much to develop the dynamical theory of heat." According to the *Athenaeum* in 1882, Siemens's suggestions were "unanimously approved" by the Physics Section of the British Association a few days later. Joule, who was sixty-four at the time and had another seven years to live, seems to have been the only man so commemorated during his lifetime.

QUANTIZATION, E = MC^2 AND EINSTEIN

Accompanying the new nineteenth-century vista on heat were insights into its relation to matter and radiation that opened a fresh crisis. Gases, according to the "kinetic" theory, comprise large numbers of rapidly moving colliding molecules. As remade by Clausius and Maxwell in the 1850s it was an attractive—indeed, a correct—picture; then something went horribly wrong. From the Newtonian mechanics of colliding bodies, Maxwell deduced a statistical theorem—later called the equipartition theorem—flatly contrary to experiment. Two numbers, 1.408 for the measured value of a certain ratio, 1.333 for its calculated value, failed to agree. So upset was he that in a lecture at Oxford in 1860 he said that this one discrepancy "overturned the whole theory." That was too extreme, for the kinetic theory, and the wider science of statistical mechanics, continued to flourish and bring new discoveries, but there was truth in it. Equipartition of energy ran on, an ever-worsening riddle, until finally in 1900 Max Planck, by focusing not on molecules but on radiation, found the answer. Energy is discontinuous. It comes in quantized packets $E = h\nu$, where h is a constant ("Planck's constant") and n the frequency of radiation. Expressed in joule-seconds, h has the exceedingly small value $6.67\text{x}10^{-34}$. Quantum effects are chiefly manifested at or below the atomic level.

The forty years from equipartition to Planck deserve more thought than they ordinarily get. The theorem itself is one hard to match for sheer mathematical surprise. It concerns statistical averages within dynamical systems. As first put by Maxwell, it had two parts: (1) if two or more sets of distinct molecules constantly collide, the mean translational energies of the individual ones in each set will be equal, (2) if the collisions produce rotation, the mean translational *and rotational* energies will also be equal. This is equipartition. Mathematically, the result seemed impregnable; the snag was this. Gases have two specific heats, one at constant pressure, the other at constant volume. Calculating their ratio C_p/C_v from the new formula was easy; by 1859 the value was already well known experimentally, having been central to both acoustics and Carnot's theory. Data and hypothesis were excellent; only the outcome was wrong.

Worse was to come. Boltzmann in 1872 made the same weird statistical equality hold for *every* mode in a dynamical system. It must, for example, apply to any internal motions that molecules might have. Assuming, as most physicists did by then, that the sharp lines seen in the spectra of chemical elements originate in just such internal motions, any calculation now of C_p/C_v would yield a figure even lower than 1.333. Worse yet, as Maxwell shatteringly remarked to one student, equipartition must apply to solids and liquids as well as gases: "Boltzmann has proved too much."

Why, given the scope of the calamity, was insight so long coming? The answer is that the evidence as it stood was all negative. Progress requires, data sufficiently rich and cohesive to give shape to theoretical musings. Hints came in 1887 when the Russian physicist V. A. Michelson began applying equipartition to a new and different branch of physical inquiry, the radiation emitted from a heated "black body"; but even there, the data were too sparse. Only through an elaborate interplay of measurement and speculation could a crisis so profound be settled.

Gradually it was, as Wilhelm Wien and others determined the form of the radiation, and Planck guessed and calculated his way to the new law. Found, it was like a magic key. Einstein with dazzling originality applied it to the photoelectric effect. Bohr, after Rutherford's discovery of the nucleus, used it to unlock the secrets of the atom. The worry about gases evaporated, for quantization limits which modes of energy affect specific heats. As for that later worry of Maxwell's about solids, trouble became triumph, a meeting ground between theory and experiment. Among many advances in the art of experimentation after 1870, one especially was cryogenics. One by one all the gases were liquefied, culminating with hydrogen in 1898 and helium in 1908, respectively at 19.2 K and 4.2 K above absolute zero. Hence came accurate measurements near zero of the specific heats of solids, started by James Dewar. They fell off rapidly with temperature. It was Einstein in 1907 who first saw why—because the modes of vibrational energy in solids are quantized. More exact calculations by Peter Debye, and by Max Born and Theodore von Kármán, both in 1912, made C at low temperatures vary as T^3. It was a result at once beautifully confirmed by F. A. Lindemann (Lord Cherwell) in experiments with Walther Nernst in Berlin.

From Planck's h to Bohr's atom to the austere reorganization of thought required in the 1920s to create quantum mechanics is a drama in itself. Crucial to its denouement was Heisenberg's "uncertainty principle," a natural limit first on knowing simultaneously the position x and momentum p of a body ($\Delta x \, \Delta p \sim h$) and second, no less profound, on energy. The energy E of a dynamical system can never be known exactly: measured over a time interval Δt it has an uncertainty $\Delta E \sim h/\Delta t$. But whether quantum uncertainty is an epistemological truth (a limit on knowledge) or an ontological one (a limit in nature) is an endless, uncertain debate.

If the $E = h\nu$ of quantum theory is one beguilingly simple later advance in the concept of energy, the supremely famous mass-energy law $E = mc^2$ is another. This and its import for nuclear energy are usually credited solely to Einstein. The truth is more interesting.

It begins with J. J. Thomson in 1881 calculating the motion of an electric charge on Maxwell's electromagnetic theory, a theme Maxwell had barely touched. To the charge's ordinary mass Thomson found it necessary to add an "electromagnetic mass" connected with electrical energy stored in the surrounding field. It was like a ship moving through water—some of the water is dragged along, adding to the ship's apparent mass. Had the calculation been perfect, Thomson would have found $E = mc^2$, as Einstein and others did years later. As it was, Thomson's formula yields after a trivial conversion the near-miss $E = \frac{3}{4} mc^2$.

After Thomson nothing particular transpired, until a number of people, of whom Joseph Larmor in 1895 was perhaps first, entertained what historians now call the "electromagnetic worldview." Instead of basing electromagnetism on dynamics, as many had tried, could one base dynamics on electromagnetism? How far this thought would have gone is unclear had it not been for another even greater leap of Thomson's, this time experimental: his discovery of the electron. In 1897, by combining measurements of two kinds, he proved that cathode rays must consist of rapidly moving charged objects with a mass roughly one one-thousandth of the hydrogen atom's. Here was revolution—the first subatomic particle. Hardly less revolutionary was his conjecture that its mass was *all* electromagnetic. A shadow ship plowing through an electromagnetic sea: That was Thomson's vision of the electron.

The argument then took two forms: one electromagnetic, the other based on radioactivity. The first was tied to a prediction of Maxwell's in 1865, proved experimentally in 1900 by Petr Lebedev, that light and other electromagnetic radiations exert pressure. This was the line Einstein would follow, but it was Henri Poincaré who, in 1900 at a widely publicized meeting in honor of H. A. Lorentz, first got the point. The Maxwell pressure corresponds to a momentum, given by the density of electromagnetic energy in space divided by the velocity c of light. Momentum is mass times velocity. Unite the two and it is natural to think that a volume of space containing radiant energy E has associated with it a mass $m = E/c^2$. Meanwhile, as if from nowhere, had come radioactivity—in Rutherford's phrase, "a Tom Tiddler's ground" where anything might turn up. In 1903 Pierre Curie and his student Albert Laborde discovered that a speck of radium placed in a calorimeter emitted a continuous flux of heat sufficient to maintain it at 1.5°C above the surroundings. Their data translated into a power output of 100W/kg. Where was the power coming from? Was this finally an illimitable source of energy, nature's own perpetual-motion machine? It was here that Einstein, in a concise phrase, carried the argument to its limit. In his own derivation of $E = mc^2$ in 1905, five years after Poincaré's observation, he remarked that any body emitting radiation should lose weight.

The events of World War II, and that side of Einstein that made credit for scientific ideas in his universe a thing more blessed to receive than to give, have created a belief that atomic energy somehow began with him. In truth the student of early twentieth-century literature notices how soon after radioactivity there arose among people who knew nothing of $E = mc^2$ or Einstein a conviction that here was a mighty new source of energy. Geologists hailed it with relief, for as the Irish physicist John Joly demonstrated in 1903, it provided an escape from Kelvin's devastating thermodynamic limit on the age of Earth. But not only they, and not always with relief. In Jack London's nightmare futurist vision *The Iron Heel*, published in 1908, atomic gloom permeates the plot. No less interesting is this, from a 1907 work London may have read that remains among the most illuminating popularizations of the events of that era, R. K. Duncan's *The New Knowledge*: "March, 1903, … was a date to which, in all probability, the men of the future will after refer as the veritable beginning of the larger powers and energies that they will control. It was in March, 1903, that Curie and Laborde announced the heat-emitting power of radium. The fact was simple of demonstration and unquestionable. They discovered that a radium compound continuously emits heat without combustion, or change in its molecular structure.… It is all just as surprising as though Curie had discovered a red-hot stove which required no fuel to maintain it on heat." But there *was* fuel: the fuel was mass.

The highest (and most puzzling) of Einstein's offerings is not $E = mc^2$, where his claim is short, but the spectacular reinterpretation of gravitational energy in the new theory he advanced in 1915 under the name general relativity. Owing much to a brilliant paper of 1908 by his former teacher Hermann Minkowski, this, of course, is not the popular fable that everything is relative. It is a theory of relations in which two paired quantities—mass-energy and space-time—are each joined through the velocity of light mass-energy via $E = mc^2$, space-time via a theorem akin to Pythagoras's where time multiplied by c enters as a fourth dimension, not additively but by subtraction $-(ct)^2$. When in this context Einstein came to develop a theory of gravity to replace Newton's, he met two surprises, both due ultimately to the curious—indeed deeply mysterious—fact that in gravitation, unlike any other force of nature, mass enters in two ways. Newton's law of acceleration $F = ma$ makes it the receptacle of inertia; his law of gravity $F = GMm/r^2$ makes it the origin of the force. If, therefore, as with electromagnetism, the energy is disseminated through space, it will have the extraordinary effect of making gravitation a source for itself. Consider Earth. Its mass is 6×10^{21} tons; the mass of the gravitational field around it, computed from $E = mc^2$, though tiny by comparison, has the by-human-standards enormous value of 4×10^{10} tons. It, too, exerts an attraction. For denser bodies this "mass of the field" is far greater. Because of it any relativistic theory of gravitation has to be nonlinear.

Even deeper, drawn from his much-discussed but too little understood "falling elevator" argument, was Einstein's substituting for Newton's equivalence between two kinds of mass an equivalence between two kinds of acceleration. How innocent it seems and how stunning was the transformation it led to. The theory could be rewritten in terms not of mass-energy variables but of space-time variables. Gravitation could be represented as warping space-time—in just the right nonlinear way. For this one form, energy had been transmuted into geometry. But was the geometrization limited to gravity, or might it extend to energy of every kind, beginning with electromagnetic? That great failed vision of the geometrization of physics would charm and frustrate Einstein and many others for the next forty years, and it retains its appeal.

Looking back over more than a century since energy began as concept and term, it is impossible not to be struck both by the magnitude of the advance and the strenuousness of the puzzles that remain. In the design of engines, the practice of quantum mechanics, the understanding of elementary particles, the application of relativity, detailed prescriptions and laws exist that work perfectly over the range of need. Beyond lies doubt. To physicists with their conviction that beauty and simplicity are the keys to theory, it is distressing that two of the most beautifully simple conceptions proposed in physics, J. J. Thomson's that all mass is electromagnetic, Einstein's that all energy is geometric, both manifestly verging on truth, have both failed, with nothing either as simple or as beautiful arising since to take their place. Then there are issues such as quantum uncertainty and the problems of dissipation, reversibility, and the arrow of time, where to a certain depth everything is clear, but deeper (and the depth seems inescapable), everything is murky and obscure—questions upon which physicists and philosophers, such as Milton's theologians, sit apart in high debate and find no end in wandering mazes lost.

C. W. F. Everitt

BIBLIOGRAPHY

Born, M. (1951). *Atomic Physics*, 5th ed. New York: Hafner.

Brush, S. G. (1976). *The Kind of Motion We Call Heat.* Amsterdam: North Holland.

Brush, S. G. (1983). *Statistical Physics and the Atomic Theory of Matter.* Princeton, NJ: Princeton University Press.

Cantor, G. (1991). *Michael Faraday: Sandemanian and Scientist: A Study of Science and Religion in the Nineteenth Century.* London: Macmillan.

Cardwell, D. S. L. (1971). *From Watt to Clasuius.* Ithaca, NY: Cornell University Press.

Cardwell, D. S. L. (1989). *James P. Joule.* Manchester, Eng.: Manchester University Press.

Carnot, S. (1960). *Reflections on the Motive Power of Fire by Sadi Carnot; and Other Papers on the Second Law of Thermodynamics, by E. Clapeyron and R. Clausius,* ed. E. Mendoza. New York: Dover.

Duncan, R. K. (1907). *The New Knowledge: A Popular Account of the New Physics and the New Chemistry in Their Relation to the New Theory of Matter.* London: Hodder and Stoughton.

Elkana, Y. (1974). *The Discovery of the Conservation of Energy.* London: Hutchinson Educational.

Everitt, C. W. F. (1975). *James Clerk Maxwell: Physicist and Natural Philosopher.* New York: Charles Scribner's Sons.

Ewing, J. A. (1894). *The Steam-Engine and Other Heat-Engines.* Cambridge, Eng.: Cambridge University Press.

Farey, J. (1827). *A Treatise on the Steam Engine.* London: Longman, Rees, Orme, Brown, and Green.

Harman, P. (1998). *The Natural Philosophy of James Clerk Maxwell.* New York: Cambridge University Press.

Jammer, M. (1957). *Concepts of Force.* Cambridge: Harvard University Press.

Joly, J. (1909). *Radioactivity and Geology: An Account of the Influence of Radioactive Energy on Terrestrial History.* London: A. Constable.

Kuhn, T. S. (1977). "Energy Conservation as an Example of Simultaneous Discovery." In *The Essential Tension.* Chicago: University of Chicago Press.

Kuhn, T. S. (1978). *Black-Body Theory and the Quantum Discontinuity, 1894–1912.* New York: Oxford University Press.

Lorentz, H. A.; Einstein A.; Minkowski, H.; Weyl, H. (1923). *The Principle of Relativity: A Collection of Original Memoirs on the Special and General Theory of Relativity, with Notes by A. Sommerfeld.* London: Dover.

Miller, A. I. (1987). "A précis of Edmund Whittaker's 'The Relativity Theory of Poincaré and Lorentz.'" *Archives Internationales d'Histoire de Science* 37:93–103.

Pais, A. (1982). *Subtle Is the Lord: The Science and the Life of Albert Einstein.* New York: Oxford University Press.

Rankine, W. J. M. (1881). "On the General Law of the Transformation of Energy." In *Miscellaneous Scientific Papers: From the Transactions and Proceedings of the Royal and Other Scientific and Philosophical Societies, and the Scientific Journals; with a Memoir of the Author* by P. G. Tait, ed. W. J. Millar. London: C. Griffin.

Rankine, W. J. M. (1881). "Outlines of the Science of Energetics." In *Miscellaneous Scientific Papers: From the Transactions and Proceedings of the Royal and Other Scientific and Philosophical Societies, and the Scientific Journals; with a Memoir of the Author by P. G. Tait*, ed. W. J. Millar. London: C. Griffin.

Smith, C., and Wise, N. (1889). *Energy and Empire : A Biographical Study of Lord Kelvin.* New York: Cambridge University Press.

Thomson, W. (1882–1911). "An Account of Carnot's Theory on the Motive Power of Heat." In *Mathematical and Physical Papers: Collected from Different Scientific Periodicals from May 1841 to the Present Time, in 6 Vols.*, vol. 1, pp. 156–164. Cambridge, Eng.: Cambridge University Press.

Thomson, W. (1882–1911). "On a Universal Tendency in Nature to the Dissipation of Mechanical Energy." In *Mathematical and Physical Papers: Collected from Different Scientific Periodicals from May 1841 to the Present Time, in 6 Vols.*, vol. 1, pp. 511–514. Cambridge, Eng.: Cambridge University Press.

Thomson, W. and Tait, P. G. (1888–1890). *Treatise on Natural Philosophy.* Cambridge, Eng.: Cambridge University Press.

Truesdell, C. (1980). *The Tragicomical History of Thermodynamics, 1822–1854.* New York : Springer-Verlag.

Whittaker, E. T. (1951–1953). *A History of the Theories of Aether and Electricity*, rev. ed. New York: T. Nelson.

Williams, L. P. (1965). *Michael Faraday.* New York: Basic Books.

SEEBECK, THOMAS JOHANN (1770–1831)

Thomas Johann Seebeck, German physician, physicist, and chemist, was born April 9, 1770, in the Hanseatic town of Revel (now Tallinn Estonia). Seebeck studied medicine in Berlin and Goettingen, where in 1802 he took a physician's doctor degree. He lived in Jena (Thuringia) between 1802 and 1810 as an independent scholar not practicing medicine. In the Jena period Seebeck met poet Johann Wolfgang Goethe, who carried on his own studies on chromatics. From 1806 untill 1818 Seebeck consulted and informed Goethe about his optical observations. He discovered entoptical colors in 1813. These are color figures which originate inside glass volume after cooling. The phenomenon of entoptical colors was broadly documented by Goethe in his treatises on theory of colors. Seebeck became a member of the Academy of Sciences in Berlin in 1814. After stays in Bayreuth and Nuernberg (Bavaria) he moved to Berlin in 1818. Goethe hoped that Seebeck, in the prestigious position of an academician, would help to lend respectability to his theory of colors. Instead, the contact between the two men loosened in the 1820s, as Seebeck seems to have distanced himself cautiously from Goethe's theory. In Berlin Seebeck devoted most of his time to experiments measuring the influence of heat on magnetism and electricity. Seebeck died December 10, 1831, in Berlin.

Seebeck's outstanding scientific achievement was the discovery of one of the three classical thermoelectric effects, which are the Seebeck, the Peltier, and the Thomson effects. Seebeck's discovery was the first, dating from 1822–1823, followed by that of Jean-Charles-Athanase Peltier in 1832 and that of William Thomson in 1854. Seebeck observed that an electric current in a closed circuit comprised different metallic components if he heated the junctions of the components to different temperatures. He noted that the effect increases linearly with the applied temperature difference and that it crucially depends on the choice of materials. Seebeck tested most of the available metallic materials for thermoelectricity. His studies were further systematized by the French physicist

Antoine-Cesar Becquerel (1788–1878). It has become common after Becquerel to classify materials according to their so-called thermoelectric power. The difference in thermoelectric power of the applied materials determines the strength of thermoelectricity. Since the Seebeck effect can only be observed if two dissimilar conductors are employed, one obtains therewith only access to differences of values of thermoelectric power. Absolute values can be ascertained through the Thomson effect, which states that in a single homogeneous material heat develops (or is absorbed) in case electric current flows parallel (or antiparallel) to a temperature gradient. The strength of this effect is expressed by the Thomson coefficient, which is a temperature-dependent quantity varying from material to material. The absolute value of the thermoelectric power of a material is given by the temperature integral of the Thomson coefficient multiplied by the inverse temperature.

Seebeck understood that his effect might be used for precision measurements of temperature differences, and indeed it is exploited for this purpose in modern thermoelements. The basic device in these thermoelements is an electric circuit comprising two metallic components of different thermoelectric power, a thermocouple. The temperature difference to which the thermocouple is exposed is measured through the voltage built up in the circuit, according to Seebeck. Thermoelements with various combinations of materials for their thermocouples are used in the temperature range from –200°C to 3,000°C. their simplicity and robustness makethem an almost universal element in the fields of temperature recording and temperature regulation. In addition to their higher precision as compared to conventional thermometers, these thermoelements have the additional advantage that they can be introduced into apertures fitting to the thinnest technologically realizable wires. Thermoelements also have a much smaller heat capacity than conventional thermometers (e.g. mercury thermometers) and therefore interfere much less with the body to be measured.

Another application of the Seebeck effect is to be found in detectors of small quantities of heat radiation. These sensitive detectors comprise a thermopile, a pile of thermocouples (small pieces of two different metals connected in V form and put into series). Half of the junctions of the thermopile are shielded within the detector, whereas the other half are exposed to external heat radiation which is recorded through a voltmeter in the thermopile circuit.

Barabara Flume-Gorczyca

BIBLIOGRAPHY

Goethe, J. W. (1949). *Naturwissenschaftliche Schriften* Vol. 16. Zurich: Artemis.

Hecker, M. (1924). *Jahrbuch der Goethe-Gesellschaft* 178. Weimar.

Seebeck, T. J. (1822). *Denkschriften der Berliner Akademie*.

Seebeck, T. J. (1826). *Poggendorff's Annalen* VI:133.

Streit, H. (1901–1902). *Programme des Progymnasiums zu Schlawe*.

SEISMIC ENERGY

Seismic energy is energy in the form of elastic waves that are propagated through the earth's interior or along its surface, or have the form of standing waves (free oscillations) in the planet as a whole. The source of seismic waves may be either natural—an earthquake—or artificial, for example, a chemical or nuclear explosion, or a weight dropped on the earth's surface. The energy source, whatever it is, imposes a strain—a change in shape and volume—on the surrounding earth material, which resists the strain because of its elastic properties. As the strained material tries to recover its original shape and volume, oscillations are produced, which are transferred to material slightly farther away from the source. In this way, seismic waves are produced that spread out from the source in all directions. In detail, the mechanics of seismic sources can be quite complex.

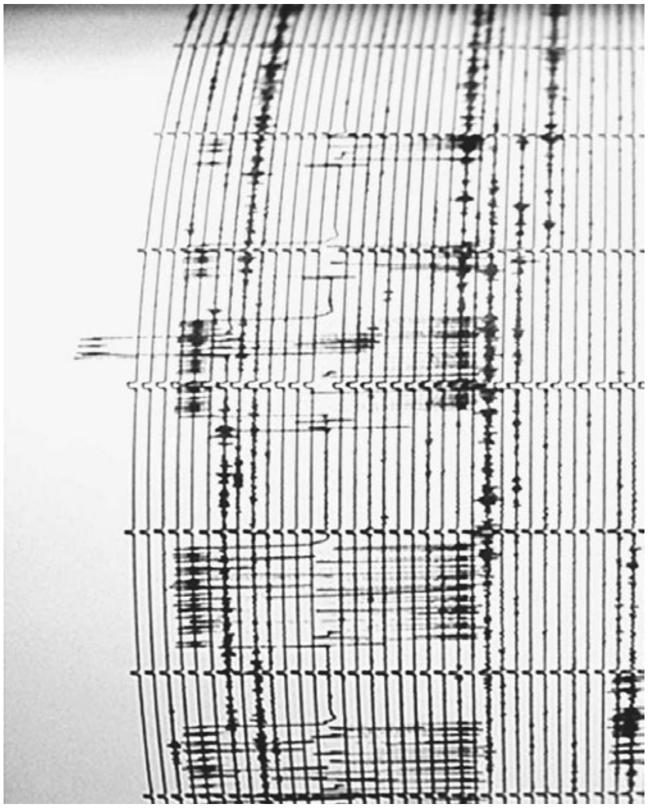
Several types of seismic waves are recognized. Longitudinal waves, also called compressional waves or P waves, consist of oscillations back and forth in the direction of wave propagation. Transverse waves, also called shear waves or S waves, consist of oscillations at right angles to the direction of wave propagation. Because S waves depend upon the rigidity of materials for their existence, they cannot propagate through a fluid medium. L waves, so-called from their long wavelengths, propagate along the free surface of the earth, or within layers near the surface. They are called surface waves, in contrast to the body waves (P and S waves) that travel through the interior body of the earth. Free oscillations of the earth, which exist in several modes, are standing seismic waves produced by very large earthquakes.

EARTHQUAKES

Although earthquakes may be caused by natural events such as rockfalls or the movement of magma in volcanic regions, most are caused by a sudden release of energy at a fault. Before being released in the forms of heat and seismic waves, this energy was stored as elastic strain—that is, recoverable distortions of the size or shape of a volume of rock. Strain develops in response to tectonic forces acting in the lithosphere, the relatively rigid uppermost 100 kilometers (60 miles) or so of the earth. This may be likened to the energy stored in a stretched rubber band, which then is released when the band snaps. The ultimate source of the tectonic forces, and therefore of the energy released by earthquakes, is to be sought in the heat of the deeper portions of the earth. This heat is thought to be a combination of heat produced by natural radioactivity, of residual heat from the earth's accretion, and of heat generated by the separation of iron into the earth's core.

At some point during the accumulation of strain, a certain volume of rock suddenly lurches back toward its original, unstrained condition. The level of stress, which has accumulated along with the strain, decreases rapidly, the rock slips a certain distance along a fracture plane (fault), and seismic waves propagate away from the source region. At sufficiently large distances, the source appears to be a single point within the earth, called the hypocenter (or focus) of the earthquake. The epicenter is a point on the earth's surface located directly above the hypocenter.

The total energy of the earthquake is partitioned into heat and seismic-wave energy. The latter has been estimated to account for only one-sixth to one-quarter of the total energy for tectonic earthquakes, and much less—perhaps only 1 percent—for rockfalls and explosions. The energy of the seismic waves is responsible for the destruction wrought by large earthquakes, either directly through ground shaking, or indirectly by inducing ground failure. The size of a tectonic earthquake is a function of the area of the fault that has slipped, the average amount of slip, and the rigidity of the rock in the vicinity of the fault. The product of these three factors is called the moment of the earthquake, and is related to the more familiar concept of earthquake magnitude.



Seismograph recording seismic activity of Volcano Island, Philippines. (Corbis Corporation)

Earthquake magnitude is a term coined in the 1930s by Charles Richter, a seismologist at the California Institute of Technology. Richter based his original magnitude scale on the amplitudes of seismic waves as recorded on a common type of seismograph then in use in southern California, after correction for the distance to the epicenter. Subsequently, many other methods of measuring earthquake magnitude were developed, not all of which gave results that closely coincided with those of Richter's original method. Today the method that gives the most consistent results across the entire range of magnitudes is based on the earthquake moment, which can be found by analyzing the wavelengths and amplitudes of the seismic waves. Moment is then converted into a magnitude number, known as moment magnitude. On this scale, the very largest earthquakes have magnitudes about 9.5, while the smallest have values of magnitude that are negative. In principle, seismic-wave energy and earthquake moment are directly proportional: $E = AM_0$. In this equation E is energy, M_0 is moment, and A is a constant with an approximate value of

Workers conduct seismic explosions in search of oil and gas deposits on private islands in Bob Marshall Wilderness Area, Montana. (Corbis Corporation)

5×10^{-5} if E and M_0 are expressed in the same units. For example, the moment magnitude of the great Alaska earthquake of 1964 has been estimated at 9.2. This corresponds to a seismic moment of nearly 10^{23} joules, or seismic-wave energy of 5×10^{18} joules. The energy generated by this one earthquake was ten times greater than the amount of seismic-wave energy generated in the entire earth in an average year, which is estimated to be 5×10^{17} joules. That is why high-magnitude earthquakes are the most destructive of natural phenomena. Each whole step on the moment magnitude scale corresponds to an increase of about thirty-two times in seismic-wave energy. Seismologists cannot yet predict when an earthquake of a given size will occur on a given segment of a given fault, but many years of observation have revealed that the frequency of occurrence of earthquakes decreases by about a factor of ten for every whole step increase in magnitude. Thus, in a typical year southern California might experience about a hundred earthquakes in the magnitude range 3 to 4, about ten in the range 4 to 5, and perhaps one of magnitude 5 or greater.

USE OF SEISMIC ENERGY

Since about the beginning of the twentieth century, seismologists have made use of natural seismic waves to investigate the internal structure of the earth. Because of changes in the density and elasticity of earth materials with depth and, to a lesser extent, horizontally, seismic waves are refracted, reflected, and dispersed in fairly complex ways as they propagate through and around the earth. Analysis of free oscillations also provides valuable information about the state of the earth's interior. It is primarily through these techniques that we know that the earth is subdivided into a crust, mantle, outer core, and inner core, as well as details about each of these internal regions.

Artificially generated seismic waves also may be used to explore below the earth's surface, for either scientific or commercial purposes. Seismic energy produced by explosions, weight drops, vibrators, or, at sea, by electrical sparkers and air guns is used extensively to explore for geologic structures that may contain petroleum. The most common method makes use of P waves that are recorded as they arrive back at the earth's surface after having been reflected from boundaries separating rocks of differing properties. Intensive computer-assisted processing converts the raw data so collected into a detailed picture of the subsurface region. Similar techniques are used on a smaller scale in civil engineering projects to find the depth to bedrock beneath soil cover, to investigate the quality of the bedrock, or to find the depth to a groundwater aquifer.

Charles K. Scharnberger

See also: Energy Intensity Trends; Geography and Energy Use; Geothermal Energy; Thermal Energy.

BIBLIOGRAPHY

Bolt, B. A. (1982). *Inside the Earth*. San Francisco: W. H. Freeman.

Bolt, B. A. (1993). *Earthquakes*. New York: W. H. Freeman.

Howell, B. F., Jr. (1990). *An Introduction to Seismological Research*. Cambridge, Eng.: Cambridge University Press.

Kearey, P., and Brooks, M. (1991). *An Introduction to Geophysical Exploration*. Oxford, Eng.: Blackwell Scientific Publications.

Richter, C. F. (1958). *Elementary Seismology*. San Francisco: W. H. Freeman.

SEMICONDUCTORS

See: Electricity

SHIPS

From the earliest times, ships and boats were propelled by human power and wind power. In the nineteenth century very large and efficient sailing ships (many 70 to 100 m long) were transporting passengers and cargo to ports all over the world. In the Western world most of the cargoes were carried by sailing ships until the 1890s. By the middle of the nineteenth century, with the development of reliable steam engines, the designers of ships began to use coal-burning steam power plants to propel ships, using paddle wheels and then propellers. Other power plants were developed in the late nineteenth century and during the twentieth century, including steam turbines, spark-ignition engines, diesel engines, gas turbines, and nuclear power plants. These ship power plants were developed from their land-based counterparts to operate in the severe ocean environment. A relatively new power source, the fuel cell, is being developed to provide electrical power for underwater vessels and also is being evaluated for future applications in surface ships.

High-speed ships and large-displacement slower ships can have power requirements of 30,000 to 75,000 kW (approximately 40,000 to 100,000 hp) for propulsion. Some nuclear-powered aircraft carriers have power plants that develop 200,000 kW. In addition to the power required to propel the vessel, some energy and power have to be provided for various ship services, such as heating, lighting, and cooling; this is often described as the "hotel load." Power also is required to refrigerate the cargo in some vessels and to pump oil in tankers.

SHIP RESISTANCE

Experiments by William Froude in the nineteenth century indicated that ship resistance has two main components: frictional, and wave-making resistance or drag. Frictional resistance is caused by the movement of the underwater hull (wetted surface area) through the viscous fluid, water. Wave-making drag is the component of resistance resulting from the energy expended on the waves generated by the moving ship. Short, bluff ships have less wetted surface area than long, slender ships for the same displacement and therefore have lower frictional drag. The long, slender ships, however, tend to generate smaller waves and thus have lower wave-making drag at high ship speeds, where wave-making drag predominates. Placing bulbous bows on large commercial ships such as tankers has been shown to reduce wave-making drag. In addition to frictional and wave-making drag, there is form drag induced by the complex fluid flow around the hull near the stern of many ships, and there is also the wind resistance of the structure of the ship above the sea surface.

Information on ship resistance has been determined from large numbers of tests on scale models of ships and from full-size ships, and compilations of these experimental results have been published. For a new and innovative hull form the usual procedure is to construct a scale model of the ship and then to conduct resistance tests in a special test facility (towing tank). Alternatively, analytical methods can provide estimates of ship resistance for a range of different hull shapes. Computer programs have been written based on these theoretical analyses and have been used with success for many ship designs, including racing sailboats.

Methods for reducing wave-making drag are often introduced for ships that operate at high speed. Lifting the hull partly or completely out of the water by using hydrodynamic lift or air pressure has been investigated for high-speed ship designs. Fast but short-range ferries have been built on these principles using planing hulls, air-cushion designs, and hydrofoil configurations. In air cushion and surface-effect ships, 15 to 30 percent of the total power has to be supplied to the fans that provide the air pressure to lift the hulls.

Another concern is the ability of a ship to operate in storms. The pounding of waves on the hull increases the resistance; this can cause structural damage, and the violent ship motions can harm the passengers, cargo, and crew of the ship. Some designers have proposed special configurations that can maintain the ship speed in storms.

PROPULSORS

Propellers are the predominant propulsive devices driving ships, although water jets are now used in some high-speed ships. An experimental installation in a small ship of a magnetohydrodynamic propulsor has been tested, but it achieved rather low propulsive efficiency. Fish-like propulsion also has been examined for possible application to ships and underwater vehicles.

In the middle of the nineteenth century some ship designers began to replace paddle wheels with propellers (or screws). A propeller has a series of identical blades placed around a hub that is driven by the engine. W. J. M. Rankine presented a theoretical model of the ideal propeller in his classic paper "On the Mechanical Principles of the Action of Propellers" (1865). He demonstrated that the efficiency of an ideal frictionless propeller to produce a specified thrust was improved as the propeller diameter was increased. There are, of course, practical limitations to the propeller diameter, depending on the ship geometry and to some extent on the fabrication capabilities of the propeller manufacturers. Designers have developed the shapes of the propeller blades to achieve good efficiency.

Methods for defining the details of the geometry of the propeller and its blading to achieve high efficiency are well established and are continuously being improved. Digital computer programs have been developed that are widely used to provide the design and analysis of propulsors. These programs were originally used to design conventional propellers, and the methods have been extended to apply to propellers in ducts and also to the pumps for waterjets.

To produce thrust it is necessary to apply torque to the propeller; this results in a swirling flow downstream of the propeller from the reaction of this torque acting on the water. The swirling flow is not used to produce thrust and is therefore a source of energy loss. Several methods have been developed to raise the propulsive efficiency by canceling or removing this swirling flow. Stationary blades (stators) upstream or downstream of the propeller have been installed in a few propellers for this purpose, and counterrotating screws have been used. An additional device downstream of the propeller that rotates freely, called a Grim wheel or a vane wheel, has been used in some applications. The inner portion of the vane wheel absorbs the swirling energy from the propeller wake while the outer portion acts as an additional propeller and produces thrust.

In the 1930s ducts or nozzles surrounding highly loaded propellers were introduced. Experiments had shown that propeller efficiency was improved, and many of these devices (often referred to as Kort nozzles) have been used in tugs and fishing boats. Theoretical analysis of ducted propellers has shown that the efficiency can be improved compared to conventional propellers when some of the thrust is produced by the fluid flow around the nozzle.

The designer is expected to develop a propeller design to minimize the power required by the power

Ship Type	*Large Tanker*	*Container Ship*	*Fast Ferry*	*Navy Destroyer*	*Aircraft Carrier*	*Submarine*
Length, m	334	290	45	142	330	110
Displacement, t	344,000	82,000	200	9,000	1000,000	7,000
Operating Speed, kt	15.3	24	44	10-32	10-30+	32
Range, n.miles	27,000	21,000	125	6,000	Unlimited	Unlimited
Power Plant Type	Large Diesel	Large Diesel	2 Gas Turbines	4 Gas Turbines	2 Nuclear Reactors	Nuclear Reactor
Propulsion Power, kW	25,000	38,000	8,400	78,000	194,000	26,000
Ship Service Power, kW	3,000	3,000	100	3,000	8,000	2,000
Fuel	Heavy Oil	Heavy Oil	Light Diesel	Light Diesel	Nuclear	Nuclear
Fuel Load, t	8,400	6,600	7.5	1,00		

Table 1.
Representative Ship and Power Data for Various Ship Types

plant to propel the ship at the design speed. At this speed the effective power, P_e, is the product of ship resistance and ship speed and would be determined by model tests, correlations of experimental results, or computed using theoretical predictions, discussed earlier. The shaft power developed by the engine, P_s, is the power required to drive the propeller through the transmission and shafting. The propulsive efficiency or propulsive coefficient (PC) is the ratio of effective power to shaft power: $PC = P_e/P_s$. The propulsive coefficient is usually between 60 and 75 percent for most ship designs.

The propulsor, in a steady situation, has to provide the thrust equivalent to the resistance of the ship. The prediction of this balance between ship resistance and propeller thrust is a complex process because the water flow in the stern region of the ship interacts with the flow through the propeller. The ship resistance may be modified by the action of the propeller, while the propeller efficiency usually is influenced by the disturbed flow regime near the stern of the ship.

The phenomenon of propeller cavitation often occurs at high ship speed and on highly loaded propellers. It provides a very serious limitation to propeller performance because it can damage the propeller and ship components near it. In addition, the efficiency may be reduced and the phenomenon can cause noise and vibration. Cavitation occurs when the local pressure of the water on the blade surface is reduced below the vapor pressure of the water passing over the blade surface. The high velocity of the water passing over regions of the propeller blades creates this reduction in pressure. The water boils when the local pressure becomes very low, and pockets of water vapor are formed at the blade surface. These vapor pockets and their subsequent collapse cause the problems associated with cavitation. Modern methods of propeller design have allowed designers to predict the onset of cavitation and to define the performance limitations.

WATERJETS

Waterjets have been developed for application to high-speed ships. The waterjet has an inlet usually on the side or bottom of the ship in the region of the stern, which allows water to flow into a water pump. The pressure of the water is raised in the pump, and the water is expelled as a jet to produce the desired thrust. The direction of the jet flow can be controlled to provide maneuvering forces, eliminating the need for rudders. The propulsive coefficient of modern waterjets usually is lower than propellers for the same application, but they have been used where severe cavitation would occur in conventional propellers.

SHIP POWER PLANTS

Most ships are powered by thermal power plants, including diesel engines, gas turbines, and nuclear systems. Merchant ships usually have diesel engines, while gas turbines or a combination of diesel engines and gas turbines often power naval vessels. Some of the larger ships in the U.S. Navy have nuclear power plants.

Large, low-speed diesel engines are remarkably efficient and convert about 45 percent of the energy of the fuel-air reaction into power. In addition, the cooling water and engine exhaust can be used to generate steam to provide heat and power for the ship. The steam has been used in some designs to generate power through steam turbines.

Improvements also have been made to the gas turbine for naval applications. An intercooled recuperative (ICR) gas turbine has been designed to improve the fuel consumption of naval power plants. The engine has a recuperator to take the heat that would otherwise be wasted in the exhaust and transfers it to the air entering the combustor. The new engine is expected to save about 30 percent of the fuel consumed, compared to the simple gas turbine. The ICR engine is, however, larger and more expensive than the simple gas turbine.

In the early steam ships the paddle wheels and propellers were directly connected to reciprocating steam engines. The large low-speed diesel engines of today are able to operate efficiently at the low propeller speeds and usually are directly connected by shafts to the propellers. Other ship power plants, such as steam turbines, gas turbines, and medium-speed diesel engines, operate most efficiently at higher rotational speeds than propellers. Speed converters have to be provided to reduce the rotational speed from the high engine speed to the low propeller speed. Mechanical reduction gearboxes are generally used to provide this conversion. Electric drives, consisting of high-speed electric generators driven by steam turbines coupled to low-speed electric motors, were used during World War II, when there was a shortage of gearboxes. A modern version of electric drive sometimes is used in cruise ships and also is being evaluated for the U.S. Navy.

Submarines, submersibles, and other underwater vehicles have a difficult operational problem, namely the absence of atmospheric oxygen, which is necessary for all thermal engines except nuclear power plants. At about the beginning of the twentieth century rechargeable batteries were developed, and these batteries, coupled to electric motors and propellers, were used to propel submarines under the sea. Spark ignition engines and later diesel engines provided the power for operation at the surface and to recharge the batteries. Since 1985, stored oxygen has been used in some thermal power plants and in fuel cells for experimental underwater vehicles. Nuclear power plants, which do not require oxygen for power generation, were introduced in 1954 to power U.S. military submarines. Other nations also have developed nuclear power for naval submarines.

ENVIRONMENTAL PROBLEMS

Two of the main environmental concerns are the atmospheric pollutants generated by ship's engines, and the possibility of damaging oil spills from accidents and ship operations.

Commercial ships have traditionally used the least expensive and therefore the most polluting oil fuels. Large diesel engines can now operate successfully with the lowest grade of fuel oils, and these fuels have a relatively high sulfur content. As a result the harmful SO_x by-products are released to the atmosphere. In addition, the combustion process in diesel engines produces harmful NOX pollutants. International, national, and local restrictions have encouraged engine manufacturers to develop remedial methods to reduce pollution.

The reduction of SO_x is best achieved by reducing the sulfur content of the fuel oil. Ship engineers are required to replace high-sulfur fuel with expensive higher-grade oil fuel as their ships approach coastlines where strict pollution restrictions apply.

Higher-grade fuels also reduce NO_x production in diesel engines, but more stringent methods are required to satisfy pollution regulations in some areas (e.g., California). Experiments conducted on diesel engines indicate that the concentration of NO_x in the exhaust gases increases as the engine power is reduced as ships approach coasts and harbors. Two approaches have been evaluated to improve the situation. First, changes to the combustion process have been tried by such measures as recirculating some of the exhaust and by the addition of water to the fuel to form an emulsion. Tests on engines have shown that about a 25 percent reduction in NO_x can be attained with such methods. The second approach has involved attempts to clean up the exhaust by catalytic reactions. Catalytic converters, as used in automobiles, have not been successful up to now, because diesel engines operate with too much oxygen in the exhaust products. An alternative approach using ammonia vapor as the catalyst has been shown to reduce the NO_x in the exhaust by more than 80 percent. This has been termed the selective catalytic reduction (SCR) method; it relies on the excess oxygen in the exhaust products to react with the ammonia and the NO_x to produce harmless water vapor and nitrogen. Careful measurement of the pollutants in the exhaust gases ahead of the ammonia injector has to be provided so the optimum quantity of ammonia can be injected.

In the last 30 years of the twentieth century, there were very damaging oil spills from the grounding of tankers filled with crude oil. In addition, there have been many smaller accidents, as well as routine operations, that have resulted in significant amounts of oil being released to the environment.

The International Maritime Organization (IMO) has issued a series of rules to tanker designers in an attempt to minimize the outflow of oil after accidental side or bottom damage. New tankers are required to have double hulls or other structural innovations to minimize tanker spills.

When cargo ships and fishing boats are involved in accidents, there are various measures that have to be carried out to limit the flow of fuel oil into the water. Accidents in harbors and close to shore are treated with great care. Oil booms would be placed around the vessel to prevent the spread of oil, and skimmers would be brought to the area to collect the oil released to the environment. In extreme cases the fuel oil cargo may be burned when the oil could not be pumped out or when a stranded vessel could not be refloated.

Routine operations of many ships have resulted in oil pollution. Cleaning up minor spills on deck and in the engine room is now treated very carefully. Fuel-oil and crude-oil tanks are cleaned from time to time. In the past the polluted oil-water mixture was dumped overboard, but now it must be pumped ashore for treatment.

A. Douglas Carmichael
Clifford A. Whitecomb

See also: Aerodynamics; Diesel Cycle Engines; Diesel Fuel; Engines; Environmental Problems and Energy Use; Nuclear Fission Fuel; Thermodynamics; Transportation, Evolution of Energy Use and; Turbines, Gas; Turbines, Steam; Waves.

BIBLIOGRAPHY

Giles, D. L. (1997). "Faster Ships for the Future." *Scientific American* 277(4):126–131.

Rankine, W. J. M. (1865). "On the Mechanical Principles of the Action of Propellers." *Trans Institute of Naval Architects* 6:13–39.

van Manen, J. D., and van Oossanen, P. (1988). "Propulsion." In *Principles of Naval Architecture, Vol. 2*, ed. E. V. Lewis. Jersey City, NJ: The Society of Naval Architects and Marine Engineers.

van Manen, J. D., and van Oossanen, P. (1988). "Resistance." In *Principles of Naval Architecture, Vol.2*, ed. E. V. Lewis. Jersey City, NJ: The Society of Naval Architects and Marine Engineers.

SIEMENS, ERNST WERNER VON (1816–1892)

Werner Siemens, German inventor, engineer, and entrepreneur was born on December 13, 1816, in Lenthe near Hanover, Lower Saxony, as the fourth child of fourteen children. His father was a tenant farmer. Siemens's education was first undertaken by his grandmother Deichmann, and it continued under a series of home tutors. In 1823 the family moved to Mecklenburg where the father took over the running of the estate of Menzendorf. From Easter 1832 until Easter 1834, Siemens attended the Katharineum, a gymnasium (secondary school) in the Hanseatic town of Lübeck, which he left without final examination. Because his interests were mathematics and natural sciences, rather than classical languages, he took some private lessons in mathematics to improve his education.

The family's precarious financial situation prompted Siemens to choose a military career as a way to advance his education in mathematics, physics and chemistry. In 1835 after basic training at the Artillery Corps in Magdeburg, he moved to Berlin where he stayed for three years as an officer candidate and a student of the Prussian Artillery School. There he was able to study under the guidance of the physicist Gustav Magnus. The death of both parents in 1840 put Siemens, as the oldest son, in charge of his nine surviving siblings. He took his brother Wilhelm to Magdeburg, where he continued his service in an artillery regiment. Passing a term of confinement in fortress, because of assisting a duel, Siemens smuggled in chemicals and experimented with galvanoplasty, a procedure discovered by Moriz Hermann Jacobi. He successfully developed a galvanic method of gold and silver plating. His brother Wilhelm went to England, patented and sold the discovery. From Siemens's electrochemical interests originated also a method of production of ozone and of electrolytic production of clean copper.

TELEGRAPHY

Siemens met the precision engineer Johann Georg Halske at Magnus's Institute of Physics in 1846. Together the two men founded "Siemens & Halske Telegraph Construction Co" in 1847 to exploit the new technology of electric telegraphy, which had been established three years earlier with the first public telegraph line of Samuel Morse. Siemens improved the dial-telegraph, invented the method for a seamless insulation of copper wire with gutta-percha, and discovered the multiple telegraphic use of copper conductors. The first long-line telegraph line, which was 660 km long and laid by Siemens & Halske between Berlin and Frankfurt was inaugurated on March 28, 1849, the day at which the first German national parliament in Frankfurt passed a constitution. His industrial career ensured, Siemens retired from the army. In 1851 Siemens & Halske commenced installation of a telegraph line between Moscow and St. Petersburg. For the Russian project Siemens put his brother Carl in charge. In 1855 the telegraph network for European Russia was installed, and through a Polish line a crucial link with Western Europe was established.

To compete worldwide, Siemens founded several industry subsidiary companies: in St. Petersburg, in London, in France, and in Vienna. In these endeavors, Siemens placced great trust in family members. After cousin Johann Georg helped to establish the company financially, Siemens incorporated other relatives: brother Friedrich and nephew Georg, later the first director of Deutsche Bank. In 1868 Siemens Brothers Co. of London started a cooperation with the Indo-European Telegraph Co. to lay mainly overhead cables through Southern Russia and Persia, effectively joining London and Calcutta. An improved printer telegraph with a punched tape synchronized with current pulses of an inductor with a double T armature operated from 1870 on this line.

Between 1874 and 1880 Siemens Brothers laid transatlantic cables using their own cable-steamer *Faraday*.

ELECTRIC GENERATORS AND MACHINES

The only source of electric energy available until the 1830s was in the form of galvanic elements. The situation changed in 1831 with the discovery of the law of electromagnetic induction by Michael Faraday. The law states that an electric current is generated in a conducting circuit if the latter is exposed to a fluctuating magnetic field. As a consequence of Faraday's discovery, numerous engineers—among them Siemens—designed and improved electric generators. The basic idea behind the first generators was to move wire coils into and out of the magnetic field of a permanent magnet, thereby generating an electric current in the coils according to the induction law. A hand-turned generator with magnets revolving around coils was constructed as early as 1832 in Paris. A year later, the reversed principle to rotate the coils in a fixed field of a permanent magnet was realized in a generator constructed in London. Around 1850, industrial production of electric generators was set up in several European countries. In 1856, Siemens developed a double T armature, cylinder magnet constructed on the basis of Faraday's law of induction to provide a source of electricity for telegraphs cheaper than Zinc galvanic batteries. Though more powerful than galvanic elements, the first generators did not yet provide sufficient electric energy for heavy mechanical work. The limitation was imposed through the rather modest size of the available permanent magnets. To overcome this shortcoming Siemens proposed in 1866, almost simultaneously and independently of some others (Soren Hjorth, Anyos Jedlik, Alfred Vaarley, and Charles Wheatstone) to substitute an electromagnetic effective magnet for the permanent magnet. Siemens is credited with having worked out the principle underlying the new type of generator: a rotating coil inside a stationary coil, both forming a single circuit connected to the external consumer circuit. By starting the rotation of the movable coil, a weak current is induced through a residual magnetic field delivered from the iron armature of the apparatus. The current in the stationary coil causes an increase of the magnetic field, which in turn gives rise to an increased current output from the rotating coil. A self-enhancing process of current generation is so initiated. Siemens called this scheme of a self-excited generator the "dynamo electrical principle." Its formulation, presented to the Academy of Science on January 17, 1867, and subsequent demonstration of the electric light from the dynamo-machine moved through a steam-machine at the factory side in Markgrafenstrasse, Berlin, laid the groundwork for the modern technology of electric generators. This discovery is considered Siemens's most important contribution to science and engineering.

Siemens constructed the first electric railway shown at the Berlin Trade Fair of 1879 and the first electrically-operated lift in 1880, and the first electric trams began operating in Berlin in 1881. Siemens received many honors for his work: an 1860 honorary doctorate from the University of Berlin, an 1873 membership in the Royal Academy of Science, and an 1888 knighthood from Emperor Friedrich III. Siemens died in Berlin, Charlottenburg, on December 6, 1892.

Barbara Flume-Gorczyca

See also: Magnetism and Magnets.

BIBLIOGRAPHY

Feldenkirchen, W. (1992). *Werner von Siemens*. Zurich: Piper München.

Kundt, A. (1979). *Physiker über Physiker*, Vol. 2. Berlin: Akademie-Verlag.

Siemens, W. (1881). *Gesamelte Abhandlungen und Vorträge*. Berlin: Springer.

Weiher, S. von. (1975). *Werner von Siemens*. Goettingen: Musterschmidt.

SMEATON, JOHN (1724–1792)

John Smeaton was one of the first great British civil engineers, the first to use the title "Civil Engineer" (seeing himself as a professional ranking alongside doctors and lawyers), and the first to achieve distinction as an engineering scientist.

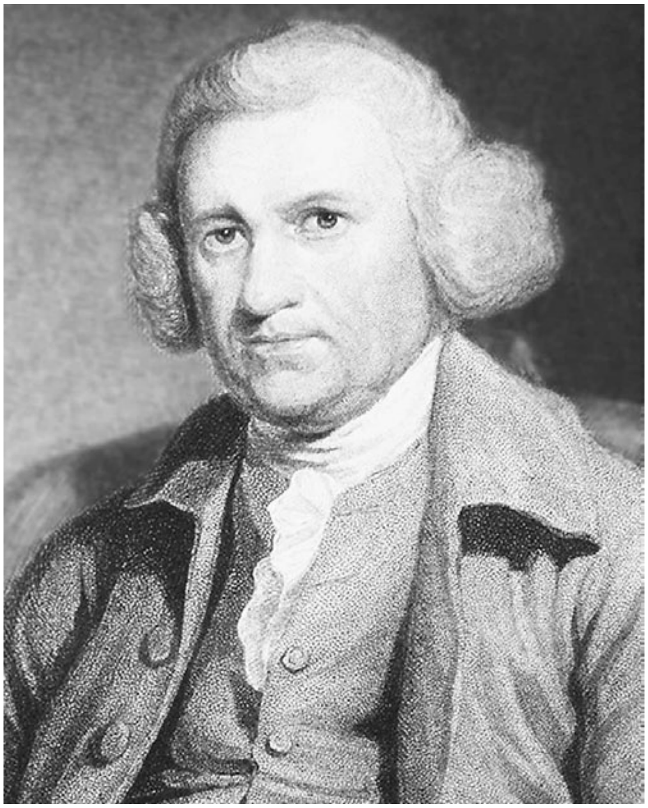
Smeaton was born on June, 8, 1724 at Austhorpe Lodge, near Leeds. He was the son of a successful Leeds lawyer, and from an early age Smeaton developed an interest in mechanics (which his father indulged by providing a workshop), building a working model of a steam engine being erected at a nearby colliery (coal mine). It was the desire of Smeaton's father that his son should succeed him, and at the age of sixteen Smeaton was employed in his father's office. In an attempt to wean him from his mechanical pursuits and partly to give him a good legal education, Smeaton was sent to London. In London, Smeaton's desire to pursue a mechanical career was not suppressed, and his father finally agreed to finance his training as a philosophical instrument maker, to make instruments or apparatus for study of natural philosophy or the physical sciences. The young Smeaton did not however, live the life of a workman, but attended meetings of the Royal Society and was often in the company of educated people. He was soon able to set up as a mathematical instrument maker (making instruments used by draftsmen, navigators, or land surveyors) in London, and at the age of twenty-six presented a paper before the Royal Society on improvements in the marine compass. More papers quickly followed, and in 1753 he was elected a fellow.

In 1759 Smeaton presented to the Royal Society a paper entitled "An Experimental Enquiry concerning the natural Power of Water and Wind to turn mills, and other machines, depending on a circular motion." This paper was the result of a series of experiments carried out in 1752 and 1753, Smeaton having delayed its publication until he had put his deductions into practice. At the time there was a lively debate as to the merits of undershot and overshot waterwheels, but little published data to substantiate various claims. Smeaton's solution to this problem depended solely on experiments made with working model waterwheels. He wanted to compare the "power" of the water delivered to the wheel with the useful "effect" produced by the wheel.

Smeaton's elegant experimental technique enabled him to deal with both hydraulic and mechanical friction losses, allowing him to calculate water velocity at the wheel and thereby determine an "effective" or "virtual" head. Smeaton's experimental apparatus was a brilliant device that enabled him to measure the efficiency of the waterwheel, alone rather than the overall efficiency of the experiment. Smeaton was able to conclusively show that a water-wheel when driven by the weight of water alone, is about twice as efficient as when driven by the impulse of water. This demonstration ensured that British mills, wherever possible, from then on would be fitted with overshot or breastshot waterwheels, rather than undershot.

Smeaton also turned his attention to windmills, but with inconclusive results. This may have been because there was no clear-cut issue to be resolved as was the case with the waterwheel regarding which form was the best type. However, he did produce useful guidelines on the construction of windmills. Smeaton was careful to distinguish the circumstances in which a model test rig differs from the full-size machine, cautioning, "otherwise a model is more apt to lead us from the truth than towards it." Smeaton's experiments demonstrated that work done by a prime mover was a good measure of its performance. In 1765, seeing a need for a practical measure, he fixed the power of a horse working an eight-hour day as 22,000/ft. lb/min.

Britain in the mid-eighteenth century experienced rapid expansion in the number and scale of public works. This enabled Smeaton, in 1753, to give up instrument-making and become a consulting civil engineer, building both wind and water mills, canals, bridges and, from 1756 to 1759, the Eddystone lighthouse. In preparing his designs, Smeaton adopted a scientific approach. It was his practice to prepare a report and have the work carried out by a resident engineer.


John Smeaton. (Library of Congress)

In 1766 Smeaton designed an atmospheric steam-pumping engine for the New River Company; however, when the engine was set to work, its performance was less than Smeaton had expected. Realizing his knowledge of steam engines was deficient, he approached the problem in the same manner as he did his study of water and wind power—by deciding to carry out a systematic, practical study. His first action was to have drawn up a list of all steam engines working in the district around Newcastle. Of these, fifteen were chosen and a set of engine tests made. Smeaton was the first to gather such comprehensive data, from which he calculated for each engine the "great product per minute" or power, and "effect per minute of one bushel of coal per hour" or performance.

Smeaton constructed on the grounds of his house, a small experimental atmospheric engine with a ten-inch cylinder. Beginning in 1770 he made some 130 tests over two years. All relevant measurements were made, including temperature, pressure—both internal and barometric, evaporation of water per bushel of coal, etc. Smeaton's method of testing was to adjust the engine to good performance and take measurements, then alter one of the parameters and take further measurements. By this means he was able to optimize valve timing, piston loading, and size of injector nozzle. He also carried out experiments to test the evaporative powers of various types of coal.

From the knowledge gained, Smeaton drew up a table for the proportions of parts for Newcomen-type engines. Although Smeaton's experiments added nothing to the invention of the Newcomen engine, establishing proper proportions for engines enabled him to built more efficient and powerful engines. The great leap forward in the operation of atmospheric steam engines came with Watt's separate condenser.

Smeaton was a born mechanic and incessant experimentor, but a man of simple tastes and wants. He limited his professional engagements in order to devote a certain portion of his time to scientific investigations. One of Smeaton's rules was not to trust deductions drawn from theory when there was an opportunity for actual experiment. In 1771 Smeaton founded a club for engineers, which later came to be call the "Smeatonian Society."

In 1756 Smeaton married Ann Jenkinson, the daughter of a merchant tailor and freeman of the city of York. They had two daughters. Ann Smeaton fell ill and died in 1784. Smeaton continued working until 1791, when he retired to Austhorpe Lodge to prepare for publication descriptions of his various engineering works. While walking in his gardens he suffered a stroke and died six weeks later, on October 28, 1792, in his sixty-ninth year.

Robert Sier

See also: Newcomen, Thomas; Watt, James.

BIBLIOGRAPHY

Farey, J. (1827). *A Treatise on the Steam Engine*. London: Longmans.

Ress, A. (1813). *The Cyclopedia; or Universal Dictionary of Arts, Science and Literature*. London: Longman.

Smeaton, J. (1759). "An Experimental Enquiry Concerning the Natural Power of Water and Wind to Turn Mills and Other Machinery, Depending on a Circular Motion." *Philosophical Transactions of the Royal Society* 51:100–174.

Smeaton, J. (1791). *A Narrative of the Building and a Description of the Construction of the Eddystone Lighthouse*. London.

Smeaton, J. (1797, 1812). *Reports of the Late John Smeaton, FRS Made on Various Occasions in the Course of His Employment as a Civil Engineer*, Vol. 1 (1797), Vol. 2–3 (1812). London: Longman, Hurst, Rees, Orme, and Brown.

Smiles, S. (1904). *Lives of the Engineers. John Smeaton*. London: John Murrey.

Wilson, P. N. (1955). "The Waterwheels of John Smeaton." *Transactions of the Newcomen Society* 30:25–43.

SMOG

See: Air Pollution

SOLAR ENERGY

Energy from the sun is abundant and renewable. It is also the principal factor that has enabled and shaped life on our planet. The sun is directly or indirectly responsible for nearly all the energy on earth, except for radioactive decay heat from the earth's core, ocean tides associated with the gravitational attraction of Earth's moon, and the energy available from nuclear fission and fusion. Located approximately ninety-three million miles from Earth, the sun, an average-size star, belongs to the class of dwarf yellow stars whose members are more numerous than those of any other class. The energy radiated into outer space by the sun (solar radiation) is fueled by a fusion reaction in the sun's central core where the temperature is estimated to be about 10 million degrees. At this temperature the corresponding motion of matter is so violent that all atoms and molecules are reduced to fast-moving atomic nuclei and stripped electrons collectively known as a plasma. The nuclei collide frequently and energetically, producing fusion reactions of the type that occur in thermonuclear explosions.

While about two-thirds of the elements found on earth have been shown to be present in the sun, the most abundant element is hydrogen, constituting about 80 percent of the sun's mass (approximately two trillion trillion million kilograms). When hydrogen nuclei (i.e., protons) collide in the sun's core, they may fuse and create helium nuclei with four nucleons (two protons and two neutrons). Roughly 20 percent of the sun's mass is in the form of helium. Also created in each fusion reaction are two neutrinos, high-energy particles having no net electrical charge, which escape into outer space, and high-energy gamma radiation that interacts strongly with the sun's matter surrounding its core. As this radiation streams outward from the core, it collides with and transfers energy to nuclei and electrons, and heats the mass of the sun so that it achieves a surface temperature of several thousand degrees

Celsius. The energy distribution of the radiation emitted by this surface is fairly close to that of a classical "black body" (i.e., a perfect emitter of radiation) at a temperature of 5,500°C, with much of the energy radiated in the visible portion of the electromagnetic spectrum. Energy is also emitted in the infrared, ultraviolet and x-ray portions of the spectrum (Figure 1).

The sun radiates energy uniformly in all directions, and at a distance of ninety-three million miles, Earth's disk intercepts only four parts in ten billion of the total energy radiated by the sun. Nevertheless, this very small fraction is what sustains life on Earth and, on an annual basis, is more than ten thousand times larger than all the energy currently used by Earth's human inhabitants. Total human energy use is less than 0.01 percent of the 1.5 billion billion kilowatt-hours (kWh) of energy per year the sun delivers to Earth. A kilowatt hour is one thousand watt hours and is the energy unit shown on your electric bill. A one-hundred-watt light bulb left on for ten hours will use one kWh of electric energy.

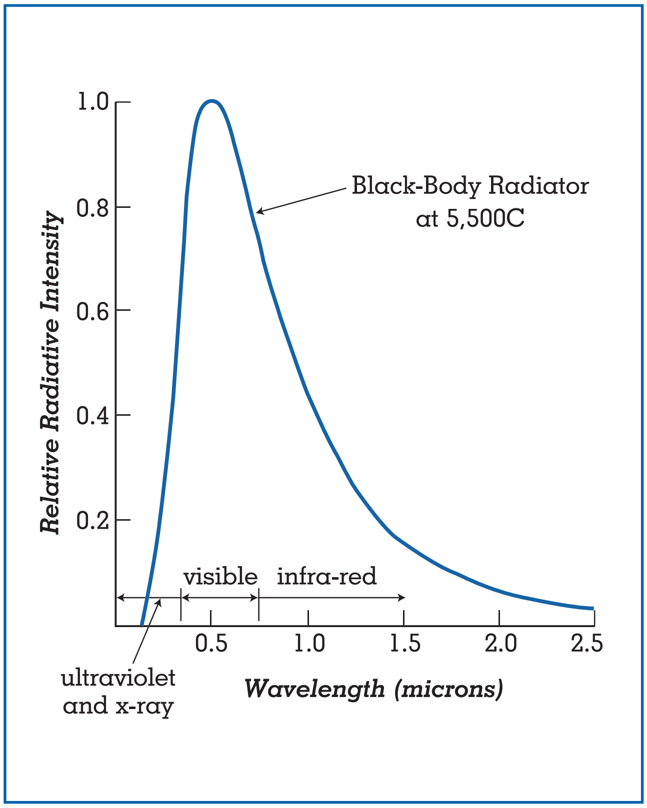


Figure 1.
Relative radiative intensity as a function of wavelength.

AVAILABILITY OF SOLAR ENERGY

While the amount of energy radiated by the sun does vary slightly due to sunspot activity, this variation is negligible compared to the energy released by the sun's basic radiative process. As a result, the amount of energy received at the outer boundary of Earth's atmosphere is called the Solar Constant because it varies so little. This number, averaged over Earth's orbit around the sun, is 1,367 watts per square meter (W/m^2) on a surface perpendicular to the sun's rays. If, on average, Earth were closer to the sun than ninety-three million miles, this number would be larger; if it were farther from the sun, it would be smaller. In fact, Earth's orbit about the sun is not circular, but elliptical. As a result, the Solar Constant increases and decreases by about 3 percent from its average value at various times during the year. In the northern hemisphere the highest value is in the winter and the lowest is in the summer.

A non-negligible fraction of the solar radiation incident on the earth is lost by reflection from the top of the atmosphere and tops of clouds back into outer space. For the radiation penetrating the earth's atmosphere, some of the incident energy is lost due to scattering or absorption by air molecules, clouds, dust and aerosols. The radiation that reaches the earth's surface directly with negligible direction change and scattering in the atmosphere is called "direct" or "beam" radiation. The scattered radiation that reaches the earth's surface is called "diffuse" radiation. Some of the radiation may reach a receiver after reflection from the ground, and is called "albedo." The total radiation received by a surface can have all three components, which is called "global radiation."

The amount of solar radiation that reaches any point on the ground is extremely variable. As it passes through the atmosphere, 25 to 50 percent of the incident energy is lost due to reflection, scattering or absorption. Even on a cloud-free day about 30 percent is lost, and only 70 percent of 1,367 W/m^2, or 960 W/m^2, is available at the earth's surface. One must also take into account the earth's rotation and the resultant day-night (diurnal) cycle. If the sun shines 50 percent of the time (twelve hours per day, every day) on a one square meter surface, that surface receives no more than (960 W/m^2) × (12 hours/day) × (365 days/year) =

4,200 kWh of solar energy per year. Since, on average, the sun actually shines less than twelve hours per day at any location, the maximum solar radiation a site can receive is closer to 2,600 kWh per square meter per year. To put this number into perspective, the average person on earth uses about 18,000 kWh of all forms of energy each year.

HISTORY

For centuries the idea of using the sun's heat and light has stimulated human imagination and inventiveness. The ancient Greeks, Romans, and Chinese used passive solar architectural techniques to heat, cool, and provide light to some of their buildings. For example, in 100 C.E., Pliny the Younger built a summer home in northern Italy that incorporated thin sheets of transparent mica as windows in one room. The room got warmer than the others and saved on short supplies of wood. To conserve firewood, the Romans heated their public baths by running the water over black tiles exposed to the sun. By the sixth century, sunrooms on private houses and public building were so common that the Justinian Code introduced "sun rights" to ensure access to the sun.

In more recent times, Joseph Priestly used concentrated sunlight to heat mercuric oxide and collected the resulting gas, thus discovering oxygen. In 1872, a solar distillation plant was built in Chile that provided six thousand gallons of fresh water from salt water daily for a mining operation. In 1878, at an exhibition in Paris, concentrated sunlight was used to create steam for a small steam engine that ran a printing press.

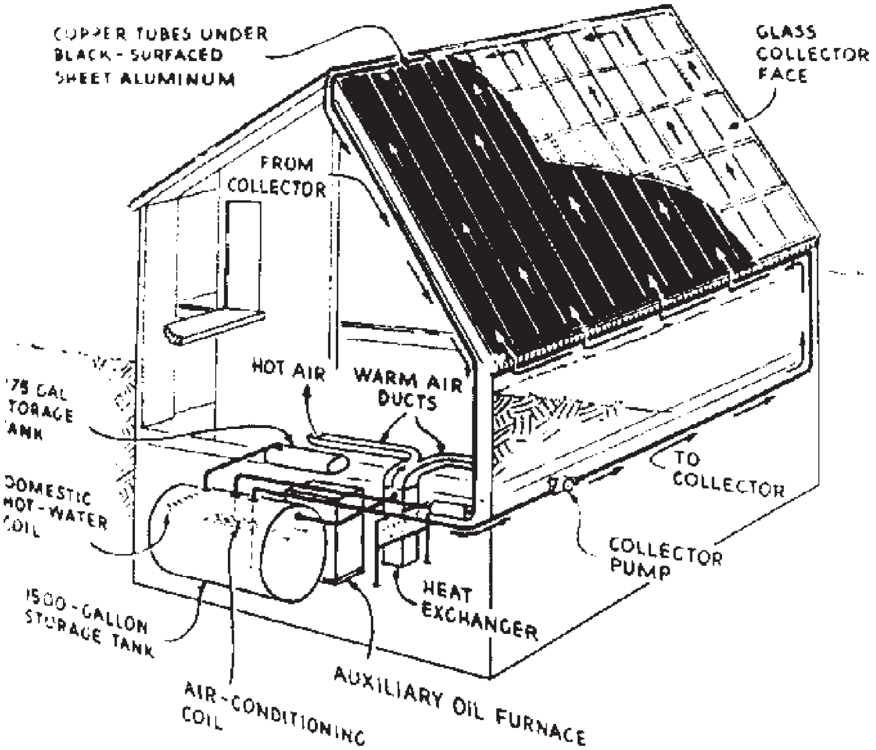
Many further solar energy developments and demonstrations took place in the first half of the twentieth century. However, only one solar technology survived the commercial competition with cheap fossil fuels. That exception was solar water heaters that were widely used in Japan, Israel, Australia and Florida, where electricity was expensive and before low-cost natural gas became available.

Also, passive solar, or climate responsive, buildings were in such demand by 1947 as a result of energy shortages during World War II, that the Libbey-Owens-Ford Glass Company published a book entitled *Your Solar House*, which profiled forty-nine solar architects. In the mid-1950s, architect Frank Bridgers designed the world's first commercial office building using passive design and solar water heating. This solar system has since been operating continuously, and is now in the U.S. National Historic Register as the world's first solar-heated office building.

It was the oil embargo of 1973 that refocused attention on use of solar energy, when the cost of fossil fuels (coal, oil and natural gas) increased dramatically and people began to appreciate that fossil fuels were a depletable resource. More recently, concerns about energy security (the dependence on other countries for one's energy supply), and the local and global environmental impacts of fossil fuel use, have led to the growing realization that we cannot project today's energy system, largely dependent on fossil fuels, into the long-term future. In the years following the oil embargo, much thought and effort has gone into rethinking and developing new energy options. As a result, a growing number of people in all parts of the world believe that renewable (i.e., non- depletable) energy in its various forms (all derived directly or indirectly from solar energy except for geothermal and tidal energy) will be the basis for a new sustainable energy system that will evolve in the twenty-first century.

USING SOLAR ENERGY

Solar energy can be used in many direct and indirect ways. It can be used directly to provide heat and electricity. Indirectly, non-uniform heating of Earth's surface by the sun results in the movement of large masses of air (i.e., wind). The energy associated with the wind's motion, its kinetic energy, can be tapped by special turbines designed for that purpose. Winds also create ocean waves, and techniques for tapping the kinetic energy in waves are under active development. Hydropower is another indirect form of solar energy, in that the sun's radiant energy heats water at Earth's surface, water evaporates, and it eventually returns to Earth as rainfall. This rainfall creates the river flows on which hydropower depends. In addition, sunlight is essential to photosynthesis, the process by which plants synthesize complex organic materials from carbon dioxide, water and inorganic salts. The solar energy used to drive this process is captured in the resulting biomass material and can be released by burning the biomass, gasifying it to produce hydrogen and other combustible gases, or converting it to liquid fuels for use in transportation and other applications.



A 1958 diagram of sun-heated home designed by the Massachusetts Institute of Technology showing air ducts and tanks. (Library of Congress)

In principle, energy can also be extracted from the kinetic energy of large currents of water circulating beneath the ocean's surface. These currents are also driven by the sun's energy, much as are wind currents, but in this case it is a slower-moving, much more dense fluid. Research into practical use of this energy resource is in its very earliest stages, but is looking increasingly promising.

Finally, there are renewable sources of energy that are not direct or indirect forms of solar energy. These include geothermal energy in the form of hot water or steam derived from reservoirs below the surface of Earth (hydrothermal energy), hot dry rock, and extremely hot liquid rock (magma). These resources derive their energy content from the large amount of heat generated by radioactive decay of elements in Earth's core. Energy can also be tapped from tidal flows resulting from Earth's gravitational interaction with the moon.

SOLAR HEATING AND COOLING

Solar heating and cooling systems are classified either as passive or active. Passive systems make coordinated use of traditional building elements such as insulation, south-facing glass, and massive floors and walls to naturally provide for the heating, cooling and lighting needs of the occupants. They do not require pumps to circulate liquids through pipes, or fans to circulate air through ducts. Careful design is the key. Active solar heating and cooling systems circulate liquids or air through pipes or other channels to move the necessary heat to where it can be used, and utilize other components to collect, store and control the energy from the solar energy source. Systems that combine both passive and active features are called hybrid systems. Solar heating systems are used primarily for hot water heating, interior space heating, and industrial and agricultur-

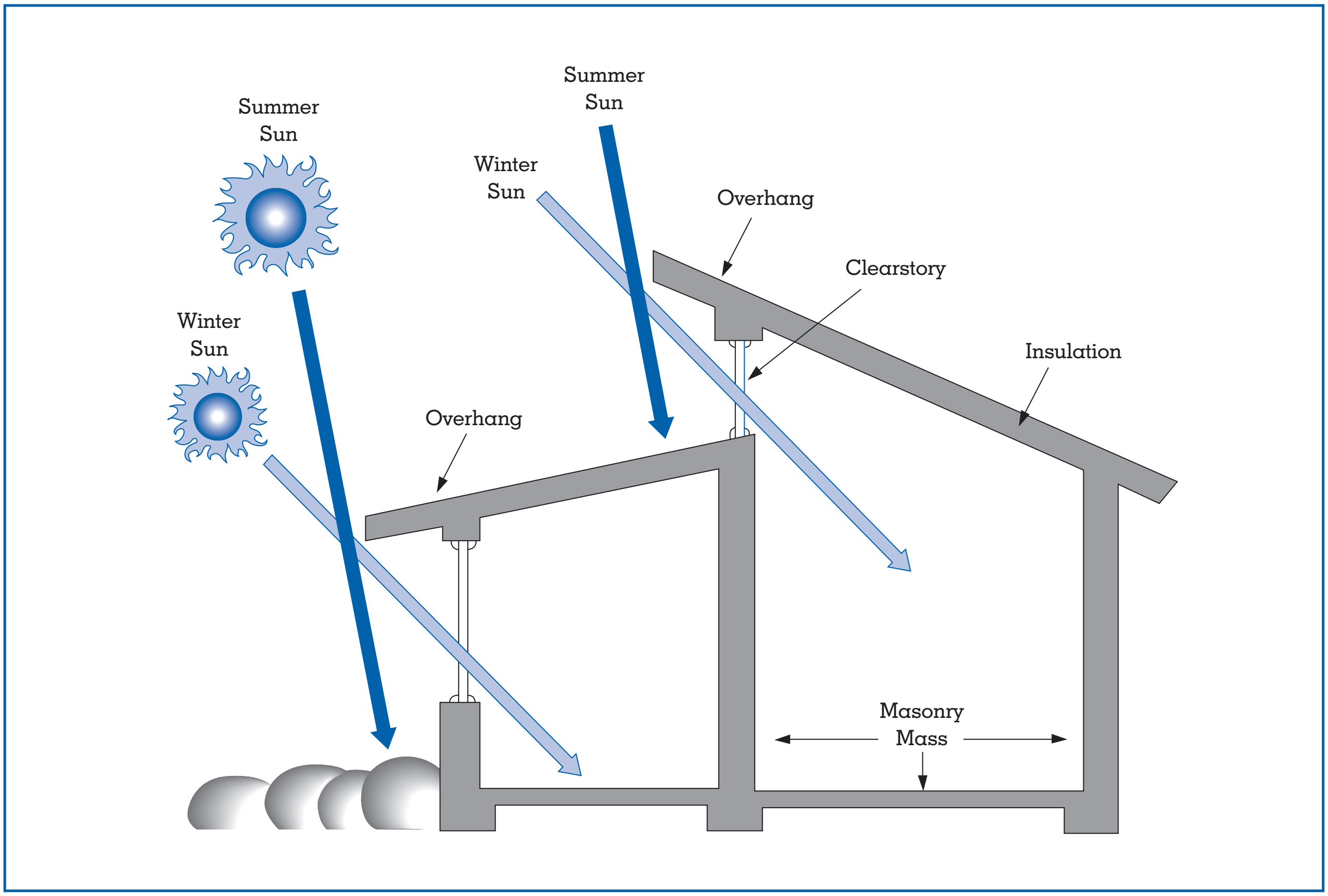


Figure 2.
Direct Gain.

al process heat applications. Solar cooling systems use solar heat to energize heat-driven refrigeration systems or to recycle dessicant-cooling systems. They are most often combined with solar hot water and space heating systems to provide year-round use of the solar energy system.

PASSIVE SYSTEMS

There are several types of passive system designs: direct gain, thermal storage, solar greenhouse, roof pond, and convective loop. The simplest is the direct gain design (Figure 2), which in the northern hemisphere is an expanse of south-facing glass (usually double glazed). The energy in the sunlight entering directly through the windows is absorbed, converted to heat, and stored in the thermal mass of the floors and walls. A thermal storage wall system consists of a massive sunlight-absorbing wall behind south-facing double glazing. The wall may contain water or masonry to store the energy during the day and release it during the night. Another type of passive system is the solar greenhouse (or sunspace), which combines direct gain in the greenhouse and a thermal storage wall between the greenhouse and the rest of the house. Solar energy provides the heat for the greenhouse and a good share of the energy for heating the living space in the house.

A fourth type of passive solar heating and cooling system is the roof pond, in which containers of water are used to collect and store the sun's energy. Movable insulating panels are used to control the gain and loss of heat. In winter, the insulation is opened during the day to allow collection of the sun's heat; at night the insulation is closed to minimize loss of heat that is to be released to the house. In summer, the insulation is closed during the day to block the sun, and opened to the sky at night to provide radiative cooling.

Solar collector panels. (Digital Stock)

The fifth type of passive system is the natural convective loop, in which the collector is placed below the living space and the hot air that is created rises to provide heat where it is needed. This same principle is used in passive solar hot water heating systems known as thermosiphons. The storage tank is placed above the collector. Water is heated in the collector, becomes less dense, and rises (convects) into the storage tank. Colder water in the storage tank is displaced and moves down to the collector where it is heated to continue the cycle.

More than one million residences and twenty thousand commercial buildings across the United States now employ passive solar design.

ACTIVE SYSTEMS

An active solar heating and cooling system consists of a solar energy collector (flat plate or concentrating), a storage component to supply heat when the sun is not shining, a heat distribution system, controls, and a back-up energy source to supply heat when the sun is not shining and the storage system is depleted.

The simplest type of active solar collector, the flat-plate collector, is a plate with a black surface under glass that absorbs visible solar radiation, heats up, and reradiates in the infrared portion of the electromagnetic spectrum. Because the glass is transparent to visible but not to infrared radiation, the plate gets increasingly hotter until thermal losses equal the solar heat gain. This is the same process that occurs in a car on a sunny day, and in the atmosphere, where gases such as carbon dioxide and methane play the role of the glass. It is called the greenhouse effect. Careful design can limit these losses and produce moderately high temperatures at the plate—as much as 160°F (71°C). The plate transfers its heat to circulating air or water, which can then be used for space heating or for producing hot water for use in homes, businesses or swimming pools.

Higher temperatures—up to 350°F (177°C)—can be achieved in evacuated tube collectors that encase both the cylindrical absorber surface and the tubes carrying the circulating fluid in a larger tube containing a vacuum which serves as highly efficient insulation. Still another high-temperature system is the parabolic-trough collector that uses curved reflecting surfaces to focus sunlight on a receiver tube placed along the focal line of the curved surface. Such concentrating systems use only direct sunlight, and can achieve temperatures as high as 750°F (400°C), but do require tracking of the sun.

Water heating accounts for approximately 25 percent of the total energy used in a typical single-family home. An estimated 1.5 million residential and commercial solar water-heating systems have been installed in the U.S. In Tokyo—there are nearly that many buildings with solar water heating. Large numbers of solar water heaters have also been installed in Greece, Israel, and other countries. More than 300,000 swimming pools are heated by solar energy in the United States.

SOLAR THERMAL POWER SYSTEMS

A solar thermal power system, increasingly referred to as concentrating solar power, to differentiate it from solar heating of residential/commercial air and water, tracks the sun and concentrates direct sunlight to create heat that is transferred to water to create steam, which is then used to generate electricity. There are three types of solar thermal power systems: troughs that concentrate sunlight along the focal axes of parabolic collectors (the most mature form of solar thermal power technology), power towers with a central receiver surrounded by a field of concentrating mirrors, and dish-engine systems that use radar-type reflecting dishes to focus sunlight on a heat-driven engine placed at the dish's focal point. Solar ponds are also a form of solar thermal power technology that does not require tracking.

Trough Systems

Trough systems currently account for more than 90 percent of the world's solar electric capacity. They use parabolic reflectors in long trough configurations to focus and concentrate sunlight (up to one hundred times) on oil-filled glass tubes placed along the trough's focal line. The dark oil absorbs the solar radiation, gets hot (up to 750°F (400°C)), and transfers its heat to water, creating steam that is fed into a conventional turbine-generator. Troughs are modular and can be grouped together to create large amounts of heat or power. Troughs operating in the Mojave Desert in the United States have been feeding up to 354 MW of electrical energy reliably into the Southern California power grid since the early 1990s, usually with minimal maintenance. While not currently cost competitive with fossil fuel-powered sources of electricity (trough electricity costs are about 12 cents per kWh), construction of the trough systems was encouraged by generous power purchase agreements for renewable electricity issued by the State of California, and Federal tax incentives.

Power Towers

Power towers, also known as central receivers, have a large field of mirrors, called heliostats, surrounding a fixed receiver mounted on a tall tower. Each of the heliostats independently tracks the sun and focuses sunlight on the receiver where it heats a fluid (water or air) to a very high temperature (1,200°F (650°C)). The fluid is then allowed to expand through a turbine connected to an electric generator. Another version heats a nitrate salt to 1,050°F (565°C), which provides a means for storage of thermal energy for use even when the sun is not shining. The heat stored in this salt, which becomes molten at these elevated temperatures, is then transferred to water to produce steam that drives a turbine-generator. Ten megawatt electrical power towers have been built and tested in the United States, one using water as the material to be heated in the receiver, and one using nitrate salt. A smaller unit that heats air has also been operated at the Platforma Solar test site in Almeria, Spain. While there are no commercial power tower units operating today, technical feasibility has been established and, in commercial production, electricity production costs are projected to be well below ten cents per kWh. At such costs, power towers will begin to compete with more traditional fossil fuel-powered generating systems. It is believed that power towers will be practical in sizes up to 200 MW (electrical).

Dish-Engine Systems

A dish-engine system uses a dish-shaped parabolic reflector, or a collection of smaller mirrors in the shape of a dish, to focus the sun's rays onto a receiv-

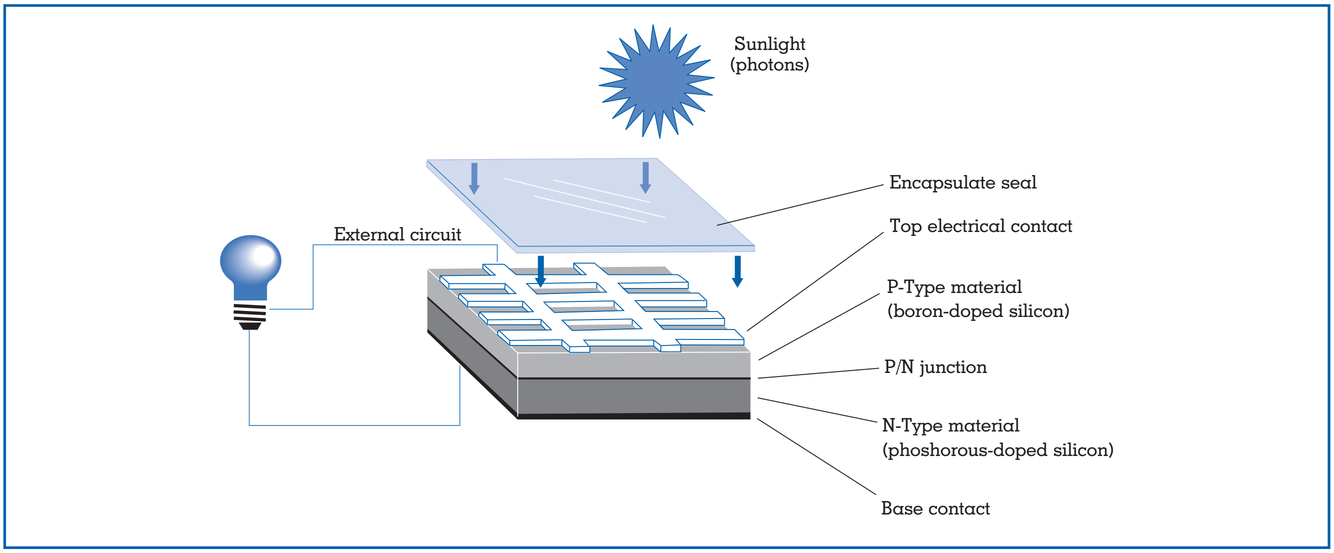


Figure 3.
Basic PV cell construction.

er mounted above the dish at its focal point. Two-axis tracking is used to maximize solar energy capture, and most often a Stirling engine (a sealed engine that can be driven by any source of external heat) is mounted on the receiver. Such systems are not yet commercial, but have been tested in sizes ranging from 5 to 50 kW (electrical). Conversion efficiencies of up to 29 percent have been achieved. Dish-engine systems are also modular, and can be used individually or in large assemblies.

Power-tower and dish-engine systems have higher concentration ratios than troughs, and therefore the potential to achieve higher efficiencies and lower costs for converting heat to electricity. All three systems can be hybridized (i.e., backed up by a fossil fuel system for supplying the heat), an important feature that allows the system to produce power whenever it is needed and not just when the sun is shining (dispatchability).

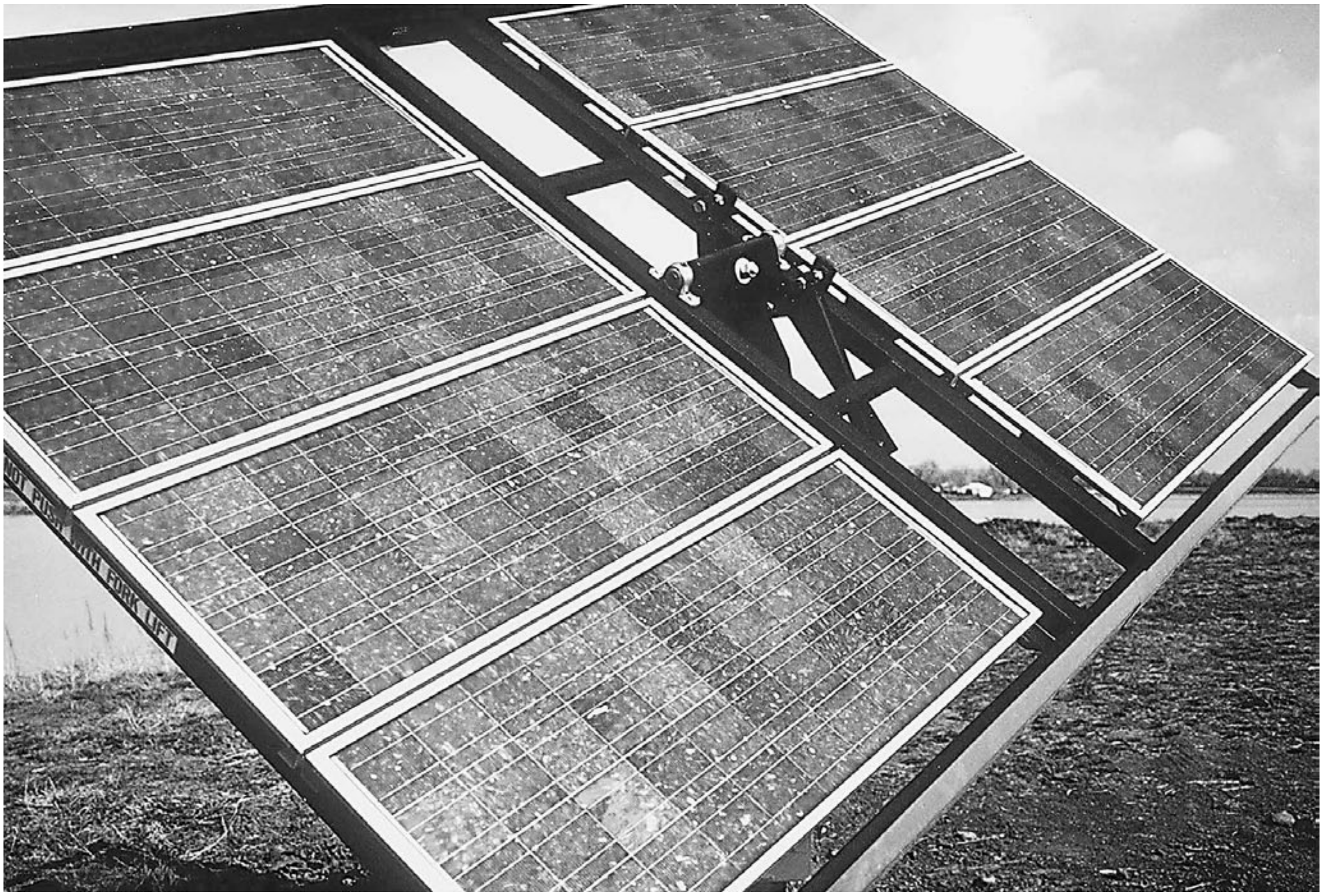
Solar Ponds

A solar pond does not concentrate solar radiation, but collects solar energy in the pond's water by absorbing both the direct and diffuse components of sunlight. Solar ponds contain salt in high concentrations near the bottom, with decreasing concentrations closer to the surface. This variation in concentration, known as a salt-density gradient, suppresses the natural tendency of hot water to rise, thus allowing the heated water to remain in the bottom layers of the pond while the surface layers stay relatively cool. Temperature differences between the bottom and top layers are sufficient to drive an organic Rankine-cycle engine that uses a volatile organic substance as the working fluid instead of steam. Temperatures of 90°C are routinely achieved in the pond bottom, and solar ponds are sufficiently large to provide some degree of energy storage.

The largest solar pond in the United States is a tenth of an acre experimental facility in El Paso, Texas, which has been operating reliably since 1986. The pond runs a 70 kW (electrical) turbine-generator and a 5,000 gallon per day desalination unit, while also providing process heat to an adjacent food processing plant. The potential of solar ponds to provide fresh water, process heat and electricity, especially for island communities and coastal desert regions, appears promising, but has not been fully investigated.

PHOTOVOLTAICS

Photovoltaics (photo for light, voltaic for electricity), often abbreviated as PV, is the direct conversion of sunlight to electricity. Considered by many to be the most promising of the renewable electric technologies in the long term, it is an attractive alternative to conventional sources of electricity for many reasons: it is silent and non-polluting, requires no special



Two huge solar energy panels track a photovoltaic system at the National Renewable Energy Lab in Golden, Colorado. (U.S. Department of Energy)

training to operate, is modular and versatile, has no moving parts, and is highly reliable and largely maintenance-free. It can be installed almost anywhere, uses both direct and diffuse radiation, and can be incorporated into a variety of building products (roofing tiles and shingles, facades, overhangs, awnings, windows, atriums).

The photoelectric effect (the creation of an electrical current when light shines on a photosensitive material connected in an electrical circuit) was first observed in 1839 by the French scientist Edward Becquerel. More than one hundred years went by before researchers in the United States Bell Laboratories developed the first modern PV cell in 1954. Four years later, PV was used to power a satellite in space and has provided reliable electric power for space exploration ever since.

The 1960s brought the first terrestrial applications for PV. At that time the technology was very expensive, with PV collectors, called modules, costing upwards of $1,000 per peak watt. The term "peak watts" is used to characterize the power output of a PV module when incident solar radiation is at its peak. Nevertheless, PV was a preferred choice in remote locations where no other form of power production was feasible. Over the past three decades, steady advances in technology and manufacturing have brought the price of modules down more than 200-fold, to $4 per peak watt. Further reductions in module cost to $1 to $2 per peak watt are expected within the next decade.

A photovoltaic cell (often called a solar cell) consists of layers of semiconductor materials with different electronic properties. In most of today's solar cells the semiconductor is silicon, an abundant element in the earth's crust. By doping (i.e., chemically introducing impurity elements) most of the silicon with boron to give it a positive or p-type electrical character, and doping a thin layer on the front of the cell with phosphorus to give it a negative or n-type character, a transition region between the two types

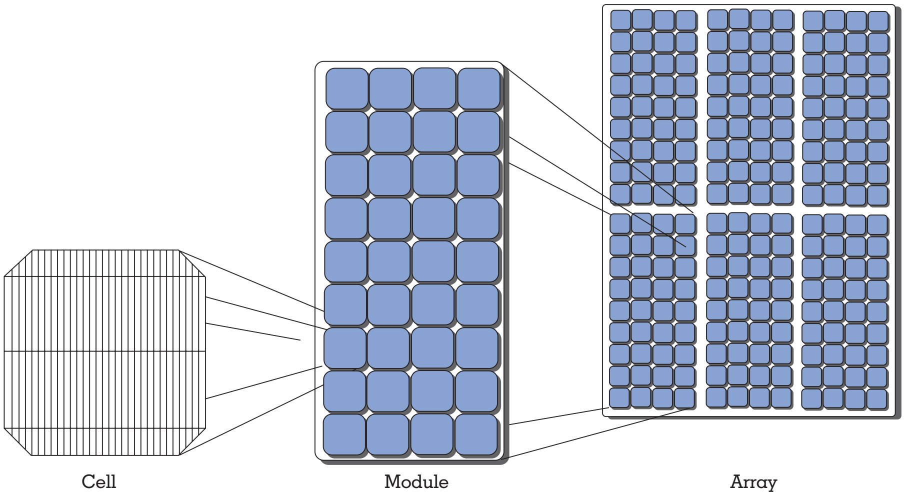


Figure 4.
PV cell, module, and array.

of doped silicon is formed that contains an electric field. This transition region is called a junction.Light consists of energy packets called photons. When light is incident on the cell, some of the photons are absorbed in the region of the junction and energy is transferred to the semiconductor, freeing electrons in the silicon. If the photons have enough energy, the electrons will be able to overcome the opposing electric field at the junction and move freely through the silicon and into an external circuit. As these electrons stimulate current flow in the external circuit, their energy can be converted into useful work (Figure 3).

Photovoltaic systems for specific applications are produced by connecting individual modules in series and parallel to provide the desired voltage and current (Figure 4). Each module is constructed of individual solar cells also connected in series and parallel. Modules are typically available in ratings from a few peak watts to 250 peak watts.

Commercially available PV systems most often include modules made from single-crystal or polycrystalline silicon or from thin layers of amorphous (non-crystalline) silicon. The thin-film modules use considerably less semiconductor material but have lower efficiencies for converting sunlight to direct-current electricity. Cells and modules made from other thin-film PV materials such as copper-indium-diselenide and cadmium telluride are under active development and are beginning to enter the market. Most experts consider thin-film technology to be the future of the PV industry because of the reduced material requirements, the reduced energy required to manufacture thin film devices, and the ability to manufacture thin films on a mass-production basis.

Amorphous silicon modules experience a conversion efficiency loss of about 10 percent when initially exposed to sunlight, but then stabilize at the reduced figure. The mechanism for this reduction is being actively investigated, but is still not well understood. Individual modules made with other PV materials do not exhibit such loss of conversion efficiency, but combinations of modules in arrays do exhibit systematic reductions in power output over their lifetimes. Estimated at about 1 percent per year on average, based on data to date, these reductions are most likely associated with deteriorating electrical connections and non-module electrical components.

Complete PV systems require other components in addition to the PV modules. This “balance-of-system” generally includes a support structure for the modules to orient them properly to the sun, an inverter to convert direct-current electricity to alternating-current electricity, a storage system for electrical energy (usually batteries), an electronic charge regulator to prevent damage to the batteries by protecting against overcharging by the PV array and excessive discharge by the electrical load, and related wiring and safety features (Figure 5).

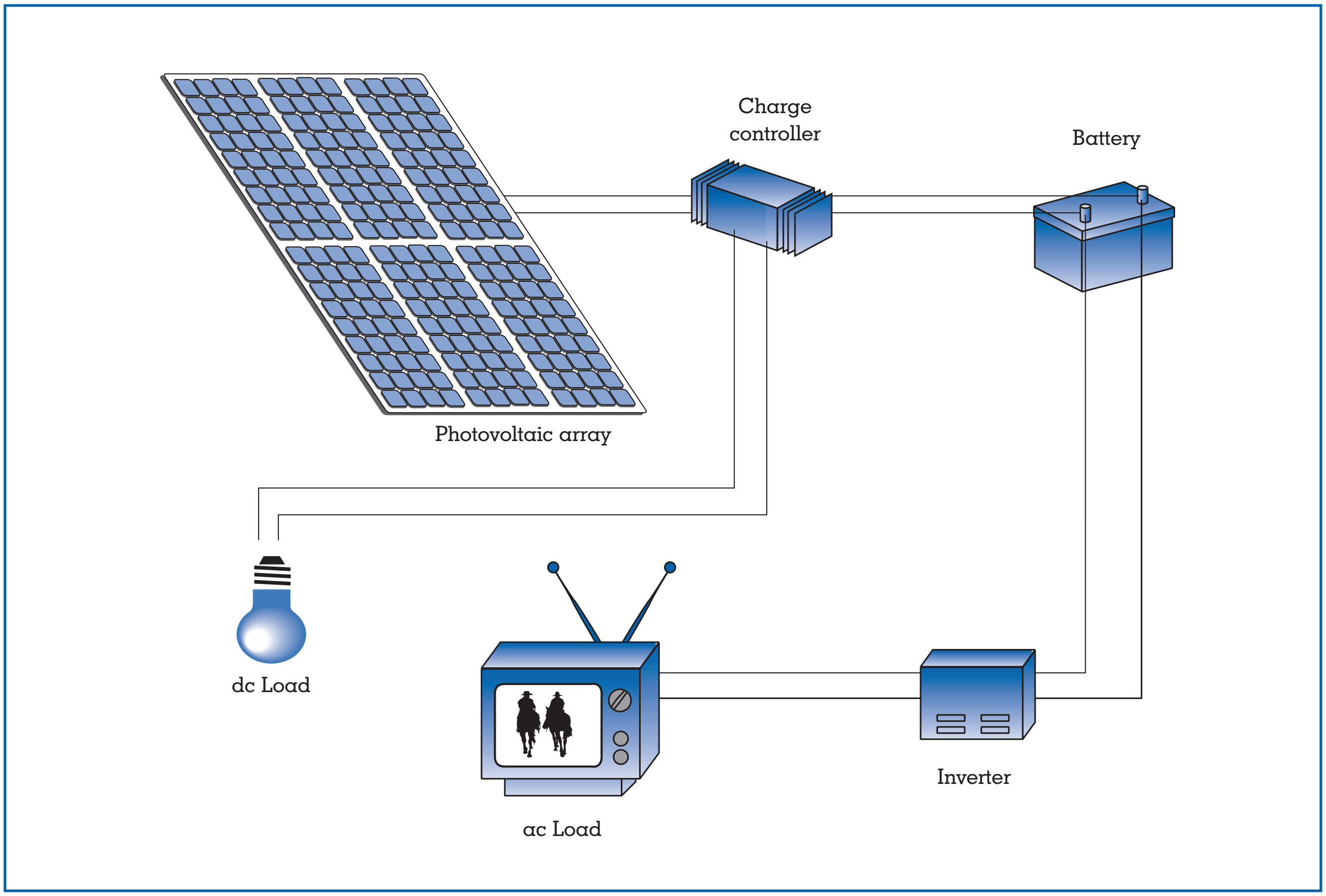


Figure 5.
Components of a typical off-grid PV system.

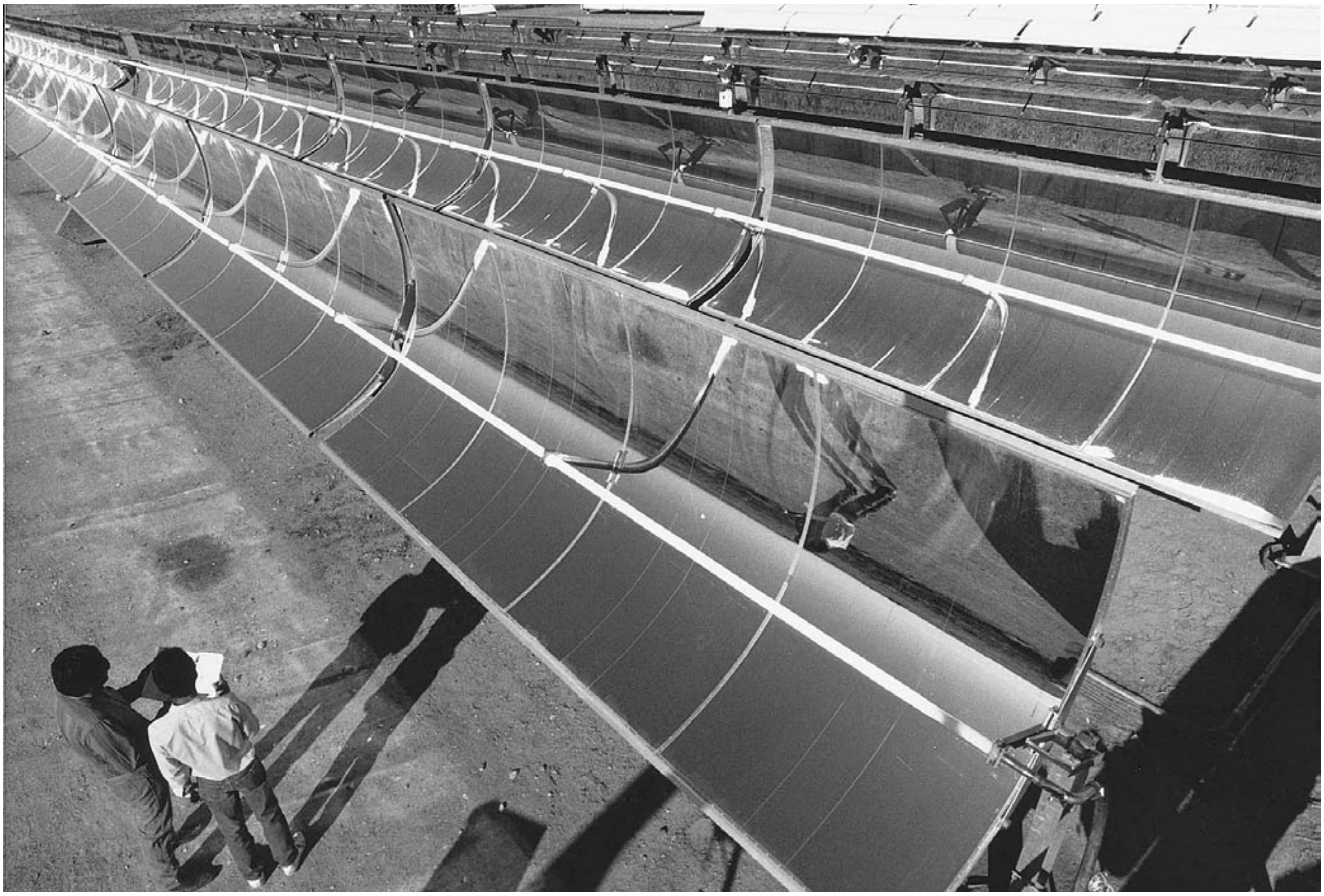
The output of a solar cell can be increased significantly by concentrating the incident sunlight onto a small area. Commercially available concentrator systems use low-cost lenses, mirrors or troughs in conjunction with tracking systems to focus sunlight on a small, high efficiency but expensive solar cell. One-axis trackers follow the sun during the day; two-axis trackers track daily and yearly variations in the sun's position. The higher cost of the cell and the associated tracking equipment, and the inability to use diffuse radiation, are offset by the higher conversion efficiencies achieved.

The cost of PV electricity is largely determined by four factors: cost of the PV modules, the efficiency of converting sunlight to electricity, the balance-of-system costs, and the lifetime of the systems. There are many combinations of these factors that provide about the same cost of electricity. Modules are now available that are designed to last at least thirty years.

Another scheme that has been proposed for use of photovoltaics is the solar satellite power system. Such a system would place large individual PV arrays (up to ten megawatts peak power) in synchronous orbit around the earth, and beam the collected and converted solar energy through the atmosphere in the form of microwaves to large receiving antennas on earth. The microwave generators would be powered by the PV electricity. The microwave energy collected at the surface would then be converted back to electricity for broad distribution. While of some interest, there are major issues associated with cost, environmental impact, and energy security.

PHOTOVOLTAIC APPLICATIONS

By its very nature PV can be used wherever electricity is needed and sunlight is available. For many years PV systems have been the preferred power sources for



Researchers at Sandia National Labs use a trough of parabolic mirrors to focus sunlight on a long glass tube in order to detoxify water. (U.S. Department of Energy)

buildings or other facilities in remote areas not served by the utility grid. As the cost of PV systems continues to decline, there is growing consensus that distributed PV systems on buildings that are grid-connected may be the first application to reach widespread commercialization. Typical off-grid applications for PV today include communications (microwave repeaters, emergency call boxes), cathodic protection (pipelines, bridges), lighting (billboards, highway signs, security lighting, parking lots), monitoring equipment (meteorological information, water quality, pipeline systems), warning signals (navigational beacons, railroad signs, power plant stacks), and remote loads (village power, parks and campgrounds, water pumping and irrigation, vacation cabins). On-grid applications include support for utility transmission and distribution (T&D) systems, where the benefits derive from avoiding or deferring T&D system investments, and from improving the quality of electrical service.

More than 200,000 homes in the U.S. currently use some type of PV technology, and more than 10,000 U.S. homes are powered entirely by PV. A large number of homes in Japan, where consumer electricity rates are three times higher than in the U.S., have roof-integrated PV systems. Half a million PV systems have been installed in developing countries by the World Bank. More than 10,000 have been installed in Sri Lanka, 60,000 in Indonesia, 150,000 in Kenya, 85,000 in Zimbabwe, 40,000 in Mexico, and 100,000 in China. Global use of PV is expected to grow rapidly in the coming decades.

OTHER DIRECT APPLICATIONS OF SOLAR ENERGY

In addition to the applications already discussed, solar energy can also be used for solar cooking (use of a reflecting surface to focus sunlight on pots used for food preparation), distillation (use of solar heat to evaporate and desalinate seawater), purification (use of concentrated ultraviolet radiation to kill organisms in contaminated water), and agriculture (use of solar heat to dry crops and grains, space and water heating, heating of commercial greenhouses, and use of PV electricity to power water pumping and other remote facilities). Many other applications will also become feasible as PV costs continue to decline.

Production of PV modules is still relatively small, but has been growing at a steady and significant rate. Global production in 1998 reached 150 megawatts peak. As costs come down and new manufacturing facilities are placed in operation, this number will grow rapidly.

SOLAR ENERGY'S POTENTIAL AND PROBLEMS

The amount of solar energy available on earth is many times greater than all the energy collectively used by people around the world. When one takes into account all the various forms in which solar energy is directly or indirectly available, along with energy available from geothermal and nuclear sources, it becomes clear that there is no global shortage of energy. The only shortage is that of low-cost energy. In principle there is no reason that solar energy in its various forms could not supply all the world's energy. To give but one example, all the electricity used by the United States (a little more than 3 trillion kWh) could be generated by commercially available PV modules covering half of a 100 mile by 100 mile area in the Nevada desert.

Greater use of solar and other forms of renewable energy is increasingly seen as the long-term response to growing demand for electricity and transportation fuels in developed and developing countries. For example, electricity derived from renewable resources can be used to electrolyze water and create hydrogen, which can then be used in fuel cells to provide electricity or power electric vehicles. As a result of the issues raised by the oil embargo of 1973, and increasing awareness of the local and global environmental damage associated with use of fossil fuels, many people now believe that the world must undergo a transition to a clean and sustainable energy system during the twenty-first century. There is great hope that renewable energy will be the basis for that new, sustainable system.

This is not to suggest that widespread use of solar energy is not facing major barriers. For many applications, today's high cost of solar energy systems is an important limitation. Nevertheless, significant progress has been and will continue to be made in improving system performance and reducing associated energy costs. A particular challenge to solar energy research and development efforts is that incident solar energy is not highly concentrated and is intermittent. The former necessitates use of concentrator systems, or large collection areas that can have impacts on ecosystems due to land use and disturbances during the construction stage of large-scale power plants. Centralized, multi-megawatt power facilities can also result in significant visual impacts.

The fact that solar energy is an intermittent energy resource means that energy storage systems (e.g., batteries, ultracapacitors, flywheels, and even hydrogen) will be required if solar energy is to be utilized widely. In addition, a variety of toxic chemicals are used in the manufacture of PV cells; however, studies of the risks associated with their manufacture and disposal indicate little threat to surroundings and the environment.

CONCLUSIONS

Solar energy offers a clean, sustainable alternative to continued use of fossil fuels. In its various forms it is already providing useful amounts of energy on a global basis, and will provide steadily increasing amounts in the twenty-first century, especially as developing countries require more energy to improve their economies.

In February 1979 the Carter Administration released a 30-federal agency study entitled *Domestic Policy Review of Solar Energy*. This study found that, with increasing oil prices and "comprehensive and aggressive initiatives at the Federal, State and local levels," renewable energy sources "could provide about 20 percent of the nation's energy by the year 2000." This did not occur, as oil prices actually dropped in the 1980s, Federal funding for renewables was reduced, and, for a period of time, energy issues largely disappeared from public view. Today, renewable energy sources supply about 12 percent of U.S. electricity and 8 percent of total U.S. energy consumption.

Continuing research and development efforts by the public and private sectors have significantly reduced the costs of solar energy, and further significant reductions are expected. Together with growing public concern about global climate change, it is anticipated that the 21st century will see widespread deployment of solar energy and other renewable energy systems, in both developing and developed countries.

Allan R. Hoffman

See also: Solar Energy, Historical Evolution of the Use of.

BIBLIOGRAPHY

Brinkworth, B. J. (1972). *Solar Energy for Man*. New York: John Wiley & Sons.

Center for Renewable Energy and Sustainable Technology (CREST). <http://www.solstice.crest.org/>

Daniels, F. (1972). *Direct Use of the Sun's Energy*. New Haven, CT: Yale University Press.

Energy Efficiency and Renewable Energy Network (EREN) home page U.S. Department of energy, Energy Efficiency and Renewable Energy Network. August 9, 2000 <http://www.eren.doe.gov/>.

Flavin, C., and Lenssen, N. (1994). *Power Surge: Guide to the Coming Energy Revolution*. New York: W.W. Norton & Company.

International Energy Agency. (1997). *Energy Technologies for the 21st Century*. Paris: OECD Publications.

International Energy Agency. (1998). *Benign Energy?: The Environmental Implications of Renewables*. Paris: OECD Publications.

International Energy Agency. (1999). *The Evolving Renewable Energy Market*. Paris: OECD Publications.

International Solar Energy Society (ISES) web site International Solar Energy Society. August 9, 2000 <http://www.ises.org/>.

Johansson, T. B.; Kelly, H.; Reddy, A. K. N.; and Williams, R. H., eds. (1993). *Renewable Energy: Sources for Fuels and Electricity*. Washington, DC: Island Press.

Marion, W., and Wilcox, S. (1994). *Solar Radiation Data Manual for Flat-Plate and Concentrating Collectors*. Golden, CO: National Renewable Energy Laboratory.

National Renewable energy Laboratory (NREL) Home Page. National Renewable Energy Laboratory. August 9, 2000 <http://www.nrel.gov/>.

Passive Solar Industries Council. <http://www.psic.org/>

Pegg, D. *CADDET Renewable Energy—International demonstration projects* (July 24, 2000). International Energy Agency, Center for the Analysis and Dissemination of Demonstrated Energy Technologies: Renewable. August 9, 2000 <http://www.caddet-re.org/>.

President's Committee of Advisors on Science and Technology, Panel on Energy Research and Development. (1997). *Federal Energy Research and Development for the Challenges of the Twenty-First Century*. Washington, DC: U.S. Government Printing Office.

U.S. Congress, Office of Technology Assessment. (1995). *Renewing Our Energy Future*. Washington, DC: U.S. Government Printing Office.

U.S. Department of Energy, Energy Information Administration. (1999). *Renewable Energy Annual 1998*. Washington, DC: U.S. Government Printing Office.

U.S. Solar Energy Industries Association. <http://www.seia.org/>.

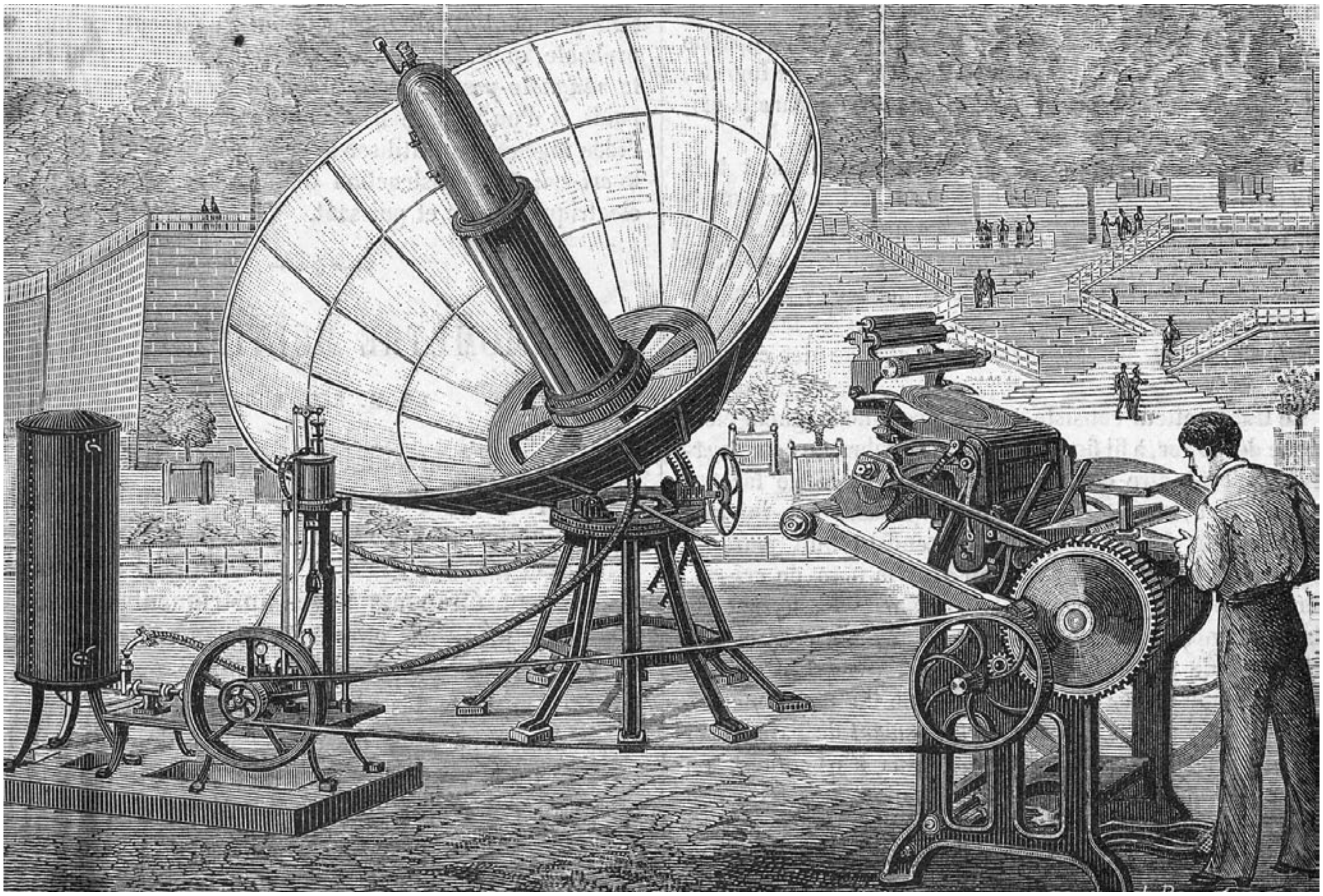
Welcome to the American Solar Energy Society American solar Energy Society. August 9, 2000 <http://www.ases.org/>.

SOLAR ENERGY, HISTORICAL EVOLUTION OF THE USE OF

Every ninety-five minutes an object the size of a bus circles Earth. It represents an unparalleled achievement in space observation technology and is called the Hubble Space Telescope (HST). Since its deployment in 1990, the HST has brought images to the astronomy community that traditionally had been reserved for expensive long-range space missions by unmanned craft. One of the essential technological developments that make the HST such an attractive investment is its use of one "natural resource" found in space, the sun. The HST utilizes photovoltaic technology to capture the sun's rays and convert them directly into electricity. This procedure is commonly referred to active solar generation.

Active solar generation is nothing new to the space industry. Photovoltaics have been used to power satellites since the late 1950s and are an essential element to the new international space station. Closer to home, active solar technologies are used on houses and buildings to generate heat, hot water, and electricity. Spacecraft use of the sun is indeed a technological marvel. However, strategic use of the sun for heating and cooling is not a new discovery. In fact, 2,400 years ago the Greek philosopher Socrates observed:

> Now in houses with a south aspect, the sun's rays penetrate into the porticos in winter, but in the sum-



The operation of a solar-powered printing press. Experiment conducting in the Garden of Tuileries, Paris, on Aug 6, 1882, for the festival of "L'Union Francaises of Jeuenesse." (Corbis Corporation)

> mer, 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 winter winds. (Xenophon, 1994, chap. 2, sec. 8)

Socrates has described the strategic placement of a home to best utilize the rotation of Earth, to receive maximum sunlight during the winter and minimal sunlight in summer. The strategic utilization of the seasonal course of the sun in home design is called passive solar architecture. Passive solar use is best described as the greenhouse effect. This simple phenomenon is illustrated by the experience of returning to your car on a sunny, cool day and finding it heated. Hubble and Socrates represent opposite ends of the solar historical spectrum. However, solar technologies have historically been met with mixed results. While the technology has consistently proved itself to be useful and environmentally friendly, it has often been pushed into the shadows by cheaper, less environmentally friendly energy technologies.

Historically the most common problems associated with solar technologies are efficiency and economics compared to competing energy technologies. During the Industrial Revolution coal was relatively cheap and, for the most part, readily available in most of the world. These facts, which still hold true today, make coal a powerful competitor against alternative power technologies. However, solar technologies have been able to survive in niche markets long enough to increase efficiency and lower costs. Although sunlight is free and also readily available, solar technologies have experienced little success in large-scale applications. Solar technologies also have been able to take advantage of advances in silicon technology. An initial niche market for solar technology was the colonies of Europe.

EARLY APPLICATIONS: SOLAR THERMAL

One of the earliest forays in the use of solar technologies for power generation was in 1861. French engineer Auguste Mouchout utilized solar thermal technology by focusing the sun's rays to create steam, which in turn was used to power engines. Mouchout's design incorporated a water-filled cauldron surrounded by a polished metal dish that focused the sun's rays on the cauldron, creating steam. Mouchout was granted the first solar patent for an elementary one-half horsepower steam engine powered by a solar reflective dish. Several of Mouchout's solar motors were deployed in Algeria, then a French protectorate, where coal was expensive due to transportation costs. One-half horsepower may not seem like much, but it was a considerable advance in alternative power generating. At the time, the major source of energy was man and draft animals, the steam engine was in its infancy, and water power was not very portable. Mouchout was able to expand his design to a larger scale, but the project was abandoned due to new innovations in coal transportation that made coal a more efficient investment.

Solar thermal technology survived Mouchout and was further developed by Swiss-born engineer John Ericsson. In 1870 Ericsson created a variation of Mouchout's design that used a trough instead of a dish to reflect the Sun's rays. Ericsson's basic design would later become the standard for modern-day parabolic troughs. He allegedly refined his design to the point where it would be able to compete actively with conventional generation techniques of the time but he died without leaving any plans or blueprints.

Englishman William Adams designed another interesting variation of Mouchout's work in 1878. Adams's design used mirrors that could be moved to track sunlight and focus the rays on a stationary cauldron. In theory, a larger number of mirrors would generate increased heat levels and greater steam output for larger engines. Adams was able to power a two-and-a-half-horsepower engine during daylight hours and viewed his invention as an alternative fuel source in tropical countries. Although Adams was more interested in proving that it could be done than in developing the technology, his basic design would become the prototype for contemporary large-scale centralized solar plants called "power towers" that are currently in operation in the United States and other countries.

CONTEMPORARY USES: SOLAR THERMAL

Solar thermal technology varied little in design from 1870 to 1970. Advances were made in the materials, such as lower-boiling-point fluids and higher-efficiency motors, but the basic design of using mirrors to focus sunlight remained the same. The aforementioned introductory ventures in the use of solar technology provided the intellectual foundation for growth and extensive technological development of solar use in the twentieth century. These pioneers helped to establish solar as a viable alternative to conventional power generation. However, many of the same barriers resurfaced, preventing widespread implementation. Many of the previously mentioned designs survived by gaining support within environmental communities as environmentally friendly alternatives to fossil fuels, as well as a safer alternative to nuclear power. In addition, solar technologies were able to establish themselves within new niche markets and to gain further interest during the oil crisis of the 1970s.

As previously mentioned, both Ericsson's and Adams's original designs were upgraded and implemented as modern solar power generating facilities. Contemporary versions of the solar trough and power tower were constructed during the 1980s in the Mojave Desert of California by LUZ International Ltd. Currently the corporate successor of LUZ, which went bankrupt in 1991, partly due to dependence on a federal subsidy (tax credits) not renewed in 1991, has nine solar thermal plants in California and has generated more energy from solar thermal than any other company in the world.

EARLY APPLICATIONS: PHOTOVOLTAICS

Use of solar panels or photovoltaics (PVs) is another popular way to generate solar electricity. The space program is perhaps the most recognized user of PVs and is responsible for most of the advancements in PVs. Many people are familiar with PVs through small applications such as calculators and perhaps solar water heaters, but early forays in PV experimentation were little more than noted side observations in non-PV experiments.

In 1839 Alexandre-Edmund Becquerel, a French experimental physicist, did the earliest recorded experiments with the photovoltaic effect. Becquerel discovered the photovoltaic effect while experimenting with an electrolytic cell made up of two metal

Buckminster Fuller stands in front of a solar-powered geodesic dome house. (AP/Wide World Photos)

electrodes placed in an electricity-conducting solution. When exposed to sunlight, the generation increased. Between 1873 and 1883 several scientists observed the photoconductivity of selenium. American inventor Charles Fritts described the first experimental selenium solar cells in 1883. Notable scientists such Albert Einstein and Wilhelm Hallwachs conducted many experiments with various elements to explore it further. In 1921 Einstein won the Nobel Prize in Physics for his theories explaining the photoelectric effect.

Many elements were found to experience the photoelectric effect. Germanium, copper, selenium, and cuprous oxide comprised many of the early experimental cells. In 1953 Bell Laboratories scientists Calvin Fuller and Gerald Pearson were conducting

Parabolic solar energy concentrating dishes sit in a row at the power station at White Cliffs, Australia. The plant, operated by Solar Research Corp. of Melbourne, is Australia's first experimental solar station. (Corbis Corporation)

experiments that brought the silicon transistor from theory to working device. When they exposed their transistors to lamplight, they produced a significant amount of electricity. Their colleague Darryl Chaplin was looking for better ways to provide power for electrical equipment in remote locations. After a year of refinements Bell Laboratories presented the first working silicon solar cell, which would later become the core of contemporary PV applications.

CONTEMPORARY USE: PHOTOVOLTAICS

The utility industry has experimented with large-scale central station PV plants but concluded that in most cases they are not cost-effective due to their perceived low efficiency and high capital costs. However, PV has proven useful in niche markets within the utility system as remote power stations, peak generation applications, and in a distributed generation scheme. Distributed generation is the inverse of central generation. Instead of constructing large generating facilities energy has to be sent to consumers, energy is generated where it is used. Highway signs are a good example of distributed generation. They use solar panels and batteries on the signs to collect and store the power used to illuminate the signs at night. Because very energy-efficient LED lighting replaced incandescent bulbs, these signs could be powered much more economically by PVs than by traditional gasoline-powered generators. Perhaps the most prolific use of PVs is in the space program. The sun, the only natural energy source available in space, is the logical choice for powering space vehicles. Photovolttaics are safer and longer-lasting energy sources than nuclear power is.

An instance where large-scale grid-connected PV generation has occurred is the Sacramento Municipal Utility District (SMUD). SMUD built several dozen "solar buildings" and placed PV systems directly on their clients' homes. SMUD purchased one megawatt of distributed PV systems each year until the year 2000.

SMUD's approach introduced another distributed application concept for PV, called building integrated photovoltaics or BIPV. Currently Japan has one of the largest BIPV programs, started in 1996. The simplest application of BIPV is the installation of solar panels directly on a building. While this method is perhaps the easiest, it does not reflect the best use of PV technology. The best potential for the growth of PV is for PV to double as building materials. Examples include exterior insulation cover, roofing shingles, windows, and skylights that incorporate solar cells. Incorporating PV into construction materials is very complex and requires joint cooperation among traditionally separate entities of the building sector. Although no actual products exist, many of the major manufacturers began experimenting with prototypes in the late 1990s and hope to introduce cost-effective products on the market by 2005. Additional uses for PV include desalinization of seawater for fresh water, solar water heaters, and solar ovens.

IMPLEMENTATION AND BARRIERS

Solar energy is used worldwide in many applications, from niche markets in developed countries to primary village power in rural and developing communities. Its attractiveness can be attributed to several factors:

- Solar power works well in small facilities, and it may be comparatively cost-effective when all factors (transmission, fuel transportation costs, service life, etc.) are taken into consideration. In many instances solar energy is the *only* viable energy option in the region.
- Solar distributed energy systems provide security and flexibility during peak use and outages associated with environmental disasters. In addition, programs that allow consumers to sell surplus energy they generate back to the utility, called “net metering,” will make residential solar technologies more attractive.
- Solar power is environmentally friendly and will help mitigate global environmental concerns associated with climate change.

These claims have been the primary call to arms for many solar advocates and industries since the 1970s, but solar technologies have not enjoyed widespread use except within niche markets, especially among developed countries. According to a U.S. Department of Energy report by the Energy Information Administration, solar technologies have only experienced modest sector growth domestically because domestic solar thermal technologies face stiff competition from fossil-fueled energy generation. PV growth faces similar challenges, since deregulation is likely to lower the cost of electric power from natural gas and coal.

Because of the wide range of ideological leanings of various administrations, direct subsidies and federally funded research and development have fluctuated wildly since the late 1970s. In 1980 the United States spent $260 million on solar research and development, but then the amount fell to $38 million by 1990. Between 1990 and 1994 support doubled, reaching $78 million.

Because of the problems associated with nuclear power and burning fossil fuels, solar technologies may again be viewed as a viable alternative. However, solar technology faces several challenges

First. the energy expended in the production of PV panels is considerable. Second, solar cells become less efficient over time (degradation is slow, but eventually they lose most of their conductivity). Third, contemporary solar cells use mercury in their construction and present toxic waste disposal problems. Finally, solar technologies are part-time power sources. Solar thermal technology requires full, direct sunlight, and PVs do not work at night (although batteries and supplementary generating applications such as wind power can be combined to provide full-time generation).

The need for long-term clean power will be a controversial but necessary issue in the twenty-first century. The debate will focus on the deregulation of the power industry and the needs of consumers, coupled with what is best for the environment. For certain applications, solar technology has proven to be efficient and reliable. The question remains whether solar will be wholly embraced as energy options for a sustainable future in a climate of cheap fossil fuels and less than enthusiastic public support.

J. Bernard Moore

See also: Batteries; Becquerel, Alexandre-Edmund; Einstein, Albert; Electric Power, Generation of; Fuel Cells; Heat and Heating; Heat Transfer; Power; Renewable Energy; Spacecraft Energy Systems; Solar Energy; Thermal Energy; Water Heating.

BIBLIOGRAPHY

Berger, J. J. (1997). *Charging Ahead: The Business of Renewable Energy and What It Means for America*. New York: Henry Holt.

Flavin, C., and Lenssen, N. (1994). *Power Surge: Guide to the Coming Energy Revolution*. New York: W. W. Norton.

Fitzgerald, M. (July 14, 2000). *The History of PV.* Science Communications, Inc. August 9, 2000 <http://www.pvpower.com/pvhistory.html>

Hoffman, J. S.; Well, J. B.; and Guiney, W. T. (1998). "Transforming the Market for Solar Water Heaters: A New Model to Build a Permanent Sales Force." *Renewable Energy Policy Project* No. 4. <http://www.repp.org>.

Maycock, P. (1999). "1998 World PV Cell Shipments Up 20.5 Percent to 151.7 MW." *PV News* 18:2.

National Renewable Energy Laboratory. (1996). *Photovoltaics: Advancing Toward the Millenium*. DOE/GO-100095-241, DE96000486.

Perlin, J. (1999). "Saved by the Space Race." *Solar Today* (July-August).

Smith, C. (1995). "Revisiting Solar Power's Past." *Tech-nology Review* 98(5):38–47.

U.S. Department of Energy, Energy Information Administration. (1996). *Annual Energy Outlook 1997 with Projections to 2015*. Washington, DC: U.S. Government Printing Office.

Wan, Yih-huei. (1996). *Net Metering Programs*. NREL/SP-460-21651. Golden, CO: National Renewable Energy Laboratory.

Xenophon. (1994). *Memorabilia*, tr. A. L. Bonnette. Ithaca, NY: Cornell University Press.

SPACECRAFT ENERGY SYSTEMS

Energy systems in space technology are devices that convert one kind of energy into another to ensure the functioning of automated and piloted satellites, interplanetary probes, and other kinds of spacecraft. Multiple functions of any spacecraft require two distinctly different energy sources: propulsion for launch and maneuvers, and electricity supply to power the onboard equipment.

PROPULSION SYSTEMS

Propulsion generates kinetic energy to facilitate motion of spacecraft. Although propulsion is provided by various types of devices, the most common propulsion system of modern space technology is a rocket engine, a device that propels a rocket by a force of reaction. In its simplest form, the rocket engine could be described as a container (combustion chamber) with an opening at one end (exhaust). Burning of combustible chemicals inside the container generates a large amount of a compressed gas—a propulsive mass—that is ejected through the exhaust. According to the laws of mechanics, the force of pressure that ejects the gas creates the reaction force, called thrust, that pushes the container in the direction opposite to the exhaust. The International System of Measurements designates thrust in a unit of force called a newton (N= 0.098 kg[f] = 0.216 lb[f]), while more often it is measured in tons, kilograms, or pounds. The thrust is greater if a bell-shaped nozzle is installed at the engine's exhaust. The nozzle's specially designed profile (the so-called Laval nozzle) creates a strong pressure differential at the exhaust. This differential facilitates expansion of the ejected gas with a substantially higher ejection velocity, thus releasing greater kinetic energy.

The energy efficiency of a rocket engine is determined by specific impulse, commonly measured in seconds (although scientifically correct designations are N s/kg, or m/s). Specific impulse is the time in which a rocket engine consumes the amount of propellant equal to the engine's thrust. For example, compare two engines with an equal thrust of 100 kg each. If the first engine consumed 100 kg of propellant in 20 seconds, and the second one consumed the same amount in 30 seconds, the second engine generated greater kinetic energy per kilogram of propellant—it worked longer. Thus the higher specific impulse indicates the more efficient engine.

The most common chemical rocket engines use gaseous propulsive mass produced through the burning of special chemicals called propellants. Thermal energy released in this process accelerates and ejects the propulsive mass, converting itself into kinetic energy of motion. Solid and liquid propellant motors are the two major types of chemical rocket engines. The solid rocket motor (SRM) burns propellant that consists of fuel and oxidizer mixed together in a solid-state compound. Solid propellant is stored inside a combustion chamber of the SRM, making the engine relatively simple and reliable due to a lack of moving parts (see Figure 1). However, because of lower specific impulse and difficulty controlling the final speed, since the SRM usually cannot be shut down and restarted until all propellant is consumed,

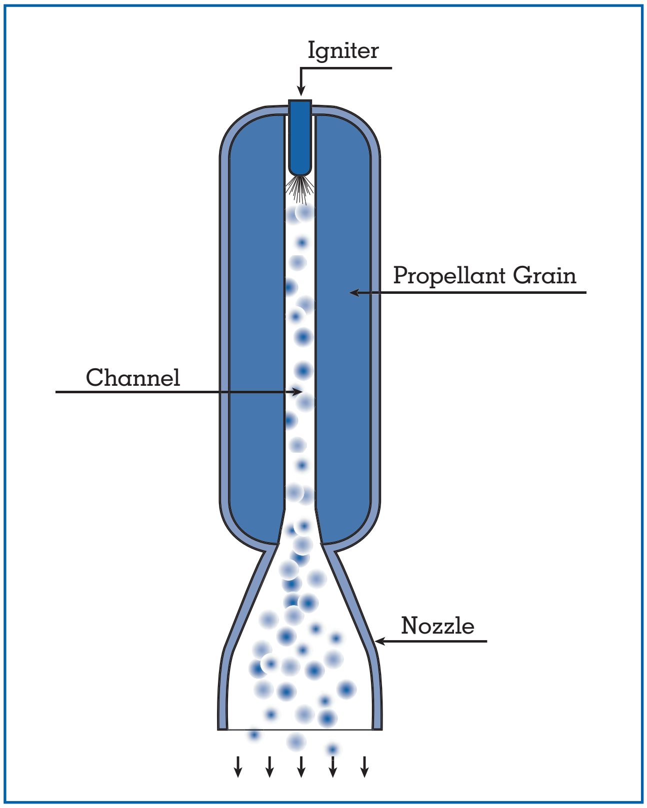


Figure 1.
Solid rocket motor.

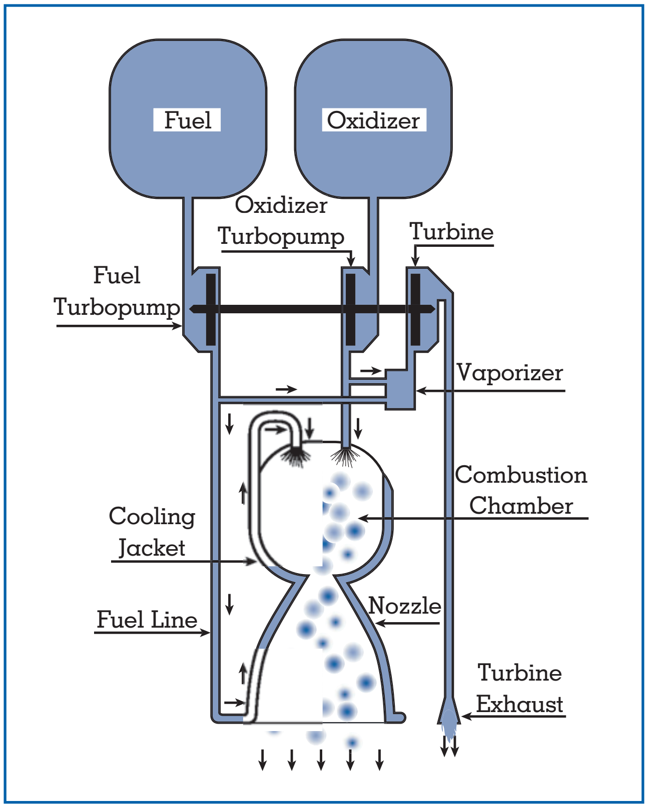


Figure 2.
Liquid propellant rocket engine.

the solid rocket is primarily for short duration but powerful impulse.

A liquid propellant rocket engine (LPRE) generates propulsive mass through burning of two liquid chemicals—fuel and oxidizer—which are pumped into a combustion chamber from separate tanks. LRPE has a very complex design: It contains numerous components with a large number of moving parts. The combustion chamber and nozzle have intricate structures that allow them to withstand high internal pressure while being lightweight. One of the propellant components is driven between the thin double walls of the chamber and nozzle to cool them and prevent hot exhaust from burning through (see Figure 2). LPRE complexity is compensated by advantages of better energy efficiency (higher specific impulse); precision control over burn time, thrust and final speed; and the possibility to restart the engine.

There are two major types of liquid propellants: a storable propellant (that does not boil under a normal temperature) and a cryogenic propellant (a liquefied gas under very low temperature). Storable propellant engines are simpler and more reliable. They are useful for spacecraft that require multiple propulsion impulses and a long stay in space. Although more complex and demanding, cryogenic engines are regarded as best for launch vehicles due to their superior energy efficiency.

In each space mission, the maximum amount of energy is used for spacecraft launching. This is because the spacecraft had to be accelerated to a velocity of 28,500 km/h (17,700 miles/h) in fewer than 10 minutes, while overcoming forces of gravity and atmospheric drag. Due to the enormous energy requirements for this task, propellant accounts for up to 92 percent of the total mass of modern launch vehicles, leaving no more than 4 percent for the useful payload (the rest is the “dry mass” of a rocket). That dictates the importance of energy efficiency—better engines allow for the launching of heavier payloads with less propellant. Although the overall efficiency of a launch vehicle depends on many fac-

Designation & Country	*Number of Stages & Propellant*	*Launch Mass (LM)*	*Payload Mass in low orbit (PM)*	*PM/LM Ratio*	*Comments*
Delta II 7925 USA	I - liquid cryogenic + 9 solid boosters II - liquid storable	230 tons (506,000 lb)	5,089 kg (11,220 lb)	2.21%	One of the newer models in a family of SLVs converted from a ballistic missile. Operational since 1990.
Soyuz-U Russia	I - 4 liquid cryogenic boosters II - liquid cryogenic III - liquid cryogenic	310 tons (682,000 lb)	7,000 kg (15,400 lb)	2.25%	One of the models in a family of SLVs converted from a ballistic missile. Operational since 1973.
Atlas II AS USA	I - liquid cryogenic + 4 solid boosters II - liquid cryogenic	234 tons (516,000 lb)	8,640 kg (19,000 lb)	3.69%	One of the newer models in a family of SLVs converted from a ballistic missile. Operational since 1993.
CZ-2E China	I - 4 liquid storable boosters II - liquid storable III - liquid storable	464 tons (1,021,000 lb)	8,800 kg (19,360 lb)	1.90%	One of the newer models in a family of SLVs converted from a ballistic missile. Operational since 1990.
Arian-44L Arianspace European Consortium	I - liquid storable + 4 liquid storable boosters II - liquid storable III - liquid cryogenic	470 tons (1,034,000 lb)	9,600 kg (21,120 lb)	2.04%	One of the newer models in a family of the purpose-built SLVs. Operational since 1989.
Zenit-2 Russia Ukraine	I - liquid cryogenic II - liquid cryogenic	459 tons (1,010,000 lb)	13,740 kg (30,300 lb)	2.99%	Purpose-built SLV. A modified version is intended for Sea Launch international project. Operational since 1985.
Proton-K Russia	I - liquid storable II - liquid storable III - liquid storable	689 tons (1,516,000 lb)	20,000 kg (44,000 lb)	2.90%	One of the models in a family of SLVs converted from a ballistic missile. Operational since 1971.
Titan IV USA	I - liquid storable + 2 solid boosters II - liquid storable III - liquid cryogenic	860 tons (1,900,000 lb)	21,640 kg (47,600 lb)	2.51%	One of the newer models in a family of SLVs converted from a ballistic missile. Operational since 1989.
Space Shuttle USA	I - 2 solid boosters II - liquid cryogenic	2,040 tons (4,500,000 lb)	29,500 kg+ orbiter 94,000 kg (65,000 + 207,000 lb)	1.45%	The world's only operational reusable launch vehicle (RLV). Purpose-built. Operational since 1981.
Venture Star (project) USA	Single Stage-to-Orbit liquid cryogenic	993.6 tons (2,186,000 lb)	20,500 kg + RLV 117,000 kg (45,100 + 257,000 lb)	2.06%	Future US reusable launch vehicle (RLV). Under development. Expected to start operations in 2005.

Table 1.
Space Launch Vehicles (SLVs)

tors, it could be roughly estimated by a simple percentage ratio between the rocket's maximum payload mass (PM) and of its total mass at launch (LM). The greater percentage would indicate a more efficient launch vehicle (see Table 1).

Energy efficiency is one of the major factors that determines the launch vehicle layout and choice of engines. Currently, about 98 percent of all spacecraft worldwide are launched by liquid-propellant rockets. While the purpose-built launchers mostly use cryogenic engines to ensure the best performance, many current space rockets utilize storable propellants. This is because the majority were developed from military ballistic missiles. Although not ideal from the energy efficiency standpoint, such conversion is less expensive than developing a brand-new launch vehicle. Solid-propellant engines are primarily used as auxiliary boosters for some liquid propellant rockets. Only a few purely solid-propellant space launch vehicles exist today in the world because they are capable of placing fewer than 1,000 kg (2,200 lb) into low orbit, while liquid-propellant rockets can launch much more—up to 29,500 kg (64,900 lb).

Some unique launchers were even more powerful. Such was *Saturn V*—a gigantic rocket capable of launching a payload of 140,000 kg (308,000 lb) that allowed American astronauts to land on the moon. Russian *N-1* and *Energia* rockets were capable of launching 95,000 kg (209,000 lb) and 105,000 kg (231,000 lb), respectively. All these launch vehicles are not in use anymore due to their enormous cost. These giants were created for specific national prestige goals, such as manned lunar expedition, and for launching of the expected massive civilian and military payloads. The shifts in political and economic priorities in the space programs of the United States and Russia have left these rockets without any feasible payload assignments. At the same time, the cost of the launch vehicles' preservation along with their production and maintenance facilities for possible future use was proven to be prohibitive and unjustifiable economically.

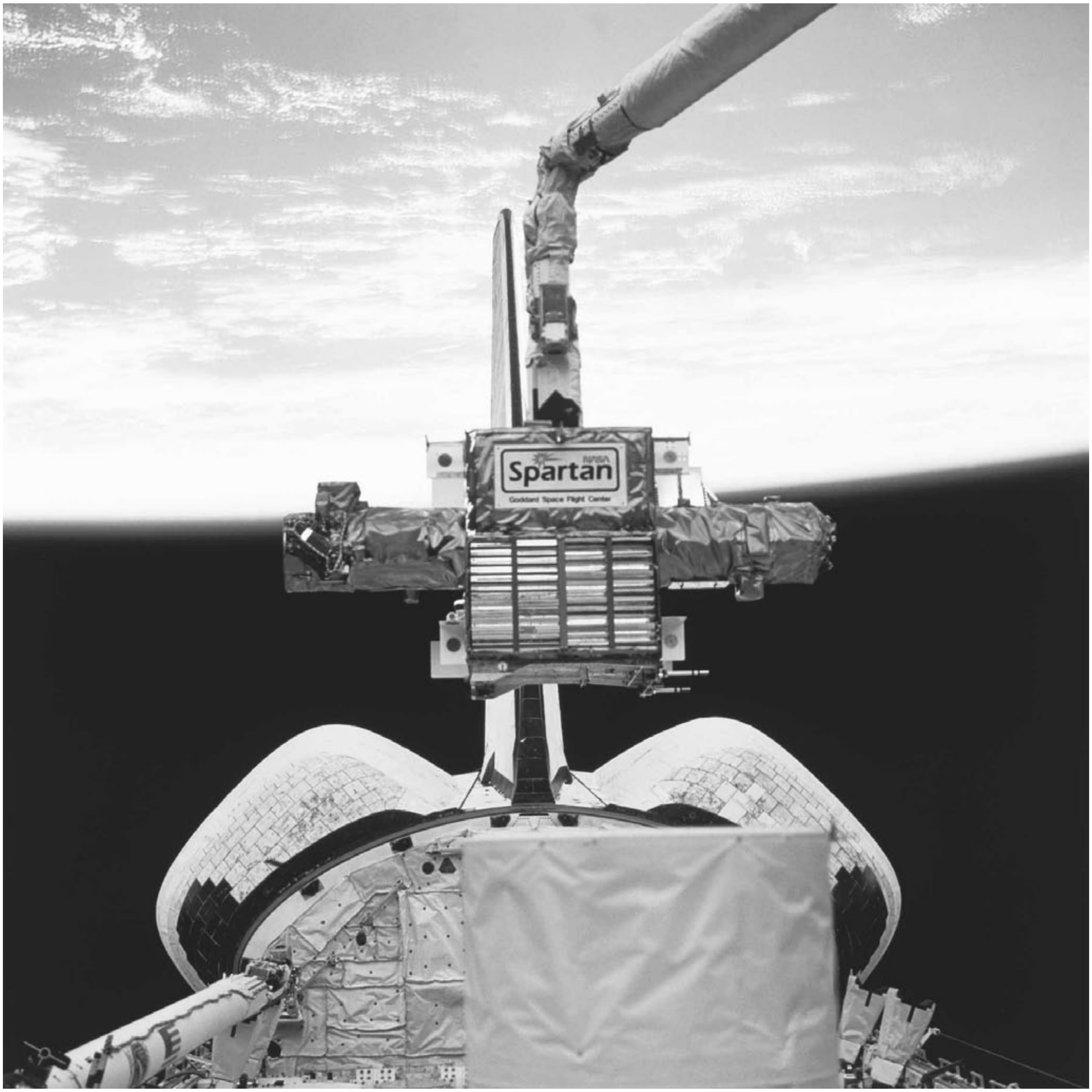
Space launchers usually consist of several stages, which are the individual rockets stacked together. This is another method to increase efficiency—the stages are ejected as soon as their propellant is spent, and a rocket does not have to carry empty tanks. The number of stages also depends on engines' effectiveness. Liquid-propellant rockets usually consist of two to three stages, while solid ones have to have three to five stages.

Rocket engines are also used for maneuvers in space. Some operations, such as a onetime transfer of a satellite from lower to higher orbit, could be performed by a solid-propellant engine. Yet many complex maneuvers, such as rendezvous and docking with another spacecraft, require multiple engine firings and variable power impulses. Hence modern spacecraft are equipped with an assortment of attitude control engines that usually use liquid storable propellant.

A good illustration of how all these engines work together in a single system is the American Space Shuttle. Two very large solid rocket boosters (SRBs) represent the first stage of the launch vehicle. The SRBs are the most powerful rocket engines in existence: Each produces more than 1,500 tons (15.3 MN, or 3,310,000 lb) of thrust for 122 seconds. The second stage consists of a large external tank with cryogenic propellant (liquid hydrogen and liquid oxygen) and three Space Shuttle main engines (SSMEs) mounted on the orbiter. Each SSME produces more than 170 tons (1.73 MN, or 375,000 lb) of thrust at sea level. At launch, SRBs and SSMEs ignite just a few seconds apart and are used to break through the most dense layers of atmosphere. After a separation of the SRBs at the altitude of about 40 km (25 mi), the Space Shuttle continues its acceleration with the SSMEs only. Just before the Space Shuttle reaches orbital velocity (520 seconds after launch), the SSMEs are shut down and the external tank is ejected. Now two smaller engines of the orbital maneuvering system (OMS) aboard the orbiter provide the final acceleration. The OMS engines are also used for orbital maneuvers and for deceleration before the spacecraft reentry. Unlike the SSMEs, they utilize storable liquid propellant, and each produces thrust of 3 tons (30.6 kN, or 6,700 lb). Very fine Space Shuttle orbital maneuvering is performed by the reaction control system (RCS), which consists of 44 small liquid-propellant engines assembled in three arrays.

The immense advantages to humanity of modern space technology come at a hefty price. High risks and environmental hazards of rocketry, uniqueness of spacecraft equipment, and the necessity for specialized launching facilities make space technology very expensive. Today the average cost to launch just 1 kg (2.2 lb) of a payload in orbit could be as high as $10,000. Because of that, only six countries—the United States, Russia, China, Japan, India, and Israel—and the France-dominated European Space Agency (ESA) maintain independent space launching capabilities.

One of the seemingly obvious ways to cut down the cost is a reusable launch vehicle, such as the Space Shuttle. NASA hoped that the enormous cost of the Space Shuttle development would have been compensated by the expected tenfold reduction in the price of putting a payload in orbit. Unfortunately, that goal was not achieved. Despite the Shuttle's unique capabilities, the cost of its payload delivery



The Space Shuttle *Discovery*'s robotic arm deploys the SPARTAN-201 satellite. (Corbis-Bettmann)

appeared to be at least two times higher than that of expendable rockets. The Shuttle's maintenance after each mission turned out to be much more expensive than had been expected. As a reusable spacecraft, the orbiter has to spend energy carrying into space additional massive components that are not essential for payload deployment, such as the main engines, wings, and heat protection. Thus the Shuttle's efficiency factor is two times lower than of the best expendable launchers. The very large size of the orbiter and the necessity to accommodate a crew even for routine satellite launches substantially increase the overall cost of Space Shuttle operations. Projected high expenses recently forced Russia and

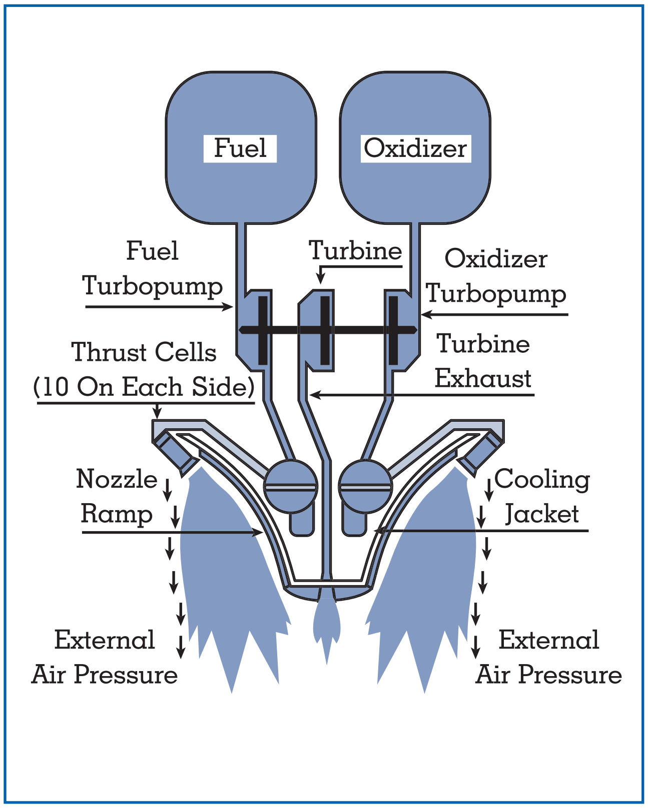


Figure 3.
Aerospike rocket engine.

the ESA to abandon their own nearly completed space shuttle programs.

The numerous projects of the reusable launch vehicles that have appeared from the mid-1980s through the 1990s are intended to overcome the Space Shuttle drawbacks. Some vehicles are designed as conventional rockets that would use parachutes, helicopter rotors, or rocket engines for landing. Others retain the airplane configuration but are planned to be more compact. Regardless of configuration, almost all these projects rely on new types of rocket engines that have yet to be developed.

The most advanced among today's projects is the future American reusable launch system known as *Venture Star*. It utilizes two new concepts that are expected to boost efficiency and cut down the cost of a payload. The first is the single stage-to-orbit concept—the launch vehicle carries the entire propellant load and does not have any expendable parts. The latter is planned to be achieved by a revolutionary rocket engine design called *Aerospike*. The bell-shaped nozzle of a conventional rocket engine controls the expansion of exhaust gases and influences the engine's thrust and efficiency. Since the optimal gas expansion is different at various altitudes, the lower and upper stages of conventional rockets have engines with nozzles of different shapes. In *Aerospike*, an array of combustion chambers fire along the curved central body, called a ramp or a spike. The spike plays the role of an inner wall of a nozzle, while the outer wall is formed by the pressure of the incoming air flow. Changes in atmospheric pressure upon the vehicle's ascent immediately change the exhaust expansion, allowing the engine to perform optimally at any altitude (see Figure 3). The vehicle will take off vertically but land like an airplane, using its lifting body configuration. The *Venture Star* is planned to be about 30 percent smaller than the Space Shuttle. The *Venture Star*'s payload will be 20,500 kg (45,100 lb), compared of the Shuttle's 29,500 kg (64,900 lb). If successful, the *Venture Star* is to make its first flight in the first decade of the twenty-first century. Meanwhile, NASA prepares to fly a scaled-down technology demonstrator, *X-33*, which will test the *Venture Star* design.

Another method to increase efficiency is to launch a rocket eastward close to Earth's equatorial plane. It would be additionally accelerated by Earth's rotation, achieving greater payload capability without any extra energy consumption. Small rockets could be launched from an airplane, which makes building expensive launching facilities unnecessary. This concept is currently utilized by the American Pegasus launch system. Larger rockets could take off from floating launchers, such as the San Marco platform, which had been used by Italy for the American Scout rockets in the 1960s. The end of the Cold War allowed the use of existing military technology for this purpose. For example, the Russian Navy successfully launched a small satellite by a ballistic missile from a submerged strategic submarine in 1998. The largest floating launch platform was built in 1998 in Norway for the Sea Launch international consortium. It operates from the United States, making commercial launches of heavy rockets jointly built by Russia and Ukraine.

In their quest for energy efficiency, some designers are seeking radically new methods of launching spacecraft. For example, leaving the energy source on the ground instead of being carried by a spacecraft would make the launching more efficient. One such method is firing a satellite from a large cannon—something that revives a space travel idea from Jules Verne's science fiction classic *From the Earth to the Moon* (1865). In his joint experiments with the U.S. Navy in the late 1960s, the Canadian ballistics expert Gerald V. Bull proved that small satellites could indeed survive tremendous shock and acceleration in a cannon barrel and achieve orbital velocity. Instead of conventional cannons, very long chemical accelerators with numerous combustion chambers along the barrel are considered. In this case, the pressure of expanding gas would not decrease, but continuously be built behind a projectile, providing it with greater velocity. Among variations of that concept are electromagnetic accelerators, where the force of chemical explosion is replaced by electromagnetic energy. A typical example is a coil gun, which accelerates a projectile by magnetic fields created in sequence by electric coils wrapped around a barrel. Unlike cannons, electromagnetic accelerators promise a smoother rate of acceleration and better control over the projectile's velocity. They also are much "friendlier" to the environment in terms of lack of excessive noise and chemical exhaust pollution. Although attractive, such launchers would require consumption of enormous amounts of electricity, as well as construction of large and expensive facilities. Experiments with chemical and electromagnetic accelerators have been under way in the United States, Russia, and some other countries for many years.

The choice of exotic propulsion systems is even broader in open space, where the lack of gravity and atmospheric drag allows a wide selection of low-thrust but economical engines. For example, thermal energy produced by a controlled chain reaction in a nuclear reactor can heat a propulsive mass to a higher extent than in chemical rocket engines. Higher temperature means greater ejection velocity and higher specific impulse. This concept of a nuclear rocket engine was thoroughly researched in the United States and the USSR from the 1960s to the 1980s, and working prototypes were ground-tested in both countries. For instance, the American NERVA (nuclear engine for rocket vehicle application) engine used liquid hydrogen as a propulsive mass and had a sea-level thrust of 29.4 tons (300 kN, or 64,680 lb). Figure 4 shows a principal design of the nuclear rocket engine with a solid-state reactor.

The heating elements in modern nuclear reactors are made of solid radioactive materials, but theoretically they could exist in a colloidal (semiliquid) or gaseous state. That would allow a radical increase in propulsive mass temperature and ejection velocity. Research has shown that the colloidal and gaseous-state nuclear reactors would respectively produce specific impulses from 20 percent to 50 percent higher than the solid-state reactor. Some very advanced concepts of gaseous-state reactors promise even greater efficiency.

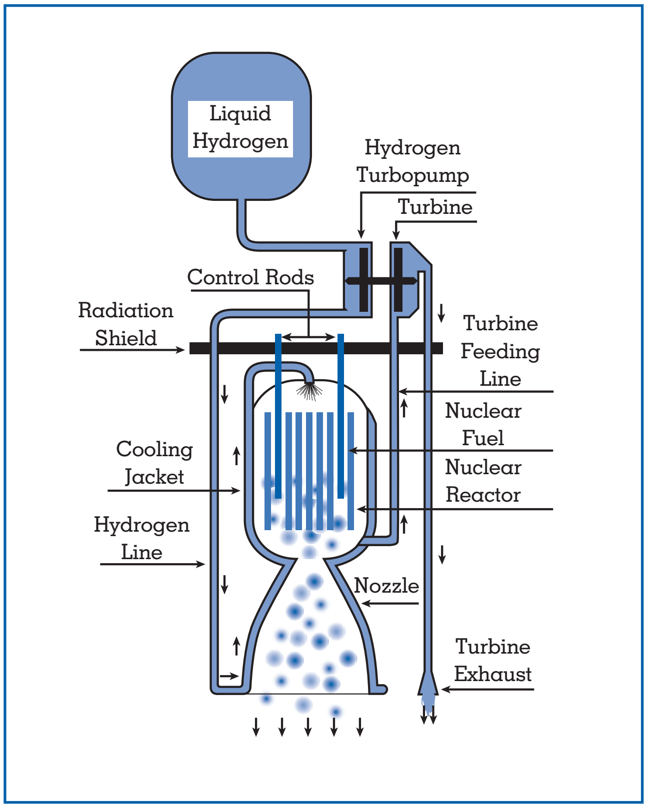


Figure 4.
Nuclear rocket engine.

Another attractive idea is to use electric power for propulsion. Having many design variations, the electric rocket engines (ERE) could roughly be divided into three groups:

- thermal electric engines, which eject propulsive mass heated by the electricity;
- magneto-hydrodynamic (plasmic) engines, which convert propulsive mass into plasma and accelerate it by a magnetic field;
- ion engines, where an ionized propulsive mass is ejected by an electrostatic field.

Type	*Maximum Thrust*	*Continuous Firing Time*	*Specific Impulse*	*Comments*
Chemical Solid Propellant Rocket Engines	up to 1,500,000 kg	dozens of seconds	270 seconds	Widely used in launch vehicles and spacecraft as auxiliary boosters and, sometimes, for main propulsion.
Chemical Liquid Propellant Rocket Engines	up to 800,000 kg	dozens of minutes	388 seconds	Main type of propulsion system of modern space technology.
Nuclear Solid State Rocket Engines	up to 30,000 kg	dozens of minutes	700 - 900 seconds	Under development. Prototypes were ground tested in the 1960s - 1980s.
Electric Rocket Engines	up to 0.15 kg	thousands of hours	1,200 - 25,000 seconds	Under development. Used to stabilize some satellites and as a main propulsion of a planetary probe.

Table 2.
Space Propulsion Systems

Experiments showed that the EREs can utilize a variety of chemical elements for propellant: from metals (cesium, rubidium, mercury) to gases (hydrogen, argon, xenon). The main advantage of these engines is in unsurpassed efficiency. They produce very low thrust but can continuously work for thousands of hours, consuming very little propellant due to their unparalleled specific impulse. In open space, slow but steady acceleration eventually produces the same result as a powerful but short-duration impulse—the spacecraft achieves a desired velocity. It is just a matter of time. Yet the low power and high electricity consumption so far remain the main difficulties in the ERE development.

The first experiments with the thermal electric engine were conducted in Russia in 1929 by its inventor, Valentin P. Glushko, who later became a world-famous authority in rocket propulsion. For more than forty years, the United States and Russia have devoted many resources to research and development of various kinds of EREs. First tested in space by the Russians in 1964, these engines have found some limited applications in modern space technology. For more than two decades Russian weather and communication satellites have regularly used electric rocket engines for orbital stabilization. The first spacecraft to employ ERE for main propulsion was the American asteroid exploration probe *Deep Space 1*, launched in 1998. The performance of its ion engine is illustrated by the fact that after 2.5 months of continuous work it increased the spacecraft velocity by 2,415 km/h while consuming only 11.5 kg (25.3 lb) of propellant (xenon).

Table 2 compares some of the modern propulsion systems.

POWER SUPPLY SYSTEMS

Power supply systems provide electricity to feed the onboard spacecraft equipment and can be classified by the primary energy sources: chemical, solar, and nuclear.

Dry cells (batteries) and fuel cells are the main chemical electricity sources. Dry cells consist of two electrodes, made of different metals, placed into a solid electrolyte. The latter facilitates an oxidation process and a flow of electrons between electrodes, directly converting chemical energy into electricity. Various metal combinations in electrodes determine different characteristics of the dry cells. For example, nickel-cadmium cells have low output but can work for several years. On the other hand, silver-zinc cells are more powerful but with a much shorter life span. Therefore, the use of a particular type of dry cell is determined by the spacecraft mission profile. Usually these are the short missions with low electricity consumption. Dry cells are simple and reliable, since they lack moving parts. Their major drawbacks are the bulk and deterioration of current stability during the life span.

In fuel cells, electricity is generated by the same chemical process of oxidation between liquid fuel and oxidizer in the presence of a catalytic agent and electrolyte. While various chemical solutions could be utilized, the most common fuel cells use the oxygen/hydrogen combination. The oxygen/hydrogen fuel cell consists of a chamber with three compartments separated by electrodes (nickel plates with tiny holes and a catalytic agent). The middle chamber is filled with electrolyte. Oxygen and hydrogen (fuel) are pumped into the side chambers. As a result of a chemical reaction, the cell produces electricity, heat, and water. Fuel cells provide 75 percent higher output than the best dry cells, and their electric current does not deteriorate with time. At the same time they are much bulkier and more complex in design, with numerous moving parts. Their life time is limited by the supply of fuel and oxidizer onboard a spacecraft. Fuel cells usually are used for short-duration missions with high electricity consumption. They were installed aboard the American *Gemini* and *Apollo* manned spacecraft and later—on the Space Shuttle.

Solar energy systems utilize an unlimited natural source of energy—the sun. Solar arrays directly convert visual light into electricity. Such devices are the most widely used sources of electric power in modern space technology. They consist of a large number of specially developed semiconducting platters (cells) that develop electrical current when exposed to light. Silicon semiconducting cells are most common, while the newest and more expensive gallium and cadmium cells are more efficient. The cells are connected and assembled in arrays on the body of the spacecraft, or on specially deployed external panels. The panels are usually made of a number of hinged plates that are stacked during the launch and open after the spacecraft reaches orbit. However, in the late 1980s a new type of panel—based on a thin, flexible film—was introduced, making solar arrays lighter in weight and easier to store and deploy.

Solar array efficiency depends on its orientation to the sun. The best result is achieved if the array faces the sun at an angle of 90° to 85°. Precise orientation can be achieved either through the rotation of the whole spacecraft, or the rotation of the solar arrays only. To ensure constant electricity flow regardless of changes in spacecraft orientation, solar arrays charge a secondary power source—the rechargeable booster batteries—which in turn power the spacecraft equipment.

Large solar arrays can produce high power, since the total electrical output depends on the overall area of panels. The capability to work for several years with stable current makes solar panels more effective than chemical sources for long-duration space missions. Spacecraft power supplies could be upgraded if solar arrays are enlarged, or replaced in flight. Such operations were performed by cosmonauts and astronauts aboard the Russian *Salyut* and *Mir* space stations and the American Hubble Space Telescope. At the same time, the solar arrays and related equipment are bulky and massive. Large solar arrays are susceptible to erosion by space radiation, micrometeorites, and damage from space debris. Large solar arrays also have limited applications for low-flying satellites (due to increase in atmospheric drag) and for deep space probes, since their efficiency diminishes at greater distances from the sun. To solve the latter problem, NASA recently pioneered a new concept—the refractive solar concentrator array. The solar panel in this case contains a layer of tiny lenses that concentrate solar light on the cells and compensate for energy losses in distant space. The first spacecraft to test this new design was *Deep Space 1.*

Modern nuclear electric systems are represented by two different types of devices: radioisotope generators and nuclear reactors. Radioisotope generators utilize the natural decay of radioactive materials, which could last for several years. The simplest among radioisotope generators—the so-called nuclear battery—develops electricity in electrodes directly exposed to isotope radiation. This is a low output but very compact device that could be used as an auxiliary power supply. Unlike nuclear batteries, thermal radioisotope generators produce electricity from the heat emitted by the isotope via static thermal converters of two major types. The most common are the radioisotope thermoelectric generators (RTGs), where electricity appears in arrays of semiconductors heated at one end and cooled at the other. RTGs have been intensively studied in the United States and Russia for almost four decades and have reached a high level of sophistication and reliability. The early models were flight-tested by both countries in the mid-1960s. Since then, the United States has used RTGs on more than two dozen military and scientific space missions, mostly for deep-

Type	*Prime Energy Source*	*Advantages*	*Disadvantages*	*Comments*
Dry Cells	Chemical.	-Simplicity.	-Low output; -Short life-span; -Bulk; -Current deterioration.	In use for short duration space missions. Various designs provide different life-span and output.
Fuel Cells	Chemical.	-High output; -Stable current; -Water supply as a sub-product.	-Complex design; -Bulk; -Life-span is limited by onboard fuel supply.	Used in some piloted spacecraft for short duration missions.
Solar Arrays	Visual solar radiation.	-Long life-span; -High output; -Stable current.	-Require precision Sun orientation; -Bulk; -Require booster batteries for electricity storage; -Ineffective in low-flying and deep space spacecraft.	The most widely used power supply system in modern spacecraft.
Atomic Batteries	Radioisotope decay.	-Long life-span; -Stable current; -Simplicity; -Compactness.	-Very low output; -Emit harmful radiation; -High cost.	Used as auxiliary power sources in some spacecraft.
Radioisotope Generators	Thermal from radioisotope decay.	-Long life-span; -Stable current; -Compactness.	-Low output; -Require additional devices to produce electricity; -Emit harmful radiation; -High cost.	Used for some long-duration missions, mostly deep-space probes.
Nuclear Reactors	Thermal from nuclear chain reaction.	-High output; -Long life-span; -Stable current; -Compactness.	-Complex design; -Require additional devices to produce electricity; -Emit harmful radiation; -High cost.	Used for some space missions with high power consumption. May produce radioactive pollution of the environment in case of disintegration at launch, or de-orbiting.
Solar Concentrators	Thermal from visual solar radiation.	-High output; -Long life-span; -Stable current.	-Complex design; -Require additional devices for electricity production and storage; -Bulk; -Require precision Sun orientation.	Not used. Considered as prime power supply systems for large space stations.

Table 3.
Power Supply Systems

space and planetary probes. Automated scientific stations installed by astronauts on the moon, *Viking* landers on Mars, *Pioneer* missions to Jupiter and Saturn, and *Voyager* explorations of the outer planets of the solar system all owe their many years of uninterrupted power supplies to radioisotope generators. The *Cassini* probe, launched toward Saturn in 1997, was the latest spacecraft equipped with RTGs. However, since the 1980s closer attention has been paid to the development of the more complex thermionic generators, since they are considered more efficient than RTGs. In thermionic converters electricity is developed in a chain of metal electrodes enclosed in a sealed container and exposed to a high temperature. The electrodes are placed very close to each other (up to 0.01 mm), with a special dielectric solution between them.

Radioisotope generators cannot produce high output, but their main advantages are compactness, self-containment, and unsurpassed durability. The lack of moving parts makes them highly reliable. At the same time, radioisotope generators are expensive and require protection of the onboard electronic equipment and ground personnel from harmful radiation. The use of radioisotope generators is justifiable in small, long-duration, and low-power-consumption spacecraft.

Unlike radioisotope generators, nuclear reactors utilize the much more intense process of nuclear chain reaction. Since this process is controlled in the reactor, the energy output could be regulated depending on the system's requirements. It actually could produce twice its nominal power, if necessary. Nuclear reactors can provide greater electrical output than radioisotope generators using the same types of thermal converters. This output is comparable to that of fuel cells and solar arrays, while nuclear reactors are more durable and compact.

Reactor-based electrical systems were studied in the United States and the USSR for many years. After testing in space of its first nuclear reactor, SNAP-10A, in 1965, the United States concentrated mostly on developing RTGs. On the contrary, the Russian engineers put more effort into researching nuclear reactors. The Soviet Union was the first nation to start regular use of spaceborne nuclear reactors, called Buk, with thermoelectric converters. That system powered the military ocean radar surveillance and targeting satellite US-A ("US" stands for universal satellite), better known in the West as RORSAT. The satellite's size limitations, low orbit, and radar's high power consumption made the nuclear reactor the only possible energy source. A total of thirty-one such satellites were launched from 1970 to 1987. At the end of their life span, the satellites eject the reactors to a higher "storage" orbit, were they will exist for several hundred years. By the time of their natural reentry, the nuclear material should decay beyond the danger of atmospheric contamination. Unfortunately, that did not always work: Three times the reactors accidentally reentered along with satellites and caused air and water pollution in Canada and the Pacific. Later the USSR developed a series of more advanced reactors with thermionic converters and successfully flight-tested them in 1987. In 1992 one of those reactors, Topaz 2, was purchased by the United States to study its possible military application.

Future spacecraft probably will require much more power than today. This problem could be solved by dynamic thermal converters, in which a high-temperature gas spins the electrical generator's turbine. Obviously, such a generator would depend on a supply of the gas-producing material onboard a spacecraft. In this respect most effective would be a closed-cycle generator, where the used gas is not discharged into space, but continuously recycled through a special radiator. Theoretically, dynamic converters can produce 45 percent higher output than static ones. As a source of thermal energy, dynamic generators may use nuclear reactors, or solar concentrators—large parabolic mirrors that produce a high temperature zone in a focal point. Unlike solar arrays, the mirrors are less susceptible to erosion and theoretically could work for a longer time. Solar concentrators have been studied in the United States and Russia but have never been used in space flight.

Table 3 compares various spacecraft power supply systems.

Peter A. Gorin

See also: Aviation Fuel; Batteries; Engines; Fuel Cells; Fuel Cell Vehicles; Military Energy Use, Historical Aspects of; Rocket Propellants; Storage Technology.

BIBLIOGRAPHY

Gatland, K., ed. (1984). *The Illustrated Encyclopedia of Space Technology*, 4th ed. London: Salamander Books.

Rycroft, M., ed. (1990). *The Cambridge Encyclopedia of Space*. Cambridge, Eng.: Cambridge University Press.

SPACE HEATING

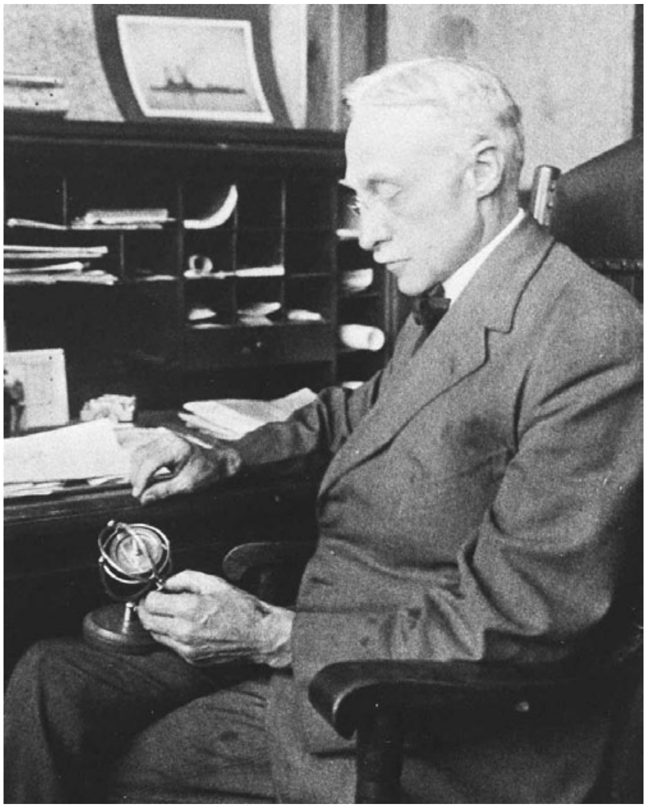
See: Heat and Heating

SPEED AND ENERGY CONSUMPTION

See: Automobile Performance

SPERRY, ELMER AMBROSE (1860–1930)

Elmer Ambrose Sperry was born near Cortland, New York, on October 12, 1860. Sperry, whose mostly absent father was a farmer and itinerant worker and whose mother died hours after his birth, was raised by his aunt Helen until she married, as well as by his paternal grandparents. The young Sperry often studied the many water-driven mills, a steam-driven mill, and the railway station in Blodgett Mills. Later, he also became a familiar presence at the mills, factories, blacksmithy, machine shop, printing press, foundry, railway yard, and pottery of the larger Cortland.



Elmer Ambrose Sperry. (Library of Congress)

A turning point for Sperry was his attendance at the Philadelphia Centennial exposition in 1876. Sperry was particularly fascinated by the Machinery Hall, observing for hours, for example, a Jacquard loom incorporating an early version of card punch automatic control.

The YMCA greatly influenced Sperry, who avidly read the latest scientific magazines, including the *Official Gazette of the U.S. Patent Office*, in its periodical room. The organization also sponsored lectures given by professors at the Cortland Normal School, which Sperry attended, as well as lectures by leaders in many fields, including Alexander Graham Bell. While growing up, he was active in the Baptist church in Cortland. After moving to Chicago in 1882, Sperry became heavily involved at the First Baptist Church where Zula Goodman, whom he later married, played the organ. That marriage, beginning June 28, 1887, was long and happy and produced three sons and a daughter.

In 1883, Sperry began his first business, the Sperry Electric Light, Motor, and Car Brake Company. This company, which pursued dynamo-driven arc lighting during a time of the rapid development of incandescent lighting and alternating current, folded in 1887. Out of the ashes of the former company, Sperry formed the Sperry Electric Company, whose principal assets were Sperry's arc-light patents. His attention, however, was divided between this company and the Elmer A. Sperry Company, a research and development enterprise formed in 1888 and almost entirely owned by Sperry, in which he was freed from engineering and business matters. It was the Elmer A. Sperry Company that allowed him to develop and patent his ideas, with the aid of about thirty workmen in the shop. Sperry developed a series of labor-saving devices, including a highly successful electric locomotive for mine haulage (introduced in 1891), to ease miners' work and increase mine productivity. Sperry also created the Sperry Electric Mining Machine Company (1889; sold in 1892) and the Sperry Electric Railway Company (1892). In addition, a separate company was developed (reorganized in 1900 as Goodman

Manufacturing Company and in which Sperry acquired a large block of stock) that bought mining industry patents, as well as manufacturing several products Sperry's company had originally manufactured. Throughout his life, Sperry stayed involved in policy decisions and technical consulting as a vice president of the Goodman Company, which became a good source of income for Sperry to apply to his future invention projects.

Sperry's method of invention was to examine a field, determine market need, and focus on creating an innovative device to meet that need. He was amazingly visually oriented rather than mathematically inclined. When an invention was patented and ready to market, its patents were often used as the principal capital assets of a specially formed company. Sperry then relegated management of its manufacture and sales to others, freeing him to concentrate on the next invention. Sperry invented his streetcar during the early 1890s, assigning the resulting patents to General Electric in 1895. With its invention, Sperry combined his in-depth grasp of both mechanical and electrical engineering. By 1898, he had developed a superior electric car. Running on a battery he created, it could drive a reputed 87 miles per charge, compared to around 30 miles for other electric cars. The jigs and tools to turn out "Sperry Electric Carriages" in lots of 100 were being readied; instead, in 1900, the American Bicycle Company acquired some of the key patents related to Sperry's electric car, using these to protect its own lines of steam, petroleum, and electric automobiles. Heartened by his success with the battery, Sperry then focused on electrochemistry, spurred in part by the harnessing of electrical power at Niagara Falls in 1895, which made electric power cheaply available. By 1904, Sperry and Clinton P. Townsend had refined an electrolysis process to the point where Elon Huntington Hooker decided to purchase the patents for the process—the beginning of the Hooker Electrochemical Company of Niagara Falls, which became a major U.S. chemical manufacturer.

Sperry turned his attention to the American Can Company's huge amount of scrap metal remaining after round can tops were pressed from square sheets. He and his colleagues refined American Can's electrolytic detinning process to deal with this scrap so that absolutely pure tin powder resulted. In 1907 and 1908, Sperry was involved in patent interference cases concerning this detinning process. The outcome was that the Goldschmidt Detinning Company, formed by American Can in 1908 using the detinning process pioneered by the German Goldschmidt company, bought Sperry's patents, and in 1909 sold the rights to these patents to same Goldschmidt company in Germany (attesting to their value). Meanwhile Sperry's interest since 1907 had been moving into the nascent and promising area of gyroscopic technology.

After attempting unsuccessfully to convince the automobile industry of the safety value of his car gyrostabilizer, Sperry continued his concentration, begun in late 1907, on a gyrostabilizer for ships, seeking to improve the one developed by Ernst Otto Schlick in Germany. By May 1908, Sperry had completed his invention and filed a patent application, marking the beginning of Sperry's most famous period—as inventor of an *active* (as opposed to Schlick's *passive*) ship gyrostabilizer. It was active in that an actuator (later called a sensor), consisting of a pendulum that, upon sensing small rolls of the ship, activated the switches or valves or other means that controlled the motor; the motor then stimulated the much less sensitive gyro to overcome its greater inertia (compared to the pendulum) and begin precession *before* the gyro actually began to feel the torque of the ship's roll. The pendulum system incorporated a feedback control mechanism, one of many Sperry had developed, and would continue to develop, as essential components of his inventions—making Sperry a pioneer in the area that came to be known as cybernetics.

By 1911, Sperry had in addition invented and developed a gyrocompass that was installed for testing on one of the U.S. Navy's state-of-the-art Dreadnoughts. After refining its design, Sperry fulfilled his company's Navy contract by installing the gyrocompass on three more of the Navy's Dreadnoughts and by June 1912 on two submarines. One refinement was a mechanical analog computer that automatically corrected compass errors.

Between 1909 and 1912, aircraft invention was burgeoning, but lack of adequate stabilization resulted in many lost lives. Sperry, his brilliant son Lawrence (who would die in 1924 in a crash while flying across the English Channel), and Sperry's hand-picked engineers developed an airplane automatic stabilizer, redesigned it into the main component of an automatically controlled missile during World War I, and after the war transformed it again into an automatic pilot system. During the war, when the Sperry Gyroscope Company (formed in 1910) became almost exclusively devoted to wartime development, several other inventions, such as a superior high-intensity searchlight, were introduced. Also introduced and installed on several Navy battleships was a highly effective improved gunfire control system, whose basic component was the gyrocompass.

Following the war, the gyrocompass was adapted for merchant-marine vessels, providing unmatched ease of use and maintenance, far outstripping the magnetic compass in navigational economies and safety. By 1932, the Sperry gyrocompass had been installed on more than 1,000 merchant-marine ships. The gyrostabilizer had also been installed on a few huge ocean liners. Another major achievement during this time was the automatic pilot developed for ships, incorporating the most sophisticated guidance and control mechanisms Sperry had yet invented. Nicknamed "Metal Mike," the gyropilot was installed in around 400 merchant-marine ships worldwide by 1932. In addition, Sperry developed and marketed a helm indicator and course recorder that proved valuable adjuncts to the gyropilot and gyrocompass.

Sperry's output continued undiminished until his wife died on March 11, 1930, after which his health declined until he died on June 16, 1930.

Sperry applied for his first patent—a dynamo-electric machine—when he was twenty. His final patent application—for a variable pitch propeller—was submitted in 1930, shortly after his death. In total, 355 of his over 400 patent applications matured as actual granted patents.

Robin B. Goodale

See also: Batteries; Electric Vehicles.

BIBLIOGRAPHY

Hughes, T. P. (1993). *Elmer Sperry: Inventor and Engineer.* Baltimore: Johns Hopkins University Press.

SPRING

See: Elastic Energy

STANFORD LINEAR ACCELERATOR CENTER

See: National Energy Laboratories

STEAM ENGINES

A steam engine is a device that converts a portion of the heat energy absorbed by a liquid, after it has vaporized, into mechanical work.

For a classic reciprocating steam engine, pressurized steam from a boiler is admitted into cylinder, driving a piston attached to a connecting rod and crankshaft. Steam admission is typically cut off in midstroke. This cutoff steam continues to expand against the piston until the stroke ends, converting more heat energy into work.

The most efficient performance is achieved when the engine expands steam to achieve the greatest absolute difference between steam admission and exhaust temperatures. *Superheaters* are devices in the boiler that raise the temperature of steam above that of the water from which it was generated. Efficiency is improved because this added temperature, or superheat, permits greater expansion (and thus temperature drop) before the steam begins condensing and is exhausted. Further efficiency gains can be realized by exhausting the steam not to the atmosphere but rather to a separate apparatus called a *condenser*. This is a sealed chamber cooled by air or water. As entering steam cools, it condenses, and its volume is greatly reduced, producing a partial vacuum. Exhausting into this partial vacuum permits yet further expansion and temperature drop.

Steam *turbines*, which generate more than 80 percent of the world's electric power, differ from steam engines in that steam drives blades and not pistons. Steam turbines expand pressurized steam through nozzles that accelerate the steam at the expense of heat energy and pressure. Work is created by transferring a portion of steam velocity to blades, buckets, or nozzles affixed to a rotor to move at high speeds. Steam turbines are relatively compact in relation to steam engines. Turbines are generally operated at higher pressure, resulting in greater output for the size. Their rotary nature also permits continuous admission, expansion, and exhaust so that power production is uninterrupted. Steam engines, by contrast, only develop maximum output during the period before cutoff and none at all during the exhaust period.

Steam power was a cornerstone of the Industrial Revolution. Steam freed humans from limits imposed by natural energy sources as muscle, water, wind, and sun. Human and beast-of-burden muscle power was limited. Most humans turning a crank affixed to an electric generator cannot sustain enough effort to illuminate a 150-watt lightbulb for more than a few minutes. Water was a fairly reliable energy source, but only obtainable along rivers capable of sustaining suitable dams. Windmills were also an early energy source. Although effective, they lacked a means of storing energy, and were only useful for jobs that could be profitably performed intermittently, such as grinding grain.

Starting in eighteenth-century England, steam engines suddenly made concentrated power available on demand anywhere. Earliest uses included pumping water from iron and coal mines, driving iron mill furnaces and hammers, and, a bit later, propelling locomotives to and from the mines. In a synergistic relationship, steam engines made iron and coal affordable, which in turn lowered the cost of building and operating steam engines. Previously costly iron products soon became affordable to all, and engine-driven industry and transportation spread rapidly across the globe.

PREDECESSORS OF THE MODERN STEAM ENGINE

In 1606 Gioavanni Battista della Porta of Naples performed two experiments that formed the basis for steam engines. He proved that steam could develop pressure by expelling water from a sealed container. In the next experiment a steam-filled flask was immersed neck down in water. As the cooling steam condensed, water flowed upward into the flask, proving that a vacuum had been created.

By 1698 Captain Thomas Savery exploited these principles with a patented steam pump. Alternately admitting pressurized steam into a vessel forced water upward through a one-way (check) valve, while the vacuum created by cooling the vessel drew water from below through a second valve. French inventor Denis Papin demonstrated the first steam-driven piston in 1690. His experiment consisted of a cylinder with a closed bottom and a piston. A weight attached to a cord passing over two pulleys applied upward force on the piston. Flame beneath the cylinder heated a small quantity of water, creating steam pressure to raise the piston. Condensation followed removal of heat, and the difference in air pressure on the piston top and the vacuum below caused the piston to descend, raising the weight. By 1705 Thomas Newcomen introduced his pumping engine to England. It built upon the work of Papin but was made practical by use of a separate boiler. The Newcomen engine consisted of a vertical cylinder with a piston attached to a counterweighted crossbeam. A piston pump to drain mines was attached to the crossbeam. The combined efforts of low-pressure steam and counterweight forced the piston upward, to the cylinder top. Spraying cold water into the cylinder condensed the steam, which created a partial vacuum, returning the piston to the bottom.

Both the Newcomen and the Papin pumps derived most of their power from this atmospheric force rather than from steam pressure. Newcomen engines consumed perhaps 32 pounds of coal each hour per horsepower developed, about 1 percent of the energy that burning coal emits.

MODERN STEAM ENGINES

Scottish doctor Joseph Black discovered in 1765 that applying heat to water would steadily raise temperature until reaching the boiling point, where no further temperature increase occurred. However, boiling did not commence immediately but only after a relatively large additional hidden; or latent heat was added. Learning of this discovery, Scottish instrumentmaker James Watt realized that the cold water spray in the Newcomen cylinder wastefully removed latent heat. Maintaining the cylinder at steam temperature would conserve heat and reduce energy consumption. Watt conceived of the condenser, a separate vessel cooled by water immersion and connected to the cylinder through an automatic valve. Steam entering the vessel rapidly condensed, and the vacuum created was applied to the piston, as in the Newcomen engine. Early Watt cylinders were open at the top, like the Newcomen engine. Later, Watt created the double-acting engine by enclosing both cylinder ends and connecting them to the boiler and condenser with automatic valves. Air no longer wastefully cooled the cylinder, and steam rather than air pressure pushed the piston in both directions.

While Watt engines admitted steam throughout the full stroke, he did patent the idea of cutoff to allow the steam to work expansively and more efficiently. Higher pressures were needed to exploit expansion, and it fell to other inventors, notably Richard Trevithick of Wales and Oliver Evans of the United States, to make this leap. Their design eliminated the pull by a vacuum and introduced a push by high pressure. Further efficiency gains occur by superheating the steam, heating it to a greater temperature than by which it was generated. This added temperature permits greater expansion and work to occur before the steam condenses. Watts's use of separate boiler, engine, and condenser, double-acting engines, and insulation all resulted in the first modern steam engine. He tripled efficiency over the Newcomen engine, reaching about 3 percent efficiency overall.

In 1781 Jonathan Hornblower invented the compound engine, but conflicting patent rights prevented development. Arthur Woolf revived the idea in 1804 and invented a valve mechanism often associated with compounding. The compound engine partially expands steam in one cylinder and then further expands it in one or more succeeding cylinders, each expansion being termed a stage. Compound expansion reduces temperature swings in each cylinder, improving efficiency since less energy is lost reheating metal on each piston stroke. While simple engines with condensers reached 10 to 15 percent efficiency on superheated steam, compounds with up to four stages readily reach 20 percent and in a few cases as much as 27 percent.

Reciprocating steam engine development reached its pinnacle with the development of the uniflow engine. Invented by T. J. Todd of England in 1885 and refined some twenty years later by German engineer Johann Stumpf, the uniflow features a single-inlet valve, and exhaust ports at the end of the stroke. Steam is cut off relatively quickly and allowed to expand; at the end of the stroke, the piston exposes the exhaust port. As the piston returns and covers the port, the piston it recompresses the steam remaining in the cylinder back to the admission pressure. This recompression heats the steam in much the same fashion as it cooled off while expanding; consequently little energy is lost reheating the metal. Uniflow engines are more simple and economical to build than compounds and generally approach compounds in efficiency; at least a few examples may have surpassed compounds, with one engine reportedly testing at 29 percent efficiency. Uniflow engines never became prevalent due to rapid concurrent development of the steam turbine.

Steam engine development preceded understanding of the principles governing their operation. Steam engines were widely used in locomotives, automobiles, ships, mills and factories. It was not until 1859, when the Rankine cycle (named after Scottish engineer and physicist William John Macquorn Rankine) was postulated, that steam engine design finally received a sound theoretical footing.

STEAM TURBINES

Curiously, although practical steam turbines are a more recent development than piston steam engines, two of the earliest steam devices were turbines. The earliest generally recognized steam-powered mechanism was the aeolipile, built by Hero of Alexandria in the first century C.E. It was essentially a pan-shaped boiler connected to a hollow rotating shaft on which was mounted a hollow sphere. Jets of steam exiting the sphere through two tangentially placed nozzles spun the device. The action was similar to a modern rotating lawn sprinkler. The Branca engine of 1629 consisted of a boiler fancifully shaped into a human head and torso. Steam exiting a nozzle placed at the mouth impinged upon a paddle wheel. Although efficiency and performance were very low, cog wheels provided sufficient torque for a small mortar and pestle to powder drugs.

Modern steam turbines can reach efficiencies of 40 percent. About 90 percent of newly installed electrical generating capacity is driven by steam turbines due to high efficiency and large capacity. The impulse turbine is the most basic modern turbine. An impulse stage consists of one or more stationary steam nozzles mounted in the casing and a rotating wheel with curved blades mounted along its periphery. Steam expanding through the nozzles exchanges pressure and temperature for great speed. Directing this steam against curved blades transfers momentum to the wheel. The Branca engine was an early impulse turbine. Reaction turbines exploit Newton's third law: "For every action there is an equal and opposite reaction." While Hero's aeolipile lacks blades like a modern turbine, recoil from steam did rotate the sphere, making it a reaction turbine.

The allure of producing rotary motion without intermediary crankshafts and connecting rods attracted efforts of engineers to little success until the 1880s, when two entirely different designs appeared in England and Sweden. In 1884 Charles Algernon Parsons patented the first practical reaction turbine. Unlike Hero's turbine, which had two jets, Parson's turbine had blades mounted around the periphery of fourteen pairs of rotating disks mounted on a single shaft. The area between blades was designed similarly to nozzles; steam passing through these spaces expanded and accelerated, imparting thrust to the rotating disk. Succeeding blade sets became larger, permitting steam to expand farther in each stage and is analogous to compounding in reciprocating engines. Fixed blades between the disks smoothly redirected steam to the following rotating blades. Large turbine sets can employ many more stages of expansion than reciprocating steam engines, since each stage adds only a small amount of friction, while additional cylinders soon add enough friction to offset any efficiency gains. The final stages in these machines can often effectively extract energy from steam at pressures below atmospheric. This aggressive use of compounding leads to significantly more efficient operation than for reciprocating steam engines.

Parsons adopted this multistage design to reduce turbine operating speeds to a useful level. His initial turbine was developed to produce electricity onboard ships and had an output of about 7 kW. In 1888 he designed the first steam turbine generating unit for public utility service. By the time of his death in 1931, his company manufactured turbines generating more than 50,000 kW.

In 1894 the success of his electrical generating plants led Parsons to undertake development of the first marine turbine, with the goal of building the fastest ship afloat. The prototype vessel *Turbinia* was 100 feet long. Initial trials with a single turbine directly connected to a propeller were unsatisfactory, as the turbine speed was too high for efficient operation. In 1896 the single turbine was replaced by three separate turbines, with steam flowing from the high-pressure turbine element to the intermediate element and then the low-pressure turbine element. Each turbine shaft was fitted with three propellers and she made 34½ knots, a new world's record.

In 1888 Carl G. P. de Laval invented his impulse turbine, which was based on the Pelton water wheel. Steam was accelerated to nearly the speed of sound through a nozzle or a series of nozzles. Directing the steam toward curved vanes mounted upon a disk resulted in high rotating speeds. Conceptually this is identical to Branca's turbine, but August Camille Edmond Rateau of Paris learned that wasp-waisted nozzles greatly enhanced velocity, making turbines feasible. Noticing that excessive rotating speeds were a weakness of de Laval turbines, Rateau patented a multistage impulse turbine in 1896. His machine was essentially a series of de Laval turbines. Breaking expansion into stages resulted in smaller pressure drops across each set of vanes and lower velocity for the same mechanical output. Over the years many variants of turbine arrangements have been tried and adopted. Efficiency has more than doubled since the turn of the twentieth century, and output has risen enormously. The first Parsons turbines of 1884 were perhaps 10 hp. By 1900, when the demand for electric power was growing tremendously, the largest steam turbine-generator unit produced about 1,250 kW, but advances over the next twenty years increased capacity to more than 60,000 kW. This output far exceeded what even the largest steam engines could provide, and paved the way for central generating plants with millions of turbine horsepower to be the principal electric power generators of the twentieth century.

The steady march of improvements in performance of steam turbines can be largely attributed to exploiting higher steam pressures and steam temperatures. Steam generator pressures and temperatures increased from about 175 psi and 440°F (227°C) in 1903 to 450 psi and 650°F (343°C) in 1921 and 1,200 psi and 950°F (510°C) by 1940. Since the late 1950s steam pressures of more than 3,500 psi and 1,150°F (621°C) have become available and at least one unit has reached 5,300 psi and 1,210°F (654°C). Nationwide, average plant efficiency climbed steadily, with an average of 29 percent in the late 1930s, 35.5 percent in the late 1940s, and 38.5 percent in the late 1950s. Efficiency improved little in the following decades as boiler and turbine operating temperatures and pressures had reached the limits of available material strength and durability.

By the turn of the twentieth century, reciprocating steam engines had become common items in everyday life. Transportation was dominated by steam-powered ships, locomotives, and boats. Steam was the preferred method of driving the earliest automobiles, with more steamers on U.S. roads than internal-combustion engines until about 1906. Portable and self-propelled steam engines provided power to drive agricultural and lumbering machinery on site. Most factories, shops, and mills had a steam engine that drove machinery by way of overhead shafts and leather belts. The sound of the mill steam whistle signaling the shift change set the routine in many communities. Within a few years this personal relationship with steam receded. Internal-combustion engines largely supplanted small steam engines where self-contained power was necessary, while electric motors became dominant in stationary applications.

Except for a few countries that still use steam engine locomotives, reciprocating steam engines have passed into history. They are used only rarely, more often than not as a hobby. Steam turbines, on the other hand, have continually increased in importance, as they are used to generate most of our electricity. All nuclear power plants generate steam to drive turbines, and some geothermal installations use the earth's heat similarly. Although steam turbines are still common on oceangoing ships, they are being replaced by large diesel and gas turbine installations due to high intrinsic efficiency, low first cost, and more economical labor and maintenance requirements.

Even when large diesels and gas turbines take over as prime movers, steam turbines still often have a vital role. A process called bottom cycling or combined cycling can team the steam turbine with an internal-combustion engine, which is inherently efficient due to its high temperature operation. Exhaust gases from these operating cycles are often sufficiently hot enough to operate a boiler and steam turbine. Since the steam system is driven by heat that would otherwise be discarded, overall system efficiency soars. Such bottom cycling plants, which are gaining favor in new generating facilities in the United States, are now 50 percent efficient and are likely to close in on 60 percent in the near future.

Kenneth A. Helmick

See also: Black, Joseph; Engines; Gasoline Engines; Newcomen, Thomas; Parsons, Charles Algernon; Rankine, William John Macquorn; Savery, Thomas; Trevithick, Richard; Turbines, Steam; Watt, James.

BIBLIOGRAPHY

Bentley, J. (1953). *Old Time Steam Cars*. Greenwich, CT: Fawcett Publications.

Briggs, A. (1982). *The Power of Steam*. Chicago: University of Chicago Press.

Hunter, L. C. (1985). *A History of Industrial Power in the United States, Volume Two: Steam Power*. Charlottesville: University of Virginia Press.

Rose, J. (1887). *Modern Steam Engines*. Philadelphia: Henry Carey Baird. Reprint, Bradley, IL: Lindsay, 1993.

Storer, J. D. (1969). *A Simple History of the Steam Engine*. London: John Baker.

Watkins, G. (1979). *The Steam Engine in History*. Derbyshire, Eng.: Moorland.

STEAM TURBINES

See: Turbines, Steam

STEPHENSON, GEORGE (1781–1848)

Many successful people are inspired by their parents and George Stephenson was no exception. In modern day vernacular, his father had one of the "coolest jobs in town" in the small village of Wylam, on the River Tyne, eight miles west of Newcastle, England, where George was born. Robert Stephenson raised his family in a primitive dwelling while serving as the pumping engine fireman at the local colliery, keeping the mines free of water. The company's coal wagon passed very close to the family's front door, helping George make up for his lack of formal education with up close and very personal attentiveness.

When George was eight yeas old, the mine's pumping engine broke down and work temporarily ceased, forcing his father to move the family in search of work. With the impact of engines permanently etched in his mind, the young boy began to make clay models in great detail, simulating his father's machines. Unknowingly, he was destined to study engines for the rest of his life. Defying parental wishes to take up farming, at age thirteen he joined his brother James as a picker in another mine. Working his way up the ladder at the Black Callerton mine, George began driving an engine and by seventeen he earned the position of assistant engine fireman—higher than any post his father had ever attained.

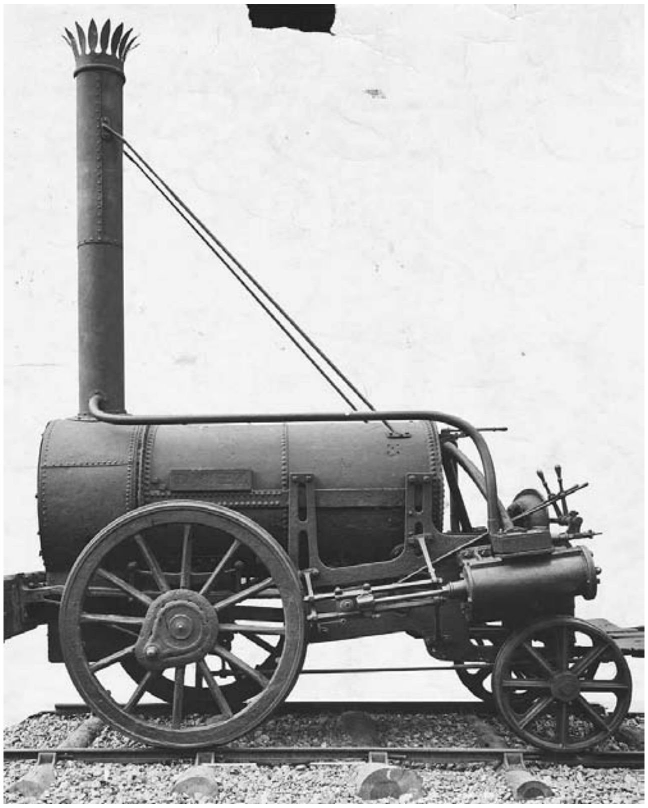
He was eventually promoted to engineman—or plugman—which required intricate engine knowledge and skills. He quickly learned to devote all his free time to the study and care of his engines, tearing down and rebuilding them to better understand their inner workings.

He became interested in stories about James Watt and Matthew Boulton but faced a significant hurdle: at the age of eighteen, he was still unable to read. He began reading lessons with a young schoolmaster, Robin Cowens, and then, to further his knowledge of engineering, was tutored in math by Andrew Robertson, a Scottish mathematician.

George earned the prestigious position of brakesman at the Dolly Pit colliery, responsible for the engine used for hauling coal out of the mines. He was married in 1802. During that year he made an unsuccessful attempt to build a perpetual motion machine. But he did gain respect in the community as an accomplished clock-mender.

In 1808, his career breakthrough came at the Killingworth High Pit, where he and two other brakesman contracted with management to work on a royalty basis, rather than a fixed weekly wage. The risk was offset by his ability to create mechanical engineering efficiencies, considered one of Stephenson's greatest strengths. Opportunity knocked in the form of a pumping engine built by John Smeaton that had been a complete failure during a period of twelve months. Approaching the investors, Stephenson made an audacious claim that he could have the engine running within a week. In fact, he did so in only three days of breakdown and reconstruction, acquiring great local renown and the true launching of his career as a self-taught "doctor of engines." Proving his flexibility, he drained an entire quarry with a unique miniature pump that he invented.

George continued his adult education with John Wigham, focusing on the areas of math and physics. In addition, he befriended an engineer, John Steele, who had apprenticed under Richard Trevithick in the building of a successful locomotive engine during 1803 to 1804.


Side view of George Stephenson's locomotive, "The Rocket." (Corbis Corporation)

Stephenson applied theories developed by others, including Ralph Allen's work on self-acting inclines, in which empty coal wagons were pulled up a track, powered by full wagons moving downhill. Other inventions included a winding-engine and new design for pumping engines, before turning his attention to the self-propelled steam engine.

In the early nineteenth century, hardly anyone imagined that passenger transportation would evolve, concentrating instead on the need to transport coal. Most of the inventiveness applied before Stephenson was to reduce track friction, allowing horses to pull more load. There were even experiments in the seventeenth century by Sir Humphrey Mackworth to build a wind-powered land-ship, which were disappointing due to the inevitable limits of nature. The first successful steam locomotive was built by French engineer Cugnot in 1763. Designed for war, its top speed was a paltry two and one-half miles per hour, with endurance of only fifteen minutes. There were numerous failures for the rest of the century, hampered by persistent attempts to operate them on roads.

It is noteworthy that William Murdock achieved significant progress in the 1780s, despite discouraging remarks by James Watt and his partner, Matthew Bolton, who thought that steam locomotion was impractical. This chiding among rivals was common and it seemed that personal credibility was sometimes as important as genius. When Trevithick did his brilliant work with locomotives, his eccentricities and irascibility deterred his acceptance, as did his unbusinesslike and unprofessional nature. Stephenson's predecessors who did earn credibility, however, were William Hedley and aide, Timothy Hackworth, at coal mines in the Newcastle area.

George Stephenson was not truly the inventor of railways or steam locomotives. But, like Thomas Newcomen and his work with the first practical atmospheric steam engine, George and his son, Robert, were the first to use concentrated imagination to turn existing work and theory into fully functional locomotive engines on rails. The Stephensons made the greatest contribution toward the building of the British railway system.

George Stephenson's first engine, named the Blucher, took ten months to complete, and first operated on the Killingworth Railway on July 25, 1814. His most notable achievement was building the Liverpool to Manchester railway, that featured thirty miles of track and sixty-three bridges. The most famous locomotive ever built—"The Rocket"—was actually designed and made by his son, Robert, although its most serious fault was reconfigured by George, who approved all concepts of design and manufacture.

Despite their growing reputation, there was great resistance at times to the Stephensons' locomotive building, most notably among farmers. Although the noise that scared animals was largely reduced, there was an economic concern to farmers—if steam trains and carriages succeeded, there would be less demand for horses and less demand to grow oats to feed them. But technological progress prevailed, even if it was often settled in court.

In addition to George's acclaimed work with locomotives, he and his son Robert designed numerous contraptions such as a scarecrow with wind-aided arms, a sundial, an oil light that burned underwater, and an automatic cradle rocker. While George Stephenson derived immediate satisfaction from his mechanical accomplishments, there is some evidence that he resented or disdained the public acclaim through his rejection of invitations to join the Royal Society or to accept Knighthood. Despite his brilliance, his relative late education kept him from feeling socially accepted. This may have started with a grudge as early as 1815, when he lost out in a patent dispute with a more privileged rival over a safety lamp that George invented.

Nonetheless, Stephenson realized that he had a great impact on society and finished his career as a traveling consultant in engineering. He was known to dine with kings, such as the King of Belgium, and richly earned his permanent place in the annals of industrial history.

Dennis R. Diehl

See also: Steam Engines; Turbines, Steam.

BIBLIOGRAPHY

Rolt, L.T.C. and Allen, J. S. (1997). *The Steam Engine of Thomas Newcomen.* Ashbourne, UK: Landmark Publishing.

Rowland, J. (1954). *George Stephenson: Creator of Britain's Railways.* Long Acre, London, Odhams Press Limited.11

STIRLING, ROBERT (1790–1878)

Robert Stirling was born on October 25, 1790, in the parish of Methvin, Perthshire, Scotland. Stirling, the son of a farmer, received a classical education followed by studies in divinity at the universities of Edinburgh and Glasgow. The Church of Scotland, after examining his knowledge of Hebrew, Greek, divinity, and church history, licensed Stirling to preach the Gospel in March 1816. Shortly after, Stirling accepted the patronage of the Duke of Portland, who proposed Stirling as a suitable candidate for the vacant post of assistant minister at Kilmarnock. Stirling was ordained on September 19, 1816.

Just eight days after his ordination, Stirling applied for a patent for "Improvements for diminishing the consumption of fuel." He had evidently carried out his experiments while he was a student at Edinburgh, but why a student of divinity should have developed an interest in fuel economy, and gone to the expense of applying for a patent to protect his invention, is something of a mystery. Stirling's patent was in several parts. The first part was for a method of diminishing the quantity of fuel used in furnaces for melting glass, while still producing the high degree of heat required. The second was for a heat exchanger suitable for use in breweries and factories where economies could be made by transferring heat from one fluid or vapor to another. The third was for an engine for moving machinery.

Stirling's engine consisted of an externally heated, hot-air engine fitted with what he termed an "economiser." The economiser absorbed and released heat to and from the enclosed air, which was shifted alternately between hot and cold spaces, with the resulting pressure changes communicated to a power piston. Stirling was the first to design an engine with a working cycle that depended on recovering waste heat for efficient operation.

Stirling's patent was based on practical experiments and a working knowledge of radiant heat. Stirling may have been influenced by the works of John Leslie, Professor of Mathematics at Edinburgh, who in 1804 published the results of his investigations into the nature of radiant heat. The general layout of Stirling's engine embodies Leslie's experiments with spaced radiation screens to reduce the transmission of heat. In a second work published in 1813, Leslie argued that heat was transmitted as a series of vibrations, and if heat were applied to one end of a cylinder of metal, then a regular descending graduation of temperature would soon be established along the whole length. Stirling seems to have perceived that here was a process that might be made reversible. If heat were transmitted in a series of vibrating layers, and if the heat source were itself moved along the layers, heat could be discharged into bands of heat at decreasing temperatures. That is to say, as the heated mass touched each layer, the temperature of the heated mass would fall. If the mass of air were to be passed back over the bands of heat, then it would regain its former temperature. This was how Stirling described the workings of his economiser, although he took care to dispel any idea of perpetual motion by pointing out the losses.

Stirling's transfer to Kilmarnock did not put a stop to his experimenting. Here he meet Thomas Morton, an inventor and manufacturer who supplied him with a workshop. Morton also built telescopes and observatories and was influential in developing Stirling's interest in constructing optical and scientific instruments. In 1818 Stirling erected an engine of his own design to drain water from a local stone quarry. This engine performed its duty well until, as a result of carelessness of the engine-man, it overheated and was rendered useless.

James Stirling, who was ten years younger than his brother Robert, initially studied for the church, taking a classical education at Edinburgh and Saint Andrews University. The pull of mechanics proved stronger than religion, and he took up an apprenticeship with Claud Girdwood and Co., of Glasgow, machine makers and iron founders. In 1824 he suggested his brother use compressed air for his engine rather than air at atmospheric pressure, and by 1827 James was influential enough to have built an experimental twenty horse-powerengine fitted with a sheet iron economiser, in which too much faith was placed, resulting in the engine failing for want of effective cooling. This engine embodied a number of improvements in the economiser, which were patented in 1827 in the joint names of Robert and James Stirling. James left Girdwood's in 1830 to become manager of the Dundee Foundry, Dundee.

At Dundee the brothers continued experimenting, directing their efforts into improving the efficiency of both cooling apparatus and economiser. The results of these experiments were patented in 1840 (the year Robert was awarded a Doctor of Divinity), and in 1843 a forty-five horsepower engine was set up to power the whole of the foundry, which it did until 1847. The trial engine successfully vindicated the Stirling brothers' claims for improved fuel efficiency, but ultimately failed as a practical engine because the available material—cast iron—could not withstand the high temperatures required for efficient operation. The development of the Stirling engine was effectively halted until improvements in metallurgy were made.

In 1837 James married Susan Hunter, the daughter of a Saint Andrews Professor; they had no children. James left Dundee in 1845 to become a successful consulting engineer in Edinburgh, where he died at the age of seventy-five. In 1819 Robert had married Jane Rankine, the daughter of a Kilmarnock merchant. They had two daughters and five sons. Robert had moved to the parish of Galston in 1823, and he remained a respected minister until his death at the age of eighty-eight, although some of his parishioners felt he devoted too much time to his mechanical pursuits.

Such was Robert Stirling's interest in science and engineering that of his sons, four became leading railway engineers in Britain and South America, and one became a church minister. His son-in-law became the general manager of an iron works. Stirling never sought to benefit from his invention; indeed, he did little to prevent infringements of his patents. Stirling's 1816 patent foresaw all possible applications of what came to be called the regenerative process, a term coined by John Ericsson, who never accepted the priority of Stirling's patent. However, Ernst Warner von Siemens, who profitably developed the regenerative furnace for smelting metals, did acknowledge Stirling as the originator.

Robert Sier

See also: Ericsson, John; Siemens, Ernst Warner von; Stirling Engines.

BIBLIOGRAPHY

Sier. R. (1987). *A History of Hot Air and Caloric Engines.* London: Argus Books.

Sier. R. (1995). *Rev. Robert Stirling D.D.* Chelmsford: L. A. Mair.

STIRLING ENGINES

The principle that makes Stirling engines possible is quite simple. When air is heated it expands, and when it is cooled it contracts. Stirling engines work by cyclically heating and cooling air (or perhaps another gas such as helium) inside a leak tight container and using the pressure changes to drive a piston. The heating and cooling process works like this: One part of the engine is kept hot while another part is kept cold. A mechanism then moves the air back and forth between the hot side and the cold side. When the air is moved to the hot side, it expands and pushes up on the piston, and when the air is moved back to the cold side, it contracts and pulls down on the piston.

While Stirling engines are conceptually quite simple, understanding how any particular engine design works is often quite difficult because there are hundreds of different mechanical configurations that can achieve the Stirling cycle. Figure 1 shows a schematic of a transparent educational demonstration engine that runs on the top of a cup of hot coffee. This engine uses a piece of foam similar to what would be used as a filter for a window air conditioning unit to "displace" the air between the hot side and the cold side. This foam displacer is carefully mounted so it does not touch the walls of the cylinder. Figure 2 shows how this particular engine achieves the Stirling cycle. In this engine, the air flows through and around the displacer from the hot side then back to the cold side, producing a power pulse during both the hot and cold portion of the cycle. Stirling engines can be mechanically quite simple since they have no valves, and no sparkplugs. This can result in extremely high reliability as there are fewer parts to fail.

It is worthwhile to compare Stirling engines to other more familiar engines and note their similarities as well as their differences. Stirling engines are a type of heat engine. They turn heat into mechanical work and in this sense they perform the same function as other well known heat engines such as gasoline, diesel, and steam engines. Like steam engines, Stirling engines are external combustion engines, since the heat is supplied to the engine from a source outside the cylinder instead of being supplied by a fuel burning inside the cylinder. Because the heat in a Stirling engine comes from outside of the engine, Stirling engines can be designed that will run on any heat source from fossil fuel heat, to geothermal heat, to sunshine. Unlike steam engines, Stirling engines do not use a boiler that might explode if not carefully monitored.

When operating on sunshine, or geothermal heat, Stirling engines obviously produce no pollution at all, but they can be exceedingly low emissions engines even when burning gasoline, diesel, or home heating oil. Unlike gasoline or diesel engines that have many thousands of start stop cycles of combustion each minute, burners in Stirling engines burn fuel continuously. It's much easier to make a continuous combustion engine burn very cleanly than one that has to start and stop. An excellent demonstration of this principle is to strike a match, let it burn for a few seconds, then blow it out. Most of the smoke is produced during the starting and stopping phases of combustion.

A BRIEF HISTORY

In the early days of the industrial revolution, steam engine explosions were a real problem. Metal fatigue

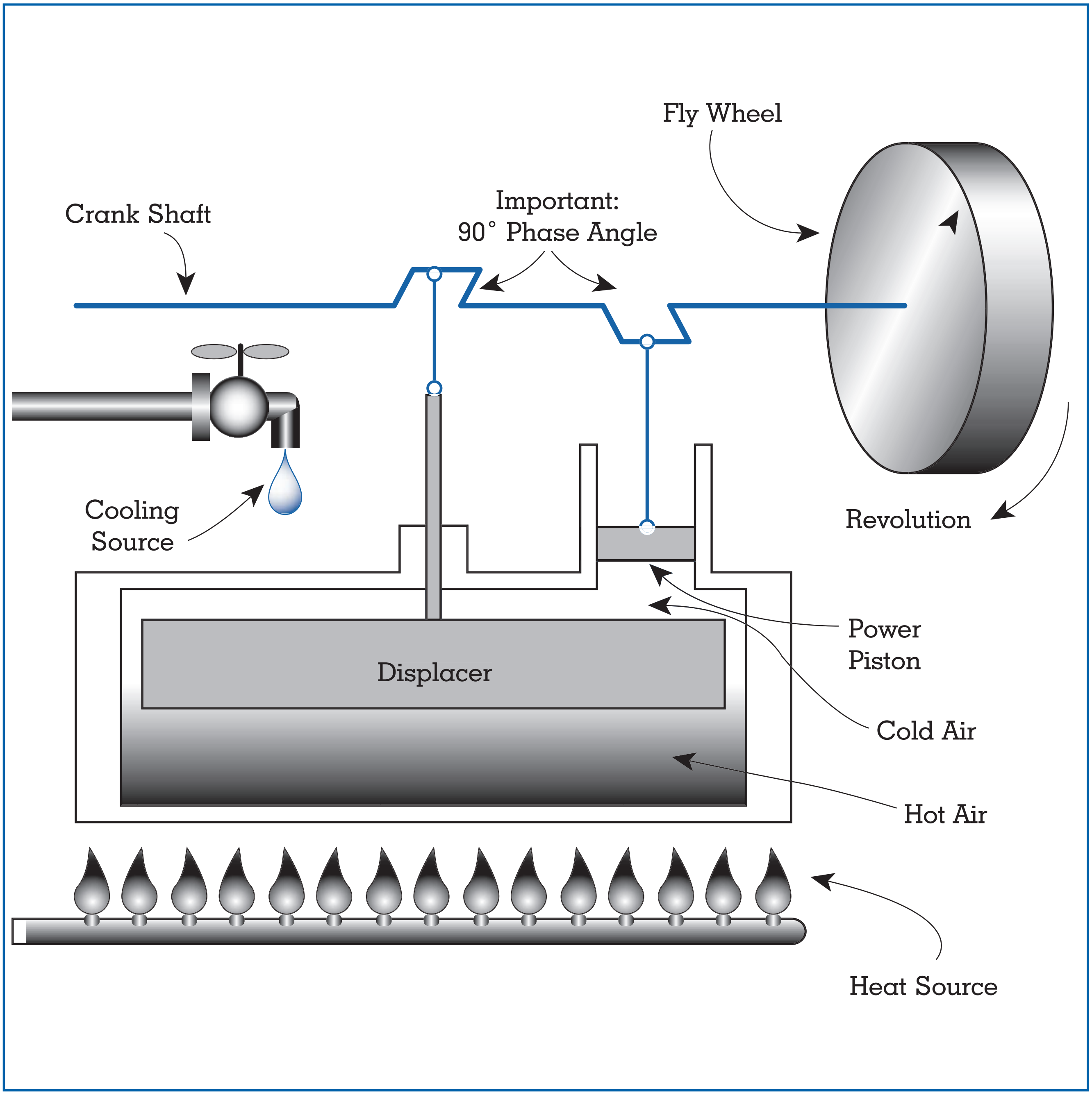


Figure 1.
Key to Stirling Engine. The air flows both through and around the porous displacer. The displacer looks like a piston but it is not.

was not well understood, and the steam engines of the day would often explode, killing and injuring people nearby. In 1816 the Reverend Robert Stirling, a minister of the Church of Scotland, invented what he called "A New Type of Hot Air Engine with Economiser" as a safe and economical alternative to steam. His engines couldn't explode, used less fuel, and put out more power than the steam engines of the day.

The engines designed by Stirling and those who followed him were very innovative engines, but there was a problem with the material that was used to build them. In a Stirling engine, the hot side of the

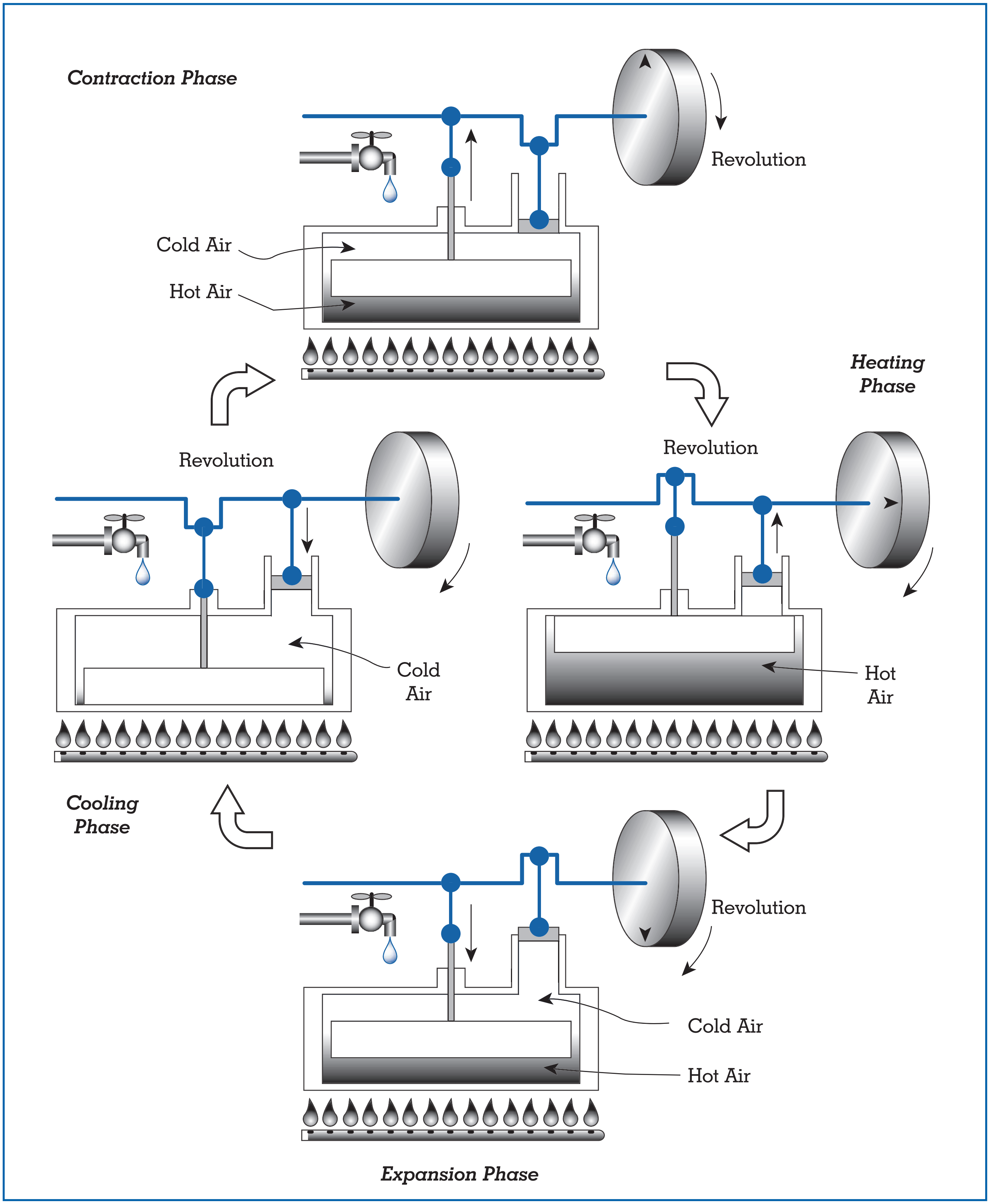


Figure 2.
Four phases of the Stirling engine power cycle.

engine heats up to the average temperature of the flame used to heat it and remains at that temperature. There is no time for the cylinder head to cool off briefly between power pulses. When Stirling built his first engines, cast iron was the only readily available material, and when the hot side of a cast iron Stirling engine was heated to almost red hot, it would oxidize fairly quickly. The result was that quite often a hole would burn through the hot side causing the engine to quit. In spite of the difficulties with materials, tens of thousands of Stirling engines were used to power water pumps, run small machines, and turn fans, from the time of their invention up until about 1915.

As electricity became more widely available in the early 1900s, and as gasoline became readily available as a fuel for automobiles, electric motors and gasoline engines began to replace Stirling engines.

REGENERATION

Robert Stirling's most important invention was probably a feature of his engines that he called an "economiser." Stirling realized that heat engines usually get their power from the force of an expanding gas that pushes up on a piston. The steam engines that he observed dumped all of their waste heat into the environment through their exhaust and the heat was lost forever. Stirling engines changed all that. Robert Stirling invented what he called an "economiser" that saved some heat from one cycle and used it again to preheat the air for the next cycle.

It worked like this: After the hot air had expanded and pushed the piston as far as the connecting rod would allow, the air still had quite a bit of heat energy left in it. Stirling's engines stored some of this waste heat by making the air flow through economiser tubes that absorbed some of the heat from the air. This precooled air was then moved to the cold part of the engine where it cooled very quickly and as it cooled it contracted, pulling down on the piston. Next the air was mechanically moved back through the preheating economiser tubes to the hot side of the engine where it was heated even further, expanding and pushing up on the piston. This type of heat storage is used in many industrial processes and today is called "regeneration." Stirling engines do not have to have regenerators to work, but well designed engines will run faster and put out more power if they have a regenerator.

CONTINUED INTEREST

In spite of the fact that the world offers many competing sources of power there are some very good reasons why interest in Stirling engines has remained strong among scientists, engineers, and public policy makers. Stirling engines can be made to run on any heat source. Every imaginable heat source from fossil fuel heat to solar energy heat can and has been used to power a Stirling engine.

Stirling engines also have the maximum theoretical possible efficiency because their power cycle (their theoretical pressure volume diagram) matches the Carnot cycle. The Carnot cycle, first described by the French physicist Sadi Carnot, determines the maximum theoretical efficiency of any heat engine operating between a hot and a cold reservoir. The Carnot efficiency formula is

$$(T(\text{hot})-T(\text{cold}))/T(\text{hot})$$

T(hot) is the temperature on the hot side of the engine. T(cold) is the temperature on the cold side of the engine. These temperatures must be measured in absolute degrees (Kelvin or Rankine).

STIRLING APPLICATIONS

Stirling engines make sense in applications that take advantage of their best features while avoiding their drawbacks. Unfortunately, there have been some extremely dedicated research efforts that apparently overlooked the critical importance of matching the right technology to the right application.

In the 1970s and 1980s a huge amount of research was done on Stirling engines for automobiles by companies such as General Motors, Ford, and Philips Electronics. The difficulty was that Stirling engines have several intrinsic characteristics that make building a good automobile Stirling engine quite difficult. Stirling engines like to run at a constant power setting, which is perfect for pumping water, but is a real challenge for the stop and go driving of an automobile.

Automobile engines need to be able to change power levels very quickly as a driver accelerates from a stop to highway speed. It is easy to design a Stirling engine power control mechanism that will change power levels efficiently, by simply turning up or down the burner. But this is a relatively slow method of changing power levels and probably is not a good way to add the power necessary to accelerate across an intersection. It's also easy to design a simple Stirling engine control device that can change power levels quickly but allows the engine to continue to consume fuel at the full power rate even while producing low amounts of power. However it seems to be quite difficult to design a power control mechanism that can change power levels both quickly and efficiently. A few research Stirling engines have done this, but they all used very complex mechanical methods for achieving their goal.

Stirling engines do not develop power immediately after the heat source is turned on. It can take a minute or longer for the hot side of the engine to get up to operating temperature and make full power available. Automobile drivers are used to having full power available almost instantly after they start their engines.

In spite of these difficulties, there are some automobile Stirling applications that make sense. Hybrid electric cars that include both batteries and a Stirling engine generator would probably be an extremely effective power system. The batteries would give the car the instant acceleration that drivers are used to, while a silent and clean running Stirling engine would give drivers the freedom to make long trips away from battery charging stations. On long trips, the hybrid car could burn either gasoline or diesel, depending on which fuel was cheaper.

To generate electricity for homes and businesses, research Stirling generators fueled by either solar energy or natural gas have been tested. They run on solar power when the sun is shining and automatically convert to clean burning natural gas at night or when the weather is cloudy.

There are no explosions inside Stirling engines, so they can be designed to be extremely quiet. The Swedish defense contractor Kockums has produced Stirling engine powered submarines for the Swedish navy that are said to be the quietest submarines in the world.

Aircraft engines operate in an environment that gets increasingly colder as the aircraft climbs to altitude, so Stirling aircraft engines, unlike any other type of aircraft engine may derive some performance benefit from climbing to altitude. The communities near airports would benefit from the extremely quiet operation that is possible. Stirling engines make sense where these conditions are met:

1. There is a premium on quiet.
2. There is a very good cooling source available.
3. Relatively slow revolutions are desired.
4. Multiple fuel capacity is desired.
5. The engine can run at a constant power output.
6. The engine does not need to change power levels quickly.
7. A warm-up period of several minutes is acceptable.

LOW TEMPERATURE DIFFERENCE ENGINES

In 1983, Ivo Kolin, a professor at the University of Zagreb in Croatia, demonstrated the first Stirling engine that would run on a heat source cooler than boiling water. After he published his work, James Senft, a mathematics professor at the University of Wisconsin, River Falls built improved engines that would run on increasingly small temperature differences, culminating in an elegant and delicate Stirling engine that would run on a temperature difference smaller than 1°F.

These delicate engines provide value as educational tools, but they immediately inspire curiosity into the possibility of generating power from one of the many sources of low temperature waste heat (less than 100°C) that are available. A quick look at the Carnot formula shows that an engine operating with a hot side at 100°C and a cold side at 23°C will have a maximum Carnot efficiency of [((373 K–296 K)/373 K) × 100] about 21 percent. If an engine could be built that achieved 25 percent of the possible 21 percent Carnot efficiency it would have about 5 percent overall Carnot efficiency.

That figure seems quite low until one realizes that calculating Carnot efficiency for an engine that uses a free heat source might not make much sense. For this type of engine it would probably be more worthwhile to first consider what types of engines can be built, then use dollars per watt as the appropriate figure of merit.

Stirling engines that run on low temperature differences tend to be rather large for the amount of power they put out. However, this may not be a significant drawback since these engines can be largely manufactured from lightweight and cheap materials such as plastics. These engines could be used for applications such as irrigation and remote water pumping.

CRYOCOOLERS

It isn't immediately obvious, but Stirling engines are a reversible device. If one end is heated while the other end is cooled, they will produce mechanical work. But if mechanical work is input into the engine by connecting an electric motor to the power output shaft, one end will get hot and the other end will get cold. In a correctly designed Stirling cooler, the cold end will get extremely cold. Stirling coolers have been built for research use that will cool to below 10 K. Cigarette pack sized Stirling coolers have been produced in large numbers for cooling infrared chips down to 80 K. These micro Stirling coolers have been used in high end night vision devices, antiaircraft missile tracking systems, and even some satellite infrared cameras.

Brent H. Van Arsdell

See also: Cogneration Technolgies; Diesel Cycle Engines; Gasoline Engines; Steam Engines; Stiring, Robert.

BIBLIOGRAPHY

Senft, J. R. (1993). *An Introduction to Stirling Engines*. River Falls, WI: Moriya Press.

Senft, J. R. (1996). *An Introduction to Low Temperature Differential Stirling Engines*. River Falls, WI: Moriya Press.

Walker, G. (1980). *Stirling Engines*. Oxford: Oxford University Press.

Walker, G.; Fauvel, O. R.; Reader, G.; Bingham, E. R. (1994). *The Stirling Alternative*. Yverdon, Switzerland: Gordon and Breach Science Publishers.

West, C. (1986). *Principles and Applications of Stirling Engines*. New York: Van Nostrand Reinhold.

STORAGE

Energy storage is having energy in reserve for future needs. While it takes millions of years to create huge energy stores in the form of coal, oil, natural gas, and uranium, humans are able to quickly manufacture energy stores such as batteries, and to build dams to store the gravitational energy of water. Although batteries and dams are very useful and important storage systems, the amount stored is very small compared to natural stores. These natural energy stores are large,

Solar panels on a house roof in Soldier's Grove, Wisconsin. (Corbis Corporation) but finite, and it is important to realize that they will not last forever.

The food a person eats is a personal energy store banking on photosynthesis. In photosynthesis, gaseous carbon dioxide, liquid water, and solar energy interact to produce solid carbohydrates, such as sugar and starch, and gaseous oxygen.

$$\text{carbon dioxide} + \text{oxygen} + \text{solar energy} \rightarrow \text{carbohydrates} + \text{oxygen}$$

Solar energy for photosynthesis is converted and stored in the molecular bonds of carbohydrate molecules. When the process is reversed by combining carbohydrates with oxygen, stored energy is released. We literally see and feel the results when wood is burned to produce heat. A flame is not produced when carbohydrates and oxygen interact in the human body, but the process is similar and appropriately called "burning." The energy we derive from food for performing our daily activities comes from burning carbohydrates.

About 80 percent of the electric energy used in the United States is derived from stored energy in coal. The stored energy has its origin in photosynthesis. Coal is the end product of the accumulation of plant matter in an oxygen-deficient environment where burning is thwarted. Formation takes millions of years. Proven reserves of coal in the United States are upwards of 500 billion tons, a reserve so great that even if coal continues to be burned at a rate of over one billion tons per year, the reserves will last for hundreds of years.

Stores of petroleum, from which gasoline is derived, are vital to industry and the transportation sector. Stores of natural gas are important to industry and the residential and commercial sectors. Both petroleum and natural gas result from the decay of animal and plant life in an oxygen-deficient environment. Solar radiation is the source of the energy stored in the molecules making up all the fossil fuels (i.e., coal, petroleum, and natural gas). The energy is released when petroleum and natural gas are burned. Economic considerations figure prominently in determining proven reserves of petroleum and natural gas. To be proven, a reserve must provide an economically competitive product. There may be a substantial quantity of an energy product in the ground, but if it cannot be recovered and sold at a profit, it is not proven. There are substantive quantities of energy stored as oil in shale deposits in the western United States, but, to date, extracting them are not economical. Proven reserves of oil and natural gas for the United States are around thirty-five billion barrels and 250 trillion cubic feet, respectively. Consumption rates of about three billion barrels per year and twenty trillion cubic feet per year suggest that they will be depleted in the early part of the twenty-first century. But economic conditions can change and new proven reserves will emerge.

An early Energizer watch battery, held against shirt cuff button for size comparison. (Corbis Corporation)

For several decades, the United States has relied on foreign sources of petroleum energy. The availability of petroleum from foreign sources is subject to political instability and changes in the economies in these countries. To guard against losses in petroleum supplies, the United States has established the Strategic Petroleum Reserve. The reserve is an emergency supply of crude oil stored in huge underground salt caverns along the coastline of the Gulf of Mexico. Upwards of one billion barrels of petroleum are stored in six sites along the coasts of Texas and Louisiana.

A battery stores electric energy. Although the concentration of energy is small compared, for example to gasoline, we see a myriad of uses of batteries in radios, cellular phones, flashlights, computers, watches, and so on. The public's demand for these portable products is ever increasing, and scientists strive to develop lighter and better batteries.

An electric power plant generates huge amounts of electric energy, but it does this only on demand by consumers because there is no economical way to store electric energy. Electric power plants are massive units that are difficult to shut down and expensive to start up. At night when consumer demand is low, electric utilities are willing to reduce the cost of electric energy in order to keep the plants operating. During the night, it is often economical to use electricity to run pumps that move water to an elevated position. Then when demand for electricity is high, the water is released to run turbogenerators at a lower level. Such a system is called a pumped storage unit. There are fourteen of these in the United States, each having an electrical output of more than 240 megawatts, which is comparable to a large electric power plant.

A nuclear power plant converts energy stored in uranium and plutonium. Uranium occurs naturally; plutonium is made through nuclear transmutations. The nucleus of a uranium atom consists of protons and neutrons that are bound together by a nuclear force. The stored energy in a uranium nucleus is associated with nuclear forces holding the nucleus together. Energy is released in a nuclear reactor by nuclear (fission) reactions initiated by interaction of neutrons with uranium nuclei. Typically, a single reaction produces two other nuclei and three neutrons. Schematically

$$\mathrm{n} + \mathrm{U} \rightarrow \mathrm{X} + \mathrm{Y} + 3\mathrm{n}.$$

Importantly, the neutrons and protons in the reacting products (n + U) are rearranged into lower energy configurations in the reaction products (X + Y + 3n), and stored energy is released. A single nuclear fission reaction releases upwards of 10,000,000 times more energy than a chemical reaction in the burning of coal. Whereas a coal-burning electric power plant producing 1,000 megawatts of electric power must burn some 10,000 tons of coal per day, a nuclear power plant producing comparable power will need an initial loading of around a hundred tons of uranium of which about twenty-five tons are changed each year.

Water freezes and ice melts at 0°C (32°F). Melting requires 334 kilojoules (kJ) of energy for each kilogram (kg) of ice turned to water. The 334 kJ of energy are stored in the one kg of water. The 334 kj of energy must be removed from the one kg of water in order to convert the water back to ice. If heat is added to water at temperatures between the freezing point (0°C) and boiling point (100°C), the temperature of the water increases 1°C for every 4200 J added to a kilogram of water. The 4,200 kJ/kg °C is called "the specific heat" of water. The specific heat is a property of all materials. Importantly, the specific heat is substantially larger for water than for nearly all other substances in nature. Heat transferred to a substance to increase the temperature is called "sensible heat." The energy added stays within the substance making the substance a reservoir of thermal energy. Because of its high specific heat, water is a superb coolant for an automobile engine or steam turbine. Fortunately, water is cheap and usually available. In a climate where the coolant in an automobile engine might freeze, antifreezes keep this from happening by lowering the freezing point of the coolant.

Solar energy systems are very popular among many for environmental reasons, and people are even willing to pay more for this "clean" energy. The biggest problem with these sources is that the time when energy is available does not necessarily match when it is needed. One way that solar system designers address this problem is by incorporating a thermal storage system. A solar-heated house or commercial building must have a thermal energy store to provide heat at night or on days when clouds block solar radiation. Solar heating systems using roof-top collectors often use water or rocks to store thermal energy. Some solar houses employ large, south-facing windows to maximize the amount of solar energy entering the house. The solar energy may warm a brick or concrete wall or floor so that it becomes a thermal energy store that provides heat when needed at night.

Joseph Priest

See also: Biological Energy Use, Ecosystem Functioning of; Conservation of Energy; Flywheels; Storage Technology.

BIBLIOGRAPHY

“Energy Conversion and Storage,” *NREL: Center for Basic Sciences home page*. Golden, CO: National Renewable Energy Laboratory, Center for Basic Sciences. <http://www.nrel.gov/basic_sciences/>.

Priest, J. (2000). *Energy: Principles, Problems, Alternatives*, 5th ed. Dubuque, IA: Kendall/Hunt Publishing.

STORAGE TECHNOLOGY

MOTIVATION

Electric utilities are expected to provide uninterrupted service. To fulfill this expectation, utilities use a range of energy storage technology so that electricity can be produced in the most economical and efficient manner. Gasoline refined from crude oil is drawn from an underground storage tank when a driver replenishes the fuel in an automobile. Imagine the difficulty meeting the demands of the drivers if the crude oil had to be refined at the time of the replenishment. The task is imposing because consumer demand varies throughout the day and throughout the year. The same is true for utilities providing electricity to consumers.

A large electric power plant produces electricity on consumer demand, and the utility must be prepared to meet demands that vary throughout the day, the week, and the year. The demand is divided roughly into three parts termed baseload, part time, and peak. Base load demand is met with large units having slow response. Usually these works are fueled by fossil fuels, such as coal and natural gas, or uranium. In some cases the electricity may come from a large hydroelectric plant. Generation of baseload electricity is the most economical of the three types. If demand were constant, only baseload electricity would be needed and the cost of electricity to consumers would be minimized. The demand labeled part time is generally provided by the large plants used for baseload. Because of the difficulty of having to start and stop large systems, electricity produced in the part time mode is some two to three times more expensive than baseload electricity. Peak demands requiring fast response are met with oil- and gas-fired turbo-generators, hydroelectricity, pumped-storage hydroelectric systems, purchases from other utilities and, in a few cases, wind-powered generators and solar-powered systems. Peak electricity may cost three to four times more than baseload electricity. Apart from the generating capacity needed to meet peak demands, the distribution system of transmission lines and associated equipment must be able to handle the electric power. Economically, it is in the interest of both the electric utility and the consumer to make an effort to “flatten out” the demand curve. To do this, a utility sometimes encourages consumers to do chores requiring electricity at night by offering a cheaper night rate. It is also in the economic interest of both consumers and utilities if the utilities can produce electric energy during the off-peak times, store the energy in some form, and then convert the stored energy to electric energy during times of greater need. In addition to using pumped storage hydroelectric systems or flywheels for this scheme, there are serious efforts for developing energy storage technology using compressed air, superconducting magnets, and batteries.

EVALUATING ENERGY STORAGE SYSTEMS: ENERGY DENSITY

Every source of energy has a certain mass and occupies a certain amount of space, and both mass and volume are important practical considerations. Clearly, some energy storage systems will be rejected simply because their physical size and cost are impractical. Accordingly, developers of energy storage systems need to know the energy density, measured in kilowatt-hours per kilogram (kWh/kg) or in kilowatt-hours per cubic meter (kWh/m^3). Determining the energy densities for a pumped hydroelectric storage system, which is a practical scheme in many areas of the United States, is straightforward and provides a basis of comparison for other energy storage technologies.

A pumped hydroelectric system makes use of the potential energy of an elevated mass of water. The potential energy (E) of a mass (m) of water elevated a height (H) is $E = mgH$, where g is the acceleration due to gravity. Dividing both sides of this equation by the mass produces $E/m = gH$, which is the energy density in kWh/kg. For water elevated a height of 100 meters, the energy density is about 1000 kWh/kg.

To measure the energy density in kWh/m^3, we relate the mass (m) to mass density (ρ) and volume (V) by $m = \rho V$. The potential energy may then be written as $E = \rho VgH$. Dividing both sides of $E = rVgH$ by the volume produces $E/V = \rho gH$, which is the energy density in kWh/m^3. For H = 100 m the energy density of water is about 0.3 kWh/m^3. Because pumped storage technology is workable, the energy density provides a benchmark for evaluating other energy storage systems.

COMPRESSED AIR ENERGY STORAGE (CAES)

Compressed air can be stored in a container until it is released to turn the blades of a fan or turbine. To determine the energy density of the compressed air, one has to make thermodynamic assumptions about the process. For reasonable assumptions, the energy density is about 2 kWh/m^3, or about 10 times the energy density of water in a pumped storage hydroelectric system. Compressed air energy storage is very attractive from energy density considerations. On the other hand, the volume required for significant amounts of energy is quite large. For example, storing a million kilowatt-hours of energy requires a volume of about 500,000 cubic meters. For comparison, a box 50 meters wide and 100 meters long (roughly the size of the playing area of a football field) and 100 meters high has a volume of 500,000 cubic meters. At this time, there are two operational compressed air energy storage facilities in the world, and both use natural underground reservoirs for the compressed air.

The compressed air energy storage unit operated by the Alabama Electric Cooperative in McIntosh, Alabama utilizes a 19-million cubic foot underground cavern. Air is released during peak periods, heated with natural gas, and expanded through a turbogenerator to produce 110 megawatts of electric power. There is sufficient energy stored to deliver power for 26 hours. A compressed air energy storage system in Huntoff, Germany employs 300,000 cubic meters of space in an underground cavity in a natural salt deposit. Electric generators in the system produce 300 megawatts and there is sufficient energy stored to operate the generators for 2 hours. The Electric Power Research Institute (EPRI) has estimated that more than 85 percent of the United States has geological characteristics that could accommodate an underground CAES reservoir. If the disparity in costs between baseline and part-time/peak load increases, the use of more underground CAES reservoirs is likely.

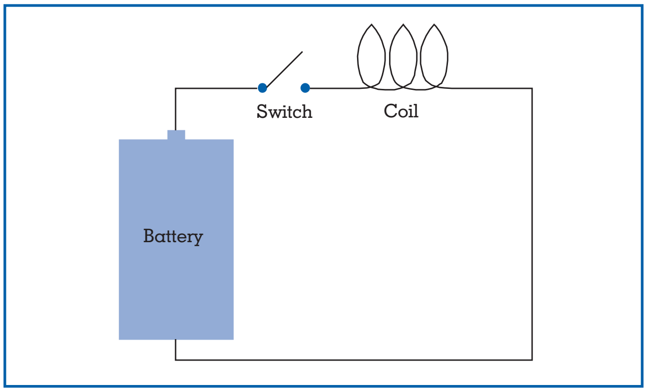


Figure 1.
An elementary electric circuit.

SUPERCONDUCTOR MAGNETIC ENERGY STORAGE (SMES)

Electric utilities can anticipate peak demands and be prepared to meet them. On the other hand, there are short-term fluctuations in the power levels that a utility cannot anticipate but must handle quickly. Superconducting magnetic energy storage (SMES) is a technology that has potential for meeting both anticipated peak demands and those requiring quick response.

Imagine an electric circuit consisting of a battery, a switch, and a coil of wire (Figure 1). There is no electric current in the coil when the switch is open. When the switch is closed, a current is initiated and a magnetic field is produced in the region around the wire. This is an example of the Oersted effect. As long as the magnetic field is changing, voltage is induced in the coil according to an example of Faraday's law. The induced voltage provides opposition to the current established by the battery, this opposition is the essence of Lenz's law. Eventually the current will stabilize at a constant value and a constant magnetic field is established. Electric forces provided by the battery do work to establish the magnetic field and this work gives rise to energy stored in the magnetic field. The energy can be recovered when the switch is opened and the current eventually falls to zero. The amount of energy stored in the magnetic field depends on the electric current and a property of the coil called the inductance. Inductance depends on the construction of the coil and the medium in which the coil is placed. Formally, the magnetic energy is given by $E = (1/2) LI^2$, where L and I represent the inductance and the electric current. For any coil, it would appear that any amount of magnetic energy could be stored by producing a sufficiently large current. However, the wire from which the coil is made has electric resistance, and the combination of current and resistance produces heat. Accordingly, the battery must continually do work and expend energy in order to maintain the current. For ordinary materials, there is a practical limit to the current that can be maintained. Despite the practical limitations imposed by heat considerations, there are many practical applications utilizing energy stored in the magnetic field of a current-carrying coil.

The electrical resistance of most conductors, metals in particular, decreases as the temperature of the conductor decreases. For some pure metals and compounds of the metals, the resistance decreases with temperature as usual, but at some critical temperature the resistance drops identically to zero. The resistance remains zero as long as the material is maintained at a temperature below the critical temperature. Such a material is termed a superconductor and the zero resistance state is termed superconductivity. Mercury, first observed to be a superconductor by Heike Kamerlingh Onnes in 1908, has a critical temperature of 4.2 K (room temperature is approximately 300 K). Since Onnes's discovery several other superconductors have been identified but until 1986 the material with the highest critical temperature, 23 K, was a compound of niobium and germanium (Nb_3Ge). In spite of requiring a very low temperature, several very important technologies involving large magnetic fields evolved, including magnetic resonance imaging for medical diagnostic purposes, and superconducting magnets for particle accelerators and nuclear fusion research. Technology for magnetic energy storage never advanced significantly because of the economics involved in producing the low temperatures required for the superconducting state of existing materials.

The optimism about inexpensive superconductivity was stimulated by two notable discoveries. In 1986 Alex Muller and Georg Bednorz discovered a new class of superconductors, ceramic in form, having a critical temperature of 35 K. In 1987 Paul Chu produced a compound that became superconducting at 94 K. While 94 K is still a very low temperature, it is easily and inexpensively attainable with liquid nitrogen (77 K). Materials with critical temperatures as high as 135 K have been found and there is speculation that someday a material may be discovered that is superconducting at room temperature. Because the critical temperatures of these superconductors are considerably higher than that of their metallic counterparts, the phenomenon has been labeled "high-temperature superconductivity" or HTS.

Although no heat is produced by current in a superconductor, the current cannot be made arbitrarily large because the superconducting state vanishes for a well-defined critical current. Because magnetic energy is proportional to the square of the current, research on HTS for magnetic energy storage focuses on developing materials with high critical current. The technology has proceeded to the development stage. For example, a superconducting magnetic energy storage system for quick response to voltage fluctuations has been field tested at Carolina Power and Light. Energy is drawn from the magnetic storage system and fed to the electrical system during moments when power drops because of faults in the electrical system.

BATTERY STORAGE

When a battery, such as the lead-acid type used in automobiles, is charged with an electric generator, an electric current stimulates chemical reactions that store energy. When a device, such as a light bulb, is connected to the battery's terminals, the chemical reactions reverse and produce an electric current in the bulb. The energy stored by a battery is relatively low. For example, a lead-acid battery has an energy density of about 0.05 kWh/kg. Nevertheless, batteries are extremely useful and they play an increasingly important role in a variety of energy storage applications for electric utilities.

In the electric power industry, batteries are used in conjunction with other ways of generating electricity. For example, a home may employ photovoltaic cells or a wind turbine for electricity. However, photovoltaic cells cannot operate without sunlight and wind turbines cannot produce electricity without wind. Batteries charged by photovoltaic cells or wind turbines in off-use times or by the electric utility at hand can provide electricity when the local system is inoperative. The electric power delivered in this mode tends to be in the range of 1 to 4 kilowatts. Batteries have been used for this purpose since the 1970s.

Electricity is delivered to a home or industry at a nominal voltage and frequency. For a home, the voltage is nominally 230 volts, which is reduced to 115 volts for most appliances, and the frequency is 60 hertz. For a variety of reasons, either (or both) the voltage and frequency may change. For example, if the demand during a peak period is excessive, the voltage level may drop. Voltage fluctuations are fairly common and sometimes abrupt. Brief fluctuations sometimes cause computers, for example, to "crash." Avoiding system crashes is important because they are costly and reduce productivity. Although the technology of quality power storage systems that can quickly correct voltage fluctuations is not widespread, there is interest in them and development is proceeding. These systems would likely be designed for a commercial establishment, an industrial complex, or a village, and would have power capability between 30 and 100 kilowatts.

A typical large coal-burning or nuclear-fueled electric power plant is capable of delivering about 1,000 megawatts of electric power, and an average home during peak usage needs about 3 kilowatts of electric power. Some 330,000 homes could be served if all the power from a large plant went to homes during peak usage. Because the power involved is significant, changes in demand during peak periods can be significant, and a utility must be prepared to meet them. One way of doing this is with storage batteries. The technology is complicated by having to change the alternating current (a.c.) voltage produced by a generator to direct current (d.c.) for charging the batteries and then changing d.c. voltage from the batteries to a.c. to meet peak demands. Nevertheless, the technology exists. The power delivered by these battery storage systems tend to be on the order of megawatts. In 1997, more than 70 megawatts of capacity was installed in ten states across the United States.

Type	*Energy* (kWh/kg)	*Power* (kW/kg)
Lead acid	0.050	0.200
Nickel cadmium	0.053	0.160
Nickel metal hydride	0.070	0.175
Sodium sulfur	0.085	0.115
Lithium iron sulfide	0.095	0.104
Zinc bromine	0.075	0.045
Long-term goal of the U.S. Advanced Battery Consortium	0.2	0.4

Table 1.
Status of Battery Technology in 1999

The low energy density of batteries is a major deterrent to more widespread use in the electric utility sector. Although progress in developing better batteries has been slow and discouraging, research continues. The status of battery development at the time of this writing is presented in Table 1. A battery subscribing to the long-term goal of the United States Advanced Battery Consortium would still have an energy density about 70 times less than that of gasoline (about 15 kWh/kg).

COOL THERMAL ENERGY STORAGE

Electric power usage peaks in summer in most areas mainly because of demand for air conditioning. If consumers could satisfy their air conditioning needs without operating energy intensive conventional systems, then the electric utility could reduce the peak demand. This can be accomplished using electricity for conventional refrigeration in non-peak periods, usually at night, to remove heat from a storage medium such as water or materials that undergo a phase change around room temperature. Then, during peak demand periods, the storage medium is tapped for air conditioning and other cooling purposes. During the 1970s and early 1980s, utilities extended cost incentives to customers for installing thermal storage systems. The financial incentive was based on the cost of constructing new power plants. Offering a rebate to customers who shifted peak demand was viewed as a very cost-effective alternative to acquiring new generating capacity, and thermal storage was seen as one of the best ways to shift demand.

There are over 1,500 cool thermal energy sytems in operation in the United States. Eighty seven percent of all systems employ ice storage because it requires smaller tanks and can be installed in a large range of configurations. It is a well-developed technology and a number of systems have been installed. For example, an ice-storage plant located near the Palmer House Hotel in Chicago cools several commercial buildings in the Loop area.

FUTURE GROWTH

Better energy storage technology benefits everyone. For consumers, batteries that are lighter and longer lasting and that recharge more rapidly will continue to accelerate the growth in cordless technology, technology that offers consumers ever greater mobility. For the utilities, more and better storage options limit the need to invest in peak capacity generation, and thereby help lower the electric rates of customers. And for industry, whose rates are discounted if demand can be shifted away from peak periods, there is a great financial incentive to install storage systems. With so many benefits from research and development, the significant storage technology advances in the twentieth century should continue well into the twenty-first century.

Joseph Priest

See also: Batteries; Capacitors and Ultracapacitors; Flywheels; Hydroelectric Energy.

BIBLIOGRAPHY

Electric Power Research Institute. (1987). *Commercial Cool Storage Design Guide*. Washington: Hemisphere Pub. Corp.

Garg, H. P.; Mullick, S. C.; and Bhargava, A. K. (1985). *Solar Thermal Energy Storage*. Boston: D. Reidel Publishing Col.

Ter-Gazarian, A. (1994). *Energy Storage for Power Systems*. Stevenage, Harts., U.K.: Peter Perigrinus Ltd.

STRATEGIC PETROLEUM RESERVE

See: Reserves and Resources

SUBSIDIES AND ENERGY COSTS

Subsidy is a transfer of money from government to an individual or a firm to stimulate undertaking a particular activity. Subsidies and subsidy-like programs are major parts of energy policies. However, such energy grants are only one example of government aid and not the most important examples. The basic economic principle that people respond favorably to financial incentives implies that such transfers stimulate the action that is aided.

More broadly, subsidy can comprise any government policy that favors an action. Such assistance includes tax provisions that lessen the burdens on some activities, and regulations designed to promote an activity. Government ownership with regular deficits, whether or not consciously planned, is another subsidy mechanism. Particularly in the United States, where the government is a major landowner, access to government property at below-market prices is another possible form of subsidy. An additional source of subvention is regulation causing the affected firms to favor (cross-subsidize) some customers.

Conversely, in principle but not in practice, completely nongovernmental subsidy could arise in energy. Private subsidies do occur, as with parental support or patronage of the arts. These illustrations suggest that the incentives to private aid are limited. More critically, only governments have the taxing, spending, and regulating powers to facilitate extensive subsidies.

THE NATURE AND EFFECTS OF SUBSIDIES

Implementation can differ widely. Aid may be freely and unconditionally available to anyone undertaking the favored activity, or restrictions may be imposed.

The awards may have an upper limit, be dependent on the income or other characteristics of the recipient, or require the beneficiary to make a contribution from its own resources. Subsidies may be awarded by holding contests to select beneficiaries.

The impact of a subsidy depends upon many factors. These include first the magnitude of the gap between costs and what can be earned in an unregulated market. To attain any goal, the subsidy must fully cover that gap. The more of an activity the government wants to encourage, the more aid is needed. The effectiveness of an outlay depends upon the elasticity (responsiveness to incentives) of production and consumption. In principle, everything from a thriving new industry to total failure can emerge.

THE ECONOMICS OF SUBSIDIES

Economic theory stresses efficiency—encouraging placing resources in their most valuable uses. An extensive theory of market failure shows when markets might fail to lead resources to their most valuable uses and when subsidies could be desirable. The failures most relevant to energy subsidies are (1) publicness, (2) cost-recovery problems, and (3) imperfect knowledge. Publicness is the simultaneous consumption of a product (or the absorption of damages) by all of society. Cost-recovery problems stem from the existence of economies of scale exist such that firms cannot price efficiently without losing money. Imperfect knowledge means that market participants may lack the knowledge needed to make the most efficient choice.

Economists differ widely in their views on the practical importance of these market failures, on what remedies are appropriate, and whether governments can satisfactorily remedy the problem. One clear principle is that the problem be identified as precisely as possible and the assistance be targeted as closely as possible to the problem. For example, if information is inadequate, providing knowledge about options would be preferable to aiding a specific option. The three justifications for subsidies set a *minimum* test of aid that many programs fail to pass.

Further problems arise in moving from principles to practice. To implement an assistance program, quantitative estimates of the magnitude of market failure must exist. Accurate measurement is a formidable problem. Determination of impacts is often difficult. Disputes usually arise about the importance of subsidy to market outcomes. Moreover, with complex tax systems, it is difficult to determine what comprises equal treatment of all activities and what is a favorable (tax-subsidy) measure.

Economists also recognize that another concern is about equity, the fairness of the distribution of income. Economists differ as much as everyone else about the proper criteria of equity. However, economic principles create considerable agreement about how best to attain any given equity goal. A primary principle is that the aid should go to people rather than specific commodities.

Several considerations are involved. First is a preference for consumer sovereignty, allowing recipients to choose how to spend. Second, being a consumer or a producer of a particular commodity is a poor indicator of whether assistance is deserved. Everyone from the poorest residents (but not the homeless) to the richest Americans are residential electricity consumers. Third, commodity-specific policies tend more to aid producers in the industry than consumers. Politicians often dispute these appraisals and institute subsidies on equity grounds.

THE PRACTICE OF ASSISTANCE

Energy subsidies are widely attacked from many different viewpoints about public policy. Thus, among economists, both severe critics of the oil companies, such as Blair, and those such as Adelman and McAvoy who stress defective government policy criticize the subsidies. Environmentalists and other supporters of active government often feel that too many programs aid the unworthy.

Others, however, suggest reorientation to more appropriate realms. The valid point is that uncorrected market failure may prevail. However, often the arguments for subsidies are misstated. Some suggest that the existence of inappropriate subsidies elsewhere justifies aid to those excluded. The preferable reaction to bad existing subsidies is their removal. Past waste cannot be recovered.

PRESERVING ENERGY INDUSTRIES

A major element of energy policy in leading industrial countries such as the United States, Germany, Britain, Japan, and France has been provision of substantial aid to existing energy industries. Part of the aid provided by governments went to meet commitments to provide pensions to retired workers. Subsidy-like measures such as trade restrictions, regulations, and forced purchase plans also were critical parts of the assistance.

Programs were instituted to preserve coal production in Western Europe and Japan. Every involved country (Britain, Germany, France, Spain, Belgium, and Japan) transferred taxpayers' money to the coal industry. The programs started in the period between the end of World War II and the resurgence of the oil supply in the late 1950s.

Germany initially had a particularly complex program. Coal sales to the steel industry were largely subsidized by direct government grants. A tax on electricity consumption partially subsidized electric-power coal use. Electricity producers were forced to contract to buy German coal and cover in higher production costs the substantial portion of their costs that the electricity tax failed to repay. In the late 1990s, a court decision invalidating the tax and the expiration of the purchase contract forced a shift to using general tax revenues to provide aid.

Every one of these programs has contracted. Costs mounted, forcing cutbacks in how much coal output was maintained. Belgium allowed its coal industry to die. Britain cut off aid, ended government ownership, and left the industry to survive on its own. France and Japan are committed to ending subvention. Germany still is willing only periodically to reduce the output target goals used to orient assistance. The nominal defenses relate to alleged national security and sovereignty issues. However, critics contend that the goal is an unwise retardation of transfer of employment from the industry.

Reliance on forcing consumers to subsidize is hardly unique to German coal. U.S. "public utility" regulation long has required unequal pricing among consumers. Historically, emphasis was on rates that caused industrial and commercial users to subsidize residential users. This has encouraged industrial customers to secure rate reforms. These are equity measures that are subject to the criticisms stated above.

STIMULATING ALTERNATIVE ENERGY

Since the energy shocks of the 1970s, other efforts to subsidize energy have arisen. One often-used justification is inadequate foresight. For example, encouraging specific forms of nonutility generation of electricity was part of 1978 U.S. federal energy legislation. That legislation gave individual states discretion about evaluating the economic attractiveness of the types of nonutility generation encompassed. These included various special sources, such as solar and wind. The most important category by far was "cogeneration." This was a new name for a long-established (but, at the time, waning) practice of simultaneously producing electricity and steam in industrial plants and at times selling some of the electricity to utilities. States anxious to promote such alternatives issued predictions of future costs upon which to base current decisions. Typically, the forecasts exceeded the prices that actually prevailed. Thus, governments acting to offset allegedly defective private insight provided forecasts to the private sector that turned out to be inaccurate.

OIL AND GAS PRICING

Complex price-control plans have been devised. Examples include cost-based prices of electricity and the multitiered pricing systems associated with U.S. oil and gas regulation. The U.S. government and many other governments are major producers of electricity. At a minimum, this involves subsidies from the failure to charge prices that fully recover costs. Beyond this, much of the electricity is from waterpower and is sold at cost (or, more precisely, a measure of costs based on economically defective regulatory accounting techniques) rather than market clearing prices. The profits that would result from such market prices thus are transferred to those lucky enough to be charged the lower price. Since the power goes into the transmission network and loses its identity, the actual users cannot be identified. Arbitrary rules are devised to determine which of those drawing from the grid are favored with low prices. Historically, access was based on the ownership of the purchasing utilities. Government and cooperative-owned utilities and thus all their customers were favored. Ultimately, the conflict of this rule with the traditional preferences for households was recognized. The rules for the Pacific Northwest, where the largest portion of federal waterpower is produced, were changed in the late 1970s so that household users, whatever the ownership of their distribution utility, received the low price.

The oil and gas price-control systems were complex devices that subdivided each fuel into numerous categories based on such considerations as the time of discovery and the difficulty of extraction. Each category was subject to different rules about initial prices; allowable price adjustments; and, in the case of gas, the duration of control. This system had the potential, realized in gas, for low-cost producers to subsidize high-cost producers. A further complication was that the operative law required charging households lower gas prices than industrial users. The result was the signing of numerous contracts at unsustainably high prices.

Further government regulation that totally changed how natural gas was distributed resolved this difficulty. Historically, pipeline companies were required to act as purchasers and resellers of gas. The new approach of the 1980s limited pipelines to selling transmission services. Customers of the pipelines purchase the gas directly from producers. By exiting from gas purchasing, the pipelines could similarly back out of unattractive contracts.

DEPLETION ALLOWANCES

Those concerned with proper pricing of energy often also allude to a special provision of the U.S. Internal Revenue Service code that gives tax favors called depletion allowances to mineral production. In particular, producers can reduce their taxable income by a percentage of total gross income (that differs both among minerals and over time for individual minerals), but the deduction is capped at 50 percent of net income without the allowance. In 1963, Stephen L. McDonald presented the first extensive analysis of these provisions and found no reason why they should exist. His updates and others have supported this conclusion. The assistance to large companies in the oil and gas sector was eliminated and aid to smaller producers was lowered in the legislation in 1975.

ENERGY RESEARCH

Market failure arguments exist for government aid to some forms of research. Since the results of research are so broadly disseminated, it may be prohibitively expensive for private investors to charge the users of their discoveries. Thus the inventors are undercompensated, and underexpenditure arises unless the government provides aid. In addition, some theorists contend (and others deny) that governments are more farsighted than private firms.

Nuclear power long has been a major element in government energy assistance. Some critics of nuclear power argue that this aid was wasted because no nuclear technologies came close to recovering the investments made and did not generate a surplus that repaid government outlays. A less severe assertion is that only the initial work on light-water reactors (the dominant technology) was justified. Some argue that even aid to that technology was extended past the time at which the private sector was ready to undertake the needed further steps.

Several alternative technologies that were heavily supported failed to become commercially viable. The most obvious case was the fast breeder reactor. Such reactors are designed to produce more fissionable material from nonfissionable uranium than is consumed. The effort was justified by fears of uranium exhaustion made moot by massive discoveries in Australia and Canada. Prior to these discoveries extensive programs to develop breeder reactors were government-supported. In addition, several different conventional reactor technologies were aided. The main ongoing nuclear effort is research to develop a means to effect controlled fusion of atoms.

The greatest success in new fossil fuel technology has a distant government-aid basis. The combustion turbine is a stationary adaptation of the jet engine. The combustion turbine's development and improvement were aided by government military aircraft programs. However, turbine manufacturers independently developed the electric-power version.

Since the decline of the U.S. coal industry from 1946 to 1960, many proposals emerged (and received government funding) to promote coal use. One type of effort sought to transform coal into a more attractive fuel. The options ranged from producing a better solid fuel to synthesizing natural gas and petroleum products. Efforts also were directed at facilitating coal use. These included design of better coal-using devices (such as a longtime dream of a viable coal-using locomotive technology). Radically new methods of generating electricity such as fuel cells and magnetohydrodynamics were advocated. Such technologies might be desirable whatever the fuel choice but also might particularly encourage coal use. Other suggestions involve noncoal alternatives such as petroleum in very heavy crude oil, tar sands, and oil shale. Once again, fears of depletion of oil and gas are used as reasons for the aid. Despite extensive government aid, these technologies failed to succeed.

South Africa massively invested in an oil and gas substitution program involving synthesis from coal; the effort was at least partially motivated by international attempts to curtail South African access to conventional sources of oil. A U.S. effort in the Carter administration of the late 1970s produced only one modest coal gassification venture that continued operation. The latest drives to find replacements for the internal-combustion engine for automobiles are behind the ambitious schedule established in 1992 U.S. energy legislation.

Government efforts to promote solar power and wind have failed to date to produce substantial results. World data for 1997 indicate that only 0.41 percent of consumption was geothermal, solar, or wind; 0.39 percent was geothermal and the others accounted for 0.02 percent. The 1999 U.S. percents are 0.06 geothermal, 0.04 solar, and 0.02 wind.

It is widely agreed among economists that market failures in energy exist. However, the record in practice has produced widespread criticism of implementaion of energy subsidies.

Richard L. Gordon

See also: Economic Externalities; Energy Economics; Governmental Intervention in Energy Markets; Industry and Business, Energy Use as a Factor of Production in; Market Imperfections; Regulation and Rates for Electricity; Supply and Demand and Energy Prices; Taxation of Energy; True Energy Costs; Utility Planning.

BIBLIOGRAPHY

Adelman, M. A. (1993). *The Economics of Petroleum Supply: Papers by M. A. Adelman, 1962–1993.* Cambridge, MA: MIT Press.

Bator, F. (1958). “The Anatomy of Market Failure.” *Quarterly Journal of Economics* 72(3):351–79.

Blair, J. M. (1976). *The Control of Oil.* New York: Pantheon Books.

Bradley, R. L., Jr. (1995). *Oil, Gas & Government: The U. S. Experience.* Lanham, MD: Rowman & Littlefield.

Cohen, L., and Noll, R. (1991). *The Technology Pork Barrel.* Washington, DC: Brookings Institution.

Gordon, R. L. (1970). *The Evolution of Energy Policy in Western Europe: the Reluctant Retreat from Coal.* New York: Praeger

Gordon, R. L. (1982). *Reforming the Regulation of Electric Utilities.* Lexington, MA: Lexington Books.

Gordon, R. L. (1987). *World Coal: Economics, Policies and Prospects.* Cambridge, Eng.: Cambridge University Press.

Heilemann, U. and Hillebrand, B. (1992). “The German Coal Market after 1992.” *Energy Journal* 13(3):141–56.

International Energy Agency. (annual since 1978). *Energy Policies and Programmes of IEA Countries.* Paris: Organisation for Economic Co-operation and Development.

MacAvoy, P. W. (1962). *Price Formation in Natural Gas Fields.* New Haven, CT: Yale University Press.

McDonald, S. L. (1963). *Federal Tax Treatment of Income from Oil and Gas.* Washington, DC: Brookings Institution.

SUPERCONDUCTING ELECTRONICS

See: Electricity

SUPERSONIC TRANSPORT

See: Aircraft

SUPPLY AND DEMAND AND ENERGY PRICES

The quantity of energy supplied is the flow of energy brought onto the market, and the quantity of energy demanded is the amount of energy purchased for a particular period of time. Quantity can be measured in terms of the number of kilowatt hours produced by an electric generator in a day, the number of barrels of oil or cubic feet of gas brought to the market in a month, or the number of tons of coal produced and sold in a year. Primary energy takes the form of fossil fuels or electricity from primary, sources including hydro, nuclear, solar, geothermal, and biomass, while secondary energy is electricity generated from fossil fuels.

DATA SOURCES

Data on the quantity of energy supplied, called energy production, are available from a variety of government, trade association, and international sources. Some of the better sources include the U.S. Energy Information Administration, the American Petroleum Institute, the International Energy Agency, and the United Nations. Secondary energy quantity is reported as net or gross. Net energy is the amount of energy produced and gross is the amount of primary energy required to produce it.

Quantities of energy demanded and supplied are reported in a bewildering variety of ways. The three typical units of measurement are energy content, volume and weight. Energy content includes British thermal units in the United States or kilocalories and kilojoules in the rest of the world, as well as coal or oil equivalents. Kilowatts are the most common units of measurement for electricity and are sometimes applied to other energy sources as well. Weights are most often expressed as short tons, long tons or metric tonnes. Volumes are expressed as barrels, cubic feet, gallons or liters.

DETERMINANTS OF ENERGY SUPPLY

The quantity of energy supplied is a function of the economic and technical variables that influence the cost of bringing the energy source to market and the price that a supplier receives for the energy source in the market. For example, in a competitive market the quantity of coal (Q_s) supplied at the mine mouth is a function of the price received for the coal (P); the price of the capital necessary to produce the coal, such as drag lines, cutting tools, and loaders (P_k); the price of labor, which includes wages, salaries, and indirect labor costs (P_l); the price of using land or any other natural resource or other factor of production (P_n); and technical variables that could include the technologies available and the geology of the deposits (T). An increasing T in this context represents better technology or more favorable geology.

Prices of other related goods also influence quantity supplied. For example, uranium production may produce vanadium as a byproduct. Thus, uranium and vanadium are complementary goods or are goods produced together. When the price of vanadium increases, uranium production becomes more profitable and will increase. Gas found with oil is called associated gas and is considered complementary to oil in production. If the price of oil goes up, drillers may look harder for oil and find more gas to produce as well.

Alternatively, goods may be substitutes in production. If the price of other minerals increase, coal producers may look for other minerals and produce less coal. If gas is nonassociated, it is found without oil and is a substitute for oil in supply rather than a complement. If the price of natural gas, a substitute good (P_s) or a good that could be produced instead of oil, increases, drillers may spend less time looking for oil and more time looking for and producing nonassociated gas, thus decreasing oil production.

Governments often interfere in energy markets and their policies may influence quantity supplied. For example, environmental regulations (E_r) that require less pollution or more safety when producing fuels decrease the quantity of fuel supplied. Such regulations in the United States include the removal of sulfur from fuels, the addition of oxygenates to gasolines in some areas of the country, and more safety in coal mines. Additional environmental regulations increase cost and should decrease quantity supplied. Aggregate market supply also depends on the number of suppliers (#S) in the industry.

We can write a general supply function as follows:

$$Q_s = f(+ P + P_c - P_s - P_l - P_k - P_n + T - E_r + \#S)$$

The sign before each variable indicates how the variable influences quantity supplied. A minus sign indicates they are inversely related and a positive sign indicates they are directly related. Thus, in the example of coal, raising the price of coal is likely to increase quantity supplied, whereas raising the price of labor is likely to decrease the quantity supplied.

For nonrenewable energy sources such as fossil fuels, expectations about the future price and interest rates influence the current quantity supplied. Expectations of higher future prices should cause less production today and more production tomorrow.

ELASTICITY OF SUPPLY

A measure of how responsive quantity supplied is to a variable (say price) is called the elasticity of supply with respect to that variable. Elasticity of supply is the percentage change in quantity divided by the percentage change in the variable in question or

$$\frac{\%\ \text{change}\ Q_5}{\%\ \text{change}\ P}$$

If the supply price elasticity of oil is 1.27, it follows that if the price of oil increases by 1 percent, the quantity of oil supplied increases by 1.27 percent. A cross elasticity of supply indicates how quantity produced is related to another price. For example, if the cross elasticity of oil supply with respect to the price of gas is 0.15, then if the price of gas increases 1 percent, the quantity of oil produced goes up 0.15 percent. Because energy production is capital-intensive, supply price elasticities are larger or more elastic in the long run than in the short run. The long run is the time it takes for producers to totally adjust to changing circumstances and allows for totally changing the capital stock. In contrast, in the short run capital stock is fixed and total adjustment does not take place. Often the short run is considered a year or less, but the exact length of time depends on the context.

Information about supply elasticities would be highly useful for those involved in energy markets, but unfortunately little is available. Carol Dahl and T. Duggan (1996) surveyed studies that use simple models to estimate energy supply or elasticities. They found estimates for the various fossil fuels and uranium in the United States and concluded that studies estimating these elasticities using reserve costs are the most promising. Such studies yielded a U.S. gas supply own-price elasticity of 0.41, a uranium supply own-price elasticity from 0.74 to 3.08, an Appalachia coal supply own-price elasticity of 0.41 to 7.90, and a U.S. oil supply own-price elasticity of 1.27. Even less is known about cross-price elasticities. Dahl and Duggan (1998) surveyed oil and gas exploration models that include cross-price elasticities for oil and gas but did not find strong statistical results from any of the models.

DETERMINANTS OF ENERGY DEMAND

Energy demand is a derived demand. Consumers and businesses demand energy not for itself but for the services that the energy can provide. A consumer may want energy for lighting, space conditioning in the form of heat in the winters and cooling in the summer, and energy to run vehicles and appliances. Businesses often have these same needs and also need energy to run motors and for process heat.

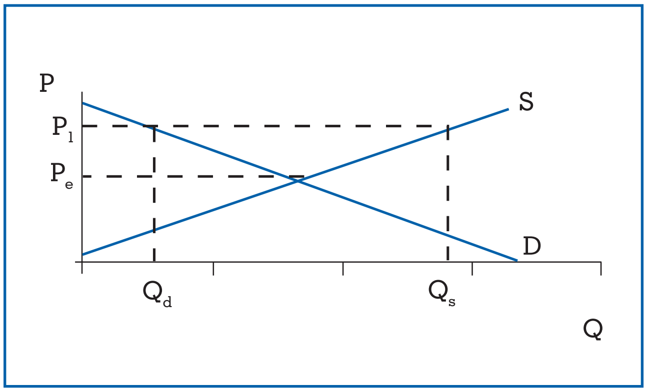


Figure 1.
Energy demand and supply.

For consumers, quantity demanded of energy (Q_{cd}) is a function of the price of energy (P), the price of other related goods, disposable income (Y), and other variables (O) such as personal preferences, lifestyle, weather, and demographic variables and, if it is aggregate demand, the number of consumers (#C). Take for example the quantity of electricity demanded by a household. If the price of electricity increases consumers may use less electricity. If the price of natural gas, a substitute for electricity in consumption (P_s), decreases, that may cause consumers to shift away from electric water heaters, clothes driers and furnaces to ones that use natural gas, thus increasing the quantity of natural gas demanded. If the price of electric appliances (P_c) increases, or decreases quantity of electricity demanded. consumers may buy less appliances and, hence, use less electricity. Increasing disposable income is likely to cause consumers to buy larger homes and more appliances increasing the quantity of electricity consumed. Interestingly, the effect of an increase in income does not have to be positive. For example, in the past as income increased, homes that heated with coal switched to cleaner fuels such as fuel oil or gas. In the developing world, kerosene is used for lighting, but as households become richer they switch to electricity. In these contexts coal and kerosene are inferior goods and their consumption decreases as income increases. We can write a general consumer energy demand function as follows:

$$Q_{cd} = f(-P + P_c - P_s \pm Y \pm O + \#C)$$

Again the signs before the variables indicate how the variables influence quantity. The sign before income (Y) is ±, since the sign would be + for a normal good and – for an inferior good. The sign before other variables is also ±, since the sign depends on what the other variable is. For example, colder weather would raise the demand for natural gas, but an aging population, which drives less than a younger one, would decrease the demand for gasoline.

For businesses the demand for energy is the demand for a factor of production. Its demand depends on the price of the energy demanded (P) as well as the price of its output (P_o), technology (T) and prices of other factors of production—land, labor, and capital—that might be substitutes (P_s) or complements (P_c) in consumption. Environmental policy (E_p) might also affect the demand for fuel. If this is aggregate business demand for energy the number of businesses is also relevant.

We can write a general business energy demand function as follows:

$$Q_d = f(+ P + P_c - P_s \pm T \pm E_p + \#B)$$

The sign on technology (T) and environmental policy (E_p) are uncertain and depend on the particular technology and policy. For example, environmental regulations requiring lower sulfur emissions favor gas over coal, while new technologies that make oil cheaper to use increase oil demand at the expense of gas and coal.

ELASTICITY OF DEMAND

The responsiveness of energy consumption to a variable can be represented by an elasticity, as was the case for energy production. Again the elasticity is measured as the percentage change in quantity over the percentage change in the variable. The demand elasticity is negative, since raising price lowers quantity demanded, and if it is less than –1, the demand is called elastic. Lowering price when demand is elastic means that quantity demanded increases by a larger percent than price falls, thus energy expenditures go up. Alternatively, raising price lowers expenditures. If the demand elasticity is between 0 and –1, then it is called inelastic. Now lowering price means that quantity increases by a smaller percent than price falls. Thus energy expenditures go down. Alternatively, raising price lowers expenditures.

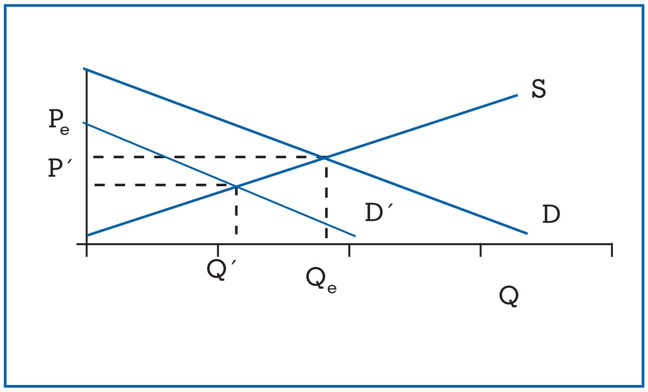


Figure 2.
Shift in demand in response to Asian crisis.

INCOME ELASTICITIES

Income elasticities are positive for normal and noninferior goods, since raising income increases total consumption of these goods. Goods with income elasticities greater than 1 are said to have income elastic demand. Suppose jet fuel has an income elasticity of 3. The in income goes up by 1 percent, the quantity of jet fuel demanded goes up by 3 percent and a larger share of income is spent on jet fuel. Such goods with income elastic demand are called luxuries and they become a larger share of spending as a person gets richer. If the income elasticity is between 0 and 1, it is called inelastic. As income increases, a smaller share of income is spent on goods with inelastic demand. Often necessities such are fuel for heating are income ineleastic.

A cross-price elasticity indicated how the quantity of one good changes when the price of another related good changes. The sign of the cross-price elasticity tells us whether goods are substitutes or complements. Take the case of a cement producer who needs a great deal of heat to produce cement. The producer's cross-price elasticity of demand for natural gas with respect to the price of the substitute good, coal, is the percentage change in natural gas demand divided by the percentage change in coal price. This elasticity for substitute goods will be positive, since an increase in the price of coal will cause an increase in the quantity of natural gas demanded. Alternatively, the cross-price elasticity of demand for complementary goods is negative. If the price of gas furnaces (a complement to gas) goes down, the cement producer may buy more gas

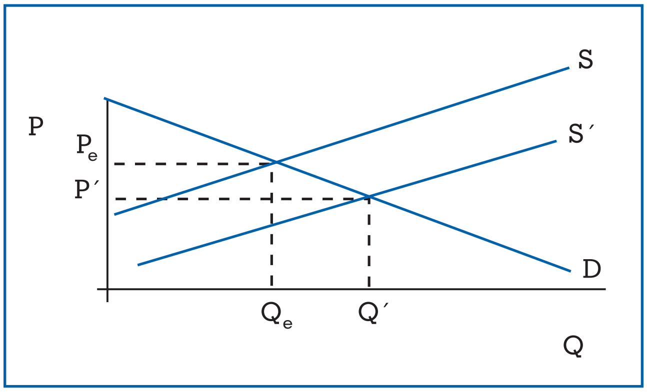


Figure 3.
Shift in supply in response to improved technology.

furnaces, instead of coal furnaces, and use more natural gas.

ESTIMATED ELASTICITIES

More statistical work has been done estimating demand elasticities. Dahl (1993a) surveyed this work for the United States. She found considerable variation in ownprice and income elasticities across studies with the most consistency for studies of residential demand and for gasoline. More products seem to be price and income inelastic and short run elasticities are more certain than long run elasticity. Short run price elasticities for a year are probably between 0 and –0.5 for energy products. Dahl (1995) surveyed transportation demand studies and concluded that the price elasticity of demand for gasoline in the United States is –0.6 and the income elasticity is just below 1. Dahl (1992, 1993b, 1994) looked at energy demand elasticities in developing countries and found that oil and price elasticities of demand are inelastic and near –0.3, while income elasticities of demand are elastic and near 1.3. Dahl (1995b) surveyed natural gas's own, cross and income elasticities in industrial countries but was not able to come to a conclusion about the magnitude of these elasticities. Studies are also often inconsistent about whether coal, oil, and electricity are substitutes or complements to natural gas.

ENERGY PRICES

Elasticities tell us how responsive are quantities demanded and supplied. In a competitive market where many buyers and sellers are competing with one another, the interactions of supply and demand determine the price in energy markets. To see how, we simplify demand and supply by holding all variables constant except price and quantity and graph both functions in Figure 1. Equilibrium in this market is at P_e where quantity demanded equals quantity supplied. If price is at P_1 quantity demanded is Q_d and quantity supplied is at Q_s. There is excess quantity supplied and there would be pressure on prices downward until quantity demanded equaled quantity supplied. Thus interactions of demand and supply determine price.

CHANGE IN DEMAND

If one of the variables held constant in the demand curve were to change, it would shift the whole demand curve, called a change in demand. For example, suppose Figure 2 represents the world market for crude oil. The Asian crisis beginning in 1997 reduced Asian income, which in turn reduced Asia's demand for oil. This decrease in demand lowered price moving along the supply curve, called a decrease in quantity supplied.

CHANGE IN SUPPLY

The 1990s have seen technical changes in finding oil such as 3D seismic and horizontal drilling that have reduced costs. These technical changes shifted the supply of oil as shown in Figure 3. This shift is called a change in supply and is the result a change in the a variable other than the own price. The increase in supply lowers price causing a movement along the demand curve called a change in quantity demanded.

MARKET SETS PRICES

Thus, as economic and political events occur along with changes in demography, preferences and technology, shifting demand and supply interact to form prices in competitive energy markets. The above discussion assumes competitive markets where consumers and producers compete to buy and to sell products, and they have to take market price as given. This is probably the case most often for buyers of energy products. However, in the case of production, sometime market power exists. For

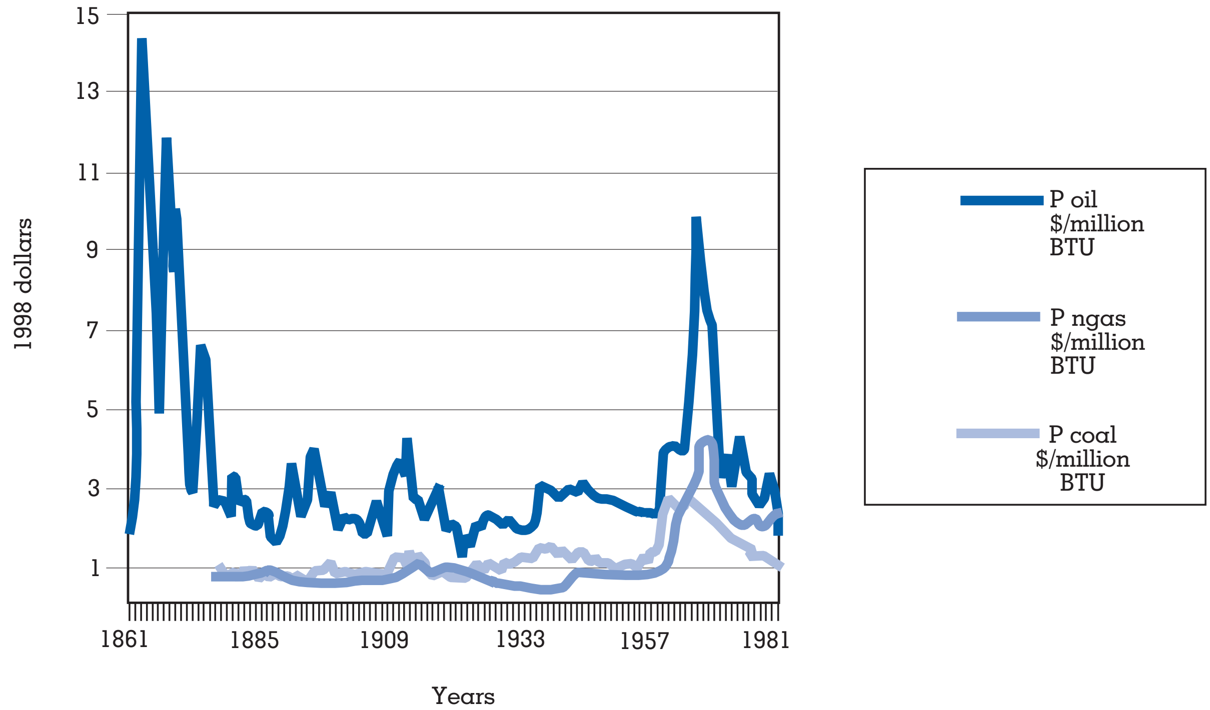


Figure 4.
U.S. prices for coal, oil, and natural gas, 1861–1981.

SOURCE: U.S. Department of Commerce, Bureau or the Census. (1975); U.S. Energy Information Administration, Department of Energy. (1998).

NOTE: Prices have been converted into Btus assuming 5,800,000 Btus per barrel of oil, 3,412,000 Btus per 1000 kWh, 1,000,000 Btus per 1000 mcf of natural gas, and 22,500,000 Btus per short ton of coal. A Btu is about ¼ of a food calorie or a kilocalorie, 1,000 Btus contain the energy content of a candy bar.

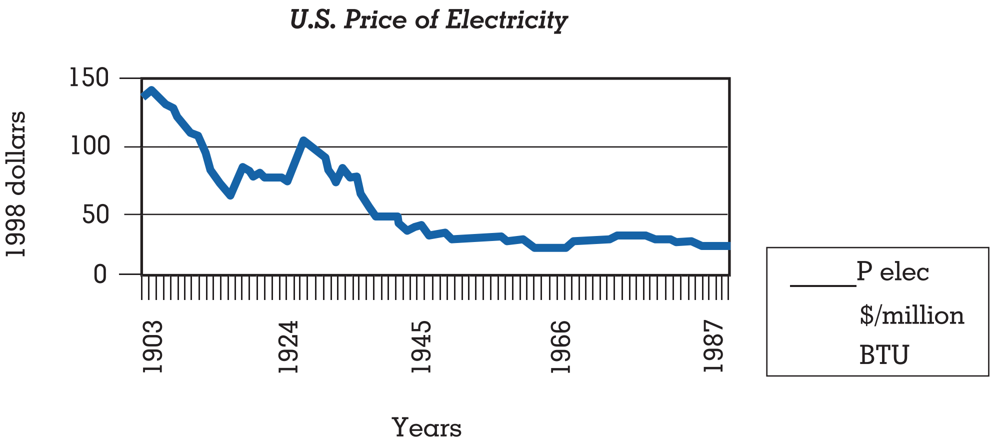


Figure 5.
U.S. Prices for Electricity, 1913–2000.

SOURCE: U.S. Department of Commerce, Bureau or the Census. (1975); U.S. Energy Information Administration, Department of Energy. (1998).

example, in the oil market, first Rockfeller, then the large multinational oil companies, and state regulatory commissions such as the Texas Railroad Commission and then OPEC have exercised pricing power. In such a case, the producer tries to influence price or quantity in order to receive higher prices and earn excess profits.

FOSSIL FUEL PRICES

Figure 4 gives an historical overview of how fossil fuel prices have changed in response to market changes. Price availability varies from series to series and is reported for the period 1861–1998 for the well head price of oil (Poil), 1880–1998 for the fob mine price of bituminous coal, and 1922-1998 for the well head price of natural gas. Figure 4 shows the volatility of oil prices as cartels formed and raised prices and then lost control as higher profits encouraged entry by other. Since the oil market is a global market these prices would be similar to those on the world market, as would prices for coal. The coal market has been reasonably competitive, and from the figure, we can see its price has been more stable than oil. However, coal price has been increased by oil price changes since oil is a substitute for coal in under-the-boiler and heating uses.

Sometimes governments interfere with markets by setting price controls. In the United States wellhead price controls were set on natural gas sold into interstate markets beginning in the early 1950s. These controls which were not completely removed until the early 1990s and sometimes caused natural gas shortages. Prices, which had been reasonably stable, became quite volatile as the price controls were increasingly relaxed throughout the 1980s and into the 1990s.

For electricity, economies of scale have existed making it more economical for one firm to produce and distribute electricity for a given market. One firm would be able to monopolize the market and earn excess profits. This has led governments to regulate electricity in the United States and to produce electricity in most of the rest of the world throughout much of the twentieth century. In such a case, electricity price is set not by the market but by the government. In Figure 5 we can see the evolution of constant dollar electricity prices in the United States from the early 1900s to the present, based on the average real consumer price of electricity (Plec). The falling real price reflects the cost reductions in producing and distributing electricity.

As the size of markets have increased and the optimal size of electricity generation units have decreased more electricity markets are being privatized and restructured to allow more competition into the markets and less government control over pricing.

Carol Dahl

See also: Subsidies and Energy Costs.

BIBLIOGRAPHY

Apostolakis, B. E. (1990). “Interfuel and Energy-Capital Complementarity in Manufacturing Industries.” *Applied Energy* 35:83–107.

Dahl, C. A. (1992). “A Survey of Energy Demand Elasticities for the Developing World.” *Journal of Energy and Development* 18(I):1–48.

Dahl, C. A. (1993a). “A Survey of Energy Demand Elasticities in Support of the Development of the NEMS” Prepared for the United States Department of Energy, Contract De-AP01-93EI23499 October 19.

Dahl, C. A. (1993b). “A Survey of Oil Demand Elasticities for Developing Countries.” *OPEC Review.* 17(4):399–419.

Dahl, C. A. (1994). “A Survey of Oil Product Demand Elasticities for Developing Countries,” *OPEC Review* 18(1):47-87.

Dahl, C. A. (1995a). “Demand for Transportation Fuels: A Survey of Demand Elasticities and Their Components” *Journal of Energy Literature* 1(2):3-27.

Dahl, C. A. (1995b). “A Survey of Natural Gas Demand Elasticities: Implications for Natural Gas Substitution in New Zealand.” Draft. CSM/IFP Petroleum Economics and Management. Golden, CO: Colorado School of Mines, Division of Economics and Business.

Dahl, C. A., and Duggan, T. (1996). “U. S. Energy Product Supply Elasticities: A Survey and Application to the U.S. Oil Market,” *Resources and Energy Economics* 18:243–263.

Dahl, C. A., and Duggan, T. (1998). “Survey of Price Elasticities from Economic Exploration Models of U.S. Oil and Gas Supply,” *Journal of Energy Finance and Development* 3(2):129–169.

U.S. Department of Commerce, Bureau or the Census. (1975). *Historical Statistics of the United States: Colonial Times to 1970.* Washington, DC: U.S. Government Printing Office.

U.S. Energy Information Administration/Department of Energy. (1998). *Annual Energy Review.* Washington, DC: U.S. Government Printing Office.

Welsh, H. E., and Gormley, J. C. (1995). “Sources of Information on the Oil Industry.” In *The New Global Oil Market: Understanding Energy Issues in the World Economy*, ed. S. Shojai. Westport, CT: Praeger.

SUSTAINABLE DEVELOPMENT AND ENERGY

Sustainable development is defined as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs" (Brundtland, 1987). Applied to energy, it is the ability of a society to continue present energy production and consumption patterns many generations into the future, with a focus on the relationship of available energy resource to the rate of resource exhaustion. A sustainable energy market requires the quantity, quality, and utility of energy to improve over time—energy becomes more available, more affordable, more usable and reliable, and cleaner over time (Bradley and Simon, 2000).

Development and population growth adversely impact the sustainability future by accelerating energy resource exhaustion and environmental degradation. For example, the greater the population growth, the greater the desire for development by the growing population, the greater the desire for short-term exploitation of energy resources for the development (less concern for future generations), and the greater the environmental degradation from the greater development and energy use.

All fossil fuels are considered unsustainable because someday they will reach a point of depletion when it becomes uneconomic to produce. Petroleum is the least sustainable because it is the most finite fossil fuel. Although levels of production are expected to begin declining no later than 2030 (U.S. production peaked in 1970), the U.S. and world reserves could be further expanded by technological advances that continue to improve discovery rates and individual well productivity. The extraction of oils found in shales (exceeds three trillion barrels of oil equivalent worldwide) and sands (reserves of at least two trillion barrels worldwide) could also significantly increase reserves. The reserves of natural gas are comparable to that of oil, but natural gas is considered a more sustainable resource since consumption rates are lower and it burns cleaner than petroleum products (more environmentally sustainable).

Coal is the least expensive and most sustainable fossil fuel energy source (world reserves of 1.2 to 1.8 trillion short tons, or five times the world's oil reserves), but many feel it is the least sustainable from an environmental perspective. Coal combustion emits far more of the pollutants sulfur dioxide, carbon monoxide, nitrogen oxides and particulates, and is the fossil fuel that emits the greatest amount of carbon dioxide—the greenhouse gas suspected to be responsible for global warming. However, advances in technology have made the conversion of all fossil fuels to energy more efficient and environmentally sustainable: fossil-fuel availability has been increasing even though consumption continues to increase; efficiency improvements have saved trillions of dollars, and vehicle and power plant emissions in 2000 were only a fraction of what they were in 1970.

Nuclear energy is more sustainable than the fossil fuels. If uranium is used in fast breeder reactors designed to produce large quantities of plutonium fuel as they produce electricity, the world resources could produce approximately 200 times the total global energy used in 1997. However, concerns about the danger of nuclear power, the high cost of building and maintaining plants, and the environmental dilemma of what to do with the nuclear wastes that result is the reason many consider it less sustainable.

Renewable energy is the most sustainable energy because the sources, such as the sun, wind and water are inexhaustible and their environmental impact minimal. But the problem with renewable resources is that most sources produce intermittently (when the sun is shining, wind is blowing). Except for hydroelectricity, which produces about 18 percent of the world's electricity, all the other renewable energy sources combined produce less than 2 percent. Advances in renewable energy technology, coupled with continued efficiency improvements in energy-using products, may be part of the energy sustainability solution, but the demand for energy from the world's ever-growing population is too great for it to be the only solution.

Over 80 percent of the world's energy consumption comes from nonrenewable sources that cannot be sustained indefinitely under current practices. If technological advances continue to make conventional energy resources plentiful and affordable for many years to come, the transition to more sustainable energy sources can be smooth and minimally disruptive.

John Zumerchik

See also: Efficiency of Energy Use; Emission Control, Power Plants; Emission Control, Vehicle; Oil and Gas, Exploration for; Oil and Gas, Production of.

BIBLIOGRAPHY

Bartlett, A. A. (1997–1998). “Reflections on Sustainability, Population Growth, and the Environment.” *Renewable Resources Journal* 15(4):6–23.

Bartlett, A. A. (2000). “An analysis of U.S. and World Oil Production Patterns Using Hubbert-Style Curves.” *Mathematical Geology* 32(1):1–17.

Bradley, R., and Simon, J. (2000). *The Triumph of Energy Sustainability*. Washington, DC: American Legislative Exchange Council.

Brundtland, G. H. (1987). *Our Common Future, World Commission on Environment and Development*. New York: Oxford University Press.

SYNTHETIC FUELS

Synthetic fuels are usually thought of as liquid fuel substitutes for gasoline and diesel fuel made from petroleum sources. In broad context, the source of these synthetics can be any feedstock containing the combustible elements carbon or hydrogen. These include coal, oil shale, peat, biomass, tar sands, and natural gas. Water could be included here because it can be decomposed to produce hydrogen gas, which has the highest heating value per unit weight of any material (excluding nuclear reactions). The conversion of natural gas is treated in a separate article, so this article will emphasize solid feedstocks.

From a practical standpoint, coal, because of its abundance, has received the most attention as a source for synthetic fuels. As early as 1807, a coal-gas system was used to light the streets of London, and until the 1930s, when less expensive and safer natural gas started to flow through newly constructed pipelines, gas piped to homes in the Eastern United States was derived from coal. Kerosene, originally a byproduct from the coking of coal for metallurgical applications, can be considered the first synthetic liquid fuel made in quantity. But once crude oil became cheap and abundant, there was little serious research on synthetic liquid fuels in the industrial world until the Energy Crisis of 1973. The main exceptions to this generalization are the important work on coal conversion in Germany, cut off from oil imports during the two World Wars, and the Sasol Process in South Africa, which produces a synthetic, waxy “crude oil” from indigenous coal deposits.

The Sasol Synfuel refinery in Secunda, South Africa, produces synthetic fuels for consumers. (Corbis Corporation)

After 1973 the United States invested heavily in synthetic fuel research and development, hoping synthetics could serve as economical substitutes for crude oil. However, coal conversion is not profitable unless the price of crude oil is over $50 per barrel, which is why the processes developed were mothballed when world crude oil prices fell in the 1980s.

CONVERSION

The major chemical difference between natural gas, crude oil, and coal is their hydrogen-to-carbon ratios. Coal is carbon-rich and hydrogen-poor, so to produce a synthetic liquid or gas from coal requires an increase in the hydrogen-to-carbon ratio. Coal’s ratio of about 0.8 has to be raised to 1.4 to 1.8 for a liquid, and to over 3 to produce a synthetic gaseous fuel. Natural gas (chiefly methane) has a ratio of 4. This can be done by either adding hydrogen or rejecting carbon.

Addition of hydrogen can involve reacting pulverized coal with hydrogen-rich liquids. More commonly, pressurized hydrogen gas is reacted with coal (hydrogenation) in the presence of a catalyst. The latter scheme also removes many of the noxious sulfurous and nitrogenous impurities in coal by converting them to gaseous hydrogen sulfide and ammonia. Carbon removal entails pyrolysis in the absence of air (coking) to produce varying amounts of gases, liquids, and char, depending on the reaction time-temperature-pressure conditions employed.

Synthetic Fuel Liquids via Gas Intermediates

Liquids can also be synthesized via an indirect scheme where the coal is first gasified in an intermediate step. The coal is pulverized and reacted with steam to produce "water gas," an equimolar mixture of carbon monoxide and hydrogen:

$$C + H_2O \rightarrow CO + H_2 \tag{1}$$

The carbon monoxide can then be further reacted with steam and/or hydrogen in the water gas shift reaction:

$$CO + H_2O \rightarrow CO_2 + H_2 \tag{2}$$

$$CO + 3H_2 \dashrightarrow CH_4 + H_2O \tag{3}$$

Combining Equation 1 and Equation 2, one can generate hydrogen and carbon dioxide:

$$C + 2H_2O \rightarrow CO_2 + 2H_2 \tag{4}$$

And adding reactions from Equation 1 and Equation 3 yields just methane:

$$C + 2H_2 \rightarrow CH_4 \tag{5}$$

These operations carry energy penalties, and the heat of combustion released when burning the methane, hydrogen, or carbon monoxide produced is less than the energy that would have been released had the coal been burned directly. To produce heavier liquids, the equimolar mixture of hydrogen and carbon monoxide (water gas, also known as synthesis gas) is the preferred feedstock. Badische Anilin pioneered the synthesis of methanol at high temperature and pressure after World War I:

$$CO + 2H_2 \rightarrow CH_3OH \tag{6}$$

Shortly thereafter, Fischer and Tropsch discovered an iron catalyst that would convert synthesis gas to a mixture of oxygenated hydrocarbons (alcohols, acids, aldehydes, and ketones) at atmospheric pressure. In the next decade Ruhrchemie developed new cobalt catalysts that could produce a mixture of hydrocarbon liquids and paraffin wax from Fischer Tropsch liquids at moderate pressure. This, plus direct hydrogenation of coal, was the basis of German synfuel capacity during World War II. Finally, in 1955, the South African Coal, Oil and Gas Co. further improved the technology and commercialized the Sasol Process as the basis for South Africa's fuel and chemicals industry. Favorable economics were possible because government mandates and subsidies dictated the use of local coal resources rather than spending scarce hard currency to import foreign crude oil.

Methanol is an excellent, high-octane motor fuel, but it was not cost-competitive with gasoline made from cheap, abundant, crude oil. It also has a lower specific energy (miles per gallon) because it is already partially oxidized. In the late 1970s, Mobil Oil announced its "Methanol to Gasoline" Process, which efficiently converts methanol to C2 to C10 hydrocarbons via a synthetic zsm-5 zeolite catalyst. This coal-methanol-gasoline technology will probably enjoy use as crude oil prices rise in the future. Since natural gas (largely methane or CH_4) is readily converted into methanol via partial oxidation with pure oxygen, this also offers an alternate to low-temperature liquefaction or pipelining as a way to utilize natural gas deposits that are in remote locations.

There is a parallel technology for the partial oxidation of methane to make ethylene and other olefins. These can then be polymerized, alkylated, and hydrotreated as need be to make hydrocarbon fuels in the gasoline and diesel fuel boiling range. Methane can also be partially oxidized to produce oxygenates, such as methyl tert-butyl ether (MBTE), that are used in reformulated gasoline as blending agents.

Direct Liquefaction of Coal

In comparison to the capital-intensive, multistep gaseous route from coal to gasoline, the direct liquefaction of coal in a single processing step is attractive. While the removal of carbon from coal to generate liquids richer in hydrogen sounds simple, there is not enough hydrogen in coal to yield much useable liquid fuel. Further, the process must have a ready market, such as a nearby smelter, to utilize the major coke fraction. Addition of hydrogen during pyrolysis (hydropyrolysis) increases liquid yields somewhat, but the direct hydrogenation of coal has always seemed a more attractive route. Nevertheless, the Office of Coal Research of the U.S. Deptartment of the Interior sponsored research on this scheme for many years. As a result, the Char-Oil-Energy Development (COED) process was developed by the FMC Corporation in the period 1965–1975. In the COED process, 50 to 60 percent of the coal feed is rejected as char, and finding a market for char is problematic. Working on subbituminous western coals, the Oil Shale Corporation developed a similar scheme (Toscoal Process), which also produced 50 percent solid char.

High-pressure hydrogenation of coal was patented by Bergius in 1918, but the liquids were of low quality. However, by the mid 1930s, plants using on the order of 500 tons per day of low-rank coals and coal pitch had been built in England and Germany. Toward the end of World War II, Germany had 12 large “coal refineries” producing over 100,000 barrels a day of motor and aviation fuels for the war effort. After the war, in spite of the abundance of crude oil in the United States, pilot plant work on coal hydrogenation was continued in by the U.S. Bureau of Mines. This work was further developed by Ashland Oil into the “H-coal” process, and a pilot plant was built in 1980. The hydrocarbon liquids produced were rather high in nitrogen and oxygen, but all heavy materials were recycled to extinction so that no pitch or char was produced.

In the 1920s, I. G. Farben discovered that certain hydrogen-rich solvents, such as tetralin, could dissolve heavy coal components. The heavy extracted liquids could more conveniently be upgraded by hydrogenation, and the extracted residue was converted to metallurgical coke and carbon electrodes. After the war a low-level research effort was undertaken in the United States on solvent-refined coal processes, with primary emphasis on producing boiler fuel. Yields of useful liquids were gradually improved by inserting additional processing steps such as hydrocracking some of the heavier fractions.

Spurred by the OPEC oil crisis, the U.S. Department of Energy encouraged major oil companies to increase efforts to produce liquid fuels from coal, tar sands and shale oil deposits. A key development was the Exxon Donor Solvent Coal Liquefaction Process (EDS). It involves reacting a coal slurry with hydrogenated recycle solvent and hydrogen. The donor solvent transfers some of its hydrogen to the coal, is distilled from the reaction products, hydrogenated, and then recycled to the process. The light ends from distillation are steam-reformed to make process hydrogen. The heavy vacuum bottoms are Flexicoked, and the coke is gasified to provide process fuel gas. Except for some residual carbon in the gasifier ash, very little of value in the coal is wasted.

Work has also continued on the solvent-refined coal + hydrocracking concept (the NTSL, or non-integrated, two-stage liquefaction process), and a pilot plant was operated by Amoco, DOE and the Electric Power Research Institute (EPRI) from 1974 to 1992.

In addition to coal, there have been extensive post-oil-crisis studies aimed at utilizing the extensive western oil shale deposits. Oil shale utilization has many problems common to coal conversion and, in addition, extensive inorganic residues must be disposed of. The products are also high in nitrogen, and this increases refining costs. Tar sands constitute a final hydrocarbon reserve of interest. Here the problems are more tractable. The sands can be extracted with hot water to produce a material similar to a very viscous crude oil. It is refined as such, and thus Great Canadian Oil Sands, Ltd. has been able to justify operation of several very large tar sands refineries at Cold Lake, Alberta. The major process involves fluid coking of the heavy bottoms from distillation.

ECONOMIC AND ENVIRONMENTAL OUTLOOK

The coal conversion efficiency to synthetic, pipeline-quality natural gas or liquid crude oil is in the 60 to 70 percent range. This means that only 60 to 70 percent of the latent heat energy in the coal can be obtained by burning the product of the conversion. However, for the lower Btu per cubic foot products of water gas and coke oven gas, conversion efficiencies can reach over 95 percent.

The reason for the poor conversion efficiency to synthetic fuels is the high energy cost in liberating hydrogen from water (thermal dissociation, electrolysis) and, when distillation does not involve water, the partial combustion needed to produce the gas (CO). The price of crude oil would have to rise to around $40 to $50 a barrel, or need government subsidies of $20 to $30 a barrel, for liquid synthetic fuel to be competitive. Because the conversion to gaseous fuels is less complex and costly than liquid fuels, the subsidy or rise in natural gas prices would not have to be as dramatic.

Besides the economic feasibility problem, synthetic fuels face significant environmental hurdles. During direct liquefaction, heavy, high-boiling polyaromatics organics are produced. Scientists are trying to eliminate these carcinogenic fractions by recycling them through the liquefaction process. Production by the indirect method is less problematic because it tends to produce fewer toxic chain hydrocarbons. There is also a significant release of solid, liquid, and gaseous residual waste that comes from the boilers, heaters and incinerators, or as part of the processing stream.

The process will adversely affect air quality by releasing nitrogen oxides, sulfur oxides, carbon monoxides and other particulates into the atmosphere. Better control of the conversion conditions and better control of emissions can make the process cleaner, yet technology cannot do anything to curb carbon emissions. Since much of the carbon in coal is converted to carbon dioxide in the synthesis process, and is not part of the synthetic fuel itself, the amount of carbon dioxide that will be released to the environment during combustion is 50 to 100 percent more than coal, and around three times more than natural gas.

Since most systems use tremendous amounts of water, the production of synthetic fuels will have a detrimental effect on water quality as well. It will require major technological advances to more effectively handle waste streams—waste-water treatment systems, sulfur recover systems and cooling towers—to make synthetic fuels an acceptable option from an environmental perspective. This emission control technology will be expensive, only adding to the economic disadvantages the synthetic fuel market already faces.

As crude oil reserves dwindle, the marketplace will either transition to the electrifying of the transportation system (electric and fuel-cell vehicles and electric railways), with the electricity being produced by coal, natural gas, nuclear and renewables, or see the development of an industry to produce liquid fuel substitutes from coal, oil shale, and tar sands. It might also turn out to be a combination of both. The transition will vary by nation and will be dictated strongly by the fuels available, the economic and technological efficiencies of competitive systems, the relative environmental impacts of each technology, and the role government takes in the marketplace.

John Zumerchik
Herman Bieber

See also: Hydrogen; Natural Gas, Processing and Conversion of.

REFERENCES

American Petroleum Institute. (1987). *Liquid Fuels from Natural Gas*. API 34–5250.

Burgt, M. J. v. d., et al. (1988). *Methane Conversion*. New York: Elsevier Science, Inc.

Crow, M. (1988). *Synthetic Fuel Technology Development in the United States*. New York: Praeger Publishing.

Haar, L., et al. (1981). *Advances in Coal Utilization*. Chicago: Institute of Gas Technology.

Hinman, G. W. (1991). “Unconventional Petroleum Resources.” In *The Energy Sourcebook*, ed. R. Howes and A. Fainberg. New York: American Institute of Physics.

Lee, S. (1996). *Applied Technology Series: Alternative Fuels*. New York: Taylor & Francis.

Storch, H. H., et al. (1951). *The Fischer Tropsch and Related Synthesis*. New York: John Wiley & Sons.

Supp, E. (1990). *How to Produce Methanol from Coal*. New York: Springer Verlag.

T

TAXATION OF ENERGY

In the United States, the federal government does not impose an energy tax or a general sales tax that is broadly applicable to energy. However, excise taxes are imposed on certain fuels, and there are a number of income tax provisions specific to the energy sector. There are three separate categories of taxes and fees that affect energy use: (1) excise taxes/fees that primarily affect energy demand; (2) income tax provisions that primarily affect energy supply by operating on the after-tax rate of return on investment; and (3) income tax provisions that primarily affect the demand for specific energy sources.

EXCISE TAXES

Excise taxes placed on specific energy sources tend to reduce the demand for these energy sources in both the short and the long run. The federal government imposes excise taxes on almost all petroleum products and coal (see Table 1). The federal government also imposes excise taxes on many transportation uses of methanol, ethanol, natural gas, and propane and imposes a fee on electricity produced from nuclear power plants.

By far the most substantial energy excise taxes imposed by the federal government are those imposed on motor fuels. Gasoline generally is taxed at a rate of 18.4 cents per gallon, and diesel fuel is taxed at a rate of 24.4 cents per gallon. Commercial aviation fuels are taxed at 4.4 cents per gallon, noncommercial aviation gasoline is taxed at 19.4 cents per gallon, and noncommercial aviation jet fuel is taxed at 21.9 cents per gallon. Diesel fuel used by commercial cargo vessels on inland or intracoastal waterways is taxed at 24.4 cents per gallon. Certain users of motor fuels are exempt from the excise tax or pay a reduced rate. For example, state and local governments and motor fuels for off-road use (such as by farmers) are exempt from gasoline and diesel taxes. Reduced rates of tax are paid on gasoline blended with ethanol (gasohol) and by providers of intercity bus service.

Receipts from these taxes are allocated to various trust funds, the largest recipients being the Highway Trust Fund (gasoline and diesel fuel taxes) and the Airport and Airway Trust Fund (aviation fuels taxes). Congress appropriates monies from these funds for surface and air transportation capital projects and operating subsidies. Each of the above tax rates includes 0.1 cent per gallon dedicated to the Leaking Underground Storage Tank Trust Fund. Monies from this fund are appropriated by Congress for remediation of environmental damage at underground fuel storage facilities.

Coal from underground mines is assessed an excise tax of $1.10 per ton, and coal from surface mines is assessed a tax of $0.55 per ton, but in each case, a constraint is added that states that the tax cannot exceed 4.4 percent of the sales price. The receipts from these taxes are allocated to the Black Lung Disability Trust Fund. Monies from this trust fund are used to pay for health benefits to coal miners.

Producers of electricity from nuclear power plants are assessed a fee of 0.1 cent per kilowatt-hour to pay for future storage of spent nuclear fuel at a federal facility. Receipts from this fee are allocated to the Nuclear Waste Trust Fund and are appropriated by Congress to cover the costs of developing and constructing a permanent storage facility.

Each of these excise taxes raises the price of the fuel for uses subject to the tax and might be expected to reduce the demand for petroleum and coal. However, to some extent the excise tax receipts col-

Item	$ Billions
Highway Trust Fund Fuel Excise Taxes	
Gasoline	21.7
Diesel fuel	7.7
Other highway fuels	1.4
Total	30.8
Noncommercial aviation fuels excise tax	0.2
Commercial aviation fuels excise tax	0.6
Inland waterways trust fund excise tax	0.1
Leaking underground storage tank trust fund excise tax	0.2
Motororboat gasoline and special fuels excise taxes	0.2
Black lung trust fund coal excise tax	0.6

Table 1
Estimated 1999 Federal Energy-Related Excise Tax Receipts
SOURCE: Joint Committee on Taxation, U.S. Congress, 1998.

lected are dedicated to trust funds to finance expenditures that directly benefit those paying the tax. To the extent that these taxes are benefit taxes, they may have no independent effect on the demand for the taxed fuel because those paying the taxes are receiving commensurate services. For example, the receipts from aviation fuels taxes generally finance airport expansion and safety projects, providing a wide set of benefits to the flying public. But the tax and the benefit of the tax are not always strongly coincident. For example, in the case of motor fuels taxes, a portion of the Highway Trust Fund is used to support mass transit projects, which are not likely to directly benefit highway users except to the extent that the projects relieve traffic congestion. In the case of electricity, the effect of the fee on overall demand is not clear, because the electricity generation industry is transitioning from a regulated to a more competitive posture and because about 20 percent of electricity is generated from nuclear power plants. Given that the proceeds of this fee are to be spent on disposal of nuclear waste, a cost of production, the fee may make the cost of electricity production more fully reflect social costs. In the case of the Black Lung Disability Trust Fund, the supported benefits are for workers, not users, and hence the tax on coal is expected to operate solely to reduce demand. This tax also is related to the cost of production such that it makes the cost of coal reflect better the social costs.

In a different context, purchasers of automobiles that do not meet specific fuel economy standards must pay a “gas guzzler” excise tax. This tax presumably decreases the demand for such vehicles and should indirectly affect demand for gasoline.

INCOME TAXATION AND ENERGY SUPPLY

Various provisions in the federal income tax treat energy producers more or less favorably than other businesses. By changing the after-tax rate of return on investments in the energy sector, the Tax Code may alter the long-run supply of specific types of energy.

In general, the income of all participants in the energy sector is subject to income tax of one form or another. Two notable exceptions arise in the generation and sale of electricity. Governmental agencies (such as the Tennessee Valley Authority, the Bonneville Power Administration, and municipally owned power companies) account for approximately 14 percent of the electricity sold in the United States. Such public power producers and suppliers are not subject to federal income tax. Electricity providers organized as cooperatives are common in rural areas, accounting for about 8 percent of electricity sales. Income from sale of electricity is not subject to income tax if the income is paid to co-op members as a patronage refund.

To compute taxable income, a taxpayer is permitted to subtract from revenues those expenses that are necessary to create current sales. Expenses related to future sales are required to be capitalized and recovered over time. Distinctions between these two categories—current and suture sales—are sometimes hard to make. In some cases the Tax Code permits energy producers to claim currently, or on an accelerated basis, expenditures that are related to future production and sales. This lowers income for current tax purposes, thereby increasing the after-tax return to investments in the energy sector compared to other investments.

The Tax Code, for example, provides special rules for the treatment of intangible drilling costs, or IDCs. These include expenditures made for wages, fuel, materials, and the like necessary for drilling wells and preparing wells for the production of oil and natural gas. IDCs also may include expenditures for the construction of derricks and tanks, for gathering pipelines, and for roads. Certain taxpayers (generally those without major refining operations) may elect to expense, rather than capitalize, IDCs incurred for domestic properties. Such acceleration of cost recovery may provide preferential treatment for investment in oil and gas production.

The Tax Code also provides special tax advantages through "percentage depletion" allowances. To determine taxable income, the Tax Code permits the owner of a mineral reserve a deduction in recognition of the fact that the mineral reserve is depleted as the mineral is extracted. Some owners of oil and natural gas properties are permitted to compute this deduction as a percentage of gross income from the property (percentage depletion is not available to owners of producing properties who also have substantial refining capacity). Generally, eligible taxpayers may deduct annually 15 percent of gross income, up to 100 percent of the net income from the property (reaching this limit means that income from the property is effectively tax-exempt). Because percentage depletion is computed without regard to the taxpayer's basis in the depletable property, the cumulative depletion deductions may exceed the amount expended by the taxpayer to develop or acquire the property. In such circumstances, an investment in elgible oil and gas properties is tax-favored compared to other investments.

There also are circumstances where income tax provisions reduce the after-tax rate of return on such investments below those available in other sectors. For example, certain types of electricity-generating property (such as those used in distributed power applications) may have useful lives that are substantially shorter than the tax lives permitted for cost recovery. In such instances, the after-tax rate of return will be negatively affected by these specific provisions.

The Tax Code allows credits against income tax liability for various types of specific energy production, including electricity produced from wind or "closed loop" biomass (essentially agricultural operations devoted to electricity generation) and oil and natural gas from "nonconventional" sources (such as natural gas from coal seams, and oil from shale). Credits also are allowed for investments in specific forms of energy production, including facilities to produce geothermal or solar power, and expenditures associated with enhanced oil recovery technologies. Credits against income tax liability directly reduce taxes paid and therefore reduce the cost of production and increase the return on investment. Accordingly, these credits would be expected to increase production of energy from the specified sources.

Item	*$ Billions*
Expensing of exploration and development costs	2.5
Excess of percentage depletion of fuels over cost depletion	5.0
Tax credit for production of non-conventional fuels	7.1
Tax credit for enhanced oil recovery costs	0.3
Tax credit for electricity production from wind and closed-loop biomass	0.4
Tax credit for investments in solar and geothermal energy	0.3
Exclusion of energy conservation subsidies	0.2

Table 2
Estimates of Selected Energy-Related Federal Tax Expenditures, FY 1999–2003
SOURCE: Joint Committee on Taxation, U.S. Congress, 1999.

In addition to specific deductions and credits, the Tax Code permits state and local governments to issue bonds on which the interest is exempt from federal income tax. This provision means that states and local governments can borrow at interest rates below those paid by private corporations. Municipally owned electricity providers often can issue tax-exempt debt; the lower interest rate may have the effect of increasing the provision of electricity by these entities.

To provide a sense of the value of some of these deviations from neutral taxation, Table 2 reports cumulative estimates over five years of selected energy-related "tax expenditures." The Congressional Budget and Impoundment Control Act defines a "tax expenditure" as a revenue loss attributable to a special inclusion, exemption, or deduction from gross income or a special credit, preferential tax rate, or deferral of tax.

INCOME TAXATION AND ENERGY DEMAND

The Tax Code contains various provisions that may affect the demand for energy from specific sources. In the transportation area, purchasers of electric vehicles may claim a credit against their income tax liability for a portion of the purchase price. Taxpayers who purchase or retrofit a vehicle that runs on alternative fuels (such as natural gas, propane, methanol, or ethanol) may claim an accelerated deduction for their expenditure. Taxpayers who commute to work by automobile and whose employer provides parking may have a portion of the value of their parking space included in taxable income (if the value of the parking space exceeds a specified amount). Alternatively, employers may provide their employees with mass transit passes and not include the full value of the benefit in the employee's taxable income. Each of these provisions is expected to reduce the demand for gasoline to some extent.

Taxpayers who receive incentive payments for installing energy-efficient equipment, such as a high-efficiency furnace, do not have to include the payment in their taxable income. This tax-favored treatment may make consumers more likely to purchase energy-saving equipment, thereby reducing demand for energy.

As mentioned above, state and local governments can borrow at relatively low interest rates by issuing tax-exempt debt, and this finance technique is used to a great extent in highway and road construction. By reducing the cost of road construction, tax-exempt debt may increase the amount of construction undertaken, thereby increasing the demand for fuels.

SUMMARY

Tax policy (including excise taxes and income tax provisions) operates in several different ways to affect energy supply and demand, for energy in general and for specific fuels. There is no consensus about the net overall effect on supply and demand, in part because these provisions operate in different sectors, in different markets, and sometimes in different directions.

Thomas A Barthold
Mark J. Mazur

See also: Capital Investment Decisions; Energy Economics; Environmental Economics; Government Agencies; Government Intervention in Energy Markets; Industry and Business, Operating Decisions and; Oil and Gas, Drilling for; Regulation and Rates for Elec-tricity; Subsidies and Energy Costs.

BIBLIOGRAPHY

Committee on the Budget, U.S. Senate. (1998). *Tax Expenditures: Compendium of Background Material on Individual Provisions*, S. Prt. 105-70. Washington, DC: U.S. Government Printing Office.

Joint Committee on Taxation, U.S. Congress. (1998). *Estimates of Federal Tax Expenditures for Fiscal Years 1999–2003*. Washington, DC: U.S. Government Printing Office.

Joint Committee on Taxation, U.S. Congress. (1999). *Schedule of Present Federal Excise Taxes (as of January 1, 1999)*. Washington, DC: U.S. Government Printing Office.

U.S. Office of Management and Budget. (1999). *Budget of the United States Government, Fiscal Year 2000*, Washington, DC: U.S. Government Printing Office.

TEMPERATURE REGULATION

See: Energy Management Control Systems

TENNESSEE VALLEY AUTHORITY

See: Government Agencies

TESLA, NIKOLA (1856–1943)

Nikola Tesla was born in 1856 in Smilja, Croatia, to parents of Serbian heritage. The region of his birth was at the time part of the Austro-Hungarian Empire. His scientific and engineering aptitude were obvious at an early age, evidenced by stories of inventions consisting of simple mechanical devices he worked on during his childhood. For this reason his parents were persuaded to allow him to pursue an education in engineering rather than becoming a Serbian Orthodox priest as they had hoped. While Tesla studied at Graz Polytechnic and at the University of Prague, he developed an interest in electrical power transmission.

Moving into the commercial world, he accepted a position with a new telephone company in Budapest, where he developed a telephone repeater system, which formed the basis for the modern loudspeaker. Tesla was promoted within the company to a position in Paris. His potential was now clear to the manager of the plant, who recommended that he move to the United States to work with Thomas Edison, then considered the premier engineering genius of the age.

Tesla arrived in New York in 1884 and was hired by Edison. Edison understood Tesla's ability but remained unconvinced by his new employee's insistence on the use of alternating current for electrical power transmission. Nevertheless, Tesla accepted his assignment to work on improving Edison's direct-current method, which was then in use. Working long hours Tesla increased the system output and asked Edison for a $50,000 bonus, which Tesla understood he was to receive. Edison refused, claiming that the bonus had only been a joke. Tesla quit Edison's employ, and thereafter relations were strained between the two men.

Tesla was awarded patents on the alternating-current motor. In the single-phase AC motor, two circuits passing current are set up diagonally opposite with respect to a circular armature. With the currents ninety degrees out of phase, the armature rotates. In the polyphase AC motor, three or more circuits, each having a different phase separation, are employed. George Westinghouse, Jr., is reported to have bought all of Tesla's related patents for $1 million plus royalties, although the exact amount remains in dispute. The two men collaborated very close by. At the World's Colombian Exposition in Chicago in 1893, Tesla's system of alternating-current power transmission was successfully employed. The Westinghouse Company later won the contract to harness the hydroelectric power of Niagara Falls, producing a viable, long-distance, electrical distribution system. When financial difficulties beset the Westinghouse Company between 1891 and 1893, Tesla agreed to give up the royalties to which he was entitled.

Tesla now devoted his energies to work in his own laboratory in New York City. His inventions were numerous. He experimented with the transmission of signals using electromagnetic waves at the same time as Marconi was developing radio. At Madison Square Garden Tesla demonstrated remote control of mechanical objects. His Tesla Coil was capable of generating extremely high voltages. It consisted of three circuits. AC power was applied to the first circuit and, by means of transformers, the voltage was stepped up in the second and third circuits, both of which possessed spark gaps. The third circuit also included a capacitor tunable for resonance, thus allowing extremely large voltages to be developed.

In 1891 Tesla became a U.S. citizen. His naturalization papers remained among his most prized possessions.

Tesla accepted an offer of land and free electricity in Colorado Springs by the local electrical company to continue his research. In 1899, he conducted experiments he considered to be of extreme importance in the conduction of electricity through the earth without the use of wires. He reported that he was able, by means of this principle, to illuminate electric lightbulbs twenty-six miles from the power source. This transmission mechanism, which Tesla explained as taking place by means of the resonant frequency of the earth, has yet to be adequately verified to the satisfaction of the scientific community. In his Colorado Springs laboratory Tesla also performed experiments to simulate lightning that were successful enough to produce a power outage and along with it a withdrawal of the local company's offer of free electricity.

Tesla left Colorado Springs in 1900 and continued his research in a new laboratory on Long Island, which was opened in 1901. However, his efforts were beset by difficulties, not least of which were financial problems, which caused closure of this facility.

His flow of innovative ideas continued unabated. These included improvements for turbines, methods for communication with life on other worlds; and an idea characterized by the press as a "death ray," which may be interpreted as a precursor to the modern laser.

In 1915 it was falsely reported that Edison and Tesla had been jointly awarded the Nobel Prize in physics and that Tesla refused the honor because of his differences with Edison. The circumstances surrounding this news remain cloudy. Nevertheless, Tesla deeply felt the hurt of not receiving this recognition. In 1917 he was persuaded to accept the Edison Medal from the American Institute of Electrical Engineers as an acknowledgment of his pioneering contributions. Among the many honors bestowed on Tesla, perhaps the most important was having an electrical unit named after him, the tesla being the unit for magnetic flux density.

Toward the end of Tesla's life, unflattering articles were written about him, and there were innuendoes that he was involved in the occult. Money problems were never far away, and he moved his residence from one hotel to another, each one cheaper than the one before. His circle of friends contracted. Feeding the pigeons that lived close to his hotel became very important to him, and he developed an almost spiritual bond with them.

Tesla died in 1943. His funeral service was held in the Cathedral of St. John the Divine. As during his lifetime, controversy was not far away. The Serb and Croat mourners sat on opposite sides of the cathedral.

Tesla remains a fascinating man because of his personal life and his engineering genius. However, from a technical point of view Tesla is most remembered for his contributions to the use of alternating-current power transmission.

James D. Allan

See also: Edison, Thomas Alva; Electricity; Electricity, History of; Electric Power, Generation of; Electric Power Transmission and Distribution Systems; Lighting.

BIBLIOGRAPHY

Corson, D. R., and Lorain P. (1962). *Introduction to Electromagnetic Fields and Waves*. San Francisco: W. H. Freeman.

Hunt, I., and Draper, W. W. (1964). *Lightning in His Hand: The Life Story of Nikola Tesla*. Denver: Sage Books.

Peat, D. (1983). *In Search of Nikola Tesla*. Bath, Eng: Ashgrove Press.

Seifer, M. J. (1996). *The Life and Times of Nikola Tesla*. Secaucus, NJ: Carol Publishing Group.

Walters, H. B. (1961). *Nikola Tesla: Giant of Electricity*. New York: Thomas Y. Crowell.

THERMAL ENERGY

Thermal energy is the sum of all the random kinetic energies of the molecules in a substance, that is, the energy in their motions. The higher the temperature, the greater the thermal energy. On the Kelvin temperature scale, thermal energy is directly proportional to temperature.

All matter is composed of molecules or, in some cases, just atoms. In gases and liquids, molecules are relatively free to move around. In a solid they are not so free to move around, but they can vibrate. Although molecules are microscopic, they do have some mass. Combining mass and speed gives them kinetic energy. Depending on the substance, the particles may interact with each other or their surroundings, in which case they would have potential energy along with kinetic energy. Summing the kinetic energy and potential energy of all the molecules gives the total energy of the substance. This energy is called internal energy because it is internal to the confines of the substance.

In a substance where the molecules may move around, the motion is random. No molecule has a definite speed or kinetic energy, but a molecule has a definite average kinetic energy that depends on the temperature. On the Kelvin temperature scale, the average kinetic energy is directly proportional to the Kelvin temperature. The sum of all the random kinetic energies of the molecules is called thermal energy, therefore, thermal energy is directly proportional to the Kelvin temperature.

Citizen Watch unveils the Eco-Drive Thermo, a wristwatch powered by the heat of the human wearing it. (Corbis Corporation)

When one puts a warm hand in contact with cold water, the hand cools and the thermal energy in the hand decreases. The water warms and the thermal energy of the water increases. The exchange of energy stops when both the hand and water come to the same temperature. While in transit, the energy is called *heat*. When two objects at the same temperature are in contact, no heat flows between either. Accordingly, there is no change in the thermal energy of either, no matter how much thermal energy in either one.

Gasoline-powered engines, diesel engines, steam turbines, and gas turbines are examples of heat engines. All heat engines work on a cyclic principle of extracting thermal energy from some source, converting some of this energy to useful work, and rejecting the remaining energy to something at a lower temperature. In an automobile engine, the ignition of a gasoline vapor-air mixture produces a gas at a temperature several hundred degrees above room temperature. The pressure of the gas forces a piston downward, doing work. The gas cools and is ejected out the exhaust at a temperature significantly lower than at the time of ignition. A heat engine converting thermal energy to work cannot function unless there is a temperature difference between the source and exhaust. The larger the temperature difference, the greater the efficiency of the engine. Usually, the lower temperature is that of the engine's surroundings and the ignition temperature is significantly higher.

If it were it is practical to have a temperature lower than what exists naturally in our environment, a heat engine could be built in which this temperature was the exhaust temperature and the temperature of the environment was the higher temperature. Heat engines extracting thermal energy from the surface water of an ocean, and rejecting thermal energy to the cooler sub-surface water, have been proposed. They would not be very efficient because the temperature difference would be small and they would not be easy to construct. The attraction is related to the huge amount of thermal energy in the oceans, which cover roughly two-thirds of the earth's surface.

Joseph Priest

BIBLIOGRAPHY

Hobson, A. (1995). *Physics: Concepts and Connections.* Englewood Cliffs, NJ: Prentice-Hall.

Serway, R. A. (1998). *Principles of Physics.* Fort Worth, TX: Saunders College Publishing.

THERMODYNAMICS

Thermodynamics, which comes from the Greek word meaning "heat motion" is the science that studies the transfer of energy among different bodies. When a typical power plant produces electricity, water is heated to produce steam by burning fuel. The steam is allowed to expand and perform work, which turns an electric generator that produces electrical energy. The efficiency of this process of energy production—how much electrical energy can be produced for burning a given amount of fuel—is determined by the laws of thermodynamics. These same laws are equally important for the heat engines that propel millions of planes, trains and automobiles around the world. In fact, nearly all energy that powers our energy intensive economy involves, in some manner, the laws of thermodynamics.

THERMODYNAMICS

Scientific descriptions of nature can be categorized as microscopic or macroscopic. A microscopic description involves a discussion of the system using the properties of, and laws governing, atoms and molecules, while in a macroscopic description, one focuses on the large scale, properties of the system. Thermodynamics is a special macroscopic theory that does not involve use of underlying microscopic laws and is, therefore, valid regardless of the form of these laws. A thermodynamic system (or just system) is a specified macroscopic system, that is, a certain specified collection of particles. The matter not included in the system is called the surroundings, and the system plus the surroundings is called the universe. In this context, the universe is not the astronomical universe of galaxies, but the system plus the rest of the matter that can interact with the system, in a measurable manner. As an example of a thermodynamic system consider a gas inside a container (can). The gas could be the system and the walls of the can plus the immediate surroundings would be the universe. Or the gas plus the walls of the can could be the system. The definition of the system is arbitrary and is up to us to define. Failure to have in mind a clear definition of the system often leads to errors in reasoning using thermodynamics principles.

Thermodynamic variables are macroscopic variables associated with the system, such as the system's energy E, volume V, temperature T, pressure P, among others. The thermodynamic variables for a system are measurable properties of the system, and their values define the state of the system. A system is in a thermodynamic equilibrium state if the values of its thermodynamic variables do not change in time. For the gas in the can system, in equilibrium, no change in the values of the temperature, pressure, volume, or any other thermodynamic variable of the gas would be observed. If the system is in an equilibrium state, then one says the system is in a particular thermodynamic state. Thermodynamic processes occur when the system changes from an initial equilibrium state to another final equilibrium state; the thermodynamic variables may undergo changes during this process. If the initial equilibrium state is labeled by i and the final equilibrium state by f then the process i→f corresponds to the energy, temperature and pressure changing from E_i, T_i, P_i to the values E_f, T_f, P_f. An important concept in thermodynamics is the idea of a reversible process. In a reversible process, the system undergoes changes where the system is at all times very close to an equilibrium state; the changes in the thermodynamic variables take place very slowly and can also be reversed by very small changes in the agents causing the process. A reversible process can be realized by changing the external conditions so slowly that the system can adjust gradually to the changes. In our can of gas a very slow compression of the system would be a reversible process. A process that is not reversible is called irreversible; a very rapid compression of the gas that would set up pressure waves and gas flows would be an irreversible process. All real processes that occur in nature are irreversible to a certain extent, but it is possible to carry out processes that are reversible for all practical purposes, and to calculate what would happen in a truly reversible process. Various types of thermodynamic processes are named depending on conditions placed on the system. For example, an isothermal process, such as the compression of a gas held at constant temperature, is a process that takes place at constant temperature while an isobaric process, such as the expansion of a gas held at constant pressure, is a process that takes place at constant pressure.

A process in a system may be associated with work W being done on the system. Work is defined the same as in mechanics—as a force times the distance through which the force acts. For example, in our can of gas it may be stirred, thereby doing work on the gas and changing, say, its temperature. In a compression or expansion of the gas, work is also done on the system. If energy is exchanged between the system and the surroundings due solely to a temperature difference, then this energy is called heat. A thermally insulated system is a system that maintains its thermodynamic state regardless of the temperature of its surroundings. A system enclosed in thermally insulated walls like the walls of a very good thermos bottle cannot sense the temperature of the surroundings. Processes that occur in thermally insulated systems are called adiabatic processes. For an insulated system, the only way to change its energy is by the performance of work; in this case the work is called adiabatic work W_a.

The first law of thermodynamics has two parts. The first part associates the change in the energy of the system in adiabatic processes to the adiabatic work done on the system:

$$DE = E_f - E_i = W_a \quad \text{(adiabatic processes).} \tag{1a}$$

In Equation 1a the symbol DE is used for the change in energy for the process i→f; the final value of the system E_f minus the initial energy of the system E_i. Thus, if the system undergoes an adiabatic compression, the work done on the system would give the change in the energy of the system using Equation 1a. Starting from a given state i, one can perform adiabatic work on the system and define the energy at other states f using Equation 1a written in the form $E_f = E_i + W_a$. In words, Equation 1a can be described by saying that energy is conserved for adiabatic processes; if work is done on the system then this work shows up in the energy of the system. Energy is neither created nor destroyed, but is transferred from the surroundings to the system, or vice versa through the work as agent.

The second part of the first law of thermodynamics arises when the requirement that the process be adiabatic is dropped; recall that this means the system is not insulated, and processes can be caused by heating and cooling. In a general process (the only assumption is that matter is not added or removed from the system), if an amount of work W is done on the system and the energy changes by DE then the heat supplied to the system Q is defined by

$$DE = E_f - E_i = W + Q \quad \text{(general processes).} \qquad (1b)$$

Notice that work is defined from mechanics (force times distance) and can be used to define the energy difference between any two states in Equation 1a. This energy difference and the work are then used to define the heat Q in Equation 1b. In words, Equation 1b says that energy is conserved and heat must be included as a form of energy; or the final energy of the system is the initial energy plus the work done on the system plus the energy added to the system by heating.

Note the following sign conventions for W and Q: the work W in Equation 1a and 1b is positive if it increases the energy of the system, and the heat Q is positive if it increases the energy of the system. If W is positive, then the surroundings do work on the system, for example, compressing the gas, but if W is negative, the system does work on the surroundings, expanding the gas. If Q is positive, the surroundings give energy to the system by means of a temperature difference, but if Q is negative, the energy passes from the system to the surroundings.

In summary, the first law of thermodynamics, Equations 1a and 1b, states that energy is conserved and the energy associated with heat must be included as a form of energy. No process i→f is possible if it violates the first law of thermodynamics; energy is always conserved in our world as dictated by Equation 1b. If Equation 1b is applied to an adiabatic process, then because Q = 0 the first part, Equation 1a is recovered, but one still needs both parts of the first law to define the quantities.

The second law of thermodynamics further restricts the types of processes that are possible in nature. The second law is particularly important in discussions of energy since it contains the theoretical limiting value for the efficiency of devices used to produce work from heat for our use.

The second law of thermodynamics also consists of two parts. The first part is used to *define* a new thermodynamic variable called entropy, denoted by S. Entropy is the measure of a system's energy that is unavailable for work.The first part of the second law says that if a reversible process i→f takes place in a system, then the entropy change of the system can be found by adding up the heat added to the system divided by the absolute temperature of the system when each small amount of heat is added:

$$DS = S_f - S_i = SUM(dQ/T) \quad \text{(reversible process).} \qquad (2a)$$

dQ is a small amount of energy added *reversibly* to the system as heat when the temperature of the system is at an absolute temperature of T. SUM means to add the individual dQ/T terms over the entire reversible process from i→f. The reversible way of adding heat to a system is to put the system in contact, through a heat conducting wall (diathermal wall), with the surroundings that are at a very slightly different temperature. The absolute temperature scale is defined so that the freezing point of water is T = 273 Kelvin or 273 K. The zero of temperature on this scale is an unattainable temperature referred to as absolute zero, although temperatures within one-millionth of a Kelvin and lower of absolute zero have been produced (Note at absolute zero all motion in the system does not cease). Notice that Equations 1a and 2a have a certain similarity since they both are used to define a relevant thermodynamic variable: energy E for Equation 1a and entropy S for Equation 2a. Equation 2a allows the definition of the entropy S

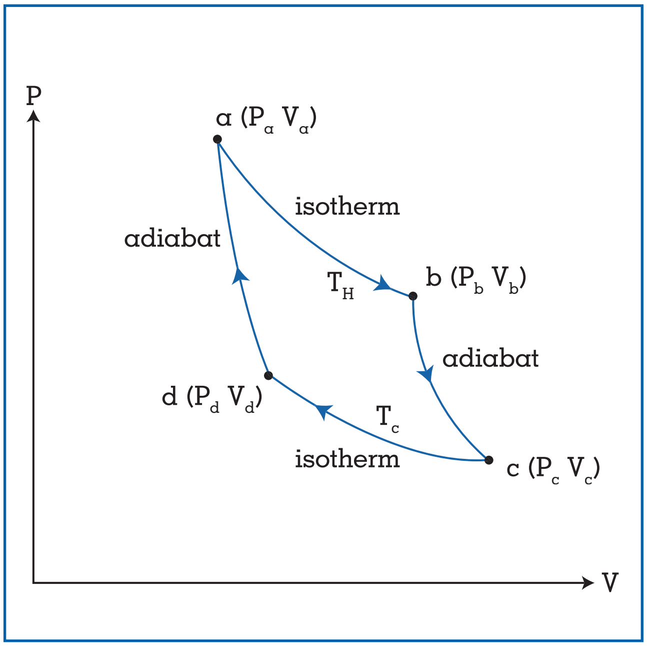


Figure 1.
Carnot cycle for a gas.

at every equilibrium state of the system by starting with the entropy at a particular state and carrying out reversible processes to other states.

The second part of the second law states that where system undergoes an adiabatic process (system surrounded by insulating walls), i→f and the process is reversible, the entropy is not changed, while when the adiabatic process is not reversible the entropy must increase:

$$DS = S_f – S_i = 0$$
(process i→f is adiabatic and reversible),

$$DS = S_f – S_i > 0$$
(process i→f is adiabatic but not reversible). (2b)

Thus, in adiabatic processes the entropy of a system must always increase or remain constant. In words, the second law of thermodynamics states that the entropy of a system that undergoes an adiabatic process can never decrease. Notice that for the system plus the surroundings, that is, the universe, all processes are adiabatic since there are no surroundings, hence in the universe the entropy can never decrease. Thus, the first law deals with the conservation of energy in any type of process, while the second law states that in adiabatic processes the entropy of the system can never decrease.

A heat engine is a device to use the flow of heat energy from a higher temperature to a lower temperature to perform useful work, often mechanical or electrical work. A steam engine heats steam and uses this heated steam and a cooler surroundings to perform mechanical work such as lifting a weight or propelling a steamship. A certain amount of heat Q is converted into a certain amount of work W per cycle of the engine.

In order to investigate heat engines, first focus on a particular cyclic process called a Carnot cycle. Figure 1 shows a Carnot cycle for our gas in the can, and shows the changes in the pressure P and the volume V for the gas during the cycle. The Carnot cycle is the process a→b→c→d→a: from a→b the system at temperature T_H undergoes an isothermal expansion from volume V_a to V_b, from b→c the system undergoes an adiabatic expansion from volume V_b to V_c while the temperature falls to the lower value T_C. Next, the system undergoes an isothermal compression c→d. Finally, the system completes the cycle by undergoing an adiabatic compression d→a, after which its state, a, is precisely the same as before the process was started.

All of the processes in the Carnot cycle are reversible, which is indicated by drawing the lines passing through equilibrium states on the diagram in Figure 1. During the isothermal expansion a→b the system will do work and absorb the energy Q_H from the surroundings that are at the temperature T_H. During the process c→d the system will have work performed on it and liberate energy Q_C to surroundings at lower temperature T_C. The reversible heat exchanges are arranged by having the system in contact with the surroundings, which are held at constant temperature during the isothermal expansion and compression. The surroundings function as a (thermal) reservoir at the temperature T_H and T_C. A reservoir is able to absorb or liberate as much energy as required without a change in its temperature. An oven held at temperature T_H functions as a reservoir. During the adiabatic expansion, the system exchanges no heat, but work is done. Note that in the Carnot cycle only two reservoirs exchange energy during the isothermal process. In other reveisible cycle one needs a series of reservoirs to exchange energy. The laws of thermodynamics may be applied to the Carnot cycle and lead to important results.

Applying the first law of thermodynamics to the Carnot cycle gives

$$DE = E_f – E_i = 0 = Q_H – Q_C – W, \qquad (3)$$

or, solving this for the work done during the cycle in terms of the heat absorbed per cycle and the heat liberated per cycle:

$$W = Q_H – Q_C. \qquad (4)$$

The system does work on the surroundings during the two expansions, and work is done on the system in the two compressions in the Carnot cycle. The net work done in the cycle is work done on the surroundings by the system and is W in Equations 3 and 4. Note that $E_f = E_i$ in the complete cycle a→a since i = f = a, and the total process is cyclic. It is easy to imagine carrying out the Carnot cycle on our can of gas example. In Equations 3 and 4, the proper signs have been used so that the quantities Q_H, Q_C, and W are all positive numbers; the system absorbs energy Q_H from the high temperature reservoir, rejects or liberates energy Q_C at the low temperature reservoir, the system performs work W on the surroundings, and the system is finally returned to the starting equilibrium state. Equation 4 is the application of the first law of thermodynamics to the Carnot engine.

Because the gas in the Carnot cycle starts and ends at the same state, the system's entropy does not change during a cycle. Now apply the second law to the universe for the case of the Carnot cycle. Because the processes are reversible, the entropy of the universe does not change by Equation 2b. This can be written:

$$DS = 0 = Q_C/T_C – Q_H/T_H \quad \text{(Carnot cycle)}, \qquad (5)$$

where the first part of the second law, Equation 2a, is used to determine the entropy change of the two reservoirs in the Carnot cycle. Solving Equation 5 for the ratio of the heat absorbed to the heat liberated, one finds the important Carnot relation connecting the energy Q_H/Q_C ratio to absolute temperature ratio:

$$Q_H/Q_C = T_H/T_C \quad \text{(Carnot cycle)}. \qquad (6)$$

This relation is used to define the absolute temperature scale in terms of energy exchanged with reservoirs in a Carnot cycle. The relations for the Carnot cycle, given by Equations 4 and 6 are general and valid for any system that is taken through a Carnot cycle and not just for our can of gas example.

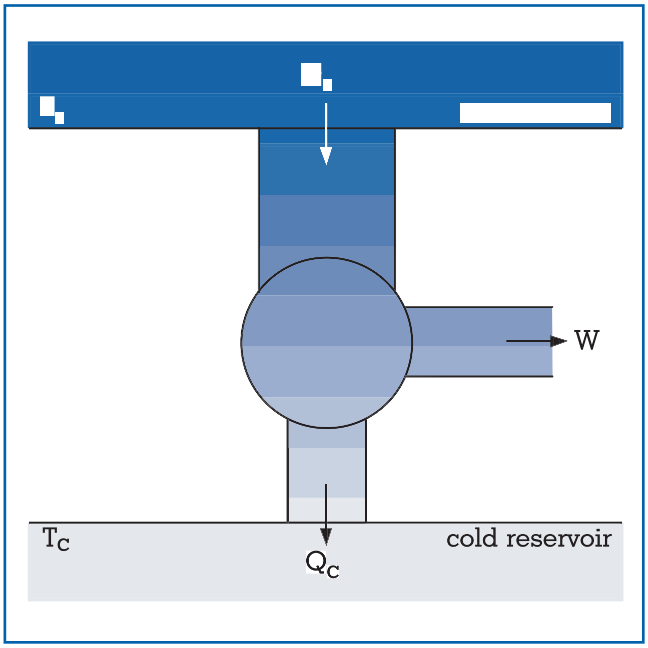


Figure 2.
Schematic of a heat engine.

Figure 1 shows the thermodynamic processes associated with the Carnot cycle. Since these thermodynamic processes produce work, this is also called a Carnot engine. Figure 2 shows another more general picture of a heat engine in thermodynamics. This engine could be a Carnot engine or some other engine. For practical purposes, an efficiency, e, of an engine is defined as the work obtained in one cycle divided by the heat absorbed; this is the what can be sold, W, divided by what must be paid for to produce W, namely Q_H:

$$e = W/Q_H \quad \text{(efficiency of a heat engine)}. \qquad (7)$$

The efficiency of the engine would be greatest if it could be arranged that W = QH, QC = 0, giving an efficiency of 1 or 100 percent efficient. In this case, all the heat absorbed at the high temperature is converted into work. This is perfectly valid from the point of view of the first law of thermodynamics, because energy is conserved. However, from the second law, Equation 5, DS = 0 = $-Q_H/T_H$, which is impossible since $Q_H = W$ is not zero and this implies the entropy of the universe would decrease.

Thus, the second law leads to the conclusion that energy cannot be absorbed from one reservoir and converted completely into work with no other changes in the universe. This is an alternate form of the second law of thermodynamics called the Kelvin-Plank form of the second law. This result is of great practical importance for discussions of energy, since without this result the conversion of heat completely into work would be possible. A steamship could sail on the ocean, absorbing energy from the ocean, and power itself with no need to carry fuel. The fact that heat must be rejected at a lower temperature means that the steamship must carry fuel to produce a higher temperature than the surroundings. The second law of thermodynamics forbids this so-called perpetual motion machine of the second kind, which is an imaginary device that converts heat completely into work with no other change in the universe.

An estimate of the efficiency of a heat engine working between two temperatures T_H and T_C can be obtained by assuming the Carnot cycle is used. By combining the results from applying the first and second laws to the Carnot cycle, the Carnot efficiency e_c may be written:

$$e_c = 1-Q_C/Q_H = 1-T_C/T_H \quad \text{(Carnot engine efficiency).} \tag{8}$$

This remarkable result shows that the efficiency of a Carnot engine is simply related to the ratio of the two absolute temperatures used in the cycle. In normal applications in a power plant, the cold temperature is around room temperature T = 300 K while the hot temperature in a power plant is around T = 600 K, and thus has an efficiency of 0.5, or 50 percent. This is approximately the maximum efficiency of a typical power plant. The heated steam in a power plant is used to drive a turbine and some such arrangement is used in most heat engines. A Carnot engine operating between 600 K and 300 K must be inefficient, only approximately 50 percent of the heat being converted to work, or the second law of thermodynamics would be violated. The actual efficiency of heat engines must be lower than the Carnot efficiency because they use different thermodynamic cycles and the processes are not reversible.

The first and second laws of thermodynamics can also be used to determine whether other processes are possible. Consider the movement of energy from a cold temperature reservoir to a hot temperature reservoir with no other changes in the universe. Such spontaneous cooling processes would violate the second law. From the first law, for such a process $Q_H = Q_C$, the heat lost by the low-temperature reservoir is gained by the high-temperature reservoir; again, the signs are chosen so that both Q_H and Q_C are positive. The change of entropy of the universe in this case is just the change in entropy of the two reservoirs, which, by Equation 2a is DS = $Q_H/T_H-Q_C/T_C = -Q_C(1/T_C-1/T_H)$ which is less than zero since Q_C and $(1/T_C-1/T_H)$ are positive. This means that the entropy of the universe would decrease in such a process. Thus, if heat flows from a cold reservoir to a hot reservoir, with no other changes in the universe, it violates the second law, which says the entropy of the universe can never decrease).

This leads to what is called the Clausius form of the second law of thermodynamics. No processes are possible whose only result is the removal of energy from one reservoir and its absorption by another reservoir at a higher temperature. On the other hand, if energy flows from the hot reservoir to the cold reservoir with no other changes in the universe, then the same arguments can be used to show that the entropy increases, or remains constant for reversible processes. Therefore, such energy flows, which are very familiar, are in agreement with the laws of thermodynamics.

Since the Earth plus the Sun form an approximately isolated system, the entropy of this system is increasing by the second law, while the energy is constant by the first law. The system (Earth plus Sun) is not in an equilibrium state since energy is being transferred from the Sun to the Earth and the Earth is radiating energy into space. The balance of energy absorbed by the Earth (gain minus loss) controls the Earth's temperature. Manmade processes occurring on Earth (burning, automobiles, energy production, etc.) lead to a continual increase in entropy and to an increase in atmospheric pollution. This increase in atmospheric pollution changes the net energy absorbed by the Earth and hence the Earth's temperature. This is the greenhouse effect or global warming. Thus, there is a connection between the entropy increase of the Earth-Sun system and global warming. The main worry is the tremendous increase in atmospheric pollution by the increasing population of the Earth. It is not known how this will end but a "heat-death" of the Earth like that which exists on Venus is a disturbing possibility. Because of the complexity of the process it is very difficult to calculate and predict if this will happen. Unfortunately, past a certain point, such processes could lead to an irreversible increase of the temperature of the Earth, a runaway greenhouse effect, even if we stopped our atmospheric pollution.

The Carnot cycle is reversible so a Carnot refrigerator or air conditioner which removes heat from the cold temperature reservoir is in essence a heat engine operated in reverse. In Figure 3 is shown a schematic diagram of a general refrigerator, where Q_C is absorbed from the low temperature reservoir, work W is done on the system, and heat Q_H is delivered to the high temperature reservoir per cycle. The signs have again been chosen so that Q_C, Q_H, and W are all positive numbers. The first law applied to the refrigerator cycle gives:

$$W = Q_H - Q_C \quad \text{(refrigerator cycle)}, \tag{9}$$

which has the same form as for an engine, Equation 5, since it is just conservation of energy for a cycle. The quantity of interest for a refrigerator is the coefficient of performance k:

$$k = Q_C/W \quad \text{(coefficient of performance)}, \tag{10}$$

which is what is valuable, Q_C, divided by what must be paid for, W. If it could be arranged so that W = 0 this would give the best value for k since cooling without cost would be realized. However, this violates the Clausius form of the second law because energy would be transferred from the cold reservoir to the hot reservoir with no other change in the universe. Therefore, the coefficient of performance of any refrigerator cycle must be finite. The application of the second law to the Carnot refrigerator cycle gives the same fundamental Carnot cycle result relating the temperature ratio to the heat ratio in Equation 6. Equations 6, 9, and 10 can be combined to obtain an expression for the value of k for the Carnot refrigerator:

$$k_c = Q_C/(Q_H - Q_C) = T_C/(T_H - T_C) \quad \text{(Carnot cycle)}. \tag{11}$$

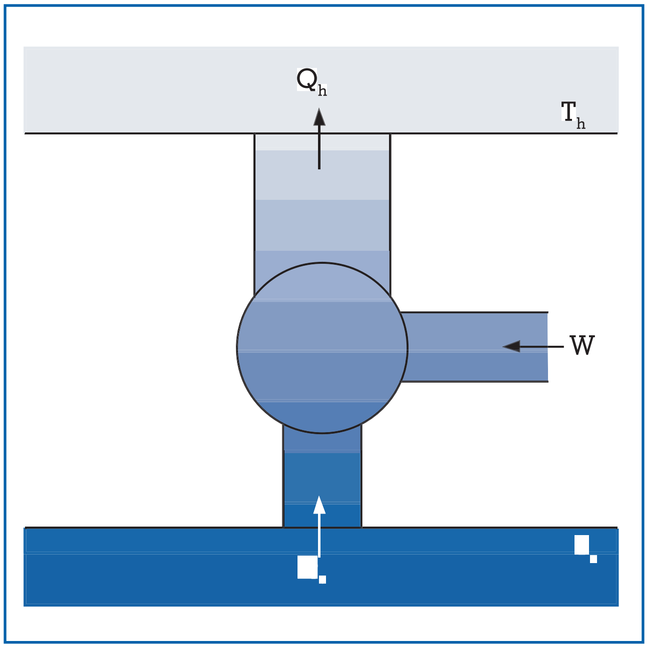


Figure 3.
Schematic of a refrigerator.

In a typical refrigerator, the cold temperature is around 273 K and the compressor raises the temperature of the gas on the high-temperature side by around 100 K so the coefficient of performance is approximately k = 273 K/100 K = 2.7. This means that the amount of heat removed is 2.7 times the amount of work done during one cycle. A heat pump cools the inside of the house in the summer and cools the outside of the house in the winter. It can be discussed as a refrigerator in both cases, but the cold temperature reservoir changes from inside to outside the house as summer changes to winter. This is accomplished by reversing the flow of the working substance in the heat pump. In the winter, as the outside cold temperature falls, the coefficient of performance falls according to Equation 11, and the heat pump produces less heating for the same amount of work. Often the heat pump is supplemented with resistive heating. As long as the coefficient of performance stays above 1, the heat pump is more efficient than simple resistive heating, which produces the same amount of heat as work. One way to make a heat pump have a larger coefficient of performance is to bury the heat exchange unit underground so that the cold temperature is higher than the ambient air temperature. With this change the heat pump can function efficiently even in cold climates.

No heat engine operating between two temperature reservoirs at temperatures T_C and T_H can be more efficient than a Carnot engine operating between these same two temperatures. To show this, first note that the engine can operate in a reversible or irreversible manner. If the engine is reversible, then it has the same efficiency as the Carnot engine since only two temperatures are involved. If the engine is irreversible, then the second law Equations 2a and 2b give

$$DS = Q_C/T_C - Q_H/T_H > 0$$

or

$$Q_C/Q_H > T_C/T_H.$$

If this is used in the definition of the efficiency of an engine,

$$e = 1 - Q_C/Q_H$$

and, compared to the efficiency of the Carnot cycle given in Equation 8, the efficiency of the engine must be less than the Carnot efficiency. This means that the work done in an irreversible engine is less for the same amount of heat absorbed from the high temperature reservoir. In the same way it follows that no refrigerator operating between two temperatures can have a greater coefficient of performance than a Carnot refrigerator. The laws of thermodynamics are of central importance for discussions of energy in our world.

John R. Ray

BIBLIOGRAPHY

Callen, H. B. (1985). *Thermodynamics: An Introduction to Thermostatics.* New York: John Wiley and Sons.

Fermi, E. (1957). *Thermodynamics.* New York: Dover Publications.

Pippard, A. B. (1966). *Elements of Classical Thermo-dynamics.* Cambridge, UK: Cambridge University Press.

Zemansky, M. W., and Dittman, R. H. (1997) *Heat and Thermodynamics*, 7th ed. New York: McGraw Hill.

THERMOSTATS

See: Energy Management Control Systems

THOMPSON, BENJAMIN (COUNT RUMFORD) (1753–1814)

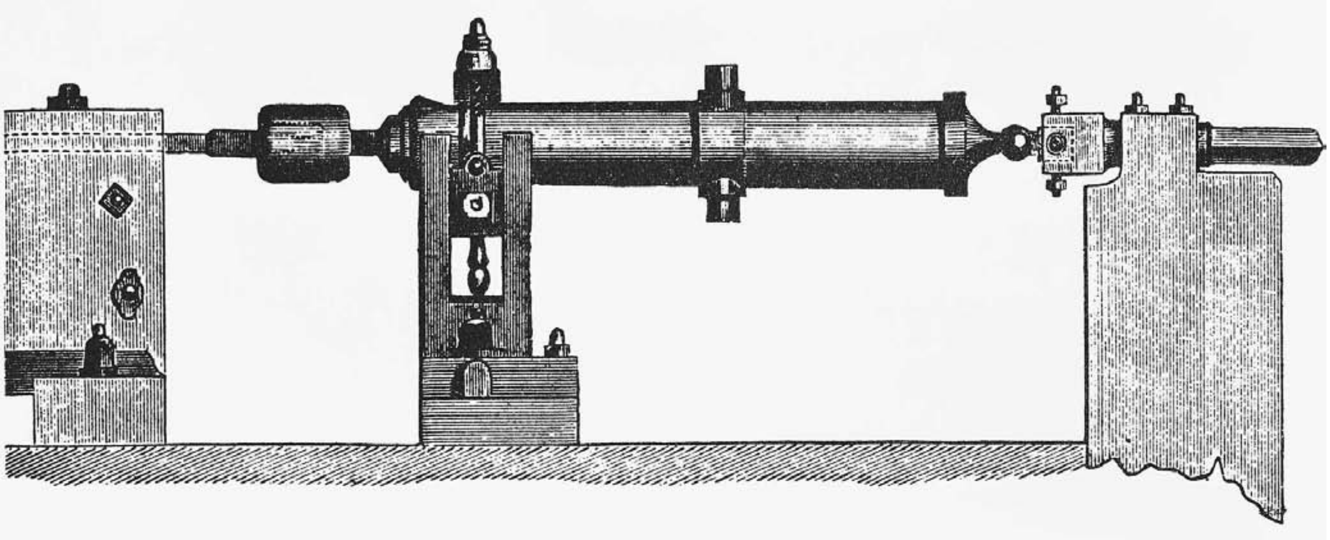
PERSONAL LIFE AND CAREER

Benjamin Thompson, born in Woburn, Massachusetts, in 1753, acquired a deep interest in books and scientific instruments as a youth, and matured into a person of great charm and intellect. In 1772 he received a position as a teacher in Concord, New Hampshire, where he soon met and rather quickly married Sarah Walker Rolfe, a wealthy widow fourteen years his senior, who had been charmed by his brilliant mind and dashing manners. Overnight his status changed to that of a country gentleman, managing his wife's estate and helping John Wentworth, the British governor of New Hampshire, with some agricultural experiments.

To reward Thompson, in 1773 Wentworth commissioned him a major in the New Hampshire militia. There Thompson acted as an informant for the British, so arousing the rancor of his fellow colonists that they planned to tar and feather him. He quietly departed for Boston, leaving his wife and baby daughter behind. He never saw his wife again, and only saw his daughter briefly many years later when he was living in France.

In Boston Thompson worked briefly for General Gage, the highest-ranking British officer in the Massachusetts Bay Colony. When the American Revolution began in 1776, his loyalty to the American cause was clearly suspect, and he left for England. Upon arriving in London his obvious talents led to rapid advancement in his career, and eventually he was appointed Undersecretary of State for the colonies. He also had a brief military career with the British Army in America, and retired in 1784 with the rank of Colonel.

In 1784 this freelance diplomat joined the court of Karl Theodor, Elector of Bavaria, and rapidly rose to become head of the ineffectual Bavarian Army. For his contributions to building up Bavaria's defensive strength, in 1793 he was made a Count of the Holy Roman Empire and took the name "Count Rumford," since that was the original name of the



Benjamin Thompson's 1798 invention for producing heat based on his idea that friction generates heat by increasing the motion of molecules. (Corbis-Bettmann)

New Hampshire town in which he had first taught school.

After returning briefly to London in 1800, Rumford was on the move again, this time to Paris, where in 1805 he married the widow of the great French chemist, Antoine Lavoisier. Rumsford's first wife had passed away by this point. His marriage to Madame Lavoisier was, however, a very stormy affair and two years later, after having ensured a handsome lifetime annuity for himself, he and his wife separated. Rumford then retired to Auteuil, outside Paris, where for the rest of his life he worked energetically and with much success on applied physics and technology.

During his life Count Rumford did much good with the funds he obtained by marrying wealthy widows, and by dazzling the leaders of three European nations with his scientific accomplishments. He established sizable prizes for outstanding scientific research, to be awarded by the American Academy of Arts and Science in Boston, and the Royal Society of London. He designed the lovely English Gardens in Munich, and supervised their construction. Finally he provided the funds to start the Royal Institution in London, which later attained great scientific prestige under the direction of Humphry Davy and Michael Faraday.

RUMFORD'S CONTRIBUTIONS TO SCIENCE

During the eighteenth century, the kinetic theory of heat had gradually lost favor and been replaced by the conception of heat as an indestructible fluid, to which Lavoisier had given the name "caloric." In 1798 Rumford, as Minister of War for the Elector of Bavaria, performed pivotal experiments in negating the existence of caloric. While watching the boring of a cannon barrel at the Munich military arsenal, Rumford was struck by the large amount of heat produced in the process. As long as the mechanical boring continued, heat continued to appear. This was hard to explain on the basis of the prevalent caloric theory. It appeared that continued boring was able to produce an inexaustible amount of caloric, which the limited amount of metal in the cannon barrel could not possibly contain.

Rumford decided to try a more controlled experiment. He placed a brass gun barrel in a wooden box containing about nineteen pounds of cold water, and used a team of horses to rotate a blunt steel borer inside the barrel. After 2.5 hours, the water boiled! As Rumford described it, "It would be difficult to describe the surprise and astonishment expressed on the countenances of the bystanders on seeing so large a quantity of cold water heated, and actually made to boil, without any fire."

Rumford suggested that anything that an isolated body can supply *without limitation* could not possibly be a material fluid. The only thing that could be communicated in this fashion was *motion*, in this case the motion of the steel borer that first produced *heat* in the form of molecular motion of the cannon molecules, which was then passed on to the water as random motion of its molecules. Therefore, according to Rumford heat was a form of random molecular motion. By this and other experiments Rumford had indeed demonstrated that a caloric fluid did not exist, but he had not yet seen the intimate connection that exists between heat, work and energy. That connection was to come later with the research of Sadi Carnot, Julius Robert Mayer, James Joule, and Hermann von Helmholtz.

In addition to this groundbreaking research on work and heat, Rumford made an extraordinary number of important contributions to applied science. He studied the insulating properties of the cloth and fur used in army uniforms, and the nutritive value of various foods and liquids. He decided that thick soup and coffee were the best sources of strength for any army in battle; and introduced the potato as a food into central Europe. He designed a large number of drip-type coffee makers and introduced the first kitchen range, the double boiler, and the pressure boiler found in modern kitchens. Rumford also designed better chimneys and steam-heating systems for houses. His continued interest in the scientific principles behind such devices made Rumford one of the world's first great applied physicists.

Rumford died suddenly at Auteuil in August 1814, leaving his entire estate to Harvard College to endow a professorship in his name. While considered a great scientist and a charming man by the best and brightest of his contemporaries, Rumford was at heart a soldier-of-fortune, and could be arrogant, obnoxious, and cruel to those he considered beneath him. For these reasons, he is less highly respected today as a man than he is as a scientist.

Joseph F. Mulligan

BIBLIOGRAPHY

Brown, S. C. (1952). "Count Rumford's Concept of Heat." *American Journal of Physics* 20:331–334.

Brown, S. C. (1953). "Count Rumford—Physicist and Technologist." *Proceedings of the American Academy of Arts and Sciences* 82:266–289.

Brown, S. C. (1962). *Count Rumford, Physicist Extra-ordinary*. Garden City, NY: Anchor Books.

Brown, S. C. (1976). "Thompson, Benjamin (Count Rumford)." In *Dictionary of Scientific Biography*, ed. C. C. Gillispie, Vol. 13, pp. 350–352. New York: Scribner.

Brown, S. C. (1979). *Benjamin Thompson, Count Rumford.* Cambridge, MA: MIT Press.

Ellis, G. E. (1871). *Memoir of Sir Benjamin Thompson, Count Rumford*. Boston: American Academy of Arts and Sciences.

Wilson, M. (1960). "Count Rumford," *Scientific American* 203 (October):158–168.

THOMSON, JOSEPH JOHN (1856–1940)

The British physicist, famous for his discovery of the electron, was born in Cheetham, near Manchester, on December 18, 1856. He first entered Owens College (later Manchester University) at the early age of fourteen. In 1876 Thomson won a scholarship in Mathematics to Trinity College, Cambridge, and remained a member of the College for the rest of his life. He became a Fellow in 1880, Lecturer in 1883 and Master in 1918, a position he held with great flair until his death on August 30, 1940.

Thomson met Rose Paget in 1889 when, as one of the first women to be allowed to conduct advanced work at Cambridge, she attended some of his lectures. They married on January 22, 1890, and had two children: a son, George, and a daughter, Joan. The marriage was a long and happy one.

Many were surprised when, at the end of 1884, Thomson was appointed Cavendish Professor of Experimental Physics in succession to Lord Rayleigh. Thomson was not yet twenty-eight, and he had almost no experience in experimentation. Nevertheless he immediately began work on the conduction of electricity through gases and single-mindedly pursued this topic until he resigned the Chair in 1919. Under Thomson's inspiration and guidance the Cavendish Laboratory became the world's foremost research institution. Seven of those who started or continued their careers there went on to win Nobel Prizes, including his own son.

J.J., as Thomson was commonly called, received the 1906 Nobel Prize in Physics in "recognition of the great merits of his theoretical and experimental investigations of the conduction of electricity by gases". He was knighted in 1908, received the Order of Merit in 1912 and was successively President of the Physical Society, the Royal Society and the Institute of Physics.

The demonstration of the independent existence of the negative electron of small mass was a watershed in the long quest to understand the nature of electricity. Many scientists had contributed to this search in the 300 years prior to Thomson's discovery: Gilbert, Franklin, Coulomb, Poisson, Galvani, Volta, Davy, Oersted, Ampère, Ohm, Faraday, Maxwell, and others. In 1874 the Irish physicist George Johnstone Stoney pointed out that, on the basis of Faraday's law of electrolysis, there exists an absolute unit of electricity, associated with each chemical bond or valency. Hermann von Helmholtz, independently, drew similar conclusions, declaring that electricity, both positive and negative, was divided into elementary portions that behaved like atoms of electricity. Stoney later suggested the name "electron" for this elementary charge.

The phenomena exhibited by the electric discharge in rarefied gases had long been known to German and British physicists. Many British physicists held that cathode rays were particles of matter, similar in size to ordinary molecules, projected from the negative pole. In contrast, most German physicists maintained that cathode rays were analogous to electric waves, getting strong support from Heinrich Hertz who in 1892 showed that the rays could pass through thin sheets of metal. That fact was difficult to reconcile with the molecule-sized particle interpretation.

Thomson was convinced that whenever a gas conducted electricity, some of its molecules split up, and it was these particles that carried electricity. Originally, he thought that the molecule was split into atoms. It was not until 1897 he realized the decomposition to be quite different from ordinary atomic dissociation. At the beginning of that year Thomson performed some experiments to test his particle theory.

Joseph John Thomson. (Library of Congress)

First he verified that cathode rays carried a negative charge of electricity and measured their deflection in magnetic and electrostatic fields. He concluded that cathode rays were charges of negative electricity carried by particles of matter. Thomson found that the ratio of the mass, m, of each of these particles to the charge, e, carried by it was independent of the gas in the discharge tube, and its value was of the order of 1/1,000 of the smallest value known at that time, namely the value for the hydrogen ion in electrolysis of solutions. He then devised a method for direct measurements of e as well as m/e, thus allowing the mass of the particles to be determined. The measurements showed that e carried the same charge as the hydrogen ion. Thus m was of the order of 1/1000 of the mass of the hydrogen atom, the smallest mass known at that time. This numerical result was perfectly adequate for the interpretation adopted and, if not at first very accurate, was soon improved by later experiments. Thomson concluded that the negative charge carrier, its mass and charge being invariable, must represent a fundamental concept of electricity, or indeed of matter in any state. With regard to its size he estimated what is now called the "classical electron radius" to be 10^{-15} m.

Thus, the search for the nature of electricity led to the discovery of the electron and the proof that it is a constituent of all atoms. These achievements gave scientists the first definite line of attack on the constitution of atoms and the structure of matter. The electron was the first of the many fundamental particles later proposed. Though Thomson was the undisputed discoverer of the electron, there were others in the hunt who came close to the prize, notably the French physicist Jean Perrin, the German physicists Emil Wiechert and Walter Kaufmann, and the Dutch physicist Pieter Zeeman. The latter had, in 1896, calculated the e/m

ratio for a vibrating "ion"—in other words a bound electron—emitting light. His result was similar to Thomson's corresponding value for a free electron.

Thomson's achievement not only produced explanations for many historically puzzling observations but also opened up new fields of science. Of these the richest is surely the electronic structure of matter, a field of unceasing development that has produced, among other things, the silicon chip, the computer and the technology that provides near-instantaneous global intercommunication and access to an unprecedented amount of information.

Leif Gerward
Christopher Cousins

BIBLIOGRAPHY

Gerward, L., and Cousins, C. (1997). "The Discovery of the Electron: A Centenary." *Physics Education* 32:219–225.

Random House Webster's Dictionary of Scientists. (1997). New York: Random House.

Thomson, J. J. (1897). "Cathode Rays." *Philosophical Magazine* 44:293–316.

Thomson, J. J. (1899). "On the Masses of the Ions in Gases at Low Pressures." *Philosophical Magazine* 48:547–567.

THOMSON, WILLIAM (LORD KELVIN) (1824–1907)

In the 1850s the Glasgow professor of natural philosophy (physics), William Thomson, and his colleague in engineering science, Macquorn Rankine, revolutionized the traditional language of mechanics with new terms such as "actual" ("kinetic" from 1862) and "potential" energy. Rankine also constructed a new "science of thermodynamics" by which engineers could evaluate the imperfections of heat engines. By the end of the decade, Thomson and Rankine had been joined by like-minded scientific reformers, most notably the Scottish natural philosophers James Clerk Maxwell, Peter Guthrie Tait, and Balfour Stewart, and the telegraph engineer Fleeming Jenkin. As a group, these physicists and engineers created a new "science of energy" intended to account for everything from the smallest particle to the largest heavenly body.

The fourth child of James and Margaret Thomson, William was born in 1824 in Belfast, then Ireland's leading industrial center, where his father taught mathematics in the politically radical Belfast Academical Institution. His mother came from a Glasgow commercial family, but died when William was just six. Encouraged throughout his early years by his father (mathematics professor at Glasgow University from 1832 until his death from cholera in 1849), Thomson received the best education then available in Britain for a future mathematical physicist. He moved easily from the broad philosophical education of Glasgow University to the intensive mathematical training offered by Cambridge University, where he came second in the mathematics examination (the "Mathematics Tripos") in 1845. Having spent some weeks in Paris acquiring experimental skills in 1846 at the age of twenty-two, Thomson was elected to the Glasgow chair of natural philosophy which turned out to be a post that he held for fifty-three years.

Thomson married Margaret Crum in 1852. She died, childless, in 1870 after a very long illness that rendered her unable to walk. William's second marriage was to Frances Blandy in 1874. Again, there were no children. She outlived him.

Through the engineering influence of his older brother James, William became increasingly committed in the late 1840s to Sadi Carnot's theory of the motive power of heat. Since its somewhat obscure publication in 1824, Carnot's theory had become better known through its analytical reformulation ten years later by the French engineer Emile Clapeyron. The theory explained the action of heat engines by analogy to waterwheels. Just as a "fall" of water drove a waterwheel, so the "fall" of heat between the high temperature of the boiler and the low temperature of the condenser drove a steam engine. This representation gave Thomson the means of formulating in 1848 an 'absolute' temperature scale (later named the "Kelvin scale" in his honor), which correlated temperature difference with work done, thereby making the scale independent of any specific working substance such as mercury or air.

In the same period, James Prescott Joule, son of a Manchester brewery owner, had been carrying out a series of experiments to determine the relationship between work done and heat produced. In the wake of Michael Faraday's electrical researches, electromagnetic engines appeared as a possible future rival to steam power, but in practice their performance failed to match the economy of the best steam engines. Attempting to explain this discrepancy, Joule's early research located the resistances to useful work in various parts of the electrical circuit including the battery. He concluded that, in order to account for all the gains and losses in a circuit, there had to be more than a mere transfer of heat from one part to another; that is, there had to be mutual conversion of heat into work according to a mechanical equivalent of heat. His investigations presupposed that no work (or its equivalent) could be truly lost in nature—only God could create or annihilate the basic building blocks of the universe.

Having met Joule for the first time at the 1847 meeting of the British Association for the Advancement of Science in Oxford, Thomson initially accepted that Joule's experiments had shown that work converted into heat. Committed to Carnot's theory of the production of work from a fall of heat, however, he could not accept the converse proposition that work had been converted into heat could simply be recovered as useful work. Therefore, he could not agree to Joule's claim for mutual convertibility. By 1848 he had appropriated from the lectures of the late Thomas Young (reprinted in the mid-1840s) the term "energy" as a synonym for *vis viva* (the term in use at the time, traditionally measured as mv^2) and its equivalent terms such as work, but as yet the term appeared only in a footnote.

Prompted by the competing investigations of Rankine and the German physicist Rudolf Clausius, Thomson finally reconciled the theories of Carnot and Joule in 1850–51. For the production of motive power, a "thermo-dynamic engine" (Thomson's new name for a heat engine) required both the transfer of heat from high to low temperature and the conversion of an amount of heat exactly equivalent to the work done. His long-delayed acceptance of Joule's proposition rested on a resolution of the problem of the irrecoverability of work lost as heat. He now claimed that work "is lost to man irrecoverably though not lost in the material world." Like Joule, he believed that God alone could create or destroy energy (that is, energy was conserved in total quantity) but, following Carnot and Clapeyron, he also held that human beings could only utilize and direct transformations of energy from higher to lower states, for example in waterwheels or heat engines. Failure to do so resulted in irrecoverable losses of useful work.

William Thomson (Lord Kelvin). (Corbis Corporation)

Thomson's "On a Universal Tendency in Nature to the Dissipation of Mechanical Energy" (1852) took the new "energy" perspective to a wide audience. In this short paper for the *Philosophical Magazine*, the term "energy" achieved public prominence for the first time, and the dual principles of conservation and dissipation of energy were made explicit: "As it is most certain that Creative Power alone can either call into existence or annihilate mechanical energy, the 'waste' referred to cannot be annihilation, but must be some transformation of energy." Two years later Thomson told the Liverpool meeting of the British Association that Joule's discovery of the conversion of work into heat by fluid friction—the experimental foundation of the new energy physics—had "led to the greatest reform that physical science has experienced since the days of Newton."

Through the British Association, Thomson and his associates offered a powerful rival reform program to that of metropolitan scientific naturalists (including T. H. Huxley and John Tyndall) who promoted a professionalized science, free from the perceived shackles of Christianity and grounded on

Darwinian evolution and the doctrine of energy conservation. In critical response to Charles Darwin's demand for a much longer time for evolution by natural selection, and in opposition to Charles Lyell's uniformitarian geology upon which Darwin's claims were grounded, Thomson deployed Fourier's conduction law (now a special case of energy dissipation) to make order-of-magnitude estimates for the earth's age. The limited time-scale of about 100 million years (later reduced) appeared to make evolution by natural selection untenable. But the new cosmogony (theory of the origin of the universe) was itself evolutionary, offering little or no comfort to strict biblical literalists in Victorian Britain.

Thomson also examined the principal source of all useful work on earth. Arguing that the sun's energy was too great to be supplied by chemical means or by a mere molten mass cooling, he at first suggested that the sun's heat was provided by vast quantities of meteors orbiting around the sun, but inside the earth's orbit. Retarded in their orbits by an ethereal medium, the meteors would progressively spiral toward the sun's surface in a cosmic vortex. As the meteors vaporized by friction, they would generate immense quantities of heat. In the early 1860s, however, he adopted Hermann Helmholtz's version of the sun's heat, whereby contraction of the body of the sun released heat over long periods. Either way, the sun's energy was finite and calculable, making possible his estimates of its limited past and future duration.

The most celebrated textual embodiment of the "science of energy" was Thomson and Tait's *Treatise on Natural Philosophy* (1867). Originally intending to treat all branches of natural philosophy, Thomson and Tait in fact produced only the first volume of the *Treatise*. Taking statics to be derivative from dynamics, they reinterpreted Newton's third law (action-reaction) as conservation of energy, with action viewed as rate of working. Fundamental to the new energy physics was the move to make extremum (maximum or minimum) conditions, rather than point forces, the theoretical foundation of dynamics. The tendency of an entire system to move from one place to another in the most economical way would determine the forces and motions of the various parts of the system. Variational principles (especially least action) thus played a central role in the new dynamics.

Throughout the 1860s Thomson and his associates (especially Jenkin, Maxwell, and Balfour Stewart) played a leading role, both in shaping the design of electrical measuring apparatus and in promoting the adoption of an absolute system of physical measurement such that all the units (including electrical resistance) of the system should bear a definite relation to the unit of work, "the great connecting link between all physical measurements." These researches were conducted in the aftermath of the failure of a number of deep-sea telegraph projects, most notably the Transatlantic Cable of 1858. They provided the scientific foundation for a dramatic expansion in British telegraphic communication around the globe in the remaining decades of the nineteenth century, when British Imperial power reached its zenith.

Elevation to the peerage of the United Kingdom in 1892 brought William Thomson the title Baron Kelvin of Largs (usually abbreviated to Lord Kelvin). He was the first British scientist to be thus honored and took the title Kelvin from the tributary of the River Clyde that flowed close to the University of Glasgow. By the closing decades of his long life, when most of his associates had passed away, Kelvin was very much the elder statesman of British science. But the science of energy had been taken up by a younger generation of physical scientists and transformed into quite different modes of scientific understanding, ranging from the "energetics" of the German physical chemist Wilhelm Ostwald (denying atoms in favor of energy) to the "radioactive" physics of Ernest Rutherford (offering new possibilities for estimating the ages of earth and sun). Resisting many of the consequences of these new conceptions, Kelvin published right up to his death at the age of eighty-three, and found at last a resting place in Westminster Abbey, not far from his hero, Sir Isaac Newton.

Crosbie Smith

See also: Carnot, Nicolas Leonard Sadi; Faraday, Michael; Fourier, Jean Baptiste Joseph; Helmholtz, Hermann von; Joule, James Prescott; Maxwell, James Clerk; Rankine, William John Macquorn.

BIBLIOGRAPHY

Smith, C., and Wise, N. (1989). *Energy and Empire. A Biographical Study of Lord Kelvin*. Cambridge: Cambridge University Press.

Smith, C. (1998). *The Science of Energy. A Cultural History of Energy Physics in Victorian Britain*. Chicago: University of Chicago Press.

Thompson, S. P. (1910). *The Life of William Thomson, Baron Kelvin of Largs*, 2 vols. London: Macmillan and Co.

Thomson, W. (1872). *Reprint of Papers on Electrostatics and Magnetism*. London: Macmillan and Co.

Thomson, W. (1882–1911). *Mathematical and Physical Papers*, 6 vols. Cambridge: Cambridge University Press.

Thomson, W. (1891-1894). *Popular Lectures and Addresses*, 3 vols. London: Macmillan and Co.

THREE MILE ISLAND

See: Environmental Problems and Energy Use; Nuclear Energy, Historical Evolution of the Use of

TIDAL ENERGY

See: Ocean Energy

TIRES

HISTORICAL DEVELOPMENT

The first commercially successful pneumatic tire was developed in 1888 in Belfast by the Scottish veterinarian John Boyd Dunlop primarily to improve the riding comfort of bicycles. Dunlop also showed, albeit qualitatively, that his air-inflated "pneumatic" took less effort to rotate than did the solid rubber tires in use at that time. His qualitative tests were the first known rolling resistance experiments on pneumatic tires. Due to this significant reduction in rolling loss, many professional cyclists in Britain and Ireland adopted air-inflated tires for their bicycles by the early 1890s. Pneumatics for the nascent automobile industry soon followed.

Tires, like everything that rolls, encounter resistance. The resistance encountered by the tire rolling across a surface is a major factor in determining the amount of energy needed to move vehicles. Since Dunlop's original efforts, a considerable number of tire design improvements have been made that have tended to cause a decrease in tire power consumption. For example: separate plies of cotton cord were introduced in the early 1900s to replace Dunlop's square-woven flax (as early tires failed from fabric fatigue before wearing out); in the late 1950s the fuel-efficient radial-ply construction was commercialized in Europe to replace the bias-ply tire; this change also improved vehicle handling and increased mileage. (The radial tire features one or more layers of reinforcing cords or plies disposed perpendicular to the two beads plus a steel belt in the tread region, while the bias construction is built-up with an even number of plies arrayed at alternating, opposing angles between beads without a belt.) There has been a trend toward using larger-diameter tires, which are more rolling-efficient for comfort, appearance and safety reasons as vehicles were downsized in the United States during the 1980s. Elimination of the inner tube by tubeless tires, the use of fewer but stronger cord plies, and the production of more dimensionally uniform tires have each made small but measurable reductions in tire energy loss, although each of these changes was instituted primarily for other reasons.

Scottish inventor John Boyd Dunlop, displays an invention he patented in 1888—the bicycle with pneumatic tires. (Corbis-Bettmann)

By the 1990s, automobile tires in the industrialized portions of the world were often taken for granted by the motoring public because of the high level of performance they routinely provide in operation. This casual attitude toward tires by a large segment of the driving public can be partially explained by the sheer number of units manufactured and sold each year. In 1996 about one billion tires of all sizes and types were marketed worldwide with a value of more than $70 billion; this includes 731 million passenger car tires and 231 million truck tires with agricultural, earthmover, aircraft, and motorcycle tires constituting the remainder. Three regions of the world dominate tire production and sales: North America, Western Europe and Japan. Among them, these mature markets are responsible for more than three-quarters of all passenger car tires and one-half of all truck tires. Bicycle tires are now largely produced in less-developed countries, so an accurate number is hard to assess.

The pneumatic tire has the geometry of a thin-walled toroidal shell. It consists of as many as fifty different materials, including natural rubber and a variety of synthetic elastomers, plus carbon black of various types, tire cord, bead wire, and many chemical compounding ingredients, such as sulfur and zinc oxide. These constituent materials are combined in different proportions to form the key components of the composite tire structure. The compliant tread of a passenger car tire, for example, provides road grip; the sidewall protects the internal cords from curb abrasion; in turn, the cords, prestressed by inflation pressure, reinforce the rubber matrix and carry the majority of applied loads; finally, the two circumferential bundles of bead wire anchor the pressurized torus securely to the rim of the wheel.

However, it is the inelastic properties of the cord and rubber components of the tire that are responsible for the heat buildup and energy dissipation that occur during each tire revolution. This loss of energy results in a drag force that impedes tire rotation. The cyclic energy dissipation of tire materials is a mixed blessing for the tire development engineer. It is required in the contact patch between tire and road to produce frictional forces that accelerate, brake, and/or corner the vehicle, but it must be minimized throughout the tire in the free-rolling condition so as not to adversely impact fuel economy.

Depending on the specific car model and service conditions, tires can consume 10 to 15 percent of the total energy in the fuel tank, and in this respect the power loss of the four tires is comparable in magnitude to the aerodynamic drag opposing the forward motion of the vehicle during urban driving at about 45 miles per hour.

	Rolling Resistance Coefficient
Radial passenger tire	.008 – .015
Bias passenger tire	.016 – .025
Radial truck tire	.005 – .007
Bias truck tire	.008 – .010
Railroad wheel (steel)	.002 – .003

Table 1.
Rolling Resistance Coefficient of Various Tire Types

Rolling resistance coefficient is defined as the nondimensional ratio of drag force retarding tire rotation to wheel load—with lower being better. For most passenger car tires freely rolling on smooth, hard surfaces, this coefficient varies between 0.01 and 0.02, but may increase to 0.03 or higher at increased speeds, at very high loads, at low-inflation pressures, and/or on soft surfaces.

Typical ranges for the rolling resistance coefficients of passenger car and truck tires in normal service (vs. a steel wheel) are given in Table 1.

Truck tires are operated at about four times the inflation pressure of passenger car tires, which principally accounts for truck tires' lower rolling resistance coefficients.

Power consumption is the product of drag force and speed—and four tires on a typical American sedan consume approximately 10 horsepower of engine output at 65 miles per hour.

The radial passenger tire introduced in North America during the 1970s had a 20 to 25 percent improvement in rolling resistance compared to the bias ply tire then in use. This was a major improvement that resulted in a 4 to 5 percent gain in vehicle fuel economy under steady-state driving conditions—that is, an approximately 5:1 ratio between rolling resistance reduction and fuel economy improvement. By the 1990s, evolutionary advances in tire design and materials continued with the general use of higher operating pressures, but an 8 to 10 percent improvement in rolling resistance now translates into only a 1 percent gain in fuel economy.

Trade-offs between rolling resistance (and therefore fuel economy) and safety occur in tire design and usage just as with vehicles. For example, the least hysteretic polymers and tread compound formulas that lower rolling resistance tend to compromise wet grip and tread wear. Also, worn-out tires with little or no remaining tread depth are much more fuel-efficient than new tires, but require greater distances to stop when braking on wet surfaces.

ENERGY FOR MANUFACTURING

The energy required to produce a tire is only about 10 percent of that consumed while in use on an automobile overcoming rolling resistance during 40,000 miles of service. The majority of the manufacturing energy expended, 60 to 70 percent, is for the mold during the vulcanization process.

At the end of its useful life, the tire, with its hydrocarbon-based constituents, is a valuable source of energy, with a higher energy value than coal. By the mid-1990s, approximately 80 percent of the tires worn out annually in the United States were being recycled, recovered, or reused in some fashion, with three-fourths of these serving as fuel for energy production in boilers and cement kilns.

Joseph D. Walter

See also: Automobile Performance; Efficiency of Energy Use; Transportation, Evolution of Energy Use and.

BIBLIOGRAPHY

Mullineux, N. (1997). *The World Tyre Industry*. London: *The Economist* Intelligence Unit.

Schuring, D. J. (1980). "The Rolling Loss of Pneumatic Tires." *Rubber Chemistry and Technology* 53:600–727.

Tabor, D. (1994). "The Rolling and Skidding of Automobile Tyres." *Physics Education* 29:301–306.

Walter, J. D. (1974). "Energy Losses in Tires." *Tire Science and Technology* 2:235–260.

TOKAMAK

See: Nuclear Energy

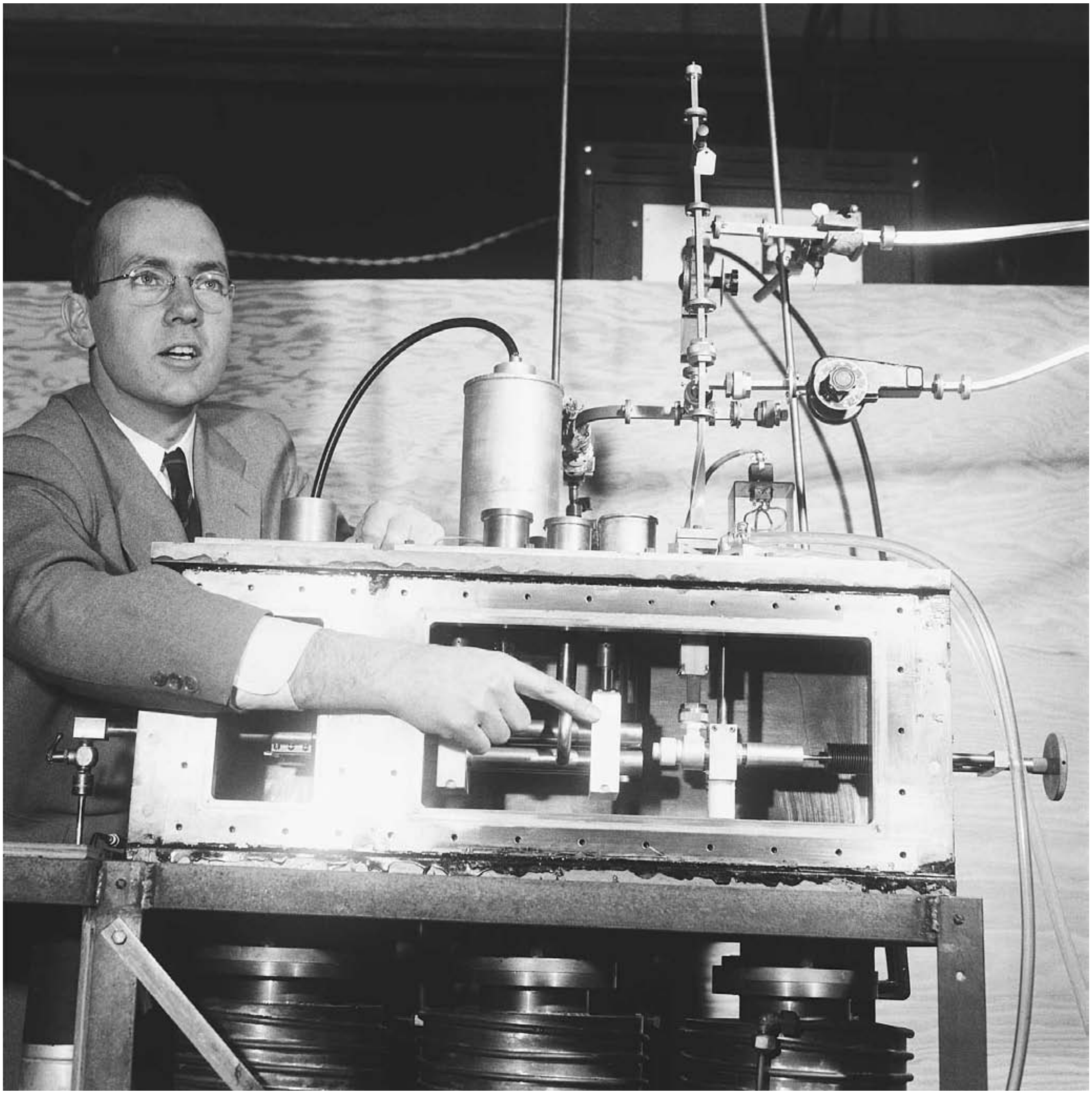
TOWNES, CHARLES HARD (1915–)

Charles Hard Townes had a long, distinguished career as a physicist and educator with years of service to the military and academic research communities. One particular line of work will ensure that his name be remembered by history—his contributions to the early development of the laser. The laser is considered by many historians to be one of the most important technological achievements of the twentieth century. As a source of energy, the laser is extremely inefficient—it devours electricity, converting only a fraction into the laser beam. Yet, due to the special properties of laser versus normal light, this technology is extraordinarily useful. A 100-watt light bulb, for example, can barely light a living room, but a pulsed, 100-watt laser can be used to cut or drill holes in metal.

Born in Greenville, South Carolina, Charles Townes early in his life showed the characteristics of both a highly intelligent individual and a scientist. As a boy, Townes studied continuously, covering daunting subjects such as Greek, Latin, and Old English. He also read every issue of Popular Mechanics. Not limiting himself as a book scholar, however, Townes plunged into the hands-on work of technical subjects. Townes's father owned a shop that he rented to a clock and watch dealer. Townes reveled in disassembling the old, broken clocks. In high school, he took shop and mechanical drawing and developed a special interest in electricity and radio. He even attempted (unsuccessfully) to build his own crystal radio.

At the age of sixteen, Townes entered Furman University and received two bachelor's degrees (modern languages and physics) in 1935. He continued his education, receiving a master's at Duke University in 1937 and a doctorate at Cal Tech in 1939. In the summer of 1939, Bell Labs hired him. Numerous lines of research were being undertaken simultaneously at Bell Labs. Most of the work done by Townes initially dealt with basic research and the transmission of telephone and television signals. Worldwide political events, however, soon changed this emphasis.

In September of 1939, Germany invaded Poland and started what would become World War II. Although the United States was still politically neu-



Charles Hard Townes operates an atomic clock. (Corbis-Bettmann)

tral at the time, in March, 1941 Townes was reassigned to work on radar bombing. Bell Labs had done some research on anti-aircraft aiming systems, and Townes and his colleagues used this knowledge to develop a way to replace the bombsight with radar targeting. Although this advance was not put into practice during the war, it helped forge the future of high-tech warfare and was influential in Townes's later civilian endeavors.

Just as the intellectual atmosphere of Bell Labs shaped Townes's career, the cosmopolitan diversity of New York City molded his personal life. He enjoyed the many theaters, museums, and restaurants there. He signed up for voice and music theory lessons at

Juilliard. And most significantly, he met Frances Brown, an activity director at the International House at Columbia University, whom he married in May 1941. The couple had four daughters. His work at Bell Labs had one downside—it often took him away from his family for trips to Florida where the radar bombing system was tested and re-tested.

In 1948, however, Townes returned to the academic world, accepting a professorship at Columbia University. There, during the early 1950s, he speculated that stimulated emission (an ambitious theory first proposed by Albert Einstein in 1917) could generate microwaves, but he also knew a population inversion was necessary. Population inversion occurs when energy is introduced, causing the majority of a material's atoms to reach an excited level rather than their normal "ground level." Edward Purcell and Robert Pound, researchers at Harvard University, had demonstrated population inversion in 1950 using a crystal composed of lithium fluoride. Using this concept as well as his own radar-related work on amplifying a signal, Townes came up with the idea for a closed cavity in which radiation would bounce back and forth, exciting more and more molecules with each pass. With that, Townes had the basic concept of the laser.

Over the next few years, Townes, with the help of graduate students Herbert Zeiger and James Gordon, calculated the precise size of the necessary cavity and built a device that used ammonia as an active medium. Success came in 1954 with the completion of the first maser, an acronym for microwave amplification by stimulated emission of radiation.

In 1957, Townes developed the equation that showed that this same process could also obtain much smaller wavelengths (in the infrared and visible light range). Townes collaborated with Arthur Schawlow, a research assistant in his laboratory from 1949 to 1951, who then moved on to become a physicist at Bell Labs where Townes was still doing consulting work. When he was a postdoctoral fellow at Columbia, Schawlow met Aurelia, Townes' younger sister, who had come there to study singing. Soon the two married.

In 1957, this team of brothers-in-law started working together on Townes's idea for an optical maser. They found atoms that they felt had the most potential, based on transitional probabilities and lifetimes. However, there was still one major problem: In the visible light portion of the electromagnetic spectrum, atoms don't remain in an excited state as long as microwaves. A new cavity seemed the most appropriate solution. In 1958, Schawlow proposed using a long, slim tube as the resonating cavity to produce a very narrow beam of radiation. With that breakthrough, Townes and Schawlow were well on their way to developing a laser. They authored a paper on the subject and submitted a patent on behalf of Bell Labs.

Others, however, were reaching similar ideas at the same time. R. Gordon Gould, a graduate student at Columbia, had come to the same conclusions. In November 1957 he wrote up his notes on a possible laser and had them notarized. Theodore Maiman, a physicist who had been working with ruby masers at Hughes Aircraft Company in Malibu, California, learned of laser research in September 1959 at a conference on quantum electronics that Townes had organized. Afterwards, he began his research on a laser using a pink ruby as an active medium.

Townes's participation in the race for the first laser lessened in the fall of 1959. He was offered and accepted a position as the director of research at the Institute for Defense Analysis in Washington, D.C. With cold war tensions throughout the world, and science playing an increasingly prominent role in government thinking and funding, Townes felt obligated and honored to serve his country in this role. He still met with and directed graduate students at Columbia on Saturdays, but his active role was dramatically reduced.

Schawlow continued working on his laser at Bell Labs. He had rejected ruby as an active medium because he felt it would not reach population inversion. By pumping the ruby with the light from a photographer's flash lamp, however, Maiman succeeded, created the world's first laser in June 1960.

Townes's academic life continued. He served as provost of MIT from 1961 to 1966. In 1964, Townes received the Nobel Prize in physics "for work in quantum electronics leading to construction of oscillators and amplifiers based on the maser-laser principle." He was named university professor at the University of California-Berkeley in 1967. There he worked for more than 20 years in astrophysics. Ironically, this field is one of many that were transformed by the laser, and Townes often used lasers in his subsequent research.

The career of Charles H. Townes and the development of the laser exemplify the technological revolution of the twentieth century. Following the second world war, the United States experienced a golden age of science. Basic research flourished with unprecedented government funding, and it provided the underlying principles for thousands of future devices.

Pioneers such as Charles Hard Townes helped steer the course for this era in the history of energy, and in doing so forged a path for scientists to come.

Karl J. Hejlik

See also: Lasers.

BIBLIOGRAPHY

Bromberg, J. L. (1992). "Amazing Light." *Invention and Technology*. Spring: 18–26.

Bromberg, J. L. (1991). *The Laser in America: 1950–1970.* Cambridge, MA: MIT Press.

Hecht, J. (1992). *Laser Pioneers*. Boston, MA: Academic Press.

Perry, T. S. (1991). "Charles H. Townes: Masers, Lasers & More." *IEEE Spectrum* 28:32.

National Academy of Sciences. (1987). *Lasers: Invention to Application*. Washington, DC: National Academy Press.

TRAFFIC FLOW MANAGEMENT

Americans living in the fifty most congested cities spend an average of thirty-three hours each year stuck in traffic. Congestion causes much more than driver aggravation: air quality suffers, vehicle idling and stop-and-go traffic reduce fuel economy by as much as 30 percent, and we lose billions of dollars in productivity. These are the consequences as the automobile does what it is designed to do—transport a highly mobile population. Continued suburban expansion, reduction in household size, increase in number of workers per household, and general changes in lifestyle have all contributed to increased travel demand and greater congestion.

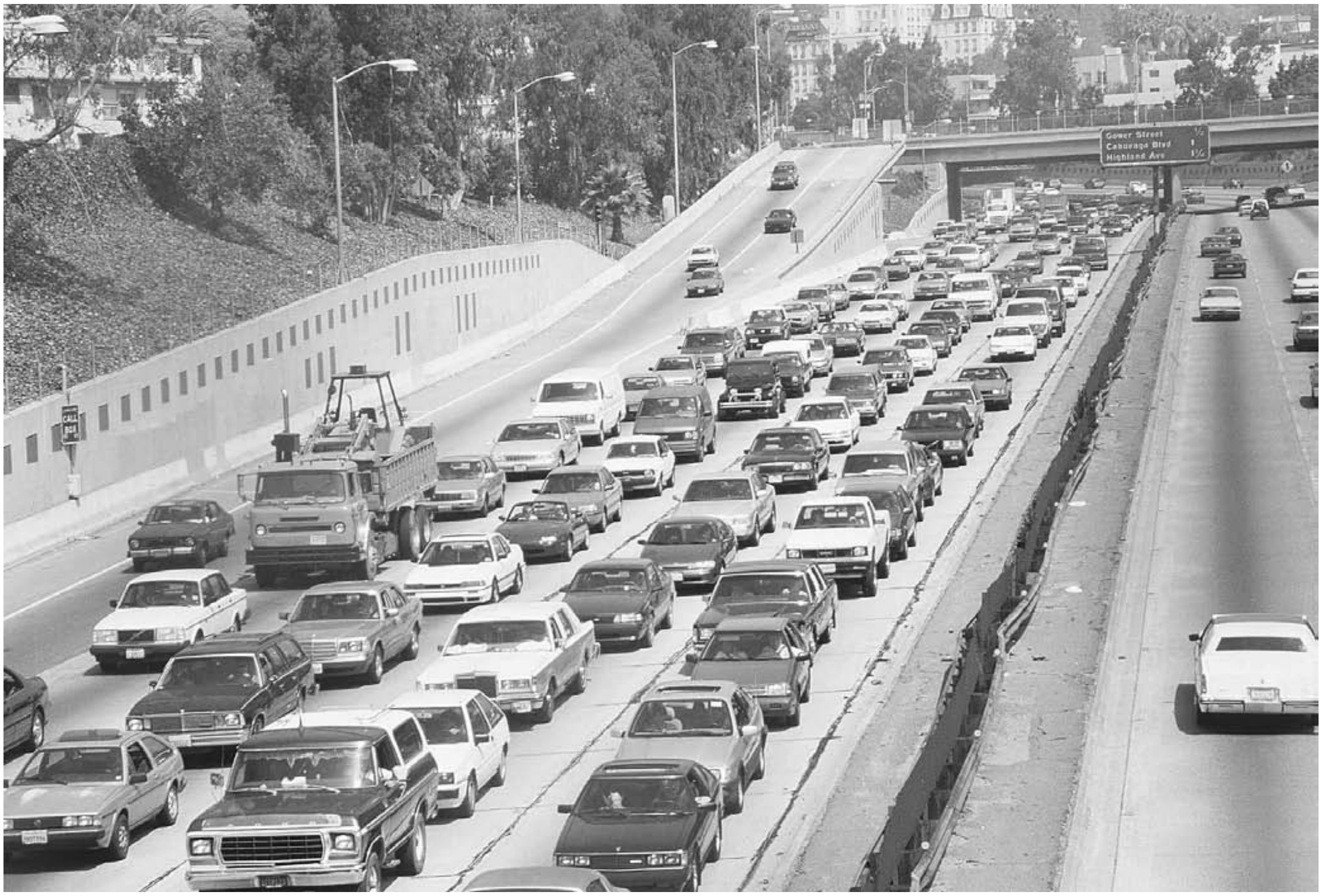
Even without congestion, from the perspective of capital utilization and energy consumption, automobile and roadway use is inefficient. First, the majority of personal transportation energy is consumed in moving personal vehicles that contain only one occupant and drive one or two hours a day. Second, the transportation infrastructure usually operates below capacity. States expend tremendous resources building highways to accommodate peak period demands (7 A.M. to 9 A.M. and 3 P.M. to 6 P.M.). Most of these lanes are not needed for the rest of the day. Rush hour demand still exceeds capacity in many places, resulting in disruption of traffic flow, stop-and-go driving conditions, a drop in freeway throughput, increased fuel consumption, increased vehicle emissions, and wasted time. Nevertheless, the personal vehicle remains the preferred means of transportation in much of the world given the personal freedom and economic opportunity (access to jobs) that it affords.

Using capital and energy more efficiently is a common goal of government, business, and the individual. State agencies have responded to congestion delay by building more capacity and/or increasing the efficiency of existing capacity. At the local level, however, there is growing opposition to the detrimental noise, air quality, and space infringements of building more and bigger highways. Federal environmental and transportation regulations have shifted from capacity expansion to a focus on congestion reduction and air quality improvement. Because mobility enhances regional economic development and productivity, shifting travel demand out of the peak period, or efficiently improving transportation supply (without necessarily adding new major facilities) still provides tremendous public benefits. Recent strategies have focused on behavior modification (transportation demand management strategies) and targeted improvements to the transportation infrastructure (transportation supply improvement strategies).

TRANSPORTATION DEMAND MANAGEMENT STRATEGIES

The objective of demand management strategies is to encourage or require drivers to reduce the frequency and length of automobile trips, to share rides, or to use alternative modes of transportation. When peak period automobile trips are shifted out of the peak, or into alternative transportation (such as carpools, mass transit, bicycling, or walking) congestion declines and the remaining automobile users benefit from improved travel times. Demand management measures include no-drive days, employer-based trip reduction programs, parking management, park and ride programs, alternative work schedules, transit fare subsidies, and public awareness programs. Demand management measures may or may not require intensive capital investment, but are usually characterized by ongoing operating costs.

Researchers have been able to identify strategies (and specific incentives and disincentives) that can



Cars in a traffic jam on the Hollywood Freeway in California. (Corbis Corporation)

change travel behavior at individual facilities. Effective strategies can be tailored for individual businesses and activities as a function of employment classification, land use, and transit service availability. The fact that many strategies are effective at reducing travel demand at individual facilities is not controversial. But it is difficult to implement regional demand management programs that are acceptable to the public without micromanagement by the government. Such regional demand reduction strategies are typically implemented in the form of regulatory mandates, economic incentives, and education campaigns.

Regulatory Mandates and Employer-Based Trip Reduction

Businesses, operating in a market economy, have little incentive to implement demand management strategies on their own. They pay for the costs of shipment delays associated with moving goods and services to the marketplace (these costs are incorporated into the selling price of the goods and services). However, the vast majority of society's congestion cost is external to marketplace price-setting. Companies sometimes implement measures to reduce employee travel to their facility when there's insufficient parking. Normally, they provide convenient automobile access because it makes it easier to attract and keep employees.

Regulatory mandates require, either directly or indirectly, that specific segments of the population change their trip-making behavior. Examples of direct regulatory mandates include: automobile bans in downtown areas, restrictions on motor vehicle idling time (i.e., heavy-duty vehicles), restricted access to airport terminals for certain types of vehicles, odd/even day gasoline rationing at retail filling stations (based on license plate numbers), restricted hours for goods delivery in urban areas, and peak hour restrictions on truck usage in downtown areas. Direct mandates have proven extremely unpopular. Although developing and rapidly growing industrial countries (e.g., Singapore, Mexico, and China) do implement such measures, none are implemented on a sustained and widespread basis in the United States.

Implementation of indirect regulatory mandates has been more common than direct mandates, primarily in the form of trip reduction ordinances implemented by local governments. These ordinances require employers to increase the average vehicle occupancy of employee vehicles during commute periods, but usually allow employers great flexibility in developing their strategies.

During the late 1980s and early 1990s, employers implemented trip reduction measures in many urban areas. Regulatory agencies developed commute vehicle occupancy goals that would result in fewer vehicle trips to the facility during the morning peak. Employers could offer their employees incentives (e.g., cash rebates for carpool participants) or impose disincentives (e.g., parking fees for employees who drove alone)to achieve these ridership goals.

The largest and most prominent experience with trip reduction ordinances in the United States was Regulation XV, adopted December 11, 1987, in the South Coast (Los Angeles) area. The regulation required employers of a hundred or more individuals to prepare and implement trip reduction plans between the home and worksite. Employers developed their own incentive programs to encourage workers to rideshare or use alternative transportation. Facilities reported progress annually, and they adjusted their plans each year until they achieved their ridership goal.

The tripmaking aspects of specific measures implemented by employers under Regulation XV were widely variable. Trip reductions depended on such local factors as employer size and location, employment and site characteristics, location of labor pool, and socioeconomic composition. A 1991 study of seventy-six facilities in the Regulation XV program found no apparent correlation between the number of incentives offered and the improvement in ridership levels. The quality, not the quantity of incentives offered was the driving force for behavioral change. Two factors had a significant effect on ridesharing: (1) use of parking incentives and disincentives coupled with transit pass or commute subsidies; and (2) management commitment coupled with the presence of an on-site transportation coordinator. A program to guarantee rides home for emergencies and last minute work/personal schedule changes was necessary but insufficient condition to encourage ridesharing.

Over the entire Los Angeles region, employer-based demand management strategies were slow to evolve. Employers were required only to develop “approvable” plans to achieve specified ridership goals with no penalty for failure to achieve the goals. A detailed study of 1,110 work sites found that the implementation of Regulation XV reduced vehicle use to participating facilities by about 5 percent during the first year of the program. The most instructive finding of this study is that the primary improvements in ridership came from increased use of carpools. All other changes were trivial: a slight increase in vanpools and compressed workweeks, a slight decrease in bicycling/walking and telecommuting, and no change in transit use. This finding suggested that reduced vehicle use can be achieved with little or no institutional change, because carpools do not require the same level of organizational effort and financial support as many other options.

Even if successfully implemented, the overall travel implications of programs similar to Regulation XV would be modest. Commute trips represent about 25 percent of daily trips, and commute trips to facilities with a hundred or more employees represent approximately 40 perent of commute trips in the Los Angeles area. Even if commute trips to affected facilities are reduced between 5 percent and 20 percent, employer-based trip-reduction strategies may yield total daily trip reductions of between 0.5 percent and 2 percent (although primarily achieved during peak periods). The costs of such initiatives are substantial. Employers often hire rideshare coordinators and provide incentives, and regulators must monitor and enforce the program. In a survey of more than 400 Los Angeles area facilities the typical cost of placing employees in carpools or transit through personalized ridesharing assistance ranged from $7.72 per employee in large firms (~10,000+ employees) to $33.91 per employee in small firms (~100 employees).

A few years after Regulation XV was implemented, when medium-sized businesses (100–250 employees) came under the regulatory requirements, the business community began exerting significantly increased political pressure on the South Coast Air Quality Management District (SCAQMD) to repeal the regulation, and took their case directly to Congress. On December 23, 1995, Congress amended Section 182(d)(1)(b) of the Clean Air Act. The new language allowed air quality planning agencies to opt out of the required employer-based programs.

Overnight, employee commute programs around the nation became voluntary, disappearing entirely from air quality management plans.

The SCAQMD undertook an eighteen-month study in which the agency encouraged voluntary rideshare efforts for medium-sized facilities. However, the exemption of facilities from employer-based trip reduction and ineffective implementation of voluntary programs yielded an increase in pollutant emissions. Despite the failure of voluntary measures, the California State Legislature permanently exempted medium-sized businesses from the Los Angeles regulations. The largest air pollution control agency with the worst air quality in the nation could not retain their commute program for medium-sized employers over public objection.

Public Information and Education

Recent behavioral shifts, such as the overall decrease in the number of smokers and the increase in residential recycling activity, suggest that ongoing media campaigns coupled with formal education programs can effectively influence human behavior. Many of California's local air pollution control districts have implemented education programs as a means of increasing public awareness of how travel behavior affects air quality. The California Air Resources Board prepared a variety of information packets for government decision-makers as well as the public. Numerous states and local air pollution control agencies have followed suit across the United States.

Education campaigns implemented in conjunction with regulatory mandates can make employer-based trip reduction strategies more efficient at both the facility and regional levels. The SCAQMD implemented an education program to support their employer trip reduction program. District staff members advised corporate representatives on cost-effective strategies implemented by other companies in the region. They recommended compressed work weeks, in-house rideshare matching, subsidized transit passes, carpool/vanpool subsidies, preferential carpool parking, flexible hours, telecommuting, bicycle lockers and showers, and company award/prize programs. Agency staff also recommended guaranteed ride home programs as a necessary supplement to successful carpooling strategies. Of the sixty-five employers that received advice and provided final cost information, about 88 percent (fifty-seven) reported a significant decrease in program implementation costs as a direct result of switching to the new options recommended by SCAQMD in 1997. The average annual cost per worksite declined from $35,616 to $16,043 (an average of $19,573 per worksite, or $70 per employee).

Programs aimed at the younger generation through grade school may achieve positive results over time. Children's programs are likely to yield a future generation that is educated about the economic, environmental, and social costs associated with transportation. Given the historic failures of regional travel programs and public resistance of pricing strategies, public awareness campaigns are becoming a major focus of U.S. regulatory agencies. It remains to be seen if these investments will prove cost-effective.

Economic Incentives

Economic incentives in transportation include monetary incentives or disincentives to the transportation consumer (i.e., vehicle operator or passenger) as encouragement to change travel behavior. Economists have long argued that monetary signals serve as the most economically efficient method to achieve changes in transportation demand. They argue that consumers will consume goods and services most efficiently when they are required to pay the full cost of the goods and services. Provided costs are not set too high, economic incentives should achieve behavioral change more efficiently than prescriptive rules. Economic incentives including congestion pricing and gasoline tax increases have received strong support from businesses, environmentalists, and local press editorials in San Francisco.

Transportation economics literature argues persuasively for the implementation of congestion pricing to improve the efficiency of the current transportation system. Area licensing schemes are also an option, where vehicle owners purchase stickers allowing the car to enter the most congested areas of the city during peak hours. Singapore, Hong Kong, and Oslo and Bergen in Norway provide examples of area license schemes.

The Environmental Defense Fund and Regional Institute of Southern California sponsored comprehensive modeling studies of pricing on transportation behavior for the Los Angeles area. The findings are as follows:

- Regional congestion pricing of $0.15 per mile may yield a vehicle miles traveled (VMT) reduction of about 5.0 percent and trip reduction of 3.8 percent;

- Regional \$3.00 per day parking charges may yield a VMT reduction of about 1.5 percent and a trip reduction of 1.8 percent;
- Regional \$0.60 per hour nonemployee parking charges may yield a VMT reduction of about 3.5 percent and trip reduction of 4.3 percent;
- Mileage and smog-based registration fees averaging \$110 per vehicle per year may yield a VMT reduction of about 0.4 percent and trip reduction of 0.7 percent.

Because of academic interest in congestion pricing theory, Congress established the Congestion Pricing Pilot Program in the Intermodal Surface Transportation Efficiency Act (ISTEA) of 1991 to fund congestion pricing studies and demonstration programs. Due to a lack of political support for congestion pricing on public roads, no significant congestion pricing programs have been implemented in the United States.

High occupancy toll (HOT) lanes are proving to be a potentially viable alternative to congestion pricing. HOT lanes allow single occupant vehicles to access new high occupancy vehicle lanes or facilities by paying a toll. HOT lane facilities are now operating in San Diego, California; Riverside, California; and Houston, Texas. Because the public perceives HOT lanes as providing new capacity and appears more accepting of tolls on these facilities, additional investigation into the consumer acceptance and economic benefits of HOT lanes will continue.

Gasoline Taxes. Studies in 1994 indicated that a \$2.00 increase per gallon in gasoline tax (raising United States prices from about \$1.00 to about \$3.00 per gallon) could yield a VMT reduction of about 8.1 percent and trip reduction of 7.6 percent. Determining the long-term effects of higher gasoline prices can be difficult. Fuel costs are a small component of the total cost of owning and operating an automobile. Research indicates that gasoline demand is relatively inelastic over the short-term and somewhat more elastic over the long-term. Significant increases in fuel price can affect short-term automobile use. However, when fuel prices rise significantly, demand for new fuel-efficient vehicles also increases. Individuals who purchase more fuel efficient vehicles can retain many of the trips and VMT without experiencing significant increases in total operating cost. Fuel price, VMT demand, and fuel intensity are interlinked, and the cumulative effect yields the change in net travel demand and net fuel consumption.

Parking Pricing. In the United States, paying for parking is the exception rather than the rule; 90 to 95 percent of auto commuters pay nothing for parking. Nationwide, employers provide 85 million free parking spaces to commuters, with a fair market value of nearly \$36 billion a year. Even in the central business district of Los Angeles, where parking fees are more common than in most areas, of the 172,000 office workers, more than 54,000 drivers park at their employer's expense.

Despite the fact that most employees pay nothing for parking, it is important to remember that there is no such thing as free parking. Even if employers do not pay an outside vendor and instead provide their own lots for employee parking, these lots must be constructed, gated, monitored, and maintained. Companies providing parking to employees at no charge pay an opportunity cost for not putting their property to more productive uses. The land could be developed and used in producing more company income or could be sold for development by others. Employers simply pass on the real and opportunity costs of parking to the consumers of the goods and services provided by the company. These consumers benefit little, if at all, from the parking provided to employees. Failure to charge employees for parking constitutes an inefficient pricing structure.

Travelers are highly sensitive to parking charges because the charges represent a large change in their out-of-pocket costs. Parking costs is one of the three most frequently cited factors (along with convenience and time saved) in the carpool decision. This responsiveness to parking prices is economically rational because motorists treat the vehicle purchase and annual insurance payments as sunk costs with respect to daily travel decisions, leaving parking costs as a large percentage of out-of-pocket expenses. For instance, typical commute trips to the Los Angeles core business district cost less than \$2.00 per day in gasoline costs. Adding the fair market value of parking to gasoline costs can increase out-of-pocket expenses to roughly \$6.00 per day. Studies show that free parking is a greater incentive to driving alone than providing free gasoline.

Various case studies lend support to the finding that parking prices are significant in affecting trip-generation and mode choice. An early study of employer-paid parking effects on mode choice, conducted in the late 1960s in the central business district of Los Angeles, examined county workers receiving employer-paid parking and federal employees paying for their own parking. The CARB (1987) study found that only 40 percent of the employees subject to parking fees drove to work alone, while 72 percent of similar employees that were not subject to parking fees drove alone. The availability of transit and alternative modes, and the amount of available free parking, influence the effectiveness of parking pricing. An analysis of thirteen employers in 1991 found that when they instituted paid parking programs the number of trips at the worksite was reduced by 20 percent. Other studies in the Los Angeles area indicate that between 19 percent and 81 percent fewer employees drive to work alone when they are required to pay for their own parking.

Employer-paid parking is changing in California because of a law requiring employers to "cash out" free parking. The 1992 California law, and subsequent regulations, require employers who provide free or subsidized offsite parking to their employees to provide a cash equivalent to those employees who do not use the subsidized parking (California Health and Safety Code 43845). The program applies to employers of more than fifty persons in areas that do not meet the state ambient air quality standards (between 3 percent and 13% of the medium and large employers in the Los Angeles area).

The California Air Resources Board (CARB) sponsored research in 1997 to evaluate eight case studies of firms that complied with California's parking cash-out. The number of drive-alone trips to these facilities dropped 17 percent after cashing out. The number of carpool participants increased by 64 percent; the number of transit riders increased by 50 percent; the number of workers arriving on foot or bicycle increased by 39 percent; and total commute trip vehicle miles of travel dropped by 12 percent. These findings are a revelation because this significant shift in travel behavior resulted from a regional policy.

Most businesses in Los Angeles have hesitated to implement employee parking pricing, despite the fact that increased parking fees can increase vehicle occupancy much more efficiently that other strategies. Most employees who receive free parking view this subsidy as a right, or fringe benefit of employment.

TRANSPORTATION SUPPLY IMPROVEMENT STRATEGIES

Transportation supply improvement strategies change the physical infrastructure or operating characteristics to improve traffic flow and decrease stop and go movements. They often require intensive capital investment and comprise bottleneck relief, capacity expansion, and construction improvements. Regional transportation plans usually include integrating high-occupancy vehicle (HOV) systems (differentiated from construction of HOV lanes). Many areas are also refocusing efforts on traditional technology-oriented means that reduce traffic congestion: signal timing optimization, rapid incident response, ramp metering, and applications of intelligent transportation system technology.

HOV Systems

Serious congestion delay often exists in the same urbanized areas that fail to meet federal air quality standards. In these areas, capacity expansion projects are usually difficult to implement. The emissions from the increased travel demand that follows corridor expansion can create additional air quality problems. One effective means of expanding capacity while restricting a corridor's travel demand and emissions growth is the addition of HOV lanes. Only carpools and transit vehicles can use these HOV lanes. The efficiency of carpool lanes as a system improvement is a function of the characteristics of the overall HOV system. Vehicles need to remain in free-flowing HOV lanes until they reach the exit for their destination or the time benefits associated with the carpooling activity are limited. Forcing carpools out of dedicated lanes, into congested lanes, and back into dedicated lanes is an impediment to carpool formation. HOV systems that ensure faster travel through congested regions are more successful in attracting users. Many urban areas are integrating HOV bypass facilities, and HOV onramps/offramps into interconnected systems.

Signal-Timing Optimization

Traffic signal-timing improvement is the most widespread congestion management practice in the United States. During the late 1970s and early 1980s, many cities and municipalities began focusing on improving signal timing as a means to reduce fuel consumption. Traffic engineers program traffic signal green, yellow, and red times at a local traffic control box located at the intersection. Signal-timing improvements can range from a simple change of a timing plan (such as increasing green time on one leg of an intersection during a peak period), to complex computer-controlled signal coordination along an

Carpool lanes, like this one on a California freeway near the Westwood neighborhood in Los Angeles, were designed to encourage more people per car. (Corbis-Bettmann)

entire transportation corridor. By linking control boxes at consecutive intersections and coordinating the green lights along a traffic corridor, vehicles moving in platoons can pass through consecutive intersections without stopping. In general, signal-timing programs reduce the number of stops, reduce hours of stop delay, and increase overall system efficiency. Consumers benefit from reduced congestion, increased safety, reduced stress, and improved response times for emergency vehicles.

Transportation engineers model signal-timing improvements using simulation programs and system optimization routines. In optimization programming, timing plans will be changed to purposely delay some vehicles, or platoons of vehicles, whenever such a delay has the potential of reducing congestion delay for many other vehicles on the road.

Numerous studies agree that signal-timing optimization programs are cost-effective. In 1986 California had an average first-year fuel use reduction of 8.6 percent from signal-timing programs. In 1980 the Federal Highway Administration's National Signal Timing Optimization Project examined the effectiveness of signal-timing improvements in eleven U.S. cities. The study found that these cities had reduced vehicle delay by more than 15,500 vehicle hours per year, stops per intersection by more than 455,000 per year, and fuel consumption by more than 10,500 gallons per year. Signal-timing optimization also has the potential to significantly decrease vehicle emissions in urban areas. Many cities do not place signal-timing improvement near the top of their annual resource priority list for transportation funding. Future programs that employ real-time signal controls (traffic responsive control strategies) will become widespread once program resources are made available by states and municipalities for the necessary technology and expertise to implement these systems.

Rapid Incident Response Programs

Roadside assistance programs detect and rapidly respond to crashes and breakdowns to clear the vehicles from the roadway as quickly as possible. Crashes often result in lane blockages that reduce capacity until the crash is completely removed from the system. Even a vehicle stalled on the shoulder of the road can create significant delays. After the incident is cleared, additional time is required for the congestion to clear. Reducing the duration of lane blockages by even a few minutes each can save thousands of hours of total congestion delay each year.

Many urban areas are now using video detection systems to monitor traffic flow. The systems, such as Autoscope, detect changes in video pixels to estimate traffic flow and vehicle speeds on freeways and major arterials. Significant differences in average speeds from one monitor to another indicate the presence of an accident between video monitoring locations. With early detection of an incident, rapid accident response teams (either roving the system or staged in specific locations along the system) can be dispatched immediately. Cellular telephones will also play an increasing role in incident detection programs as the cellular network density increases.

A recent study of the Penn-Lincoln Parkway indicated that the annual program investment of $220,000.00 for their roadside assistance program reduces congestion delay by more than 547,000 hours per year (and a benefit-cost ratio of 30:1).

Ramp Metering

Major urban areas in the United States have begun implementing ramp metering to optimize performance of major freeway systems. A ramp meter brings vehicles on the onramps entering the freeway to a complete stop during peak periods and releases vehicles to the freeway one or two at a time. Metering prevents queues of vehicles from entering the traffic stream and disrupting freeway traffic flow. Ramp meters delay the onset of congested stop-and-go activity, allowing freeways to continue to operate at efficient flow rates for extended periods. Ramp metering delays vehicles arriving at the ramps, but the net congestion savings on the freeway more than offsets the ramp delay. Although emissions are significantly lower on the freeway systems when metering maintains smooth traffic flow, the net emissions benefits of such systems are not clear at this time. The hard accelerations associated with vehicles leaving the stop bar on the ramp result in significantly elevated emissions levels. New ramp metering research projects are trying to determine the systemwide tradeoffs in emissions.

Intelligent Transportation Systems

A variety of intelligent transportation systems (ITS) with the potential to improve traffic flow and relieve congestion are currently being explored. ITS technologies range from simple delivery of traffic information into the vehicle (helping drivers optimize route choice decisions) to complete vehicle automation. Although many of the proposed ITS strategies require development of new roadway and electronic infrastructure systems, some new technologies will evolve over the near term. Drivers can expect to access, via cellular Internet connections, detailed transportation system performance data. Electronic in-vehicle maps and systems data will allow users to optimize travel decisions and avoid getting lost. New radar systems will transmit warnings from roadside systems to vehicles, giving construction zones a new level of protection. Onboard collision avoidance systems may also prevent a significant number of rear-end crashes that cause congestion.

RELATIVE EFFECTIVENESS OF DEMAND AND SUPPLY STRATEGIES

Due to political changes regarding transportation control measures (TCMs), the CARB shifted from advocating TCMs in the late 1980s to evaluating their cost effectiveness by the mid 1990s. In 1995 they studied the cost effectiveness of twenty emissions reduction strategies funded by California vehicle registration fees. The CARB analyses indicate that signal timing, purchases of new alternative fuel vehicles, and construction of bicycle facilities can be cost effective compared to many new stationary source control measures. A videophone probation interview system designed to substitute for office trips resulted in an agency cost savings, indicating that there are still technology projects that can simultaneously attain emissions reductions and save public dollars. Of the twenty measures, eleven tried to change travel demand, and nine tried to improve traffic flow or to shift drivers to new vehicles or alternative fuels. Only six of the eleven demand management measures proved cost-effective. The technology-oriented supply improvement strategies (such as signal timing) and fuel shifts fared better, with eight of the nine measures proving cost-effective.

MODELING THE EMISSIONS IMPACTS OF TRAFFIC FLOW IMPROVEMENTS

Experts believe that current emissions models overestimate emissions at low operating speeds. They predict emissions solely as a function of average speed. However, emissions are a strong function of speed/acceleration operating conditions. Modal emission rate models demonstrate that emissions in grams/second are fairly consistent when the vehicle is operating at low to moderate speed with low acceleration rates. Under these conditions, the gram/mile emissions rates are a function of the constant gram/mile rate and the amount of time the vehicle spends on the road segment in question (which is a function of average speed). Emissions can skyrocket under high-speed conditions and conditions of moderate speed with high acceleration and deceleration rates. The operating condition corresponding to the lowest emissions and reasonably efficient fuel consumption is smooth traffic flow, probably between 20 to 35 mph, with little acceleration and deceleration.

Evaluation of traffic flow improvement projects using the current emission rate modeling regime is bound to underestimate the air quality benefits. Because emissions are activity-specific, it is important to develop methods that can estimate the effect of traffic flow improvements on trip emissions in terms of changes in speed/acceleration profiles. The Georgia Institute of Technology and University of California at Riverside are currently developing emissions models that can assess the benefits of traffic flow improvements. The application of these modal emissions models is expected to show that improved signal timing may have a pronounced impact on improving emissions.

CONCLUSION

Evidence suggests that demand management initiatives have had relatively small impacts on travel behavior and fuel consumption in the United States. Direct agency intervention at the regional level has not worked as intended. During the 1990s, economic incentives designed to internalize the personal and social costs of the automobile seemed to be the most logical and promising ways of achieving changes in travel behavior. As consumers internalize the true costs of owning and operating the personal automobile, individual tripmaking decisions become more rational, increasing system efficiency. Strategies such as congestion pricing, emission fees, and even pay-as-you drive automobile insurance received a great deal of attention in state legislatures and in the popular press. Before the public widely accepts pricing strategies in the United States, regulators need to address a variety of equity dilemmas. In an era of cheap and abundant energy, it is likely that such pricing arguments will focus on potential air quality and congestion benefits rather than on energy consumption benefits.

The most successful transportation-related economic incentive to date has been the parking cash-out program in California. Limited-scale implementation has been very successful. Regional parking pricing is likely to be a viable travel demand strategy, but will be difficult to adopt. Tax codes that allow employers to provide parking as a tax-exempt employee benefit would need to change. Employers and employees need to feel that the costs of parking are real. By implementing pricing strategies that the public will support, and simultaneously continuing public education campaigns explaining the energy and environmental problems associated with owning and operating an automobile, gradual acceptance of more widespread incentives might be achieved.

Despite the implementation and performance limitations of demand management strategies, metropolitan areas have been able to significantly improve traffic flow by implementing of transportation system improvement strategies. Lane expansion projects that provide bottleneck relief have become the focus of new roadway construction. Traffic signal timing optimization, rapid incident response systems, and implementation of HOV systems and high-occupancy toll lanes have significantly enhanced system capacity and reduced congestion. These strategies are capital intensive, but the returns on investment in congestion relief and fuel and emissions savings have been significant. The public seems much more willing to expend resources to enhance capacity than to endure less personal mobility imposed by demand management strategies.

Because of continued increases in vehicle ownership, tripmaking, and mileage, many researchers question the extent to which system improvement projects can mitigate congestion growth. As congestion worsens in major urban areas, there will be a greater focus on demand management strategies. In addition, land use strategies are getting more attention. Transit trips typically account for only 3 percent of the total trips made in the United States. Only in cases where transit is available, where origin and destination land use characteristics match desired tripmaking characteristics, and where transit costs (transit time, wait time, fares, and opportunity) are lower than automobile costs, do Americans opt to commute by transit. Proper design of land use, at densities that support transit operation and provide more opportunity to walk to common destinations, can provide significant long-term changes in travel patterns that minimize congestion, fuel consumption, and motor vehicle emissions. Because land use decisions come under local jurisdiction, federal agencies are now developing guidance documents for state and local agencies.

Randall Guensler

See also: Air Pollution; Emission Control, Vehicle; Energy Management Control Systems; Government Intervention in Energy Markets.

BIBLIOGRAPHY

Adler, K.; Grant, M.; and Schroeer, W. (1998). *Emissions Reduction Potential of the Congestion Mitigation and Air Quality Improvement Program: A Preliminary Assessment*. Transportation Research Record 1641. Transportation Research Board. Washington, DC: National Research Council.

Bay Area Economic Forum. (1990). *Congestion Pricing—The Concept*. Oakland, CA: Author.

California Air Resources Board. (1991). *Employer-Based Trip Reduction: A Reasonably Available Transportation Control Measure*. Executive Office. Sacramento, CA: Author.

California Air Resources Board. (1995). *Evaluation of Selected Projects Funded by Motor Vehicle Registration Fees, Revised*. Technical Support Document. Sacramento, CA: Author.

Cambridge Systematics, Inc. (1991). *Transportation Control Measure Information Documents*. Prepared for the U.S. Environmental Protection Agency Office of Mobile Sources. Cambridge, MA: Author.

Cambridge Systematics, Inc.; COMSIS Corporation; K. T. Analytics; Deakin, E.; Harvey, G.; and Skabardonis, A. (1990). *Transportation Control Measure Information* (December Draft). Prepared for the U.S. Environmental Protection Agency, Office of Mobile Sources. Ann Arbor, MI: Author.

Chatterjee, A.; Wholley, T.; Guensler, R.; Hartgen, D.; Margiotta, R.; Miller, T.; Philpot, J.; and Stopher, P. (1997). *Improving Transportation Data for Mobile Source Emissions Estimates* (NCHRP Report 394). Transportation Research Board. Washington, DC: National Research Council.

COMSIS Corporation. (1994). *Implementing Effective Employer-Based Travel Demand Management Programs*. Institute of Transportation Engineers. Washington, DC: Author.

Dahl, C., and Sterner, T. (1991). "Analyzing Gasoline Demand Elasticities: A Survey." *Energy Economics* 13(3):203–210.

Deakin, E; Skabardonis, A; and May, A. (1986). *Traffic Signal Timing as a Transportation System Management Measure: The California Experience*. Transportation Research Record 1081. Transportation Research Board. Washingto, DC: National Research Council.

Donnell, E.; Patten, M.; and Mason J. (2000). *Evaluating a Roadside Assistance Program: Penn-Lincoln Parkway Service Patrol*. Transportation Research Record 1683. Transportation Research Board. Washington, DC: National Research Council.

Ferguson, E. (1989). *An Evaluation of Employer Ridesharing Programs in Southern California*. Atlanta, GA: Georgia Institute of Technology.

Giuliano, G. (1992). "Transportation Demand Management: Promise or Panacea?" *APA Journal* (Summer):327–334.

Giuliano, G.; Hwang, K.; Perrine, D.; and Wachs, M. (1991). *Preliminary Evaluation of Regulation XV of the South Coast Air Quality Management District*. Working Paper No. 60. The University of California Transportation Center. Berkeley, CA: University of California.

Gordon, D., and Levenson, L. (1989). *DRIVE+: A Proposal for California to Use Consumer Fees and Rebates to Reduce New Motor Vehicle Emissions and Fuel Consumption*. Berkeley, CA: Lawrence Berkeley Laboratory, Applied Science Division.

Guensler, R. (1998). *Increasing Vehicle Occupancy in the United States. L'Avenir Des Deplacements en Ville (The Future of Urban Travel)*. Lyon, France: Laboratoire d'Economie des Transports.

Guensler, R., and Sperling, D. (1994). "Congestion Pricing and Motor Vehicle Emissions: An Initial Review." In *Curbing Gridlock: Peak Period Fees to Relieve Traffic Congestion*, Vol. 2. Washington, DC: National Academy Press.

Hallmark, S; Bachman, W.; and Guensler, R. (2000). *Assessing the Impacts of Improved Signal Timing as a Transportation Control Measure Using an Activity-Specific Modeling Approach*. Transportation Research Record. Transportation Research Board. Washington, DC: National Research Council.

Hartgen, D., and Casey, M. (1990). "Using the Media to Encourage Changes in Travel Behavior" (890423). *Proceedings of the 1990 Annual Meeting of the Transportation Research Board*. Washington, DC: National Research Council.

Harvey, G., and Deakin, E. (1990). *Mobility and Emissions in the San Francisco Bay Area*. Prepared for the Metropolitan Transportation Commission. Oakland, CA: Author.

Harvey, G. (1994). "Transportation Pricing and Travel Behavior." In *Curbing Gridlock: Peak Period Fees to Relieve Traffic Congestion*, Vol. 2. Washington, DC: National Academy Press.

Haug International and LDA Consulting. (1998). *SB836 Evaluation of Rule 2202 Voluntary Replacement Measures*, Final Report. Presented to the SB 836 Oversight Committee. Los Angeles, CA: Haug International.

Hirst, E. (1978). "Transportation Energy Conservation Policies." In *Energy II: Use Conservation and Supply*, eds. P. H. Abelson and A. L. Hammond. Washington, DC: American Association for the Advancement of Science.

Kain, J. (1994). "The Impacts of Congestion Pricing on Transit and Carpool Demand and Supply." In *Curbing Gridlock: Peak Period Fees to Relieve Traffic Congestion*, Vol. 2. Washington, DC: National Academy Press.

Knapp, K.; Rao, K.; Crawford, J.; and Krammes, R. (1994). *The Use and Evaluation of Transportation Control Measures*. Research Report 1279-6. Texas Transportation Institute. Texas: Federal Highway Administration and Texas Department of Transportation.

Meyer, M. (1997). *A Toolbox for Alleviating Traffic Congestion and Enhancing Mobility*. Washington, DC: Institute of Transportation Engineers.

Mohring, H., and Anderson, D. (1996). "Congestion Costs and Congestion Pricing." In *Buying Time Symposium*. Hubert H. Humphrey Institute of Public Affairs. Minneapolis, MN: University of Minnesota.

Public's Capital, The. (1992). "The Rising Tide of User Fees." In *Governing* (April). Washington, DC: Author.

Robinson, F. (1996). "Selected Findings from the Twin Cities Congestion Pricing Study." In *Buying Time Symposium*. Hubert H. Humphrey Institute of Public Affairs. Minneapolis, MN: University of Minnesota.

Schreffler, E.; Costa, T.; and Moyer, C. (1996). *Evaluating Travel and Air Quality Cost-Effectiveness of Transportation Demand Management Projects*. Transportation Research Record 1520. Transportation Research Board. Washington, DC: National Research Council.

Shoup, D. (1997). "Evaluating the Effects of Cashing Out Employer-Paid Parking: Eight Case Studies." *Transport Policy* 4(4).

Shoup, D., and Breinholt M. (1997). "Employer-Paid Parking: A Nationwide Survey of Employers' Parking Subsidy Policies." In *The Full Costs and Benefits of Transportation*, eds. D. L. Greene, D. Jones, and M. A. Delucchi. Berlin: Springer.

Sierra Club. (1990). *Heading the Wrong Way: Redirecting California's Transportation Policies*. Sacramento, CA: Author.

South Coast Air Quality Management District. (1998). *Rule 2202—On-Road Motor Vehicle Mitigation Options Status Report*. Diamond Bar, CA: Author.

Spock, L. (1998). *Tolling Practices for Highway Facilities*. Synthesis of Highway Practice 262. National Cooperative Highway Research Program (NCHRP). Transportation Research Board, National Research Council. Washington, DC: National Academy Press.

Ullberg, C. (1991). "Parking Policy and Transportation Demand Management." Proceedings of the 84th Annual Meeting of the Air and Waste Management Association. Pittsburgh, PA.

U.S. Department of Transportation. (1997). *1995 National Personal Transportation Survey Data Files* (FHWA-PL-97-034). Federal Highway Administration. Washington, DC: Author.

U.S. Environmental Protection Agency. (1998). *Assessing the Emissions and Fuel Consumption Impacts of Intelligent Transportation Systems (ITS)*. Report Number 231-R-98-007. Washington DC: Author.

U.S. Environmental Protection Agency and U.S. Department of Transportation. (1998). *It All Adds Up to Cleaner Air*. Pilot Site Resource Kit. USEPA, Office of Mobile Sources, Transportation Air Quality Center. Ann Arbor, MI: Author.

Wachs, M. (1991). "Transportation Demand Management: Policy Implications of Recent Behavioral Research." *Journal of Planning Literature* 5(4)333–341.

Wilson, R., and Shoup, D. (1990). "Parking Subsidies and Travel Choices: Assessing the Evidence." *Transportation* 17:141–157.

Winston, C., and Shirley, C. (1998). *Alternate Route: Toward Efficient Urban Transportation*. Washington, DC: Brookings Institution Press.

Winters, P.; Cleland, F.; Pietrzyk, M.; Burris, M; and Perez, R. (2000). *Predicting Changes in Average Vehicle Ridership on the Basis of Employer Trip Reduction Plans*. Transportation Research Record 1682. Transportation Research Board. Washington, DC: National Research Council.

Zavattero, D.; Ward, J.; and Rice, D. (2000). *Analysis of Transportation Management Strategies for the 2020 Regional Transportation Plan*. Transportation Research Record 1682. Transportation Research Board. Washington, DC: National Research Council.

TRAINS

See: Freight Movement; Mass Transit

TRANSFORMERS

A transformer is an electrical component used to connect one alternating current (ac) circuit to another through the process of electromagnetic induction. The input current travels through a conductor (the primary) wound around a conductive core. The current traveling through the primary windings creates an alternating magnetic field in the core. An output conductor (the secondary) is wound around the same core so that the magnetic

Power transformer station in South Carolina. (Corbis-Bettmann)

field cuts through the secondary windings, inducing the output electrical current. For most transformers, the primary and secondary windings never come into direct electrical contact with each other. Instead, the transfer of energy from primary to secondary is accomplished solely through electromagnetic induction.

Transformers were developed through a series of scientific discoveries in the nineteenth century. Most notably, Michael Faraday showed in 1831 that a variable magnetic field could be used to create a current, thus pioneering the concept of electromagnetic induction. It was not until the 1880s that Nikola Tesla was able to use this principle to bolster his patents for a universal ac distribution network.

The majority of power transformers change the voltage from one side of the transformer to the other. The change in voltage is directly related to the number of turns each conductor (primary and secondary) makes around the transformer's core. For example, if the primary makes ten turns around the core, and the secondary makes five turns around the core, the secondary voltage will be half of the primary voltage. This type of transformer would be called a step-down transformer, since it steps down the voltage. On the contrary, if the number of turns in the primary is less than the number of turns in the secondary, the transformer will be a step-up transformer.

The power transformer also must maintain a balance of power from one side to the other. Power is the product of voltage and current. Therefore, neglecting any internal losses, if a transformer steps up the voltage by a given factor, it will also step down the current by the same factor. This has great application in the generation and transmission of electricity. It is difficult to generate extremely high voltages of electricity. Generating plants produce electricity at a relatively low voltage and high current, and step the voltage up to transmission levels through a transformer. This has the added effect of reducing transmission losses, since as the voltage increases, the current decreases, thereby reducing resistive voltage drops.

Iron (or steel) is frequently used as a transformer core because it provides adequate magnetic flux at a relatively low reluctance. In other words, iron cores, however, require additional design considerations to prevent excessive power losses and heat dissipation. The changing magnetic field inside the core creates current not only in the secondary windings of the transformer but also within the core itself. Using a laminated core constructed from small laminated plates stacked together and insulated from each other reduces these losses.

Other types of transformers include instrument, radio-frequency, wide-band, narrow-band, and electronic transformers. Each of these transformers operates similarly and is used in specific applications best suited for the transformer's design characteristics.

Brian F. Thumm

See also: Electric Motor Systems; Electric Power, Generation of; electric Power Transmission and Distribution Systems; Faraday, Michael; Magnetism and Magnets; Tesla, Nikola.

BIBLIOGRAPHY

Faraday, M. (1965). *Experimental Researches in Electricity*. New York: Dover Publications.

Fink, D. G., and Beaty, H. W., eds. (1993). *Standard Handbook for Electrical Engineers*, 13th ed. New York: McGraw-Hill.

TRANSIT

See: Mass Transit

TRANSMISSIONS

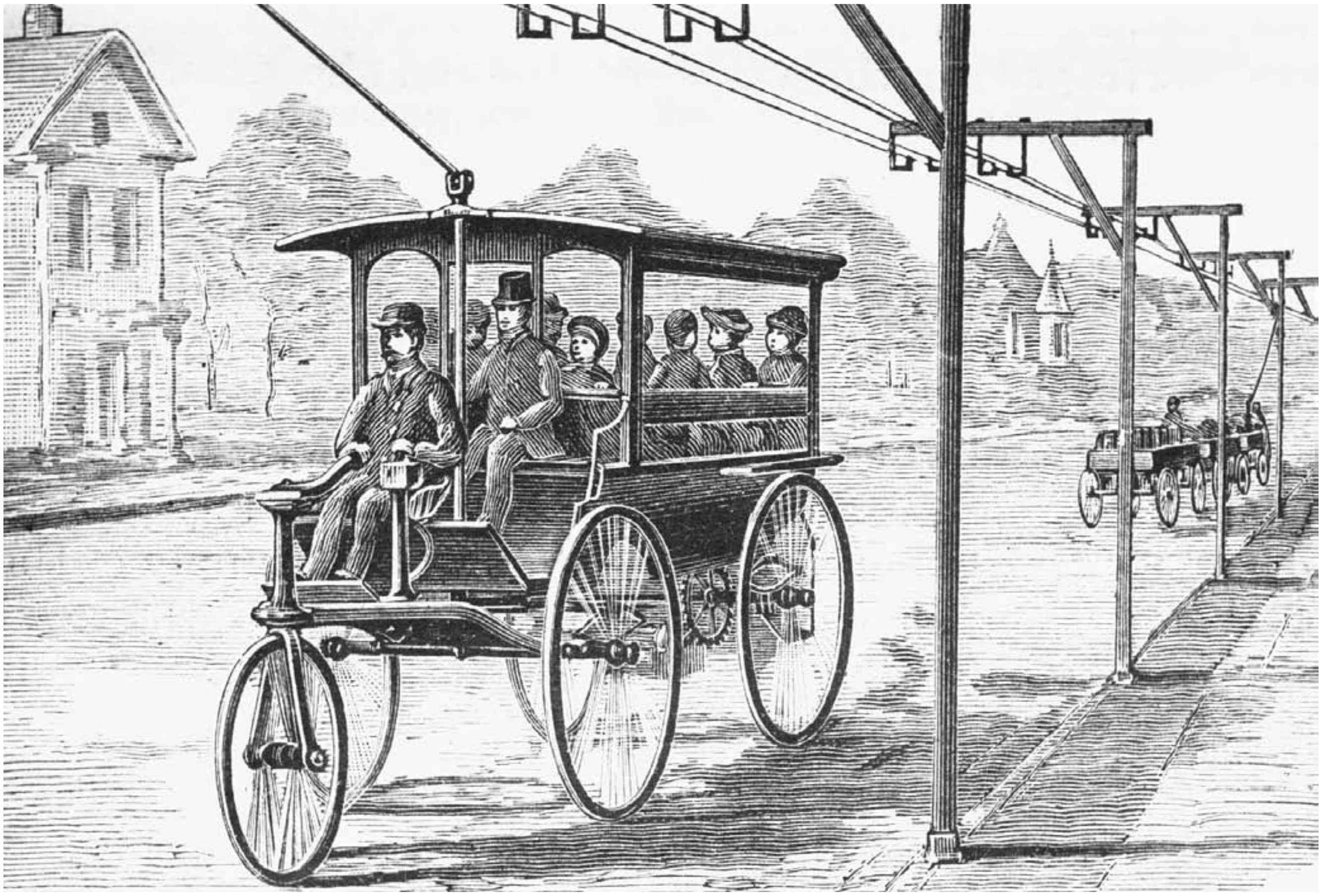
See: Drivetrains

TRANSPORTATION, EVOLUTION OF ENERGY USE AND

Transportation is energy in motion. Transportation, in a fundamental sense, is the application of energy to move goods and people over geographic distances. Freight transportation may be regarded as part of complex logistical and/or distribution systems that carry needed materials to production facilities and finished goods to customers. Transportation of passengers can serve long-distance travelers, daily commuters, and vacationers, among others. Special systems accommodate the requirements people have for mobility throughout the day. A wide variety of specific technologies and management systems are involved, and from the mid-nineteenth century, a large portion of the world's energy supply has been devoted to transportation.

When sailing vessels predominated at sea and when horses or draft animals provided basic land transport, energy demands by transport systems were small. With the advent of increasingly reliable steam power for river, lake, and ocean vessels beginning in the late 1820s, together with the spread in the 1830s of steam railways running long distances in Europe and North America, demand rose sharply in the industrialized countries for wood and coal as fuel. Coal became the fuel of choice for ocean transport as steam gradually displaced sail from the mid- to the late nineteenth century, since far more Btus per volume of fuel could be stored in the limited space of fuel bunkers.

Nearly all locomotives in Britain, and many on the Continent, used coal as well. In North America, wood—because of its plentiful supply—was the principal fuel for locomotives and for western and southern riverboats through the late 1860s. Locomotives of some railroads in Pennsylvania burned anthracite—”hard” coal. Most of the earliest coal production in the United States, starting in the 1820s, was anthracite, which was slow-burning but produced very little smoke. That virtue made anthracite the preferred type of coal for home heating and for industries within large urban centers such as Philadelphia and New York, to which the costs of transporting wood were higher than to smaller cities and towns closer to their wood supplies.



Electric propulsion street cars. (Corbis Corporation)

Railway use of wood as fuel in the United States peaked at almost 3.8 million cords per year in the 1860s and fell slowly thereafter. From the 1850s, the use of coal—preponderantly the more common type, called bituminous or “soft” coal—increased on railways and on waterways, especially in the East and Midwest as coal production in those regions accelerated. (A ton of good bituminous equaled the heating value of 1¾ to 2 cords of seasoned hardwood.) Two major factors influenced the broad conversion to coal: First, as the quantity mined increased, prices fell—from $3.00 per ton in the 1850s to about $1.00 per ton to high-volume purchasers just ten years later. Second, the country’s voracious appetite for wood—for domestic and industrial fuel, for charcoal used in ironmaking, and for construction—had denuded vast stretches of the eastern forests by 1880. In that year U.S. railroads consumed 9.5 million tons of coal, which accounted for about 90 percent of their fuel supply.

Electricity, generated primarily from coal, became widely used for transport within and near cities, as trolleycar systems proliferated after the late 1880s. The enabling developments were, first, successful forms of large-scale electric generation from steam-driven dynamos and, second, a practical method of electricity distribution to the railcars. Most such cars took their power from an overhead wire, with electrical grounding through the track. In many parts of the world, trolley systems became a fixture of urban transportation in large and medium-size cities. Trolley companies usually owned their own power-generation stations and often provided a given city with its first electrical distribution network. Trolley companies were therefore eager to sell electricity and often sponsored efforts to spread the use of electric lighting systems and other electric appliances. Elevated urban railways, powered at first by tiny, coal-fired steam locomotives in the 1880s and converted later to electricity, also helped ease urban congestion in New York and Chicago. A related development by 1900 in a few of the world's major cities was the electric-powered subway.

Especially in North America, a direct effect of the electric trolley—from about 1890 through the 1920s—was the rapid growth of suburbs. In some cases, trolley systems that extended beyond city boundaries accelerated the expansion of existing suburbs. In many other cases, new suburbs sprang up, facilitated by the cheap mobility offered by newly built trolley lines. Long-distance, electric interurban rail systems also grew at a fast rate in the United States after 1890. Throughout New England, the East, and the Midwest, and in many places in the South and West, travelers and longer-distance commuters could use electric railcars connecting small towns with medium-size and large cities, over distances of up to a hundred miles or so.

Since the mid-twentieth century, petroleum has been the predominant fuel stock for energy used in transportation on land, water, and in the air. The shift began slowly. Oil-fired boilers for ships were tried in the 1860s. A railway in Russia fired more than a hundred of its steam locomotives on oil in the 1880s. By the first decade of the twentieth century—and although coal was still the favored fuel for steam transportation throughout the world—several thousand steam locomotives in the western part of the United States as well as a few ocean vessels were fired by residual oil from a rapidly growing number of refineries or, in some cases, by light petroleum.

Regional scarcities of coal initially drove these uses. As petroleum became more abundant and as its price fell, oil became more attractive. In firing boilers, fuel oil possessed only a slight advantage over good-quality coal in Btus per unit volume. But liquid fuels were much easier to handle and store than coal. Competitive pressures kept the prices per Btu of residual oil and coal quite close.

The transition to oil for the boilers of oceangoing steamships was well along by the 1920s. A coincidental development in the early part of the century was the steam-turbine engine for ships, which mostly displaced multicylindered steam piston engines in marine applications; the turbine is more energy-efficient at constant speeds maintained over long periods of time. A decreasing number of older riverboats, towboats, lake boats, and freighters continued to use coal through the 1950s (and a rare few such steam vessels were still in use in the 1980s). As late as 1957, U.S. railroads purchased $48 million worth of bituminous coal for steam locomotive fuel—and $374 million of diesel fuel. By 1960, all use of steam locomotives on major lines in the United States and Canada had ended, and by 1970 the change to diesel locomotives was virtually complete worldwide, except in some parts of China, India, Africa, and Eastern Europe. Thus the use of coal as a transport fuel ended for the most part in the late twentieth century, except for electricity (generated from coal) used on a small percentage of the world's railway mileage.

The burgeoning growth in the use of petroleum-based fuels in transportation came with the widespread use of automobiles and trucks after 1910 and the beginning of air travel after World War I. In the 1890s, little gasoline was consumed. It was regarded as an explosively dangerous by-product of the refining process in the production of kerosene (a petroleum distillate) for lighting. But with the invention of practical automobiles in the 1890s in the United States and Europe, demand slowly grew for a high-energy fuel that could be used readily in small, internal-combustion engines. In such use, the high Btu per unit volume of gasoline was a distinct advantage. The story of the huge oil companies and gasoline distribution systems that arose in the United States and overseas in the early twentieth century is one driven entirely by the rise of the automobile.

During the first two decades of the twentieth century, several hundred firms in Britain attempted automotive production, and nearly 3,000 tried to do so in the United States. But there was one car that created much of the rising demand for gasoline: the Ford Model T. Between 1908 and 1927, 15 million Model Ts rolled out. Ford built half of the automobiles manufactured in the world in 1920. The Model T's low price and high reliability created a vast pool of willing purchasers. A result was to sweep away the economic incentives that might have supported a viable market for alternative fuels or motors for automobiles, such as electric or steam propulsion. Versus the electric, a gasoline engine provided much greater operating range; versus steam (such as the compact boiler and engine used in the temporarily popular Stanley steamer), a gasoline engine was much simpler and less expensive.

After 1910, gasoline-powered trucks became common. Their utility eventually replaced horse-drawn vehicles in farm-to-market and intracity use, though that replacement took decades. Even in the industrial world, horses provided a large but declining share of intracity freight transport through at least the 1930s and early 1940s. Highway buses became popular in the 1920s. Such buses quickly displaced the long-distance, electric interurban rail systems of the 1890s and early 1900s.

Rudolph Diesel first patented his engine in 1892, but it was not until 1910 that a successful application was made in a vessel. The first diesel engines suitable for use in trucks came in the early 1920s. Annual U.S. production of diesels did not exceed one million horsepower until 1935. Over the next two years that production doubled, largely because of railroad locomotive applications. Due to the high compression pressures a diesel engine must withstand in its cylinders, it must be built heavier than a gasoline engine of similar output. But the diesel's advantage is that it is the most energy-efficient internal-combustion engine yet developed. Marine diesel design advanced in the 1930s and during World War II, leading to greater use in larger transport vessels from the late 1940s. The demand for high-quality diesel fuel—similar to kerosene—thus expanded.

With the inception of the U.S. interstate highway system in the late 1950s, the modern long-distance highway tractor and its semitrailer evolved and became the predominant method of freight transportation for time-sensitive, high-value goods. Elsewhere in the world, different forms of the long-distance truck took over most freight transport. Today, in the United States and elsewhere, these trucks are almost universally diesel. In the United States in the 1990s, some 39 percent of intercity ton-miles of commercial freight were borne by rail, 28 percent by trucks, and 15 percent on lakes, rivers, and canals (the rest was carried by pipelines and a fraction of 1% by air). Diesel fuel powered nearly all this road, rail, and water transport.

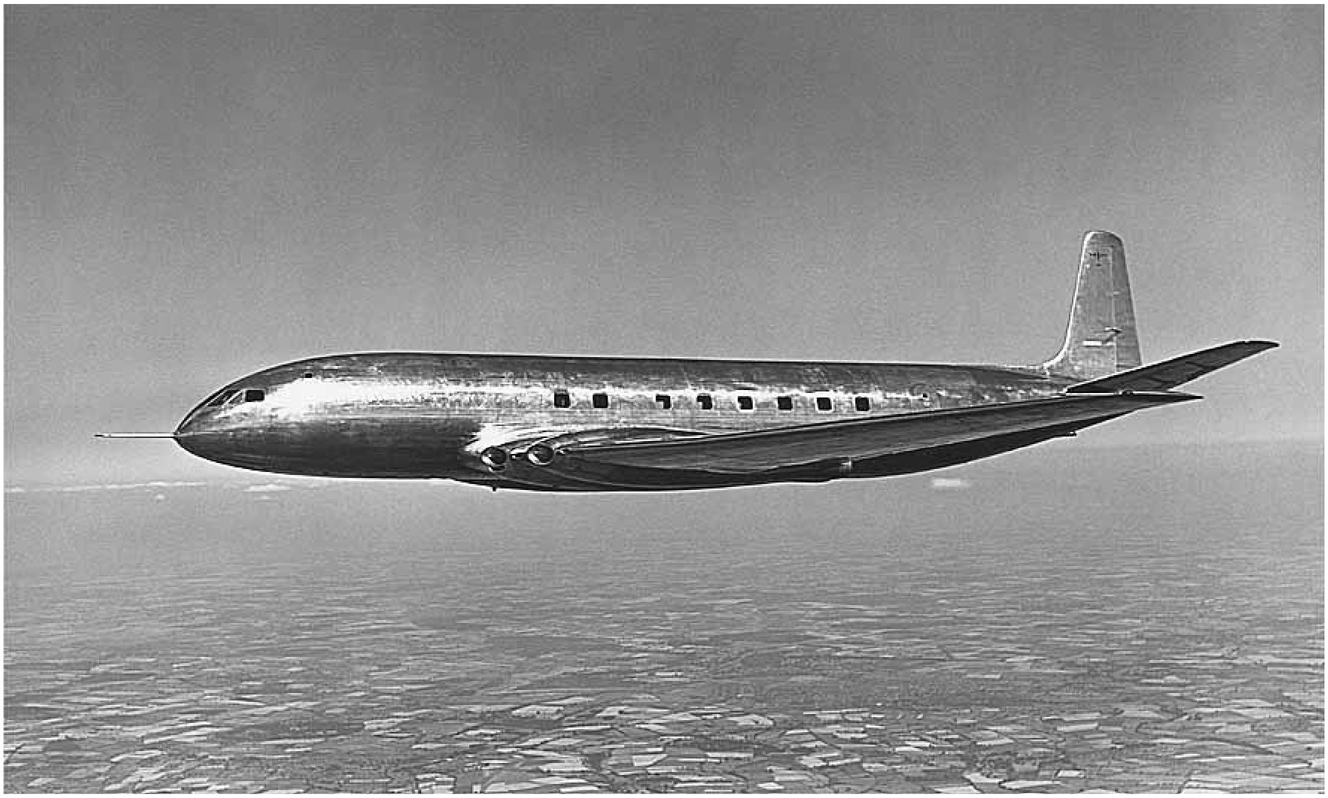
An aspect of transportation that has a heavy bearing on energy consumption is horsepower. Transport can be measured in units of *work*—tons hauled per mile, for example—but energy consumption is more proportionately related to *power*, such as tons per mile *per unit of time*. For a given load carried over a given distance, higher speed requires more power and hence more energy. Thus there is a trade-off especially important in all forms of transportation: Fuel-saving strategies often involve reduced speeds, while increased transport capacities of systems per unit of time often necessitate higher speeds and therefore require higher fuel use per ton-mile or per seat-mile.

Aviation has been powered from its beginnings by gasoline. Only spark-ignited gasoline engines proved able to provide the extremely high power-to-weight ratios needed by aircraft engines, where weight is always at a premium. During World War II, German and British engineers developed the gas-turbine engine for aircraft (known commonly as the jet engine). Such engines ran best on a variation of kerosene. In the "turbojet" form of the gas-turbine, there is no propeller; in the "turboprop" form, the central turbine shaft of the engine turns a propeller through a gearcase.

In transport applications in the 1950s, the jet's favorable power-to-weight ratio and efficiency at high altitudes resulted in a travel revolution. (The first commercial jet, the de Havilland Comet, first flew with paying passengers in 1952; the pioneering Boeing 707 came a few years later.) Speed was only one factor. Flying at more than 30,000 feet—well above the turbulent weather systems ordinarily encountered by piston-engined aircraft, which operate most efficiently up to about 20,000 feet or below—the comfort level of jets allowed millions of people to accept these new planes as long-distance transport.

Turboprop types of aircraft proved more reliable and cheaper to operate than piston-engined airplanes for short hauls and more fuel-efficient than turbojets in such use. Overall airline patronage shot up dramatically after 1960. Since the 1980s, air express has boomed as well. Thus most aviation fuel used today is well-distilled kerosene, with a small portion of antifreeze for the extremely low ambient temperatures at high altitude. High-octane gasoline powers much of the "general aviation" sector of private aircraft, where piston engines are still the norm for planes of up to four or five passengers.

During the oil shortages of the 1970s and early 1980s, researchers in the United States and Europe investigated alternative fuels for transportation. Shale oil deposits were tapped, and other projects experimented with the conversion of various grades of coal to oil (a technology developed by Germany in World War II). Two U.S. railroads (the Burlington Northern and the Chessie System) seriously considered a new-technology steam locomotive that would burn coal in a gasification furnace, thus controlling emissions.



Test flight of the De Havilland Comet 1 jet transport prototype. (Corbis-Bettmann)

Meanwhile, diesel manufacturers experimented with modified engines burning blends of finely pulverized coal mixed in fuel oil. While these engines ran reasonably well, tiny amounts of noncarbon impurities in the pulverized coal inevitably built up within the engines' cylinders, causing rough running or damage. Stabilizing oil prices from 1982 through 1999 removed most incentives for further research.

From 1980 to 1998, total U.S. petroleum consumption ranged between 15 million and 19 million barrels daily, with consumption at the end of the 1990s about the same as in the peak of 1978. Use of petroleum for U.S. transportation was slightly less than 10 million barrels per day in 1980 and was about 12 million barrels in 1997, or some 60 percent of total domestic oil consumption. That oil consumption rate supported a growth in U.S. freight of about 20 percent from 1980 to the late 1990s, amounting in 1997 to a bit less than 15,000 ton-miles of freight transported per person per year.

Travel has also multiplied. The U.S. Departments of Commerce and Transportation surveyed long-distance trips in 1977 and 1995 (with a long-distance trip defined as a round-trip of at least a hundred miles each way). Between those years, long-distance trips per person per year by automobile grew by almost 90 percent; by air, such trips nearly trebled. Increased mobility is a central feature of modern life.

All modes of transport—by road, rail, water, and air—have increased engine fuel efficiency and have instituted other fuel-saving strategies since the oil shortages of the 1970s. New turbine-engine aircraft have doubled the average fuel efficiency of earlier jets. On railroads, a gallon of diesel fuel moved 235 ton-miles of freight in 1980; in the early 1990s, a gallon moved more than 360 ton-miles. Average automobile fuel mileages have improved. Redesigned hull shapes and large, higher-efficiency diesel engines have lowered fuel costs in maritime transport.

New vehicles, propulsion systems, and fuels are on the horizon: "hybrid" automobiles (combining an electric motor, batteries, and a gasoline or diesel engine), better electric cars, greater use of compressed natural gas (CNG) and propane for urban fleet vehicles such as taxis and delivery trucks, fuel cells (using gasoline, natural gas, or propane as the feedstock for hydrogen), methanol or ethanol to supplement gasoline, and other technologies are advancing. For such technologies and fuels successfully to compete with oil in the transportation market, the techniques to liquefy and compress natural gas would need to become cheaper, new types of storage batteries must be made practical, and the necessary infrastructure to support CNG-, fuel-cell-, and electric-powered cars and trucks would need to be developed. The huge worldwide investment in petroleum extraction, refining, and distribution makes economic change in fuel technologies difficult.

In Europe, Japan, and the United States, newer high-speed electric railroad trains are being advanced. Japan has shelved its development since the 1980s of magnetic-levitation trains, but German firms continue work on such systems for possible application in Germany (a "maglev" line has been planned between Berlin and Hamburg) and elsewhere. Engine development for aircraft has recently cut the number of turbojet engines required for a long-distance, high-capacity transport airplane from three or four to two, cutting fuel consumption per seat-mile, and further fuel efficiency increases can be expected in the years ahead.

William L. Withuhn

See also: Aircraft; Air Travel; Automobile Performance; Aviation Fuel; Diesel, Rudolph; Diesel Cycle Engines; Diesel Fuel; Electric Motor Systems; Fossil Fuels; Freight Movement; Fuel Cell Vehicles; Gasoline and Additives; Gasoline Engines; Hybrid Vehicles; Kerosene; Locomotive Technology; Magnetic Levitation; Mass Transit; Methanol; Petroleum Consumption; Propellers; Railway Passenger Service; Ships; Steam Engines; Tires; Traffic Flow Management.

BIBLIOGRAPHY

Association of American Railroads. (1983). *Railroad Facts.* Washington, DC: Author.

Babcock & Wilcox. (1928). *Marine Steam.* New York: Author.

Davies, R. E. G. (1972). *Airlines of the United States Since 1914.* London: Putnam.

Davies, R. E. G. (1964). *A History of the World's Airlines.* London: Oxford University Press.

Flink, J. J. (1970). *America Adopts the Automobile, 1895–1910*, Cambridge, MA: MIT Press.

Flink, J. J. (1988). *The Automobile Age.* Cambridge, MA: MIT Press.

Grosser, M. (1978). *Diesel, the Man and the Engine.* New York: Atheneum.

Johnston, P. F. (1983). *Steam and the Sea.* Salem, MA: Peabody Museum of Salem.

Middleton, W. D. (1961). *The Interurban Era.* Milwaukee: Kalmbach Publishing Co.

Middleton, W. D. (1967). *The Time of the Trolley.* Milwaukee: Kalmbach Publishing Co.

Nevins, A., and Hill, F. E. (1957). *Ford: Expansion and Challenge, 1915–1933.* New York: Scribner.

Nevins, A., and Hill, F. E. (1954). *Ford: the Times, the Man, and the Company.* New York: Scribner.

Sternberg, E. R. (1981). *A History of Motor Truck Development.* Warrendale, PA: Society of Automotive Engineers.

U.S. Department of Transportation, Bureau of Transportation Statistics. (1999). *Pocket Guide to Transportation.* Washington, DC: Author.

U.S. War Industries Board, General Bureau of Planning and Statistics. (1918). *The World's Steamship Fuel Stations.* Washington, DC: Author.

Wakefield, E. H. (1994). *History of the Electric Automobile.* Warrendale, PA: Society of Automotive Engineers.

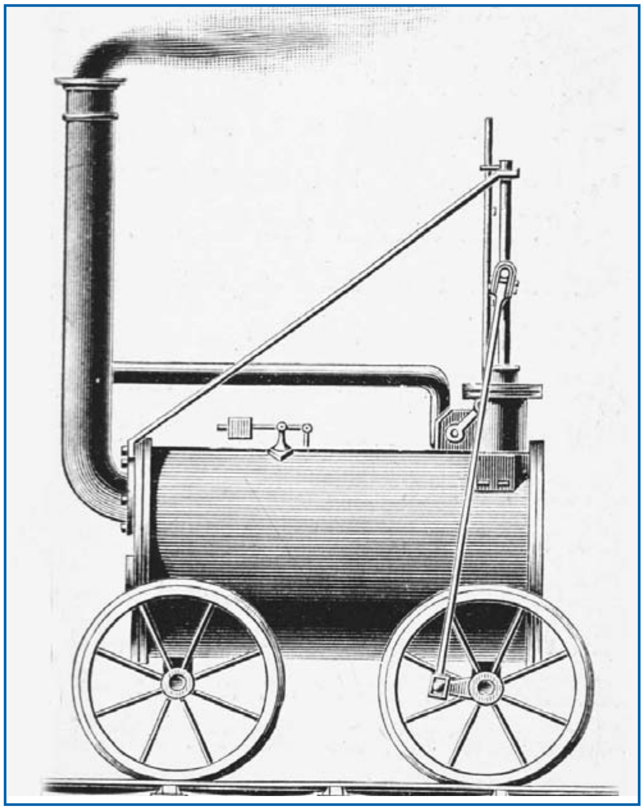
White, J. H., Jr. (1997). *American Locomotives: an Engineering History, 1830–1880*, rev. and expanded ed. Baltimore: Johns Hopkins University Press.

White, J. H., Jr. (1993). *The American Railroad Freight Car.* Baltimore: Johns Hopkins University Press.

TREVITHICK, RICHARD (1771–1833)

The father of high pressure steam power was born at the village of Illogan in Cornwall, England, on April 13, 1771, the only boy after five older sisters. His father, also Richard, was a mine "captain," as mine managers were called. In 1760 Richard, Sr., had married Anne Teague of Redruth. Her family included several more mine captains.

While the village schoolmaster thought the younger Richard disobedient, slow, obstinate and spoiled, the positive sides of these qualities emerged in independent thought, technical planning, persistence, and the loyalty of his workers and friends. His uninspiring record as a student and lack of formal training in engineering mattered less as he began work in the mines. There his talent and interest in



Richard Trevithick's locomotive around 1804. (Archive Photos, Inc.)

engineering gainied early notice. First employed as a mine engineer in 1790 at age nineteen, he was being called on as a consultant by 1792.

The steam-power systems in Trevithick's youth were massive but lightly loaded low-pressure engines. This technology was controlled by Boulton & Watt, whose business acumen had extended patents far beyond their normal expiration dates. The royalties typically took one-third of the savings in fuel over the Newcomen atmospheric engine. Another necessary evil was the expense of fuel, as coal was not locally mined. Cornish engineers worked incessantly to design and invent their way past these limitations.

Working around the Watt patents, he ensured safety while progressively raising the pressure of his systems to ten times atmospheric pressure, avoiding Watt's condenser altogether. This threat to the income of Boulton & Watt was met with a court injunction to stop the construction and operation of Trevithick's systems. Boulton & Watt had ruined other competitors, and their alarmist views of high pressure steam were largely due to lack of patent control in that emerging technology.

In 1797, Trevithick married Jane Harvey, lost his father, Richard, Sr., and made his first working models of high pressure engines. His wife was of the established engineering and foundry family of Hayle.

Trevithick's high-pressure engine models, including one with powered wheels, were the seeds for a line of improvements to steam engines and vehicles by many developers through the following decades. These departed from the preceding vehicle concepts and constructs of Murdock or Cugnot. These "puffer" engines made better use of the expansive properties of steam since they worked over a wider pressure differential. Using no condenser, Trevithick directed the exhaust up the chimney, inducing draft to increase the firing rate proportional to engine load, avoiding Watt's patents.

The first successful steam road vehicle was demonstrated climbing a hill in Camborne, Cornwall, on Christmas Eve, 1801. A new carriage, fitted with a proper coach body, was demonstrated in London in 1803. Eventually, a backer appeared with confidence and purchased a share in the developments. Samuel Homfray had Trevithick construct the first successful railway locomotive for his Penydarren Ironworks in South Wales. It was first run on February 13, 1804. Eight days later it won Homfray a wager of five hundred guineas. Homfray's existing cast-iron, horse-drawn rails were not yet suitable for continuous use. A similar locomotive was built at Newcastle-on-Tyne in 1805 to Trevithick's design. Again, the wooden rails were too light for the service. Trevithick's final showing of his locomotive concept was on a circular track at Gower Street in London where "Catch-me-who-can" gave rides for a shilling, but did not attract investment. Future developers would found an industry on Trevithick's technical generosity.

Trevithick's engineering reputation was established by the thirty high-pressure engines he built for winding, pumping and iron-rolling in Cornwall, Wales and Shropshire. His patent work and correspondence with members of the Royal Society might have brought him wider notice, but he was on a technical collision course with Watt and the engineering establishment, and conservative investors stayed away. Trevithick was left to financial alliances with indecisive or unscrupulous businessmen. While a near-fatal bout of typhus in 1811 kept him from work for months, his partner Richard Dickinson led him into bankruptcy in 1811.

Trevithick's high-pressure steam engines attracted attention, but his mechanical improvements enabling the boiler to withstand ten atmospheres of pressure were even more significant to power plant economy and practicality. He doubled the boiler efficiency. His wrought-iron boiler fired through an internal flue, the "Cornish" boiler became known worldwide. He applied the high-pressure engine to an iron-rolling mill (1805), a self-propelled barge using paddle-wheels (1805), a steam dredge (1806) and to powering a threshing machine (1812).

Some of Trevithick's other novel accomplishments and inventions include steam quarry drilling, applications for his superior iron tank construction, a "recoil engine" like Hero's Aeolipile, mining engines and mills in Peru, a one-piece cast-brass carbine for Bolivar's army, ship-raising by flotation, being the first European to cross the Isthmus of Nicaragua (from necessity), recoil-actuated gun-loading, iron ships, hydraulic dockside cranes, methods to drain Holland's lowlands, mechanical refrigeration, water-jet-propelled ships, and a portable heat storage heater.

Throughout his lifetime, Trevithick continued to measure his personal success in terms of his technical success in maintaining the fine balance of economy, utility and safety. At the end of his life he wrote, "I have been branded with folly and madness for attempting what the world calls impossibilities, and even from the great engineer, the late Mr. James Watt, who said...I deserved hanging for bringing into use the high-pressure engine. This so far has been my reward from the public; but should this be all, I shall be satisfied by the great secret pleasure and laudable pride that I feel in my own breast from having been the instrument of bringing forward and maturing new principles and new arrangements of boundless value to my country. . . . the great honour . . . far exceeds riches."

After becoming ill during a consulting stint at Dartford, Kent, England, Trevithick died, April 22, 1833, and was buried in a pauper's grave.

Continual disappointment in his business affairs kept him in relative obscurity even as his developments were reshaping the industrial world. Today, even the engineering community rarely notes his successes as the inventor of the first self-propelled road carriage we could call an automobile, and the railway locomotive.

Karl A. Petersen

See also: Watt, James.

BIBLIOGRAPHY

Boulton, I. W. Marginal notes, Science Museum copy of: Fletcher, William (1891). *Steam on Common Roads.*

Hodge, J. (1973). *Richard Trevithick: An Illustrated Life of Richard Trevithick*. Aylesbury: Shire Publications Ltd.

Rolt, L. T. C. (1960). *The Cornish Giant, The Story of Richard Trevithick, Father of the Steam Locomotive.* London: Lutterworth Press.

Trevithick, F. C. E. (1872). *Life of Richard Trevithick, with an Account of His Inventions.* London: E. & F.N. Spon.

TRIBOLOGY

The word tribology is derived from the Greek word, "tribos," which means "rubbing." Tribology covers the science of friction, lubrication and wear. Virtually any machine ever invented consists of components that must slide or roll relative to other components. Resistance to sliding and rolling caused by friction results in significant loss of energy. In fact, it has been estimated that as much as 33 percent of energy produced worldwide is used to overcome friction.

Friction and wear have many undesirable effects on individuals and society that are not widely known. Transportation vehicles have hundreds of moving parts that are subject to friction. Auto-mobiles generally wear out after about 100,000 miles of operation. Largely because of innovations to reduce friction and improve lubricants, the average lifetime (14 years in 2000) and miles of operation continue to increase. Since it requires as much energy to produce an automobile as it does to operate it for 100,000 miles, extending the life of an automobile would save considerable energy. The same applies to the steam and natural gas turbines that supply electricity and the home appliances that consume that electricity.

Friction and wear is even more important for national security. Downtime of military hardware as parts wear out and lower output power of military engines due to high friction can contribute to decreased effectiveness of the military and increases the costs of keeping highly specialized equipment in operation.

Wear of medical devices and biomaterials can affect quality of life. Wear of tooth fillings, artificial joints and heart valves can be inconvenient, costly (more frequent replacement) or even life-threatening (premature breakdowns). Wear of components can also cause accidents. Worn brakes and tires can cause automobile accidents, worn electrical cords can result in electrocution and fires and worn out seals can lead to radiation leaks at nuclear power plants.

RESEARCH

In the last few decades of the twentieth century the field of tribology has undergone a great surge in interest from industry and the scientific community. This is likely due to greater interest in conserving energy and natural resources. By lowering friction we conserve energy, and by lowering wear we increase product life, which conserves raw materials and the energy needed to turn raw materials into useful technology. However, there are many challenges to be overcome in the study of tribology.

Tribology is a surface phenomenon, and surfaces are extremely complex. As seen in Figure 1, when surfaces are sufficiently magnified we can see they are not flat, but consist of a series of peaks and valleys. In addition to being geometrically complex, surfaces are also chemically complex. Surfaces typically react with oxygen in the air to form a thin oxide film on the original surface. A thin gaseous film then generally forms on top of the oxide film. These films, which are usually so thin that they are transparent to the eye, significantly affect friction and wear.

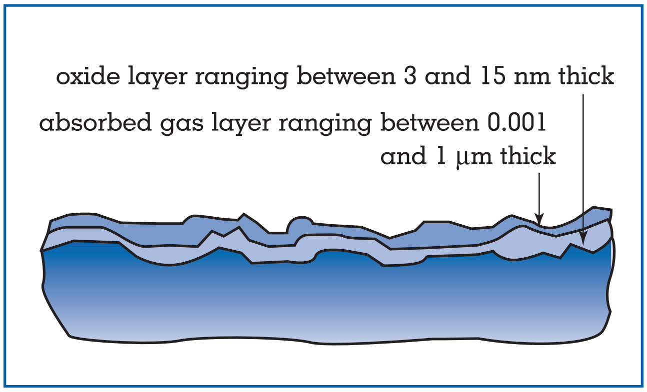


Figure 1.
Roughness of a "smooth" surface and surface films.

Besides surface topography, the hardness of the sliding materials, thickness and properties of oxide films, temperature, and type of lubricant all affect tribology. These factors overlap many fields of study, including physics, chemistry, material science and mechanical engineering. A complete understanding of tribology will require scientists and engineers from these diverse disciplines to work together. Traditionally, interdisciplinary research in tribology has not been done due to researchers being unwilling to cross the boundaries of their own discipline.

Surfaces are not easily observed during sliding. Tribologists study surfaces before and after sliding to determine chemical and geometric changes that occurred, and to infer how these changes came about and how they affected friction and wear. Because of the microscopic nature of tribology research, and the difficulty in duplicating real world conditions, it is not easy to observe and determine the cause of wear.

HISTORY OF LUBRICANTS

The most common method of minimizing friction and wear is through lubrication. The first recorded use of a lubricant was in ancient Egypt in 2400 B.C.E. Records show that they would pour a lubricant in front of a sledge being used to move a stone statue, which weighed tens of tons. The lubricant used was probably either water or animal fat. Various lubricants derived from animals and vegetables were used for thousands of years. Some examples of lubricants derived from animals are sperm oil, whale oil, and lard oil. Sperm oil was taken from a cavity in the head of sperm whales, whale oil from whale blubber, and lard from pig fat. Examples of vegetable lubricants are olive oil, groundnut oil and castor oil.

Lubricants began to receive significantly more attention from the industrial and scientific community in the mid-1800s with the introduction of mineral oils as lubricants. These proved to be effective lubricants; the demand for their use in machinery led to the development of many oil companies.

One common application for oils is the lubrication of automotive engines. When automotive oils were first introduced, their viscosities were classified as light, medium, medium heavy, heavy and extra heavy. This method of classification was too subjective and led to some engines being improperly lubricated. To remedy this, the Society of Automotive Engineers in 1912 introduced a quantitative numbering system for automotive viscosity. These numbers range from zero to fifty with lower numbers indicating lower viscosities. A general trend observed in oils is that as temperature increases, viscosity decreases. If viscosity becomes too small, the oil will be ineffective; if it is too high, the oil will not flow properly. Multigrade oils are now available that behave as lower viscosity oil at lower temperatures and higher viscosity oil at higher temperatures. For instance, a 10W-30 oil would have the viscosity of a 10-grade oil at low temperatures and the viscosity of a 30-grade oil at high temperatures. The use of multigrade oils minimizes change in viscosity with temperature.

In addition to viscosity grade, automotive oils also have a quality designation based on the API Engine Service classification system. This system was developed through the combined efforts of the American Society of Testing Materials (ASTM), the Society of Automotive Engineers (SAE), the American Petroleum Institute (API), automotive manufacturers and the United States military. API ratings describe the oil's ability to provide resistance to factors such as wear and corrosion. The original API rating was termed SA. This type of oil is suitable only for engines operated in very mild conditions. Oils with an API rating of SB are of higher quality than SA; SC has a higher API rating than SB, and so on.

In addition to liquid lubricants, solid lubricants such as graphite and grease have also been used for several centuries. Graphite was mined in the sixteenth century in Cumberland, England. It was originally called "black-lead" and was used in writing instruments. It is still used in "lead" pencils. There are numerous reports of graphite being used to lubricate machinery and silk looms in the nineteenth century. A common modern application for graphite is in the lubrication of door locks.

A specification for grease was given in 1812 by Henry Hardacre as follows: "one hundred weight of plumbago (graphite) to four hundred weight of pork lard or beef suet, mutton suet, tallow, oil, goose grease or any other kind of grease or greasy substance, but pork lard is the best, which must be well mixed together, so as to appear to the sight to be only one substance." Grease is used in many industrial applications including lubrication of gears and bearings.

CONSERVATION OF LUBRICANTS AND ENERGY

Most lubricating oils and greases use mineral oil as their base component. However, oil is a natural resource with a finite supply. Thus, it is imperative that measures such as recycling and extending the life of lubricants be taken to conserve the world's supply of oil and energy. There are approximately 2.5 billion gallons of waste oil generated each year in the United States. Only 83 million gallons are refined and reused, while one billion gallons are burned as fuel and another billion gallons or more are released into the environment. In response to these statistics, the United States Congress encouraged recycling of used oil through the Energy Policy and Conservation Act of 1975 and Resource Conservation and Recovery Act of 1976. The United States Environmental Protection Agency has stated that used oil is an environmental hazard, making the need for recycling even more critical.

Extending the interval between oil changes in automobiles and machinery is another way to conserve lubricants. Oil life is limited by depletion of oil additives, overheating, chemical contamination, and contamination of the oil by foreign particles. Using improved additives and filtering of particles, oil life can be considerably extended.

HISTORY OF FRICTION

DaVinci is well known for his famous paintings, such as the Mona Lisa, but his genius had amazing breadth. The scientific study of friction began with his research in the later part of the fifteenth century. He performed experiments in which he measured the force required to slide wood blocks over a wooden table. One of his most important findings was that the force F, required to produce sliding, increases in proportion with the applied load, L. DaVinci also defined what is today called the coefficient of friction, f, between two sliding objects as:

$$f = F/L$$

where F is friction force and L is normal load, as shown in Figure 2.

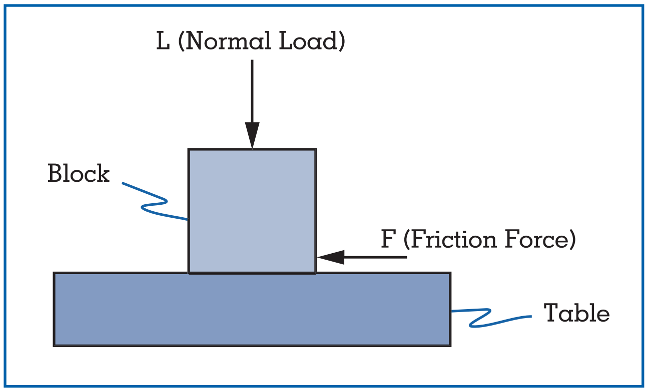


Figure 2.
The friction force between sliding objects resists movement. In this schematic, the block is being moved to the right.

DaVinci's experiments on friction also formed the basis for what today are called the first two laws of friction:

1. The force of friction, F, is directly proportional to the applied load, L.
2. The force of friction is independent of the apparent area of contact.

In general, a larger coefficient of friction between sliding bodies will require a larger force to produce sliding and, hence, more energy will be consumed. The most common method to reduce frictional energy losses is to lubricate the sliding surfaces. The reduction in coefficient of friction due to lubrication can be dramatic; typical values for dry sliding can range from 0.2 to 0.3, while typical values for lubricated sliding can range from 0.03 to 0.12.

The phenomenon of frictional heating is well known; for example, starting fires through the use of friction has been common since prehistoric times. However, frictional heating is generally a detrimental phenomenon. The temperature rise on sliding surfaces that occurs due to frictional heating generally produces a decrease in surface hardness, causing the surfaces to wear more easily, thus decreasing product life.

From an energy consumption point of view, we desire friction to be as small as possible. However, there are specific applications where friction should not be minimized. One common example is the brakes used in automobiles. If friction is too low, automobiles will not stop in a reasonable distance. However, if the friction is too high, a jerky ride will be produced. Other examples where high friction is desirable are shoe soles, nails, screws, and belt drives.

It is interesting to note that although friction has been studied for hundreds of years, there is no universal agreement on the fundamental mechanisms of how friction is produced. The two most popular theories are interlocking and adhesion. The interlocking theory states that as two rough surfaces slide over each other, the peaks on one surface "interlock" with peaks on the adjacent surface, producing resistance to sliding. The adhesion theory suggests that as the sliding surfaces contact each other, adhesive bonds form. These molecular bonds must be broken to continue sliding, which results in increased friction. The cost of not understanding the fundamentals of friction is high. Estimates have shown that as much as 0.5 percent of industrial countries' gross national products are wasted because we do not know enough about minimizing frictional losses during sliding.

FUTURE ADVANCES IN LUBRICATION

The transportation industries are being challenged with increasingly stringent government regulations that demand improved vehicle fuel economy and reduced emissions. Engine oils with greater ability to conserve energy are becoming more important as a way to lower engine friction and thus conserve natural resources. One specific approach under development is to produce engine oils with a friction-reducing additive such as molybdenum dialkyldithiocarbamate which has showed promise in developmental engine tests.

Another approach being used to conserve energy is to increase combustion temperatures in engines. This results in increased engine efficiency that produces energy savings. Combustion temperatures may eventually become so high that liquid lubricants break down and become ineffective. A possible approach for lubrication under these extreme temperatures is vapor phase lubrication, which refers to the deposition of lubricant from an atmosphere of vaporized lubricant. The vaporized lubricant is delivered in a carrier gas, such as nitrogen, to the component to be lubricated where it reacts chemically to form a solid lubricant film. These films are effective at much higher temperatures than liquid-based lubricants. Other approaches that can be used in components that operate in high operating temperatures are the use of gas bearings or magnetic bearings.

Gary C. Barber
Barbara Oakley

See also: Automobile Performance; Materials

BIBLIOGRAPHY

Booser, R. E. (1988). *Handbook of Lubrication, Theory and Practice of Tribology: Volume II.* Boca Raton: CRC Press.

Bowden, F. P., and Tabor, D. (1964). *The Friction and Lubrication of Solids, Part II.* Oxford: Clarenden Press.

Cameron, A. (1996). *Principles of Lubrication.* New York: John Wiley.

Czichos, H. (1978). *Tribology.* Amsterdam: Elsevier.

Dowson, D. (1979). *History of Tribology.* New York: Longman.

Fuller, D. D. (1984). *Theory and Practice of Lubrication for Engineers.* New York: John Wiley.

Groeneweg, M.; Hakim, N.; Barber, G. C.; and Klaus, E. (1991). "Vapor Delivered Lubrication of Diesel Engines—Cylinder Kit Rig Simulation." *Lubrication Engineering,* 47(12):1035.

Hamrock, B. J. (1994). *Fundamentals of Fluid Film Lubrication.* New York: McGraw-Hill.

Hutchings, I. M. (1992). *Tribology: Friction and Wear of Engineering Materials.* London: Edward Arnold.

Jost, H. P. (1975). "Economic Impact of Tribology." *Mechanical Engineering,* 97:26.

Ku, P. M. (1978). "Energy Conservation Through Tribology." *Tribology International* 11:153.

Ludema, K. C. (1996). *Friction, Wear, Lubrication: A Textbook in Tribology.* Boca Raton: CRC Press.

Peterson, M. B., and Winer, W. O., eds. (1980). *Wear Control Handbook.* New York: ASME.

Rabinowicz, E. (1995). *Friction and Wear of Materials.* New York: John Wiley.

Sarkar, A. D. (1980). *Friction and Wear.* New York: Academic Press.

Szeri, A. Z. (1980). *Tribology.* New York: McGraw-Hill.

Tung, S. C., and Tseregounis, S. I. (2000). "An Investigation of Tribological Characteristics of Energy-Conserving Engine Oils Using a Reciprocating Bench Test." SAE 2000-01-1781.

TRUE ENERGY COSTS

Market prices of energy often diverge from the true cost to society of consuming that energy. Two of the most common reasons for that divergence are external costs and subsidies, both of which make consumers think that energy is less expensive to society than it really is, and hence lead to more consumption of energy than would be economically optimal.

EXTERNAL AND INTERNAL COSTS

According to J. M. Griffin and H. B. Steele (1986), external costs exist when "the private calculation of costs differs from society's valuation of costs." Pollution represents an external cost because damages associated with it are borne by society as a whole, not just by the users of a particular fuel. Pollution causes external costs to the extent that the damages inflicted by the pollutant are not incorporated into the price of the fuel associated with the damages. External costs can be caused by air pollution, water pollution, toxic wastes, or any other damage to the environment not included in market prices for goods.

Some pollutants' external costs have been "internalized" because the resulting damage has already been incorporated into the price of energy by a tax on the energy source, or because an emissions-trading regime has been established to promote cost-effective control of that pollutant. The pollutants may still cause environmental damage, but these damages have been made internal to the economic calculations of consumers and firms by the taxes or emissions-trading schemes. They are thus no longer external costs.

For example, sulfur emissions from utility power plants in the United States are subject to an emissions cap and an allowance-trading system established under the Clean Air Act. An effective cap on annual sulfur dioxide emissions took effect in 2000, so no more than 8.95 million tons of SO_2 can be emitted annually. Utilities that want to build another coal plant must purchase sulfur emission allowances from others who do not need them. This system provides a market incentive for utilities to reduce their sulfur emissions as long as the cost of such reductions is less than the price of purchasing the allowances.

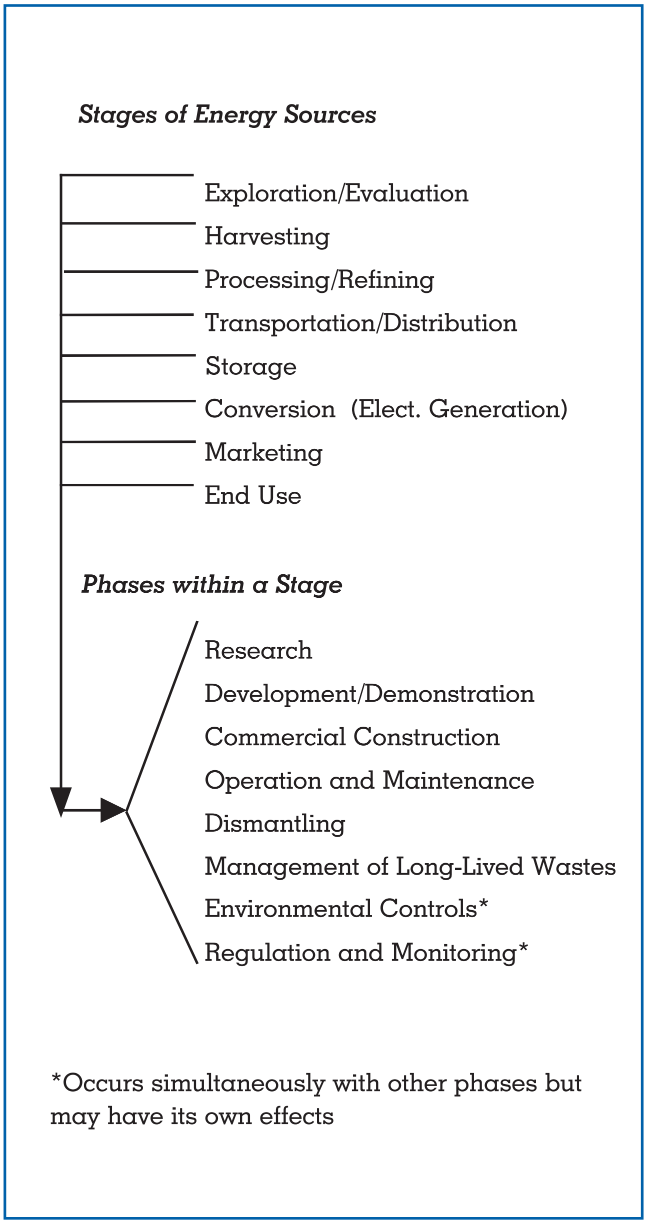


Figure 1.
Steps in energy production, processing, and use.
SOURCE: Holdren, 1981.

Subsidies

Subsidies represent an often hidden financial benefit that is given to particular energy sources by government institutions in the form of tax credits, research and development (R&D) funding, limits on liability for certain kinds of accidents, military spending to protect Middle East oil supply lines, below-market leasing fees for use of public lands, and other forms of direct and indirect government support for energy industries. Subsidies affect choices between different energy sources (which garner different levels of subsidy) and between energy-efficiency and energy supply choices (because some energy-efficiency options that would be cost-effective from society's point of view may not be adopted when energy prices are kept artificially low by the subsidies).

Subsidies can be created to reward important political constituencies, or they can promote the adoption of particular technologies. For example, an energy technology that has desirable social characteristics and low external costs might be a recipient of tax subsidies in the early stages of its development, so that it can gain a foothold against more established energy forms. This kind of support can be especially important for mass-produced technologies that have high costs when few units are being manufactured, but could achieve significant economies of scale in large volume production if the technology were widely adopted. The subsidy can provide the initial impetus that allows the technology to achieve lower costs and widespread acceptance. This rationale was the one used to justify the early subsidies of wind generation, which eventually led to the creation of an economically viable and internationally competitive U.S. wind industry.

Subsidies can also apply to the creation of new technology, through funding of research and development (R&D). The rationale for government subsidies for R&D (particularly long-term R&D) is well established in the economics literature. Companies will not fund the societally optimal level of basic R&D in new technologies, because many of the benefits of such research will flow to their competitors and to other parts of the economy (Mansfield 1982, pp. 454–455). Innovations are often easy to imitate, so which the innovators spend money on R&D, the followers can copy those innovations and avoid the risk and expense of such spending. R&D therefore has characteristics of a "public good."

Subsidies for relatively new energy technologies or fuels tend to be small in the aggregate and they tend not to have a measurable effect on the overall price of energy. When subsidies apply to widely used energy sources, however, they may create significant economic distortions, and the divergence from long-term economic efficiency may be substantial. Such distorting subsidies can sometimes exist for years because they are defended by powerful political constituencies.

EXTERNAL COSTS, SUBSIDIES, AND THE FUEL CYCLE

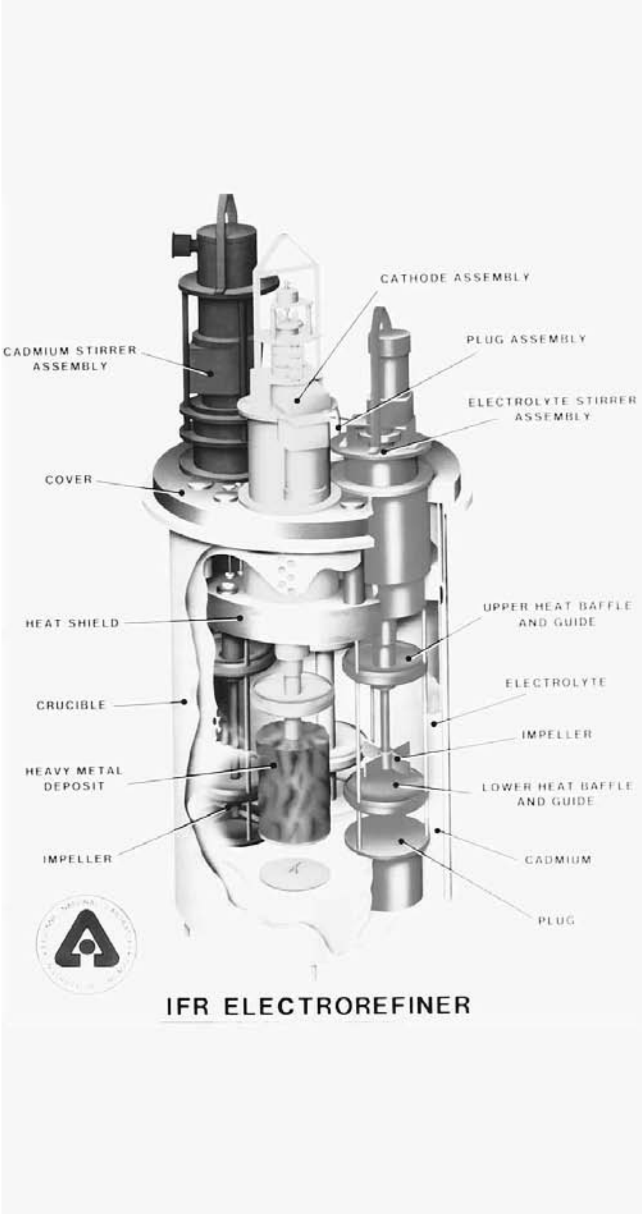
A comprehensive analysis of external costs and subsidies must treat each and every stage and phase in the process, which makes any such calculation inherently difficult. Uncertainties abound in these calculations, especially for external costs. As a result, estimates of total external costs and subsidies for different energy sectors vary widely.

External costs for fossil fuels are generally largest at the point of end use (combustion), though exploration (oil and gas drilling, mining), processing (refineries), and transportation (pipeline ruptures, tanker spills) can each contribute significant external costs in particular cases. For nuclear power, the accident risks associated with the conversion stage and the long-term issues surrounding disposal of spent fuel are the external costs that typically garner the most attention. There are also external costs from other stages of the nuclear fuel cycle, including the various effects of exploration, harvesting, and processing, as well as the risk of nuclear weapons proliferation from the spread of fissionable materials and related knowledge.

For nonfuel renewables such as hydroelectricity and wind power, external costs are most significant at the point of conversion (e.g., salmon migration blocked by dams, birds killed by wind turbine blades, noise and visual pollution from wind turbines). Construction of dams, particularly large ones, can also cause significant externalities by flooding large land areas, displacing people and wildlife, releasing methane from anaerobic decay of plant matter, and affecting local evapotranspiration rates. In contrast, generation of electricity using solar photovoltaics has few external costs, with the exception of small amounts of pollutant emissions from the manufacture and installation of the modules.

Most analyses of external costs have focused on electricity because of regulatory activities in that sector. One analysis of energy-related externalities in all sectors was conducted by Hohmeyer for Germany (Hohmeyer, 1988), but such comprehensive estimates are rare. For those analyses that have been done, estimates of external costs range from near zero (for photovoltaics and energy efficiency), to amounts that are significant relative to the market price of energy for some fossil fuels and nuclear power. The uncertainties in these calculations are typically quite large.

The most common subsidies for fossil fuels have been R&D and production incentives that have affected exploration and harvesting. R&D funding has also been important for fossil-fired, end-use technologies, such as furnaces that use a particular fossil fuel. For nuclear power, significant subsidies exist for processing/refining of fuel, R&D (which affects most stages), limitation of accident liability from operation of the plants, and management of long-lived wastes from the fuel cycle.



Cutaway diagram of an Integral Fast Refiner (IFR). It removes the short-lived by-products of nuclear fusion from the long-lived materials so the latter can be used to create new fuel. (Corbis-Bettmann)

A U.S. federal tax subsidy exists for wind generation and certain kinds of biomass-fired power plants built before December 31, 2001. Qualifying wind and biomass generators are paid 1.7 cents per kilowatt-hour generated over their first ten years of operation (the amount paid per kilowatt-hour increases over time at the rate of inflation). In the early days of wind generation there were subsidies in the United States for wind capacity installed, but these subsidies were phased out in the mid-1980s. R&D subsidies have been important for the development of new and more reliable renewable energy technologies. For energy-efficiency technologies, R&D funding and consumer rebates from electric and gas utilities have been the most important kinds of subsidies, but there has also been subsidization of installation of energy-efficiency measures in low-income housing. As the electric utility industry moves toward deregulation, states are increasingly relying on so-called “systems benefit charges” to fund subsidies for energy efficiency and renewable energy sources.

Unfortunately, the most recent estimates of energy subsidies in the United States date from the early 1990s and earlier, and such subsidies change constantly as the tax laws are modified and governmental priorities change. It is clear that total subsidies to the energy industries are in the billions to a few tens of billions of dollars each year, but the total is not known with precision.

ENERGY EFFICIENCY AND SUPPLY TECHNOLOGIES

External costs and subsidies for both energy efficiency and supply technologies must be included in any consistent comparison of the true energy costs of such technologies. Pollutant emissions from supply technologies are both direct (from the combustion of fossil fuels) and indirect (from the construction of supply equipment and the extraction, processing, and transportation of the fuel). Emissions from efficiency technologies are generally only of the indirect type. Increasing the efficiency of energy use almost always reduces emissions and other externalities.

THE TRUE COST OF ENERGY: AN EXAMPLE

While exact estimates of the magnitude of external costs and subsidies are highly dependent on particular situations, a hypothetical example can help explain how these two factors affect consumers’ decisions for purchasing energy. In round numbers, if subsidies are about $20 billion for energy supply technologies, and total direct annual energy expenditures for the United States are about $550 billion, the combined cost of delivering that energy is $570 billion per year. This total represents an increase of about four percent over direct expenditures for energy.

Including external costs associated with energy supplies would increase the cost still further. Typical estimates for external costs for conventional energy supplies are in the range of 5 to 25 percent of the delivered price of fuel (for the sake of this example, we assume that these percentages are calculated relative to the price of fuel plus subsidies). If we choose ten percent for externalities in our example, we can calculate the “true energy price” that the consumer would see if subsidies and externalities were corrected in the market price. If the average price of fuel (without any taxes) is P, then the price of fuel correcting for subsidies is $P \times 1.04$, and the price of fuel correcting for both subsidies and externalities is $P \times 1.04 \times 1.10$. So in this example, the price of fuel would be about fourteen percent higher if subsidies and externalities were correctly included in the price.

If the particular energy source in question is already taxed (say at a five percent rate), then part of the external cost is already internalized. The true cost of fuel would remain the same ($P \times 1.04 \times 1.10$) but the size of the additional tax needed to correct for the externality would be smaller (about five percent instead of ten percent).

KEY POLICY ISSUES

New subsidies often outlive the public policy purpose that they were intended to address. The better-designed subsidies contain “sunset” provisions that require explicit action to reauthorize them after a certain time. Subsidy and externality policies are often interrelated. It may be politically difficult to tax an energy source with high external costs, but much easier to subsidize a competing energy source with low external costs. Such “second best” solutions are often implemented when political considerations block the preferred option.

Another important consideration is that “getting prices right” is not the end of the story. Many market imperfections and transaction costs affecting energy use will still remain after external costs are incorporated and subsidies that do not serve a legitimate public policy purpose are removed. These imperfections, such as imperfect information, asymmetric information, information costs, misplaced incentives, and bounded rationality, may be addressed by a variety of nonenergy-price policies, including efficiency standards, incentive programs, and information programs.

Jonathon G. Koomey

See also: Economic Externalities; Market Imperfections; Subsidies and Energy Costs.

BIBLIOGRAPHY

Bezdek, R. H., and Cone, B. W. (1980). "Federal Incentives for Energy Development." *Energy* 5(5):389–406.

Brannon, G. M. (1974). *Energy Taxes and Subsidies*. Cambridge, MA: Ballinger Publishing Co.

Golove, W. H., and Eto, J. H. (1996). *Market Barriers to Energy Efficiency: A Critical Reappraisal of the Rationale for Public Policies to Promote Energy Efficiency*. Berkeley, CA: Lawrence Berkeley Laboratory. LBL-38059.

Griffin, J. M., and Steele, H. B. (1986). *Energy Economics and Policy*. Orlando, FL: Academic Press College Division.

Heede, H. R. (1985). *A Preliminary Assessment of Federal Energy Subsidies in FY1984*. Washington, DC: Testimony submitted to the Subcommittee on Energy and Agricultural Taxation, Committee on Finance, United States Senate, June 21.

Hohmeyer, O. (1988). *Social Costs of Energy Consumption: External Effects of Electricity Generation in the Federal Republic of Germany*. Berlin: Springer-Verlag.

Holdren, J. P. (1981). "Chapter V. Energy and Human Environment: The Generation and Definition of Environmental Problems." In *The European Transition from Oil: Societal Impacts and Constraints on Energy Policy*, ed. by G. T. Goodman, L. A. Kristoferson, and J. M. Hollander. London: Academic Press.

Jaffe, A. B., and Stavins, R. N. (1994). "Energy-Efficiency Investments and Public Policy." *The Energy Journal* 15(2):43.

Koomey, J. (1990). *Energy Efficiency Choices in New Office Buildings: An Investigation of Market Failures and Corrective Policies*. PhD thesis, Energy and Resources Group, University of California, Berkeley. <http://enduse.lbl.gov/Projects/EfficiencyGap.html>.

Koomey, J.; Sanstad, A. H.; and Shown, L. J. (1996). "Energy-Efficient Lighting: Market Data, Market Imperfections, and Policy Success." *Contemporary Economic Policy* 14(3):98–111.

Koplow, D. N. (1993). *Federal Energy Subsidies: Energy, Environmental, and Fiscal Impacts*. Washington, DC: The Alliance to Save Energy.

Kosmo, M. (1987). *Money to Burn? The High Costs of Energy Subsidies*. World Resources Institute.

Krause, F.; Haites, E.; Howarth, R.; and Koomey, J. (1993). *Cutting Carbon Emissions—Burden or Benefit?: The Economics of Energy-Tax and Non-Price Policies*. El Cerrito, CA: International Project for Sustainable Energy Paths.

Kushler, M. (1998). *An Updated Status Report of Public Benefit Programs in an Evolving Electric Utility Industry*. Washington, DC: American Council for an Energy-Efficient Economy.

Levine, M. D.; Hirst, E.; Koomey, J. G.; McMahon, J. E.; and Sanstad, A. H. (1994). *Energy Efficiency, Market Failures, and Government Policy*. Berkeley, CA: Lawrence Berkeley Laboratory. LBL-35376.

Mansfield, E. (1982). *Microeconomics: Theory and Applications*. New York: W. W. Norton and Co.

Ottinger, R. L.; Wooley, D. R.; Robinson, N. A.; Hodas, D. R.; Babb, S. E.; Buchanan, S. C.; Chernick, P. L.; Caverhill, E.; Krupnick, A.; Harrington, W.; Radin, S.; and Fritsche, U. (1990). *Environmental Costs of Electricity*. New York: Oceana Publications, Inc., for the Pace University Center for Environmental and Legal Studies.

Sanstad, A. H., and Howarth, R. (1994). "'Normal' Markets, Market Imperfections, and Energy Efficiency." *Energy Policy* 22(10):826–832.

U.S. Department of Energy. (1980). *Selected Federal Tax and Non-Tax Subsidies for Energy Use and Production* (Energy Policy Study, Volume 6). Washington, DC: Energy Information Administration, U.S. Department of Energy. DOE/EIA-0201/6, AR/EA/80-01.

TURBINES, GAS

The aircraft gas turbine engine, developed more than sixty years ago, uses the principle of jet reaction and the turbine engine. The engine consists of three major elements: a compressor and a turbine expander, which are connected by a common shaft; and a combustor, located between the compressor and the turbine expander. The useful work of the engine is the difference between that produced by the turbine and that required by the compressor. For the simple cycle system shown in Figure 1, about two-thirds of all the power produced by the turbine is used to drive the compressor.

Jet reaction used in the first steam-powered engine, the aeolipile, is attributed to Hero of Alexandria around the time of Christ. In his concept, a closed spherical vessel, mounted on bearings, carried steam

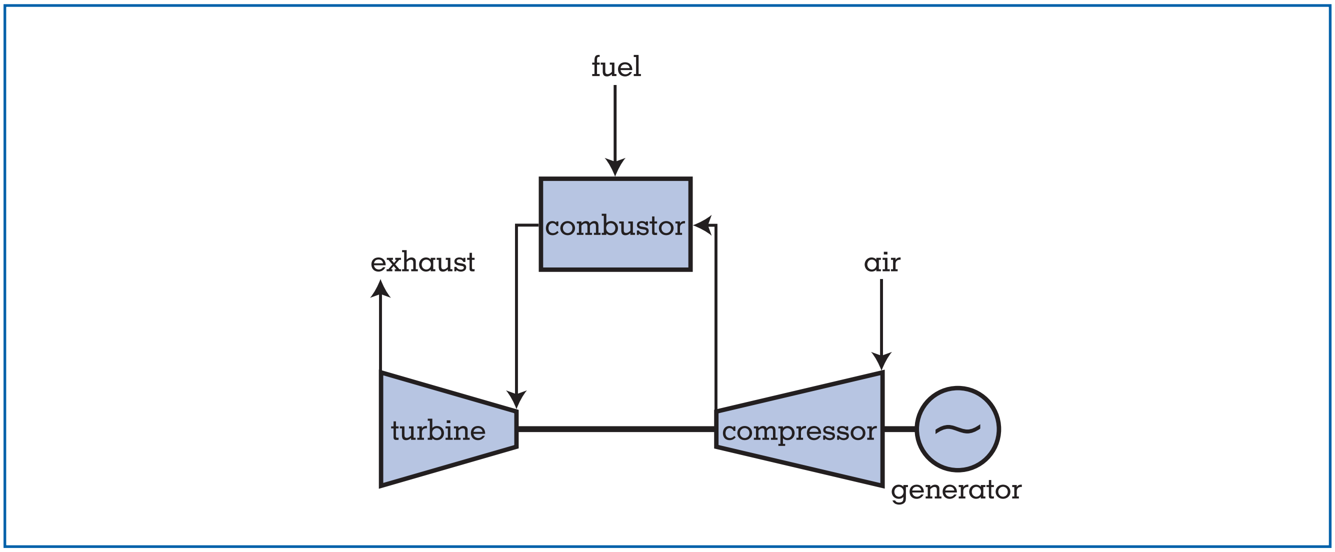


Figure 1.
Simple gas turbine cycle.

from a cauldron with one or more people discharging tangentially at the vessel's periphery, and was driven around by the reaction of steam jets. According to the literature, the first gas turbine power plant patent was awarded to John Barber, an Englishman, in 1791. Intended to operate on distilled coal, wood, or oil, it incorporated an air compressor driven through chains and gears by a turbine operated by the combustion gases. It was actually built but never worked.

There was a steady increase in the number of gas-turbine patents after Barber's disclosure. However, the attempts of the early inventors to reduce them to practice was entirely unsuccessful.

The early inventors and engineers were frustrated in their efforts to achieve a workable gas turbine engine because of the inadequate performance of the components and the available materials. However, gas turbine engine technology has advanced rapidly since the 1940s. Thrust and shaft horsepower (hp) has increased more than a hundredfold while fuel consumption has been cut by more than 50 percent. Any future advances will again depend upon improving the performance of components and finding better materials.

ORIGINS OF THE LAND-BASED GAS TURBINE

Perhaps the first approach to the modern conception of the gas-turbine power plant was that described in a patent issued to Franz Stolze in 1872. The arrangement of the Stolze plant, shown in Figure 2, consisted of a multistage axial-flow air compressor coupled directly to a multistage reaction turbine. The high-pressure air leaving the compressor was heated in an externally fired combustion chamber and thereafter expanded in the multistage turbine. The Stolze plant was tested in 1900 and 1904, but the unit was unsuccessful. The lack of success was due primarily to the inefficiency of the axial-flow compressor, which was based on aerodynamics, a science that was in its early stages of development. The lack of aerodynamic information also led Charles Algernon Parsons, the inventor of the reaction steam turbine, to abandon the development of the axial-flow compressor in 1908 after building approximately thirty compressors of that type with little success.

Several experimental gas-turbine power plants were built in France from 1903 to 1906. Those power plants operated on a cycle similar to that of a modern gas-turbine power plant. The most significant unit had a multistage centrifugal compressor that consisted essentially of a stationary casing containing a rotating impeller. Because of material limitations, the temperature of the gases entering these turbines was limited to 554°C (1,030°F). The thermal efficiency of the unit (work output divided by heat supplied) was less than 3 percent, but the unit is noteworthy, however, because

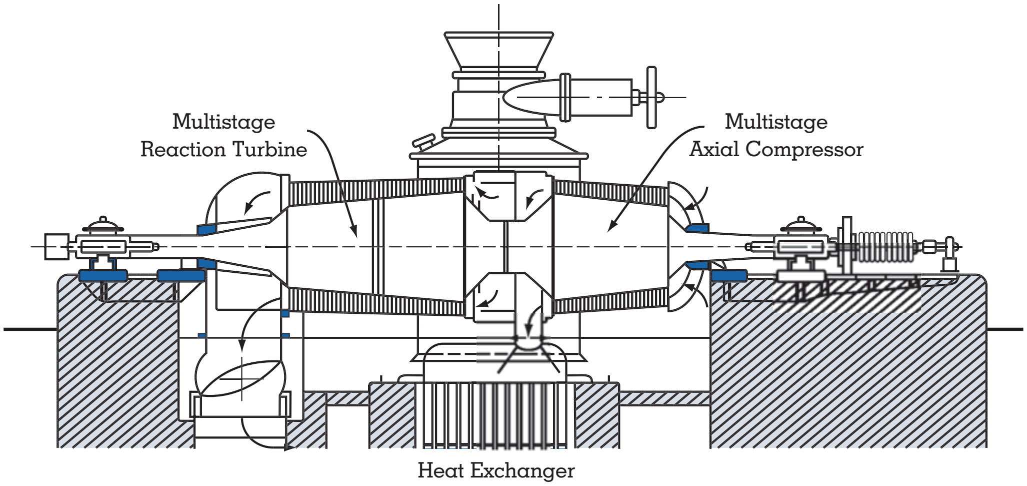


Figure 2.
The Stolze gas-turbine power plant.
SOURCE: Zipkin, 1964.

it was probably the first gas-turbine power plant to produce useful work. Its poor thermal efficiency was due to the low efficiencies of the compressors and the turbine, and also to the low turbine inlet temperature.

Brown Boveri is credited with building the first land-based gas turbine generating unit. Rated at 4 MW, it was installed in Switzerland in 1939. Leaders in the development of the aviation gas turbine during the Second World War included: Hans von Ohain (Germany), Frank Whittle (England), and Reinout Kroon (United States). Thus, combining the technology derived for aviation gas turbines with the experience of using turbines in the chemical industry, the birth of gas turbines for power generation started in the United States in 1945 with the development of a 2,000-hp gas turbine set that consisted of a compressor, twelve combustors, a turbine, and a single reduction gear. This turbine had a thermal efficiency of 18 percent. By 1949, land-based gas turbines in the United States had an output of 3.5 MW with a thermal efficiency of 26 percent at a firing temperature of 760°C (1,400°F).

COMPONENT DEVELOPMENT

Aircraft and land-based turbines have different performance criteria. Aircraft turbine engine performance is measured in terms of output, efficiency, and weight. The most significant parameter in establishing engine output (thrust or shaft hp) is turbine inlet temperature. Power output is extremely dependent on turbine inlet temperature. For example, an engine operating at 1,340°C (2,500°F) would produce more than over two and one-half times the output of an engine operating at 815°C (1,500°F). Engine efficiency, which is next in importance to engine output, is determined largely by overall engine or cycle pressure ratio. Increasing the pressure ratio from one to twenty reduces the fuel consumption by approximately 25 percent. Engine weight is affected most by turbine inlet temperatures (which acts to reduce the physical size of the engine) and by state-of-the-art materials technology that sets the temperature criteria.

During the Second World War, centrifugal compressors were used in early British and American fighter aircraft. As power requirements grew, it became clear that the axial flow compressor was more suitable for large engines because they can produce higher pressure ratios at higher efficiencies than centrifugal compressors, and a much larger flow rate is possible for a given frontal area. Today, the axial flow machine dominates the field for large power generation, and the centrifugal compressor is restricted to machines where the flow is too small to be used efficiently by axial blading.

For land-based gas turbines, the overall plant output, efficiency, emissions, and reliability are the important variables. In a gas turbine, the processes of compression, combustion, and expansion do not occur in a single component, as they do in a diesel engine. They occur in components that can be developed separately. Therefore, other technologies and components can be added as needed to the basic components, or entirely new components can be substituted.

Advanced two- and three-dimensional computer analysis methods are used today in the analyses of all critical components to verify aerodynamic, heat transfer, and mechanical performance. Additionally, the reduction of leakage paths in the compressor, as well as in the gas turbine expander, results in further plant efficiency improvements. At the compressor inlet, an advanced inlet flow design improves efficiency by reducing pressure loss. Rotor air cooler heat utilization and advanced blade and vane cooling are also used.

Several advanced turbine technologies may be applied to the gas turbine expander. These include the application of single crystal and ceramic components to increase material operating temperatures and reduce turbine cooling requirements, and active blade tip clearance control to minimize tip leakages and improve turbine efficiency. The objective of the latter scheme is to maintain large tip clearance at start-up and reduce them to minimum acceptable values when the engine has reached steady-state operating conditions.

ALTERNATIVE THERMAL CYCLES

As an alternative to raising firing temperature, overall power plant performance can be improved by modifications to the cycle. Combining a land-based simple cycle gas turbine with a steam turbine results in a combined cycle that is superior in performance to a simple gas turbine cycle or a simple cycle steam turbine considered separately. This is due to utilizing waste heat from the gas turbine cycle. By 1999, land-based sample cycle gas turbine efficiencies had improved from 18 percent to more than 42 percent, with the better combined cycles reaching 58 percent, and the ones in development likely to exceed 60 percent. Combined cycle efficiency improvements have followed the general advance in gas turbine technology reflected in the rising inlet temperature trend shown in Figure 3, which, in turn, was made possible by advances in components and materials.

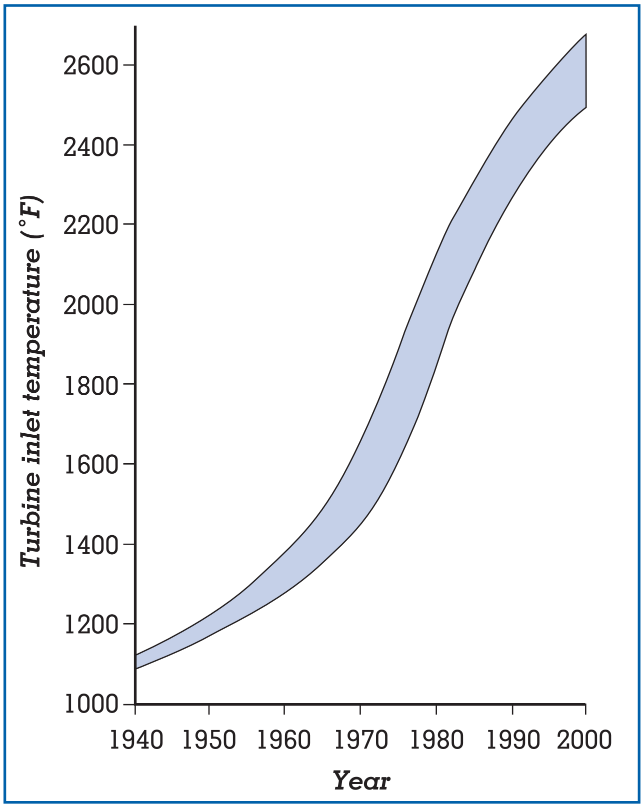


Figure 3.
Gas turbine inlet temperature trend.

Gas turbines can operate in open or closed cycles. In a simple cycle (also known as an open cycle), clean atmospheric air is continuously drawn into the compressor. Energy is added by the combustion of fuel with the air. Products of combustion are expanded through the turbine and exhausted to the atmosphere. In a closed cycle, the working fluid is continuously circulated through the compressor, turbine, and heat exchangers. The disadvantage of the closed cycle (also known as the indirect cycle), and the reason why there are only a few in operation, is the need for an external heating system. That is an expensive addition and lowers efficiency.

Most current technology gas turbine engines use air to cool the turbine vanes and rotors. This allows the turbine inlet temperature to be increased beyond the temperature at which the turbine material can be used without cooling, thus increasing the cycle efficiency and the power output. However, the cooling air itself is a detriment to cycle efficiency. By using closed-loop steam cooling, a new concept, the cooling air mixing loss mechanisms can be largely elimi-

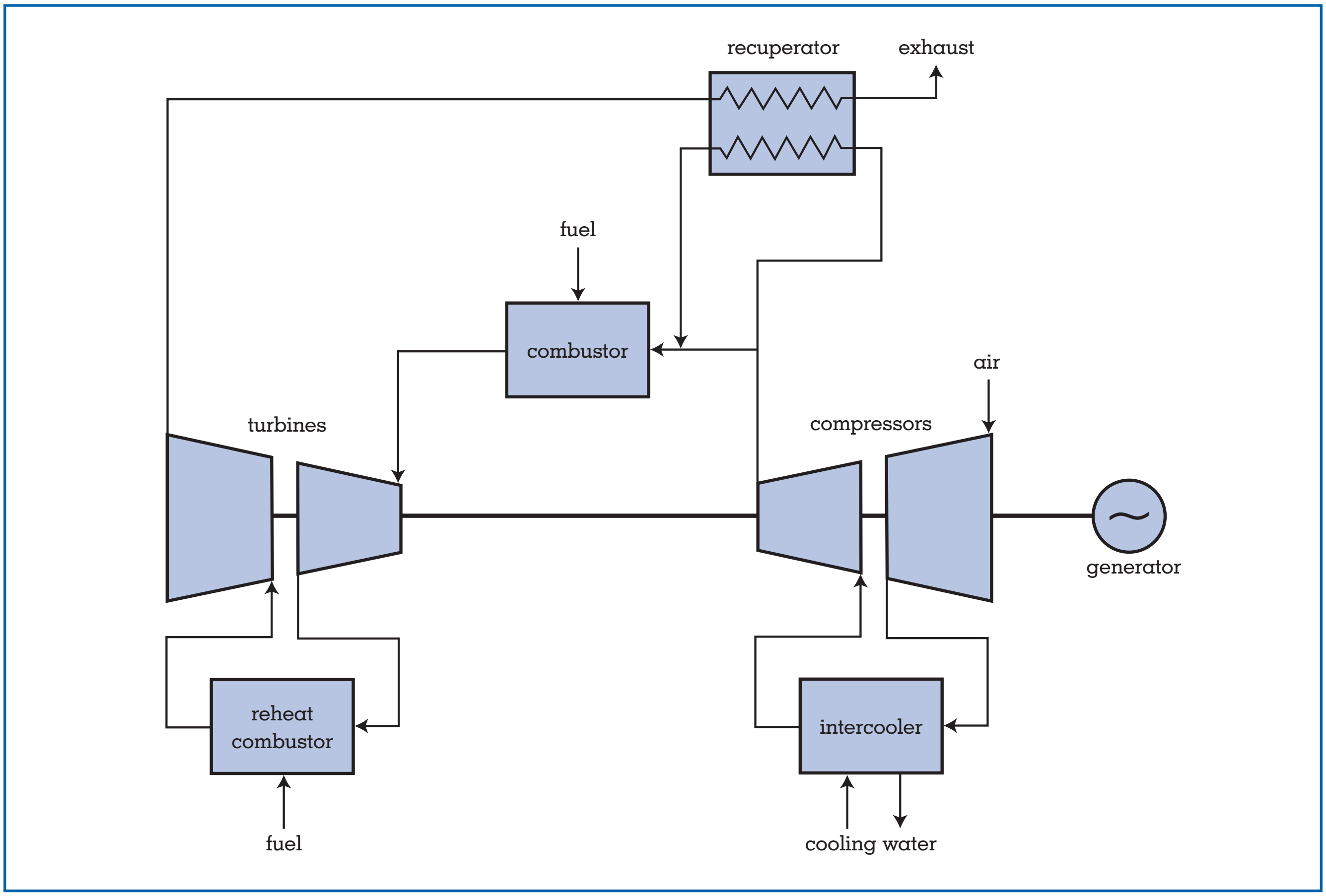


Figure 4.
Some gas turbine cycle options.
SOURCE: Maude and Kirchner, 1995.

nated, while still maintaining turbine material temperatures at an acceptable level. An additional benefit of closed-loop cooling is that more compressor delivery air is available for the lean premix combustion. The result is lower flame temperature and reduced oxides of nitrogen (NOx) emission. In combined cycles, the steam used for cooling the gas turbine hot parts is taken from the steam bottoming cycle, then returned to the bottoming cycle after it has absorbed heat in the closed-loop steam cooling system. For an advanced bottoming steam cycle, closed-loop steam cooling uses reheat steam from the exit of the high-pressure steam turbine to cool the gas turbine vane casing and rotor. The steam is passed through passageways within the vane and rotor assemblies and through the vanes and rotors themselves, then collected and sent back to the steam cycle intermediate-pressure steam turbine as hot reheat steam.

Several of the gas turbine cycle options discussed in this section (intercooling, recuperation, and reheat) are illustrated in Figure 4. These cycle options can be applied singly or in various combinations with other cycles to improve thermal efficiency. Other possible cycle concepts that are discussed include thermochemical recuperation, partial oxidation, use of a humid air turbine, and use of fuel cells.

The most typical arrangement for compressor intercooling involves removing the compressor air flow halfway through the compressor temperature rise, sending it through an air-to-water heat exchanger, and returning it to the compressor for further compression to combustor inlet pressure. The heat removed from the compressor air flow by the intercooler is rejected to the atmosphere because the heat is usually at too low a temperature to be of use to the cycle.

Another intercooling application is to spray water droplets into the compressor. As the air is compressed and increases in temperature, the water evaporates and absorbs heat. This results in a continuous cooling of the compressor. Note that for this concept the heat absorbed by the water is also rejected to the atmosphere. This water is never condensed by the cycle, but instead exhausted with the stack gases as low-pressure steam.

Compressor intercooling reduces the compressor work because it compresses the gas at a lower average temperature. The gas and steam turbines produce approximately the same output as in the nonintercooled case, so the overall cycle output is increased. However, since the compressor exit temperature is lowered, the amount of fuel that must be added to reach a given turbine inlet temperature is greater than that for the nonintercooled case. Intercooling adds output at approximately the simple cycle efficiency—the ratio of the amount of compressor work saved to the amount of extra fuel energy added is about equal to the simple cycle efficiency. Since combined cycle efficiencies are significantly greater than simple cycle efficiencies, the additional output at simple cycle efficiency for the intercooled case usually reduces the combined cycle net plant efficiency.

In recuperative cycles, turbine exhaust heat is recovered and returned to the gas turbine combustor, usually through a heat exchange between the turbine exhaust gases and the compressor exit airflow. The discharge from the compressor exit is piped to an exhaust gas-to-air heat exchanger located aft of the gas turbine. The air is then heated by the turbine exhaust and returned to the combustor. Since the resulting combustor air inlet temperature is increased above that of the nonrecuperated cycle, less fuel is required to heat the air to a given turbine inlet temperature. Because the turbine work and the compressor work are approximately the same as in the nonrecuperated cycle, the decrease in fuel flow results in an increase in thermal efficiency. For combined cycles the efficiency is also increased, because the gas turbine recovers the recuperated heat at the simple cycle efficiency, which is larger than the 30 to 35 percent thermal efficiency of a bottoming steam cycle, which recovers this heat in the nonrecuperated case. Since recuperative cycles return exhaust energy to the gas turbine, less energy is available to the steam cycle, and the resulting steam turbine output is lower than that of the baseline configuration. Even though the gas turbine output is approximately the same as in the baseline cycle (minus losses in the recuperation system), recuperative cycles carry a significant output penalty because of reduced steam turbine work, which is proportional to the amount of recuperation performed.

From a simple cycle standpoint, the combination of intercooling with recuperation eliminates the problem of the reduced combustor inlet temperature associated with intercooled cycles. The simple cycle then gets the benefit of the reduced compressor work and, at all but high pressure ratios, actually has a higher burner inlet temperature than the corresponding nonintercooled, nonrecuperated cycle. This results in a dramatic increase in the simple cycle efficiency.

Reheat gas turbines utilize a sequential combustion process in which the air is compressed, combusted, and expanded in a turbine to some pressure significantly greater than ambient, combusted again in a second combustor, and finally expanded by a second turbine to near ambient pressure. For a fixed turbine rotor inlet temperature limit, the simple cycle efficiency is increased for a reheat gas turbine compared to a nonreheat cycle operating at a pressure ratio corresponding to the second combustor's operating pressure. This is because the reheat cycle performs some of its combustion and expansion at a higher pressure ratio.

Thermochemical recuperation (TCR), also known as chemical recuperation, has been under evaluation for several years as a promising approach to increase power generation efficiencies. In a TCR power plant, a portion of the stack exhaust gas is removed from the stack, compressed, mixed with natural gas fuel, heated with exhaust heat from the gas turbine, and mixed with the air compressor discharge as it enters the combustor. As the mixture of natural gas and flue gas is heated by the gas turbine exhaust, a chemical reaction occurs between the methane in the fuel and the carbon dioxide and water in the flue gas. If this reaction occurs in the presence of a nickel-based catalyst, hydrogen and carbon monoxide are produced. For complete conversion of the methane, the effective fuel heating value is increased. Therefore, the natural gas/flue gas mixture absorbs heat thermally and chemically, resulting in a larger potential recuperation of exhaust energy than could be obtained by conventional recuperation, which recovers energy by heat alone. In fact, with full conversion of the natural gas fuel to hydrogen and carbon monoxide, up to twice the energy recuperated by the standard recuperative cycle may be recovered.

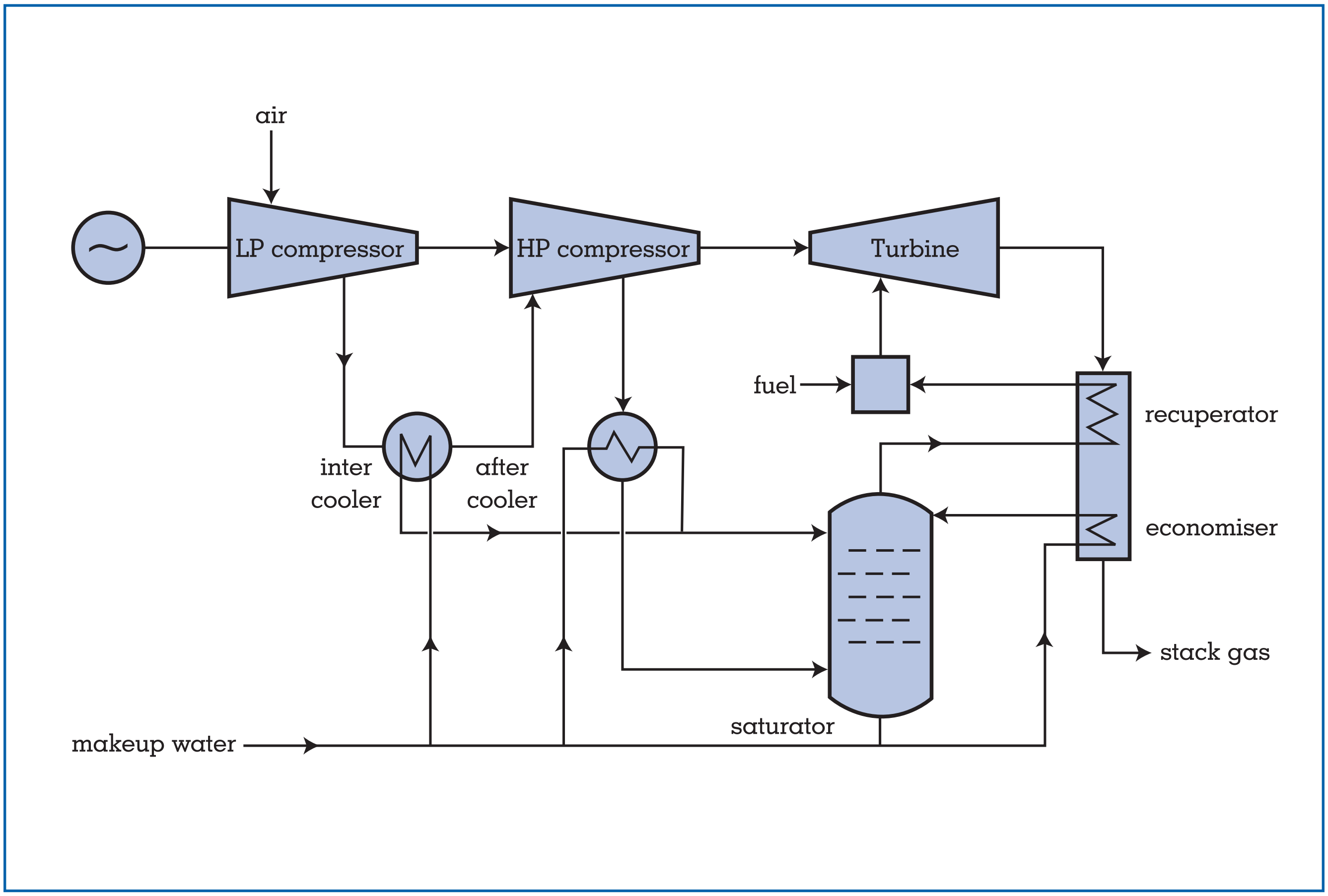


Figure 5.
HAT cycle flow diagram.
SOURCE: Morton and Rao, 1989.

Partial oxidation (PO) has also been proposed as a means to increase the performance of gas turbine power systems. PO is a commercial process used by the process industries to generate syngases from hydrocarbons, but under conditions differing from those in power generation applications. In this concept, a high-pressure, low-heating-value fuel gas is generated by partially combusting fuel with air. This fuel gas is expanded in a high-pressure turbine prior to being burned in a lower-pressure, conventional gas turbine. This process reduces the specific air requirements of the power system and increases the power output. PO has several potential advantages over conventional cycles or reheat cycles: lower specific air consumption and reduced air compressor work; increased power plant thermal efficiency; and potentially lower NOx emissions with improved combustion stability.

One difficulty associated with waste heat recovery systems is the satisfactory use of low grade heat to reduce stack losses. The humid air turbine (HAT), shown in Figure 5, uses a concept that should be considered on how to improve the use of low-grade heat. Warm waste is brought into contact with the compressor delivery air in a saturator tower to increase the moisture content and increase the total mass flow of the gases entering the turbine without significantly increasing the power demand from the compressor. Thus low-grade heat is made available for the direct production of useful power. Also, the high moisture content of the fuel gas helps to control NOx production during the combustion process.

Over a number of years, fuel cells have promised a new way to generate electricity and heat from fossil fuels using ion-exchange mechanisms. Fuel cells are categorized by operating temperature and the type of electrolyte they use. Each technology has its own operating characteristics such as size, power, density, and system configurations.

One concept, a solid oxide fuel cell (SOFC) supplied with natural gas and preheated oxidant air, produces direct-current electric power and an exhaust gas with a temperature of 850° to 925°C (1,560° to 1,700°F). In this hybrid system, the fuel cell contributes about 75 percent of the electric output, while the gas turbines contribute the remaining 25 percent. The SOFC exhaust can be expanded directly by a gas turbine with no need for an additional combustor. SOFC technology can be applied in a variety of ways to configure both power and combined heat and power (CHP) systems to operate with a range of electric generation efficiencies. An atmospheric pressure SOFC cycle, capable of economic efficiencies in the 45 to 50 percent range, can be the basis for a simple, reliable CHP system. Intergating with a gas turbine to form an atmospheric hybrid system, the simplicity of the atmospheric-pressure SOFC technology is retained, and moderately high efficiencies in the 50 to 55 percent range can be achieved in either power or CHP systems.

The pressurized hybrid cycle provides the basis for the high electric efficiency power system. Applying conventional gas turbine technology, power system efficiencies in the 55 to 60 percent range can be achieved. When the pressurized hybrid system is based on a more complex turbine cycle—such as one that is intercooled, reheated, and recuperated—electric efficiencies of 70 percent or higher are projected.

OTHER FUELS

The majority of today's turbines are fueled with natural gas or No. 2 distillate oil. Recently there has been increased interest in the burning of nonstandard liquid fuel oils or applications where fuel treatment is desirable. Gas turbines have been engineered to accommodate a wide spectrum of fuels. Over the years, units have been equipped to burn liquid fuels, including naphtha; various grades of distillate, crude oils, and residual oils; and blended, coal-derived liquids. Many of these nonstandard fuels require special provisions. For example, light fuels like naphtha require modifications to the fuel handling system to address high volatility and poor lubricity properties.

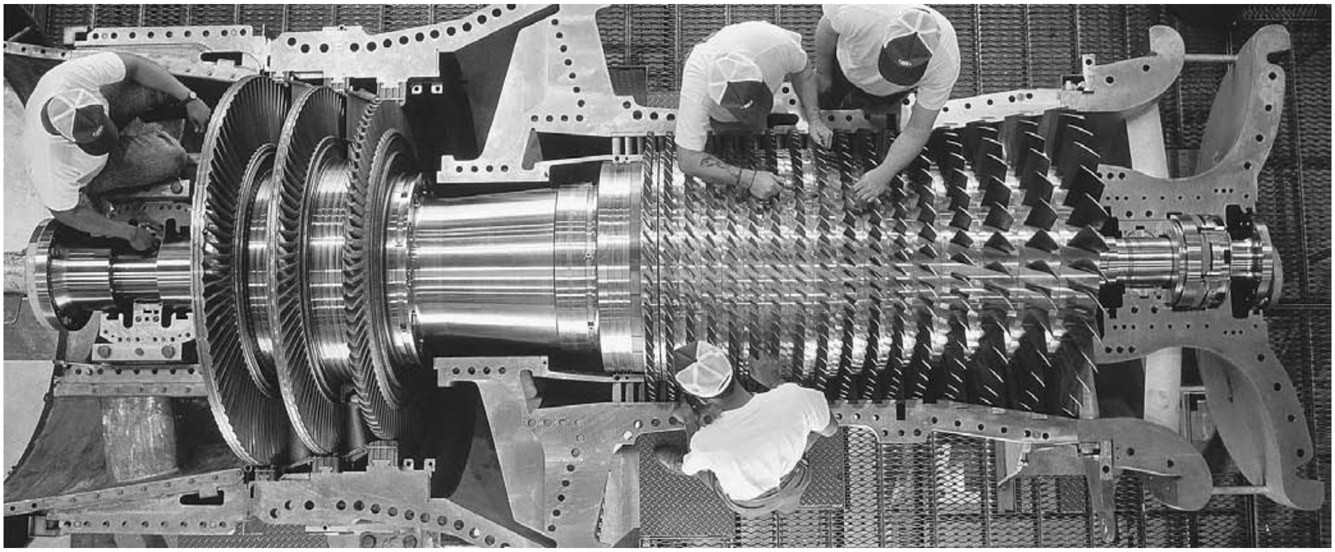
The need for heating, water washing, and the use of additives must be addressed when moving from the distillates toward the residuals. Fuel contaminants such as vanadium, sodium, potassium, and lead must be controlled to achieve acceptable turbine parts life. The same contaminants also can be introduced by the inlet air or by water/steam injection, and the combined effects from all sources must be considered.

The final decision as to which fuel type to use depends on several economic factors, including delivered price, cost of treatment, cost of modifying the fuel handling system, and increased maintenance costs associated with the grade of fuel. With careful attention paid to the fuel treatment process and the handling and operating practices at the installation, gas turbines can burn a wide range of liquid fuels. The ultimate decision on the burning of any fuel, including those fuel oils that require treatment, is generally an economic one.

Due to an estimated global recoverable reserve of more than 400 years of coal that can be used to generate electricity, future power generation systems must be designed to include coal. Today there are a number of gasification technologies available or being demonstrated at commercial capacity that are appropriate for power generation. Most of the development effort has been in the area of gasifiers that can be used in gasification combined cycle (GCC) applications. Other coal-fueled concepts under development include first- and second-generation pressurized fluidized bed (PFBC), and the indirect coal-fired case already discussed as a closed cycle.

Gasification is the reaction of fuel with oxygen, steam, and CO_2 to generate a fuel gas and low-residue carbon char. This char differentiates gasification from other technologies that burn residual char separately or apply char as a by-product. The gasification fuel gas is composed primarily of hydrogen and carbon oxides as well as methane, water vapor, and nitrogen. Generic classes of gasifiers that can be used in a GCC are moving bed, fluidized bed, and entrained flow.

GCC integrates two technologies: the manufacture of a clean-burning synthesis gas (syngas), and the use of that gas to produce electricity in a combined cycle power generation system. Primary inputs are hydrocarbon feeds (fuels), air, chemical additives, and water; primary GCC outputs are electricity (generated within the power block that contains a gas turbine, steam turbine, and a heat recovery steam generator), syngas, and sulfur by-products, waste, and flue gas. The flows



Workers assemble a 38-megawatt gas turbine at a General Electric plant in South Carolina. They build 100 turbines a year. (Corbis-Bettmann)

that integrate the subsystems include auxiliary power for an air separation unit and gasifier; air and nitrogen between an air separator and the gas turbine; heat from gasifier gas coolers to generate steam for the steam cycle; and steam to the gas turbine for NOx control.

Worldwide there are about thirty major current and planned GCC projects. Within the United States, the first GCC placed into commercial service, and operated from 1984 to 1989, was the 100-MW Cool Water coal gasification plant near Daggett, California.

Most commercial-size gasification projects have used oxygen rather than air as the oxidant for the gasifiers. Recent GCC evaluations have looked at using the gas turbine air compressor to supply air for the air separation unit. Typically this air stream is sent to a high-pressure air separator unit, which produces oxygen for gasification and high-pressure nitrogen for gas turbine NOx control. The diluent nitrogen lowers the flame temperature and therefore lowers the NOx.

Through Clean Coal Technology programs, the U.S. Department of Energy (DOE) is supporting several GCC demonstration projects that range in capacity from 100 to 262 MW. One of these plants is fueled by an advanced air-blown gasifier technology, and two of the projects will demonstrate energy-saving hot-gas cleanup systems for removal of sulfur and particulates. In Europe there are several commercial-size GCC demonstration plants in operation or under construction, ranging in size from 250 to 500 MW.

In a first-generation PFBC plant, the PFBC is used as the gas turbine combustor. For this application, the temperature to the gas turbine is limited to a bed temperature of about 870°C (1,600°F). This temperature level limits the effectiveness of this cycle as a coal-fired alternative.

In second-generation PFBC, a topping combustor is used to raise the turbine rotor inlet temperature to state-of-the-art levels. Pulverized coal is fed to a partial-gasifier unit that operates about 870° to 925°C (1,600° to 1,700°F) to produce a low heating value fuel gas and combustible char. The char is burned in the PFBC. The fuel gas, after filtration, is piped back to the gas turbine, along with the PFBC exhaust.

Fuel gas cleaning systems are being developed to fulfill two main functions: controlling the environmental emissions from the plant, and protecting downstream equipment from degraded performance. The fuel gas cleaning system also protects the gas turbine from corrosion, erosion, and deposition damage. Conventional GCC fuel gas cleaning systems, designated as cold gas cleaning, operate at temperatures of less than 315°C (600°F). Alternative technologies for fuel gas cleaning, which operate at considerably higher temperatures, are under development because of potential power plant performance benefits.

A biomass power generation industry is emerging that can provide substantial quantities of feedstock such as sugarcane residue (bagasse), sweet sorghum, rice straw, and switchgrass, as well as various wood byproducts. Today some of these residues are burned to generate power using a steam boiler and a steam turbine. However, operating problems result from ash agglomeration and deposition. Also, plant thermal efficiencies are relatively low, less than 30 percent, and in a number of plants less than 20 percent. The U.S. DOE has supported the development of biomass-fueled gasification systems that can be integrated into a combined cycle power plant and thereby obtain thermal efficiencies of greater than 40 percent.

Coal gasification is fuel flexible so that the process can use the most available feedstock at the best price. Gasifiers have successfully gasified heavy fuel oil and combinations of oil and waste gas. Other possible gasification feedstock includes petroleum coke, trash, used tires, and sewage sludge. Various combinations of feedstocks and coal have been successfully gasified.

Orimulsion is a relatively new fuel that is available for the gasification process. Orimulsion is an emulsified fuel, a mixture of natural bitumen (referred to as Orinoco-oil), water (about 30%), and a small quantity of surface active agents. Abundant Orinoco-oil reserves lie under the ground in the northern part of Venezuela.

ADVANCED GAS TURBINES

As discussed, the efficiency of a gas turbine cycle is limited by the ability of the combustion chamber and early turbine states to continuously operate at high temperatures. In 1992, the DOE started the Advanced Turbine Systems (ATS) Program that combined the resources of the government, major turbine manufactures, and universities to advance gas turbine technology and to develop systems for the twenty-first century. As pilot projects, two simple cycle industrial gas turbines are being developed for distributed generation and industrial and cogeneration markets, and two combined cycle gas turbines for use in large, baseload, central station, electric power generation markets.

DEVELOPMENT OF A HYDROGEN-FUELED GAS TURBINE

Looking to the future, the Japanese government is sponsoring the World Energy Network (WE-NET) Program through its New Energy and Industrial Technology Development Organization (NEDO). WE-NET is a twenty-eight-year global effort to define and implement technologies needed for hydrogen-based energy systems. A critical part of this effort is the development of a hydrogen-fueled gas turbine system to efficiently convert the chemical energy stored in hydrogen to electricity when hydrogen is combusted with pure oxygen. A steam cycle with reheat and recuperation was selected for the general reference system. Variations of this cycle have been examined to identify a reference system having maximum development feasibility while meeting the requirement of a minimum of 70.9 percent thermal efficiency. The strategy applied was to assess both a near-term and a long-term reference plant. The near-term plant requires moderate development based on extrapolation of current steam and gas turbine technology. In contrast, the long-term plant requires more extensive development for an additional high-pressure reheat turbine, has closed-loop steam cooling and extractive feedwater heating, and is more complex than the near-term plant.

OTHER GAS TURBINE APPLICATIONS

In addition to power generation and aircraft propulsion, gas turbine technology has been used for mechanical drive systems, gas and oil transmission pipelines, and naval propulsion. In natural gas pipelines, gas turbines provide mechanical pumping power, using the fluid being pumped as fuel. For marine applications, aero-derivative engines have been developed, yet a major disadvantage is its poor specific fuel consumption at part-load operation. For example, a naval vessel having a maximum speed of thirty knots and a cruise speed of fifteen knots loses considerable efficiency at cruising speed, since the cruise power will be only one-eighth of the maximum power (the power required being proportional to the cube of the speed). One alternative to reduce this concern is the use of a recuperator to heat the air going to the combustor with the heat from the gas turbine's exhaust. Another alternative to overcome high specific fuel consumption at part-load operation is to develop a shipboard combined cycle power plant consisting of gas turbines in conjunction with a steam turbine.

Gas turbines also have been considered for rail and road transportation. The Union Pacific successfully operated large freight trains from 1955 to 1975, several high-speed passenger trains and locomotives were built using aviation-type gas turbines. These, however, gave way to more economical diesels.

The first gas-turbine-propelled car (at 150 kW) was produced in the United Kingdom in 1950. For more than fifty years, significant efforts have been expended on automotive programs; however, diesel and gasoline engines continue to dominate. The major problem is still the poor part-load thermal efficiency of the gas turbine despite the use of a recuperated cycle with variable area nozzle guide vanes. Other problems include lack of sufficient high-temperature material development, and the relatively long acceleration time from idle to full load. For long-haul trucks, gas turbines were developed in the range of 200 to 300 kW. All used a low-pressure cycle ratio with a centrifugal compressor, turbine, and rotary heat exchanger.

A recent convergence of economic opportunities and technical issues has resulted in the emergence of a new class of gas turbine engines called microturbines. Designed for use in a recuperative cycle and pressure ratios of three to one to five to one, they can produce power in the range of 30 to 300 Kw. Initial work on this concept, which is primarily packaged today for cogeneration power generation, started in the late 1970s. In cogeneration applications that can effectively use the waste heat, overall system efficiencies can be greater than 80 percent. Manufacturers are exploring how microturbines can be integrated with fuel cells to create hybrid systems that could raise overall system efficiencies. Many issues, however, are still to be resolved with this approach, including cost and integration.

Another application of gas turbines is in a compressed air energy storage (CAES) system that allows excess base load power to be used as peaking power. It is similar to hydro-pumped storage, and the idea is to store energy during low demand periods by converting electrical power to potential energy. Rather than pumping water up a hill, CAES compresses and stores air in large underground caverns. When the power is needed, the air is allowed to expand through a series of heaters and turbines, and the released energy is then converted back to electricity. To increase output and efficiency, fuel is mixed with the air as it is released, the mixture is burned, and the energy released by combustion is available for conversion to electricity and heat recovery. This is similar to the operation of a standard gas turbine, except with CAES the compressor and turbines are separate machines that each run when most advantageous.

ROLE OF ADVANCED GAS TURBINE TECHNOLOGY

The electricity industry is in the midst of a transition from a vertically integrated and regulated monopoly to an entity in a competitive market where retail customers choose the suppliers of their electricity. The change started in 1978, when the Public Utility Regulatory Act (PURPA) made it possible for nonutility power generators to enter the wholesale market.

From various U.S. DOE sources, projections have been made that the worldwide annual energy consumption in 2020 could be 75 percent higher than it was in 1995. The combined use of fossil fuels is projected to grow faster from 1995 to 2020 than it did from 1970 to 1995. Natural gas is expected to account for 30 percent of world electricity by 2020, compared to 16 percent in 1996.

The power generation cycle of choice today and tomorrow is the combined cycle that is fueled with natural gas. Power generating technologies, regardless of the energy source, must maximize efficiency and address environmental concerns. To support the fuel mix and minimize environmental concerns, advanced coal combustion, fuel cells, biomass, compressed air energy storage, advanced turbine systems, and other technologies such as the development of a hydrogen-fueled cycle are under development.

Beyond the ATS program, the DOE is looking at several new initiatives to work on with industry. One, Vision 21, aims to virtually eliminate environmental concerns associated with coal and fossil systems while achieving 60 percent efficiency for coal-based plants, 75 percent efficiency for gas-based plants, and 85 percent for coproduction facilities. Two additional fossil cycles have been proposed that can achieve 60 percent efficiency. One incorporates a gasifier and solid oxide fuel into a combined cycle; the other adds a pyrolyzer with a pressurized fluidized bed combustor. Also under consideration is the development of a flexible midsize gas turbine. This initiative would reduce the gap between the utility-size turbines and industrial turbines that occurred during the DOE ATS program.

Ronald L. Bannister

See also: Cogeneration; Cogeneration Technologies; Locomotive Technology; Parsons, Charles Algernon; Storage; Storage Technology; Turbines, Steam; Turbines, Wind.

BIBLIOGRAPHY

Bannister, R. L.; Amos, D. J.; Scalzo, A. J.; and Datsko, S. C. (1992). "Development of an Indirect Coal-Fired High Performance Power Generation System." ASME Paper 92-GT-261.

Bannister, R. L.; Cheruvu, N. S.; Little, D. A.; and McQuiggan, G. (1995). "Development Requirements for an Advanced Gas Turbine System." *Transactions of the ASME, Journal of Engineering for Gas Turbines and Power* 117:724–733.

Bannister, R. L., and Horazak, D. A. (1998). "Preliminary Assessment of Advanced Gas Turbines for CVX." ASME Paper 98-GT-278.

Bannister, R. L.; Newby, R. A.; and Diehl, R. C. (1992). "Development of a Direct Coal-Fired Combined Cycle." *Mechanical Engineering* 114(12):64–70.

Bannister, R. L.; Newby, R. A.; and Yang, W.-C. (1999). "Final Report on the Development of a Hydrogen-Fueled Gas Turbine Cycle for Power Generation." *Transactions of the ASME, Journal of Engineering for Gas Turbines and Power* 121:38–45.

Briesch, M. S.; Bannister, R. L.; Diakunchak, I. S.; and Huber, D. J. (1995). "A Combined Cycle Designed to Achieve Greater Than 60 Percent Efficiency." *Transactions of the ASME, Journal of Engineering for Gas Turbine and Power* 117:734–741.

Cohen, H.; Rogers, G. F. C.; and Saravamuttoo, H. I. J. (1996). *Gas Turbine Theory*, 4th ed. Essex, Eng.: Longman House.

Diakunchak, I. S.; Bannister, R. L.; Huber, D. J.; and Roan, F. (1996). "Technology Development Programs for the Advanced Turbine Systems Engine." ASME Paper 96-GT-5.

Diakunchak, I. S.; Krush, M. P.; McQuiggan, G.; and Southall, L. R. (1999). "The Siemens Westinghouse Advanced Turbine Systems Program." ASME Paper 99-GT-245.

Horlock, J. H. (1992). *Combined Power Plants*. Oxford, Eng.: Pergamon Press.

Horlock, J. H. (1994). "Combined Power Plants—Past, Present and Future." *Transactions of the ASME, Journal of Engineering for Gas Turbines and Power* 117:608–616.

Horlock, J. H. (1997). *Cogeneration-Combined Heat and Power (CHP)*. Malabar, FL: Krieger.

Larson, E. D., and Williams, R. H. (1987). "Steam Injection Gas Turbines." *Transactions of the ASME, Journal of Engineering for Gas Turbines and Power* 106:731–736.

Layne, A. W., and Hoffman, P. A. (1999). "Gas Turbine Systems for the 21st Century." ASME Paper 99-GT-367.

Layne, A. W.; Samuelsen, S.; Williams, M. C.; and Hoffman, P. A. (1999). "Developmental Status of Hybrids." ASME Paper 99-GT-400.

Maude, C. W. (1993). "Advanced Power Generation: A Comparative Study of Design Options for Coal." *IEA Coal Research*. 55:15-32.

Maude, C. W., and Kirchner, A. T. (1995). "Gas Turbine Developments." *IEA Coal Research* 84:16-36.

Morton, T., and Rao, A. (1989). "Perspective for Advanced High Efficiency Cycles Using Gas Turbines." *EPRI Conference on Technologies for Producing Electricity in the Twenty-first Century* 7:1–7.

Newby, R. A., Bannister, R. L. (1994). "Advanced Hot Gas Cleaning System for Coal Gasification Processes." *Transactions of the ASME, Journal of Engineering for Gas Turbines and Power* 116:338–344.

Newby, R. A., and Bannister, R. L. (1998). "A Direct Coal-Fired Combustion Turbine Power System Based on Slagging Gasification with In-Site Gas Cleaning." *Transactions of the ASME, Journal of Engineering for Gas Turbine and Power* 120:450–454.

Newby, R. A.; Yang, W.-C.; and Bannister, R. L. "Use of Thermochemical Recuperation in Combustion Turbine Power Systems." ASME Paper 97-GT-44.

Richerson, D. W. (1997). "Ceramics for Turbine Engines." *Mechanical Engineering* 119(9):80–83.

Scalzo, A. J.; Bannister, R. L.; Decorso, M.; and Howard, G. S. (1996). Evolution of Westinghouse Heavy - Duty Power Generation and Industrial Combustion Turbines. *Transactions of the ASME, Journal of Engineering for Gas Turbines and Power* 118(2):361–330.

Scott, P. (1995). "Birth of the Jet Engine." *Mechanical Engineering* 117(1):66–71.

Touchton, G. (1996). "Gas Turbines: Leading Technology for Competitive Markets." *Global Gas Turbine News* 36(1):10–14.

Valenti, M. (1993). "Propelling Jet Turbines to New Users." *Mechanical Engineering* 115(3):68–72.

Valenti, M. (1995). "Breaking the Thermal Efficiency Barrier." *Mechanical Engineering* 117(7):86–89.

Valenti, M. (1997). "New Gas Turbine Designs Push the Envelope." *Mechanical Engineering* 119(8):58–61.

Watts, J. H. (1999). "Microturbines: A New Class of Gas Turbine Engine." *Global Gas Turbine News* 39:4–8.

Williams, M.; Vanderborgh, N.; and Appleby, J. (1997). "Progress Towards a Performance Code for Fuel Cell Power Systems." *ASME Joint Power Generation Conference* 32:423–427.

Wilson, D. G., and Korakianitis, T. (1998). *The Design of High-Efficiency Turbomachinery and Gas Turbines*, 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Wilson, J. M., and Baumgartner, H. (1999). "A New Turbine for Natural Gas Pipelines." *Mechanical Engineering* 121(5):72–74.

Zipkin, M. A. (1977). "Evolution of the Aircraft Gas Turbine Engine." *Israel Journal of Technology* 15:44–48.

TURBINES, STEAM

EVOLUTION OF AN INDUSTRY

Since the turn of the twentieth century, the steam turbine has evolved from an experimental device to the major source of electrical generation. Practical steam turbine inventions coincided with the development of direct-current electric dynamos first used to power arc-lighting street systems. In the United States, the first central station to provide electrical lighting service was Thomas Edison's Pearl Street Station in New York City in 1882. Powered by 72 kW of steam engines, it served 1,284 16-candlepower dc lamps. This installation demonstrated the feasibility of central station electricity. Initially, Edison's system needed a large number of scattered power plants because it utilized direct current, and dc transmission was uneconomical over large distances. In 1885, George Westinghouse's Union Switch & Signal Co. acquired rights to manufacture and sell a European-design transformer, and the company then developed alternating-current distribution capability to utilize its transformers, which made longer-distance transmission of electricity practical. The Westinghouse Electric Co. was formed to exploit this device. By 1900 there were numerous dc and a few ac generating stations in the United States, all with reciprocating steam engines or hydraulic turbines as prime movers. However, the ac technology quickly became a primary factor in stimulating the development of power generation.

Although they were reliable, the early steam engines were huge, heavy devices that were not very efficient. Nearly all the companies in the electric equipment business seized the opportunity to develop the steam turbine as an alternative. In 1895, Westinghouse acquired rights to manufacture reaction turbines invented and patented in 1884 by the English inventor Charles Algernon Parsons. Allis-Chalmers also acquired rights to manufacture under Parsons' patents, so early machines of these two manufacturers were quite similar. In 1887, the General Electric Co. (founded by Edison) entered into an agreement with Charles Curtis to exploit his steam turbine patent.

The Curtis and the Parsons turbine designs are based on different fundamental principles of fluid flow. The Curtis turbine has an impulse design, where the steam expands through nozzles so that it reaches a high velocity. The high-velocity, low-pressure steam jet then impacts the blades of a spinning wheel. In a reaction turbine such as the Parsons design, the steam expands as it passes through both the fixed nozzles and the rotating blades. High pressure stages are impulse blades. The high-pressure drops quickly through these stages, thus reducing the stress on the high pressure turbine casing. The many subsequent stages may be either impulse or reaction designs.

STEAM TURBINE CYCLES

The basic function of a steam turbine is to efficiently convert the stored energy of high-pressure, high-temperature steam to useful work. This is accomplished by a controlled expansion of the steam through stages consisting of stationary nozzle vanes and rotating blades (also called buckets by one major manufacturer). The size and shape of the nozzle vanes and rotating blades are such as to properly control the pressure distribution and steam velocities throughout the turbine flow path. Blading improvements have increased turbine cycle efficiency by reducing profile losses, end-wall losses, secondary flow losses, and leakage losses. Use of tapered twisted designs for longer blades reduces losses on the innermost and outmost portions of the blades.

A complete turbine generator unit could consist of several turbine elements connected in tandem on a single shaft to drive a generator. To extract as much energy from the steam as possible, as it decreases in temperature and pressure in its passage through the machine, the typical arrangement could include a high-pressure (HP), an intermediate-pressure (IP), and one or more low-pressure (LP) elements, as illustrated in Figure 1.

The HP, IP, and LP turbines may be either single-flow or double-flow designs, depending upon the volume of steam utilized. In a single-flow turbine, the total volume of steam enters at one end and exhausts at the other end. The double flow is designed so the steam enters at the center and divides. Half flows in one direction, and half in the other direction into exhausts at each end of the turbine.

The basic steam cycle for a steam turbine installation is called a Rankine cycle (named after Scottish engineer and physicist William John Macquorn Rankine). This cycle consists of a compression of liquid water, heating and evaporation in the heat source (a steam boiler or nuclear reactor), expansion of the

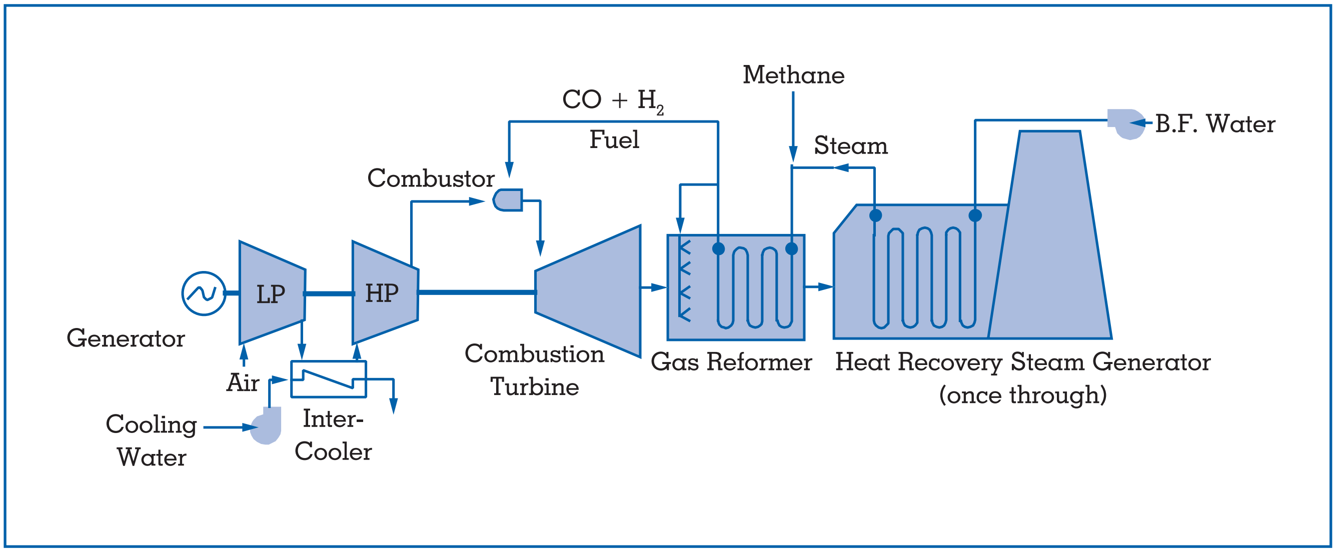


Figure 1.
Diagram of a complete turbine generator unit.

steam in the prime mover (a steam turbine), and condensation of the exhaust steam into a condenser. There is a continuous expansion of the steam, with no internal heat transfer and only one stage of heat addition. By increasing the pressure and/or temperature and decreasing the heat rejected by lowering exhaust temperature, cycle efficiency can be improved.

Weir patented a regenerative feedwater heating cycle in 1876. The regenerative Rankine cycle eliminates all or part of the external heating of the water to its boiling point. In this cycle, a small amount of expanded steam is extracted from a series of pressure zones during the expansion process of the cycle to heat water in a multiplicity of heat exchangers to a higher temperature. Theoretical and practical regenerative cycles reduce both the heat added to the cycle and the heat rejected from the cycle.

Reheat involves steam-to-steam heat exchange using steam at boiler discharge conditions. In the reheat cycle, after partially expanding through the turbine, steam returns to the reheater section of the boiler, where more heat is added. After leaving the reheater, the steam completes its expansion in the turbine. The number of reheats that are practical from a cycle efficiency and cost consideration is two.

EVOLUTION OF THE STEAM TURBINE

The first central station steam turbine in the United States was built for the Hartford Electric Light Co. in 1902. Steam conditions for this 2,000-kW unit and similar units were approximately 1.2 MPa (180 psig) and 180°C (350°F). The evolution of steam turbine power generation in the United States is summarized in Figure 2. Plotted against time are the maximum inlet steam pressure and temperature, along with plant thermal efficiency, and maximum shaft output in megawatts. The steady increase in steam turbine inlet pressure and temperature achieved an increase in plant thermal performance. From 1910 to 1920 steam turbine generators were manufactured in the 30- to 70-MW range. By 1945, the median unit sold in the United States was still only 100 MW. By 1967 the median unit had increased to 700 MW, with a peak of 1,300 MW for several fossil-fueled units placed in service in the 1970s. (A 1,300-MW unit can generate enough electricity to supply the residential needs of more than 4 million people.) During the first fifty years of the twentieth century, inlet steam pressure and temperature increased at an average rate per year of 0.3 MPa (43 psi) and 7°C (13°F), respectively. Until the early 1920s, throttle pressures were 1.4–2.1 MPa (200–300 psi), and throttle temperatures did not exceed 600°F (315°C). Above 450°F (230°C), cast steel replaced cast iron for turbine casings, valves, and so on.

Figure 3 shows the thermal performance evolution of the steam cycle as a function of material development and cycle improvements, starting in 1915. By the early 1920s, regenerative feedheating was well estab-

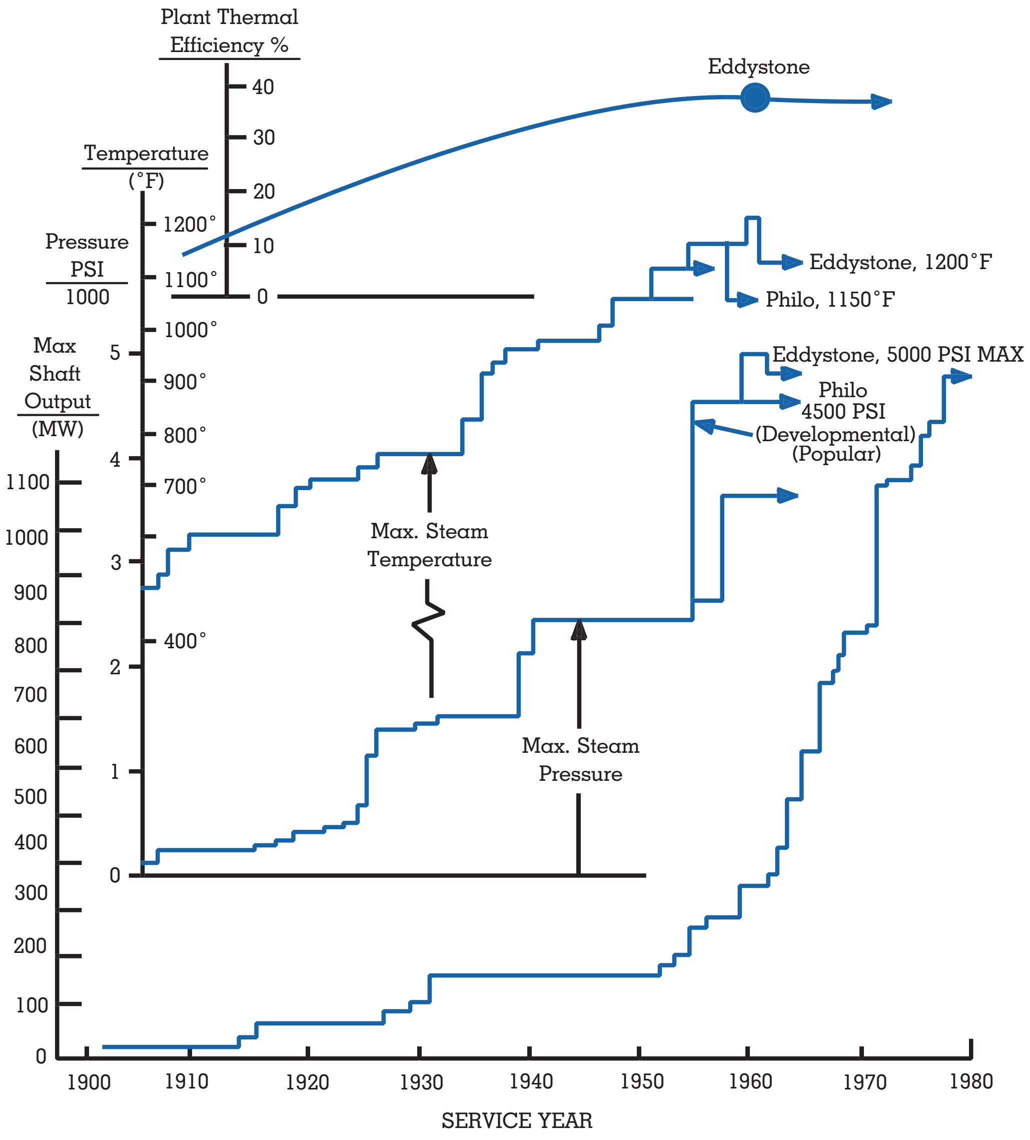
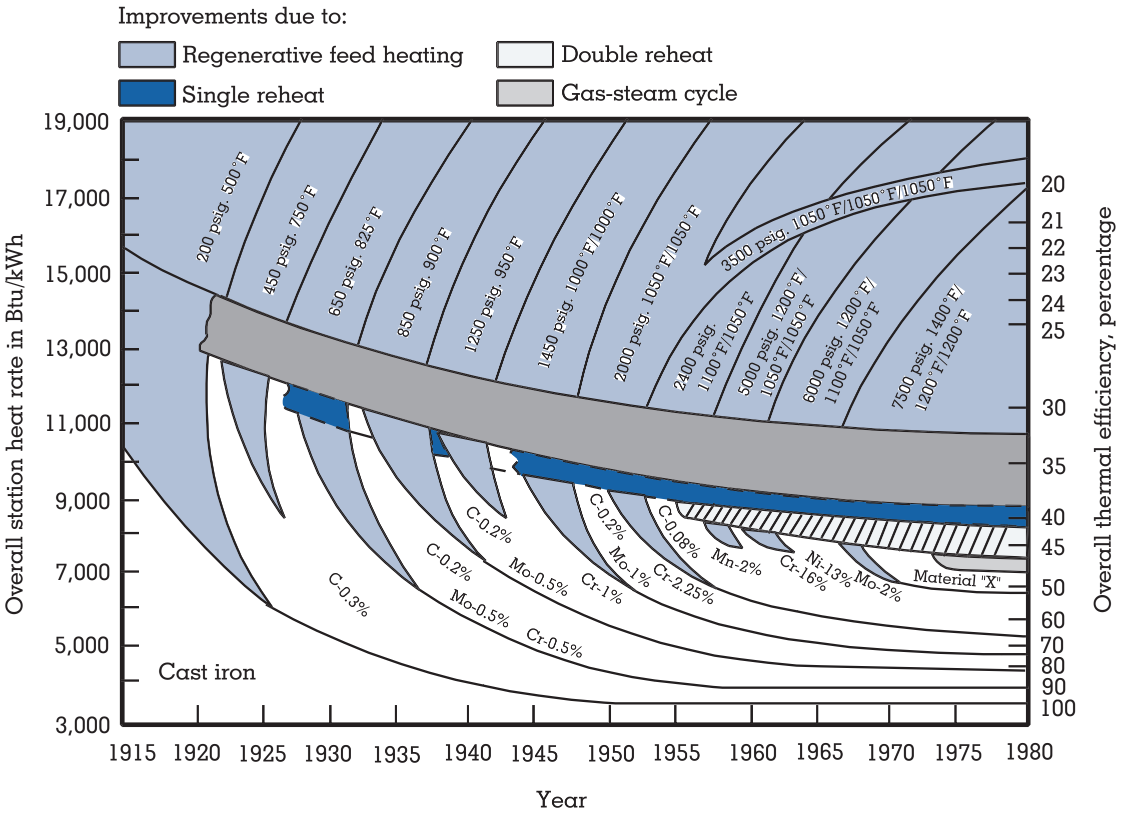


Figure 2.
The evolution of steam turbine power generation in the United States.

lished. Reheat cycles came into use in the mid-1920s. At the throttle temperature of 370°C (700°F) that was current when the pioneer 4.1- and 8.2-MPa (600- and 1,200-psi) units went into service, reheat was essential to avoid excessive moisture in the final turbine stages. As temperatures rose above 430°C (800°F) molybdenum proved effective. Using carbon-moly steels, designers pushed temperatures beyond 480°C (900°F) by the late 1930s. As a result of rising throttle temperatures, reheat fell out of use. By the late 1940s, reheat was reintroduced to improve plant efficiency, and second reheats appeared by the early 1950s.

Over the years, exhaust area was a major limitation on size. The earliest answer was the double-flow



Evolution of the steam cycle.

Figure 3.
Evolution of the Steam Cycle in the United States.

single-casing machine. Cross-compounding, introduced in the 1950s, represented a big step forward. Now the speed of the LP unit could be reduced. Thus, the last-stage diameter could be greater, yielding more exhaust annulus area.

Continued advances in metallurgy allowed inlet steam conditions to be increased, as illustrated in Figure 3. In 1959 this progress culminated with the Eddystone 1 unit of the Philadelphia Electric Co. and Combustion Engineering. With initial steam conditions of 34 MPa and 650°C (5,000 psi and 1,200°F) and two reheats of 1,050°F (570°C), Eddystone 1 had the highest steam conditions and efficiency (40%) of any electric plant in the world. The generating capacity (325 MW) was equal to the largest commercially available unit at the time. Eddystone 1 has operated for many years with throttle steam conditions of approximately 4,700 psi and 1,130°F (32 MPa, 610°C), and achieved average annual heat rates comparable to today's best units.

Figure 3, initially published in 1954, summarizes the evolution of the steam cycle from the year 1915 through the steam conditions of Eddystone 1 (average station heat rates were used) and a projection of where the industry might be by the year 1980. The relationship of operating steam pressures and temperatures to available materials in this figure indicates how the increase in pressure and temperature is dependent upon metallurgical development. The magnitudes of heat rate gains resulting from the application of various kinds of steam cycles also are shown.

Since the early 1960s, advanced steam conditions have not been pursued. In the 1960s and early 1970s there was little motivation to continue lowering heat rates of fossil-fired plants due to the expected increase in nuclear power generation for base-load application and the availability of relatively inexpensive fossil fuel. Therefore the metallurgical development required to provide material "X" for advanced steam conditions was never undertaken.

Raising inlet pressure and temperature increases the cycle's available energy and thus the ideal efficiency. However, pressure increases reduce the blade heights of the initial stages and decrease ideal efficiency, offsetting some of the ideal improvement, unless unit rating is increased commensurately. Based on potential heat rate improvement, there is no reason to raise the turbine steam conditions above 48.2 MPa, 760°/760°/593°C (7,000 psi, 1,400°/1,400°/1,100°F).

NUCLEAR POWER APPLICATIONS

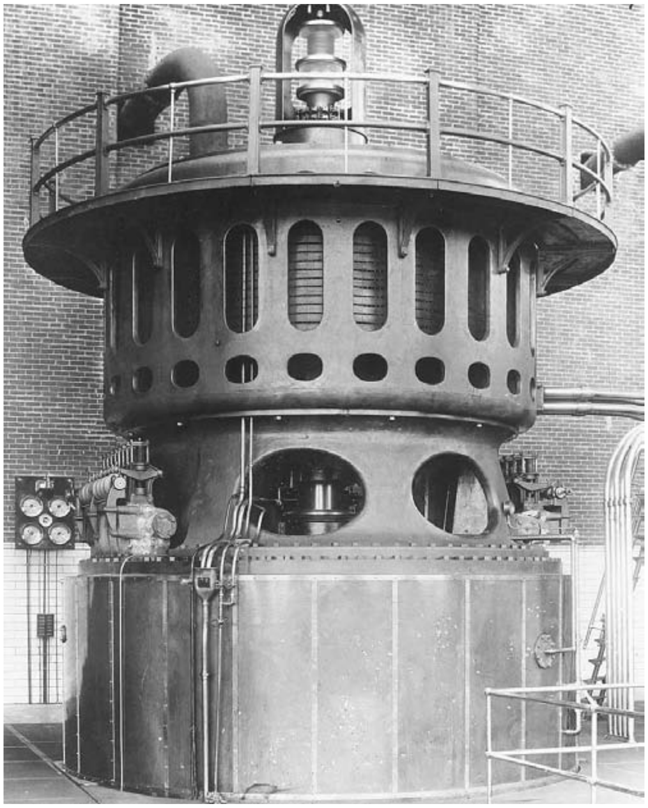
The first steam turbine generator for nuclear power application was placed in service at Duquesne Light Co.'s Shippingport Station in 1957. Initially rated at 60 MW, it had a maximum turbine capability rating of 100 MW. This was the first of a series of 1,800-rpm units, which were developed from a base design and operating experience with fossil-fuel machines dating back to the 1930s. Inlet steam conditions at maximum load were 3.8 MPa (545 psi), with a 600,000 kg (1.3 million-lb.) per-hour flow. This single-case machine had a 1-m (40-in.)-long last-row blade. Second-generation nuclear turbines introduced reheat at the crossover zones for improved thermal performance. Nuclear turbines, ranging in size from 400 to 1,350 MW, have used multiple LP exhausts.

Since moisture is a major concern for turbines designed for nuclear operation, a number of erosion-control techniques were used in the LP turbine. For example, adequate axial spacing between the stationary and the rotating blades minimizes blade erosion. Moisture is removed at all extraction points in the moisture region.

Nuclear turbines designed for use with a boiling-water reactor will be radioactive. Radioactivity could build up in the turbine because of the accumulation of corrosion products. A fuel rod rupture could result in highly radioactive materials entering the turbine. Therefore internal wire-drawing-type leakage paths, ordinarily unimportant in steam turbine design, must be eliminated as much as possible. Where it is impossible to eliminate, the surfaces forming the leakage paths should be faced with erosion-resistant materials deposited by welding.

FUTURE ROLE FOR STEAM TURBINE POWER GENERATION

In 2000, the power generation cycle of choice is the combined cycle, which integrates gas and steam cycles. The fuel of choice for new combined-cycle power generation is generally natural gas. Steam turbines designed to support the large gas turbine land-based power generation industry are in the range of 25 to 160 MW. Steam conditions are in the range of 12.4 MPa (1,800 psi) and 16.6 MPa (2,400 psi) and 538°C (1,000°F) and 565°C (1,050°F).



A Volt Curtis steam turbine, by GE from the designs of Charles G. Curtis, was installed for the Twin City Rapid Transit Company in Minneapolis and photographed in around 1900. (Corbis-Bettmann)

Large natural gas-fired combined cycles can reach a cycle efficiency of 58 percent, higher than a typical steam turbine power plant efficiency of 36 to 38 percent. During the 1990s, experience with optimizing advanced supercritical steam turbine cycles led to thermal cycle efficiencies in the 40 to 48 percent range for units rated at 400 to 900 MW that are existing or planned to be located in Japan, Denmark, Germany, and China.

In Denmark, two seawater-cooled 400 MW units, operating at a steam inlet pressure of 28.5 MPa (4,135 psi), have an efficiency of about 48 percent. Chubu Electric, in Japan, has been operating two 700 MW units since 1989 with inlet steam conditions of 31 MPa (4,500 psi). The efficiency gains of these two units is about 5 percent more than that of previous conventional plants of comparable size at 45 percent.

With an estimated 400 years of coal available for future power generation, coal-powered steam turbines are expected to continue to dominate global electricity fuel markets.

In the United States, coal had a 57 percent share of the electric power fuel market in 2000, up from 46 percent in 1970. This amounts to 430,000 MW generated by steam turbines that are fueled with coal. When considering other sources of generating steam for electric power—such as nuclear reactors, gas- or oil-fired broilers, and waste heat from gas turbines—steam turbines now comprise more than 600,000 MW of capacity, or approximately 75 percent of all generating capacity in the United States.

Since 1900, manufacturers have made many step changes in the basic design of steam turbines. New technology and materials have been developed to support the industry's elevation of steam conditions, optimization of thermal cycles and unit capacity. Steam turbines will continue to be the principal prime mover for electricity generation well into the twenty-first century.

Ronald L. Bannister

See also: Parsons, Charles Algernon; Rankine, William John Macquorn; Steam Engines; Turbines, Gas; Turbines, Wind.

BIBLIOGRAPHY

Artusa, F. A. (1967). "Turbine and Cycles for Nuclear Power Plant Applications." *Proceedings, American Power Conference* 29:280–294.

Bannister, R. L.; Silvestri, G. J., Jr.; Hizume, A.; and Fujikawa, T. (1987). "High-Temperature, Supercritical Steam Turbines." *Mechanical Engineering* 109(2):60–65.

Bennett, S. B.; and Bannister, R. L.; Silvestri, G. J., Jr.; and Parkes, J. B. (1981). "Current Day Practice in the Design and Operation of Fossil-Fired Steam Power Plants in the U.S.A." *Combustion* 52(7):17–24.

Bennett, S. B., and Bannister, R. L. (1981). "Pulverized Coal Power Plants: The Next Logical Step." *Mechanical Engineering* 103(12):18–24.

Church, E. F., Jr. (1950). *Steam Turbines.* New York: McGraw-Hill.

Keller, E. E., and Hodgkinson, F. (1936). "Development by the Westinghouse Machine Company." *Mechanical Engineering* 58(11):683–696.

Miller, E. H. (1984). "Historical Perspectives and Potential for Advanced Steam Cycles." *Electric Forum* 10(1):14–21.

Mochel, N. L. (1952). "Man, Metals, and Power." *Proceedings, American Society Testing Materials* 52:14–33.

Robinson, E. L. (1937). "Development by the General Electric Company." *Mechanical Engineering* 59(4):239–256.

Silvestri, G. J., Jr.; Bannister, R. L.; Fujikawa, T.; and Hizume, A. (1992). "Optimization of Advanced Steam Conditions Power Plants." *Transactions of the ASME, Journal of Engineering for Gas Turbine and Power* 114:612–620.

Smith, D. (1999). "Ultra-Supercritical CHP: Getting More Competitive." *Modern Power Systems* 19(1):21–30.

Somerscales, E. F. C. (1992). "The Vertical Curtis Steam Turbine." *Newcomen Society of Engineering Technology Transactions* 63:1–52.

Zink, J. (1996). "Steam Turbines Power and Industry." *Power Engineering* 100(8):24–30.

TURBINES, WIND

Harnessing the wind to do work is not new. In 3000 B.C.E. wind propelled the sail boats of ancient peoples living along the coasts of the Mediterranean Sea. The Swiss and French used wind-powered pumps in 600, which was shortly followed by windmills used to make flour from grain. By 1086, there were 5,624 water mills south of the Trent and the Severn rivers in England. Holland alone once had over 9,000. As late as the 1930s in the United States, windmills were the primary source of electricity for rural farms all across the Midwest. Although the advent of fossil fuel technologies shifted energy production from animal, wind, and water, dependence upon these finite fossil energy sources and concern over atmospheric pollutants, including carbon dioxide, are causing a resurgence of interest in early energy sources such as wind.

Wind turbines in use at the turn of the twenty-first century primarily produce electricity. In developed countries where electric grid systems connect cities, town, and rural areas, wind farms consisting of numerous wind turbines produce electricity directly to the electric grid. In developing countries, remote villages are not connected to the electric grid. Smaller wind turbines, singly or in groups, have tremendous potential to bring electricity to these remote locations without requiring the significant investment in trans-

Egg-beater shaped vertical axis wind turbine (VAWT) at the test site at Sandia Laboratories. (U.S. Department of Energy) mission lines that would be required for grid connection. Turbines range in size from less than one meter in diameter to more than 50 meters in diameter, with power ratings from 300 watts to more than one megawatt. This wide variation enables village systems to be sized to meet specific electrical demand in the kilowatt range, while large, grid-connected utility applications produce many megawatts of electricity with a few large wind turbines.

HOW WIND TURBINES WORK

Wind is the motion of the atmosphere, which is a fluid. As the wind approaches an airfoil-shaped object, the velocity changes as the fluid passes the object, creating a pressure gradient from one side of the object to the other. This pressure gradient creates a net force on one side of the object, causing it to move in the fluid. When wind hits the airfoil-shaped blade of a turbine, the lift force that is created causes the blade to rotate about the main shaft. The main shaft is connected to an electric generator. When the rotor spins due to forces from the wind, the generator creates electricity that can be fed directly into the electric grid or into a system of batteries.

Aerodynamic drag has also been used to capture energy from the wind. Drag mechanisms consist of flat or cup-shaped devices that turn the rotor. The wind simply pushes the device around the main shaft. Anemometers used to measure wind speed are often drag devices, as are traditional farm windmills.

Airplane propeller analysis relies upon the "axial momentum" theory, which is based on energy, momentum, and mass conservation laws. This theory has been applied to wind turbines as well. The power (P_w) of a fluid passing across an area perpendicular to the flow is

$$P_w = 1/2\ \rho\ A\ V_w^{\ 3}$$

where ρ is the air density, A is the disk area perpendicular to the wind, and V_w is the wind speed passing through the disk area. For instance, if the wind speed is 10 m/s and the rotor area is 1,200 m^2, the available power is 600 kW. When the wind speed doubles to 20 m/s, the available power increases to 4,800 kW. This value represents the total power available in the wind, but the turbine cannot extract all of that power. If the turbine were able to extract all the available power, the wind speed would drop to zero downwind of the rotor.

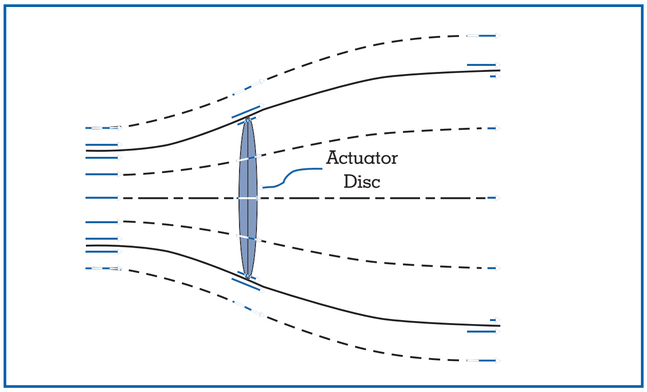


Figure 1.
Disk flow field model of an actuator.

A simple, ideal model of fluid flow through a rotor was used by both F. W. Lanchester and A. Betz (Lanchester, 1915; Betz, 1920) to study the limitation of power extracted from the wind. This "actuator disk" model is shown in Figure 1. It assumes that a perfectly frictionless fluid passes through an actuator disk, which represents the wind turbine. The fluid approaching the actuator disk slows, creating a pressure greater than atmospheric pressure. On the downwind side of the disk, the pressure drops below atmospheric pressure due to extraction of energy from the impinging fluid. As the fluid moves further downstream of the turbine, atmospheric pressure is recovered. Using axial momentum theory, Betz and Lanchester independently showed that the maximum fraction of the wind power that can be extracted is 16/27 or about 59 percent. This is known as the Lanchester/Betz limit, or, more commonly, the Betz limit, and is assumed to be an upper limit to any device that extracts kinetic energy from a fluid stream. Real wind turbines capture from 25 percent to more than 40 percent of the energy in the wind. More refined engineering analyses account for fluid friction, or viscosity, and wake rotation (Eggleston and Stoddard, 1987; Hansen and Butterfield, 1993).

TURBINE DESIGNS

There are two major types of wind turbines: horizontal-axis and vertical-axis. A wind turbine that rotates about an axis parallel to the wind is a horizontal-axis wind turbine (HAWT). Although HAWTs have not been proven clearly superior to Vertical-Axis Wind Turbines (VAWTs), they have dominated recent

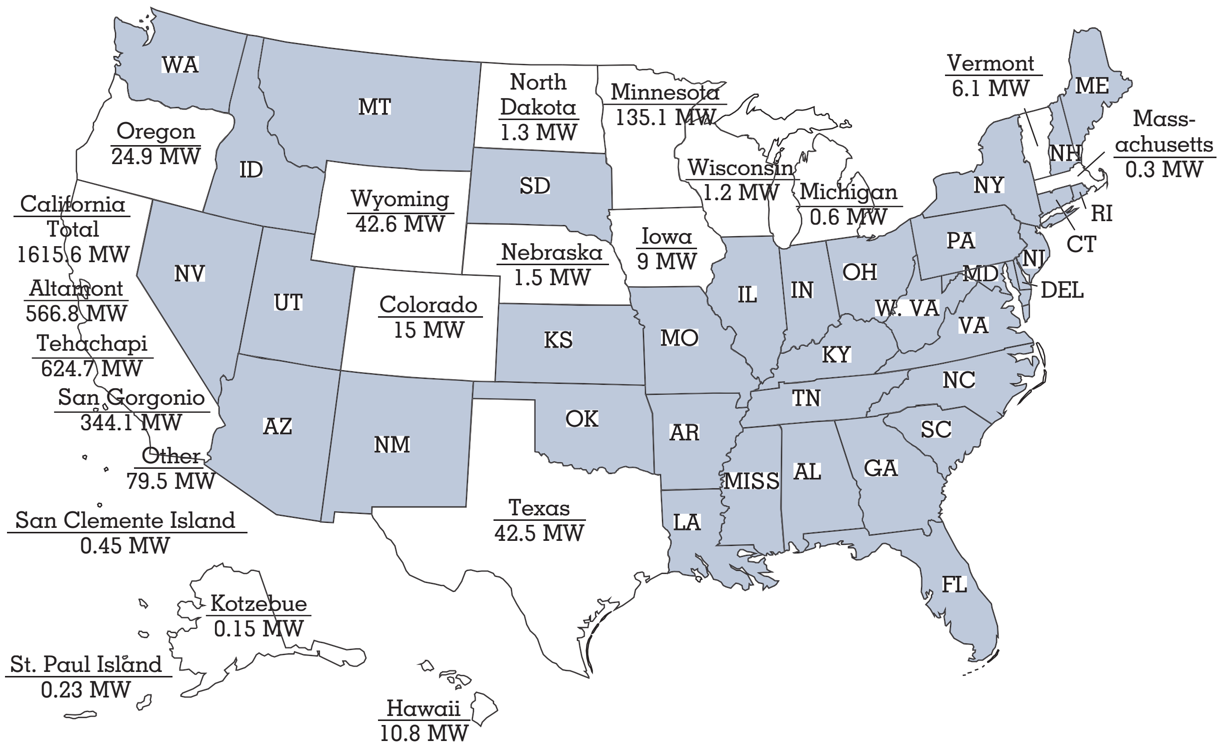


Figure 2.
U.S. capacity as of June 1999.

installations. Of all the utility-scale turbines installed today, 97 percent are HAWTs. A variety of configurations exist. HAWT rotors differ in orientation (upwind or downwind of the tower), flexibility (rigid or teetered), and number of blades (usually two or three).

Horizontal-Axis Wind Turbines (HAWTs)

An upwind turbine rotates upwind of the tower. In order to maintain the rotor's upwind position, a yaw mechanism is required to position the rotor as the wind direction shifts. In this configuration, the wind flowing toward the rotor is unobstructed over the entire rotor. Conversely, the wind flowing toward a downwind rotor is obstructed by the turbine tower over part of the rotor area. This causes fluctuating loads on each blade as it passes behind the tower, which can decrease the fatigue life of the rotor. The downwind turbine, however, aligns itself with the prevailing wind passively, eliminating the need for additional yaw drive components.

Flexibility at the rotor hub has been used to alter the load conditions on the blades. A rigid hub turbine generally has two or three blades attached to the hub. A teetered rotor, however, consists of two blades attached to the hub, which forms a hinge with the main shaft of the turbine. When the wind speed above the main shaft is higher than that below the main shaft, the rotor moves slightly downwind over the top half of the rotational cycle to reduce the loads on the blades. Although the cyclic loads are reduced by teetering the rotor, when the wind causes the rotor to exceed its maximum flexibility range hitting the teeter stops, large, transient loads can be introduced.

Although multiple-blade turbines are effective, two- and three-blade rotors are most cost-effective. Two-blade rotors commonly use teetering hinges, and all three-blade rotors are mounted on rigid hubs. Rotors up to 15 m in diameter are economically fea-

Pacific Gas & Electric engineers at verticle axis wind turbines at Altamont Pass. (Corbis-Bettmann)

sible and simplified with three blades. For large rotors exceeding 30 m in diameter, the blade weight is directly related to turbine costs. Thus, reducing the number of blades from three to two results in lower system costs with respect to the potential power available in the wind. For mid-range turbines, 15 m to 30 m in diameter, the trade off between power production and reduced cost as a result of reduced blade weight is more difficult to determine.

Most turbines are designed to rotate at a constant speed over a specific range of wind speed conditions. The generators in these turbines produce electricity compatible with the established grid system into which electricity is fed. Operating the turbine at variable rotor speeds increases the range of wind speeds over which the turbine operates. The amount of energy produced annually is increased as well. However, sophisticated power electronics is required to convert the electricity to the grid standard frequency.

Vertical-Axis Wind Turbines

A vertical-axis wind turbine (VAWT) rotates about an axis perpendicular to the wind. The design resembling an eggbeater was patented by D. G. M. Darrieus, a French inventor, in the 1920s. Because the axis of rotation is perpendicular to the ground, components such as the generator and gearbox are closer to the ground, making servicing these turbines fairly easy. Also, the turbine is not dependent upon its position relative to the wind direction. Since the blades cannot be pitched, these turbines are not self-starting. The greatest disadvantage of VAWTs is the short machine lifetime. The curved blades are susceptible to a variety of vibration problems, that lead to fatigue damage.

A modern VAWT that relies upon aerodynamic drag is known as a Savonius wind turbine. Sigurd Savonius, a Finnish inventor, developed this design in 1924. Two S-shaped panels are arranged to cup the wind and direct it between the two blades. This recirculation improves the performance of this drag device, but at best only 30 percent of the power available in the wind can be extracted.

UTILITY-SCALE APPLICATIONS

Utility-scale wind turbines range in size from 25 m to more than 50 m in diameter, with power ratings from 250 kW to 750 kW. Modern, electricity-producing wind turbines are often placed in large areas called wind farms or wind power plants. Wind farms consist of numerous, sometimes hundreds, of turbines regularly spaced on ridges or other unobstructed areas. The spacing varies depending upon the size of the turbine. Until recently California had the largest wind farms in the world. In the late 1990s, installations with capacities of 100–200 MW in the midwestern United States. Figure 2 shows the producing wind power plants in the U.S. resulting in a total capacity of 2.471 MW in 2000. This represents a small fraction of the 750 GW generating capacity in the U.S. Globally, wind power exceeded 10 GW of installed capacity in 1999. Much of the worldwide growth in wind energy use is in Europe, particularly Denmark, Germany, and Spain. To reduce pollution, many European countries subsidize wind generated electricity.

Future wind turbines will exploit the accomplishments of many years of U.S. government-supported research and development, primarily at the national laboratories, over the past twenty-five years. During the 1980s, a series of airfoils specifically designed for wind turbine applications were shown to increase annual power production by 30 percent. Detailed studies of the aerodynamics of wind turbines operating in the three-dimensional environment have led to improved models that are used for turbine design. These modeling capabilities allow turbine designers to test new concepts on paper before building prototype machines.

Several avenues for improving power production and cutting costs are being pursued by turbine designers. Building taller towers places the turbine in higher wind regimes, which increases the potential power production. In addition to taller towers, larger rotor diameters will improve power production. Increasing blade flexibility will reduce loads that reduce the fatigue life of blades, but sophisticated dynamic models are required to make such designs. Sophisticated control algorithms will monitor the turbines in order to accommodate extreme load conditions.

SMALL-SCALE APPLICATIONS

Because one-third of the world's population does not have access to electricity, many countries lacking grid systems in remote, rural areas are exploring various methods of providing citizens with access to electricity without costly grid extensions. Wind turbines are an intermittent source of electricity because the wind resource is not constant. For this reason, some form of energy storage or additional energy source is required to produce electricity on demand. Systems designed to supply entire villages or single homes are used throughout the world.

Battery systems are the most common form of energy storage. In some developing countries, battery-charging stations have been built using wind turbines. These sites simply charge batteries that people rent. The discharged batteries are exchanged for charged batteries at regular intervals. Other systems are designed for storage batteries to provide electricity during times when the wind is not blowing.

Wind turbine systems are often combined with other energy sources such as photovoltaic panels or diesel generators. Many remote areas currently rely upon diesel generators for electricity. Transportation costs limit the amount of diesel fuel that can be supplied, and diesel fuel storage poses environmental risks. By combining the generators with wind turbines, diesel fuel use is reduced. Wind and solar resources often complement each other. It is common in many areas for wind resources to be strongest in seasons when the solar resource is diminished, and vice versa. Systems that combine energy sources are called hybrid systems.

Although electricity production is the primary use of wind turbines, other applications still exist. Water pumping, desalination, and ice-making are applications that wind turbines serve. Wind turbine rotors for small-scale applications generally range in size from less than 1 m diameter to 15 m diameter, with power ratings from 300 W to 50 kW.

SITING WIND TURBINES

To obtain the best productivity from a wind turbine, it must be sited adequately. Whether establishing a wind farm with a hundred turbines or a single turbine for a home, documenting the wind conditions at a given site is a necessary step. Several databases of wind conditions have been established over the years (Elliot et al., 1987). Improvements in geographic information systems have enabled mapping wind conditions for very specific regions. Figure 3 shows the wind resource throughout the United States. Class 1 is the least energetic, while Class 7 is the most energetic. Wind turbines are generally placed in regions of Classes 3-7. This map indicates that a large part of the country has wind conditions that are conducive to wind turbine applications. These advanced mapping techniques can be used to compare one ridge to another in order to place the turbines in the most productive regions.

In addition to obtaining adequate wind resources, site selection sites for wind turbines must also consider avian populations. Several studies have been performed to determine the impact that turbines have on bird populations, with inconclusive results (Sinclair and Morrison, 1997). However, siting turbines to avoid nesting and migration patterns appears to reduce the impact that turbines have on bird mortality.

Other considerations in siting wind turbines are visual impact and noise, particularly in densely populated areas (National Wind Coordinating Committee Siting Subcommittee, 1998). Due to Europe's high population density, European wind turbine manufacturers are actively examining the potential of placing wind turbines in offshore wind farms.

ECONOMICS

Deregulation of the electric utility industry presents many uncertainties for future generation installations. Although the outcome of deregulation is unknown, ownership of the generation facilities and transmission services will most likely be distributed among distinct companies. In the face of such uncertainties, it is difficult to predict which issues regarding renewable energy will dominate. However, the benefits of integrating the utility mix with wind energy and the determination of the cost of wind energy, are issues that will be relevant to most deregulation scenarios.

Wind energy economics focuses on the fuel-saving aspects of this renewable resource, but capacity benefits and pollution reduction are important considerations as well. The capital costs are significant, but there is no annual fuel cost as is associated with fossil fuel technologies. Thus, wind energy has been used to displace fossil fuel consumption through load-matching and peak-shaving techniques. In other words, when a utility requires additional energy at peak times or peak weather conditions, wind energy is used to meet those specific needs. In addition to fuel savings, wind energy has been shown to provide capacity benefits (Billinton and Chen 1997). Studies have shown that although wind is an intermittent source, wind power plants actually produce consistent, reliable power that can be predicted. This capacity can be 15 percent to 40 percent below the installed capacity. Last, the emission-free nature of wind turbines could be exploited in a carbon-rading scenario addressing global climate change.

Two methods of determining the cost of wind energy are the Fixed Charge Rate (FCR) and the Levelized Cost of Energy (LCOE). An FCR is the rate at which revenue must be collected annually from customers in order to pay the capital cost of investment. While this incorporates the actual cost of the wind turbines, this method is not useful for comparing wind energy to other generation sources. The LCOE is used for comparison of a variety of generation technologies that may vary significantly in size and operating and investment time periods. This metric incorporates the total life cycle cost, the present value of operation and maintenance costs, and the present value of depreciation on an annual basis.

Subsidies, in the form of financing sources, and tax structures significantly impact the levelized cost of energy. The Renewable Energy Production Incentive (REPI), enacted in 1992, at $0.015/kilowatt-hour (kWh), applies only to public utilities. This tax incentive is renewed annually by Congress, making its longevity uncertain. Private owners of wind power plants are eligible for the federal Production Tax Credit (PTC), which is also $0.015/kWh. A project generally qualifies for this tax credit during its first ten years of operation. This credit is also subject to Congressional approval. Ownership by investor-owned utilities (IOUs), and internal versus external project financing, also affect the LCOE. Cost differences that vary with ownership, financing options, and tax structures can be as great as $0.02/kWh (Wiser and Kahn, 1996).

The annual LCOE for wind turbines has decreased dramatically since 1980. At that time, the LCOE was $0.35/kWh. In 1998, wind power plant projects were bid from $0.03kWh to $0.06/kWh. These numbers still exceed similar figures of merit for fossil fuel generation facilities in general. However, wind energy can become competitive when new generation capacity is required and fossil fuel costs are high, when other incentives encourage the use of clean energy sources. For example, a wind power plant installed in Minnesota was mandated in order to offset the need for storing nuclear plant wastes.

Incentives that currently encourage the use of clean energy sources may or may not survive the deregulation process. Green pricing has become a popular method for utilities to add clean energy sources, with consumers volunteering to pay the extra cost associated with renewable energy technologies. Some proposed restructuring scenarios include a Renewable Portfolio Standard (RPS), which mandates that a percentage of any utility's energy mix be comprised of renewable energy sources. Last, distributed generation systems may receive favorable status in deregulation. Small generation systems currently are not financially competitive in many areas, but through deregulation individuals or cooperatives may be able to install small systems economically.

Wind energy was the fastest growing energy technology from 1995 to 1999. This translates to an annual market value of over $1.5 billion. Supportive policies in some European countries, improved technology, and the dramatic drop in the cost of wind energy, have contributed to the growth of wind energy.

M. Maureen Hand

See also: Aerodynamics; Climate Effects; Kinetic Energy; National Energy Laboratories; Propellers; Subsidies and Energy Costs.

BIBLIOGRAPHY

Bergey, K. H. (1979). "The Lanchester-Betz Limit." *Journal of Energy* 3:382–384.

Betz, A. (1920). "Das Maximum der theoretisch möglichen Ausnützung des Windes durch Windmotoren." *Zeitschrift für das gesamte Turbinenwesen*, Heft 26.

Betz, J. A. (1966). *Introduction to the Theory of Flow Machines*. New York: Pergamon.

Billinton, R., and Chen, H. (1997). "Determination of Load Carrying Capacity Benefits of Wind Energy Conversion Systems." Proceedings of the Probabilistic Methods Applied to Power Systems 5th International Conference; September 21–25, 1997. Vancouver, BC, Canada.

DOE Wind Energy Program Home Page. U.S. Department of Energy, energy Efficiency and Renewable Energy Network. July 17, 2000. <http://www.eren.doe.gov/wind/>.

Eggleston, D. M., and Stoddard, F. S. (1987). *Wind Turbine Engineering Design*. New York: Van Nostrand Reinhold Company.

Eldridge, F. R. (1980). *Wind Machines*, 2nd ed. New York: Van Nostrand Reinhold.

Elliot D. L.; Holladay, C. G.; Barchet, W. R.; Foote, H. P.; and Sandusky, W. F. (1987). *Wind Energy Resource Atlas of the United States*, DOE/CH 10093–4, Golden, CO: Solar Energy Research Institute <http://rredc.nrel.gov/wind/pubs/atlas/>.

Gipe, P. (1993). *Wind Power for Home and Business*. Post Mills, VT: Chelsea Green Publishing Company.

Hansen, A. C., and Butterfield, C. P. (1993). "Aerodynamics of Horizontal-Axis Wind Turbines." *Annual Review of Fluid Mechanics* 25:115–49.

International Energy Agency. (1998). *1998 Wind Energy Annual Report*. NREL/BR-500-26150. Golden, CO: National Renewable Energy Laboratory.

Johnson, G. L. (1995). *Wind Energy Systems*. Englewood Cliffs, NJ: Prentice-Hall.

Lanchester, F. W. (1915). "Contributions to the Theory of Propulsion and the Screw Propeller," *Transactions of the Institution of Naval Architects* 57:98–116.

National Wind Coordinating Committee. July 17, 2000. <http://www.nationalwind.org>.

National Wind Coordinating Committee Siting Subcommittee. (1998). *Permitting of Wind Energy Facilities*. Washington, DC: National Wind Coordinating Committee.

National Wind Technology Center. National Renewable Energy Laboratory. July 17, 2000. <http://www.nrel.gov/wind/>.

Short, W.; Packey, D.; and Holt, T. (1995). *A Manual for the Economic Evaluation of Energy Efficiency and Renewable Energy Technologies* NREL/TP-462-5173. Golden, CO: National Renewable Energy Laboratory.

Sinclair, K. C., and Morrison, M. L. (1997). "Overview of the U.S. Department of Energy/National Renewable Energy Laboratory Avian Research Program." *Windpower 97 Proceedings, June 15–18, 1997, Austin, TX*. Washington, DC: American Wind Energy Association.

Spera, D. A., ed. (1994). *Wind Turbine Technology*. New York: ASME Press.

Wiser, R., and Kahn, E. (1996). *Alternative Windpower Ownership Structures: Financing Terms and Project Costs* LBNL-38921. Berkeley, CA: Lawrence Berkeley Laboratory.

U

UNITS OF ENERGY

The joule, symbol J, is the unit of energy in the science community. It is not widely used outside the science community in the United States, but it is elsewhere. For example, in the United States, the energy content of a food product is likely expressed in food calories. This energy could be expressed in joules, and that is being done in many parts of the world. A Diet Coca Cola in Australia is labeled Low Joule rather than Low Cal as in the United States. To obtain a feeling for the size of a joule, consider the following statistics:

- the sound energy that is in a whisper is about 0.01 joules,
- the kinetic energy of a 1,000-kilogram car traveling 25 meters per second (55 miles per hour) is about 300,000 joules,
- the energy from burning a barrel of oil is 6,000,000 joules,
- the annual energy use in the United States is about 100 billion billion joules, and
- the daily energy input to the Earth from the sun is about 10,000 billion billion joules.

Numerical Unit	*Power of Ten*	*Prefix*	*Symbol*
million trillion	10^{18}	exa	E
trillion	10^{12}	tera	T
billion	10^{9}	giga	G
million	10^{6}	mega	M
thousand	10^{3}	kilo	k
hundredth	10^{-2}	centi	c
thousandth	10^{-3}	milli	m
millionth	10^{-6}	micro	μ

Table 1.
Symbols of Some Numerical Units

Even if the use of units of joules were universal, the numbers involved in energy discussions would be large and cumbersome. Therefore, it is customary to use powers of ten notation and prefixes for powers of ten. For example, the energy from burning a barrel of oil is 6×10^6 joules, which can be expressed as 6 megajoules or, symbolically, 6 MJ. Table 1 summarizes the powers of ten, prefixes, and symbols usually encountered in energy considerations.

The gasoline tank on an automobile holds about 15 gallons, and the automobile can travel about 300 miles before the tank is empty. Even though the gasoline was purchased for its energy content, the driver probably did not know that the energy content of 15 gallons of gasoline is 2×10^9 J = 2 GJ. For reasons like this, there exists a variety of energy units and energy equivalents. The unit and equivalent depends on the type of energy commodity in question. A homeowner pays an electric utility for electric energy, and the

Special Unit	*Study area of main use*	*Symbol*	*Equivalent in joules*	*Other units*
Kilowatt hour	electricity	kWh	3,600,000	3413 Btu
Calorie	heat	cal	4.186	
Kilocalorie (food calorie)	heat	kcal	4,186	1000 cal
British Thermal unit	heat	Btu	1,055	252 cal
Electron volt	atoms, molecules	eV	$1.60\text{x}10^{-19}$	
Kilo-electron Volt	X-rays	keV	$1.60\text{x}10^{-16}$	1000 eV
Mega-electronVolt	nuclei, nuclear radiations	MeV	$1.60\text{x}10^{-13}$	1000 keV
Quadrillion	energy reserves	quad	$1.055\text{x}10^{21}$	10^{15} Btu
Quintillion	energy reserves	Q	$1.055\text{x}10^{21}$	10^{18} Btu

Energy equivalents

1 gallon of gasoline = 126,000 Btu
1 cubic foot of natural gas = 1030 Btu
1 pound of bituminous coal = 13,100 Btu
1 42-gallon barrel of oil = 5,800,000 Btu
1 therm = 100,000 Btu

Variations of these energy equivalents will appear in the literature. The values listed here are typical.

Table 2.
Some Energy Units and Their Equivalents

unit is likely a kilowatt-hour (kWh). A politician interested in imported oil will likely talk in terms of barrels of oil. Table 2 summarizes some special energy units and their equivalents.

Joseph Priest

BIBLIOGRAPHY

Hobson, A. (1995). *Physics: Concepts and Connections.* Englewood Cliffs, NJ: Prentice-Hall.

Serway, R. A. (1998). *Principles of Physics.* Fort Worth, TX: Saunders College Publishing.

UTILITY PLANNING

Electricity is a commodity that has captured the attention of Americans, from its creation to the present and onward to projections of the future. From an object of luxury to a mainstream "staff of life" and forward into the sophisticated world of tomorrow, electricity has become a critical necessity. The criteria have become: electricity on demand, reliably delivered, in sufficient quantities, and at the right price. At the heart of meeting this challenge is the utility planner.

Generating and delivering electricity requires a complicated infrastructure, significant funding, competitive markets, and adequate regulation. Broken into its simplest terms, utility planning is beginning with the most optimum operation of existing electrical facilities and foreseeing the future well enough to provide additional facility to "guarantee" an adequate supply of a reliable product (electricity) at an acceptable cost. Generation provides the product, transmission is the vehicle to deliver the product over long distances, and distribution is the functional method of providing the product to the individual customer.

Originally all facilities, financing, and decisions related to building an electrical system resided within a utility. In this simplistic approach, projections of load growth, established construction and planning criteria, and coordination of facilities remained within the single entity. In today's competitive unregulated marketplace, the planners, designers, generators, and transmission providers are within a complicated mix of organizational units; thereby providing a complicated network of system participants. The challenge for utility planning today is to coordinate all of the inputs for constructing and maintaining an electrical system that combines the economics efficiencies necessary for survival and the maintenance of a reliable network where generated power can be delivered to an ultimate customer on demand, in sufficient quantities, both today and in years to come.

HISTORY

The history of the interconnected high-voltage electric power system has roots in the early decades of the twentieth century. Moving power from generating plants remote from load centers began as utilities and customers realized that costs associated with many small generating entities placed strategically at concentrated load points would soon escalate to unacceptable values. Likewise, the many advantages of the new power source would be restricted to cities and other high-density areas. The efficiencies of larger generating units utilizing fuels of choice and availability soon became the economic choice for resources thus developed the need to "push" power across long distances.

There were many examples of early interconnected electric transmission systems extending across state territories and beyond state boundaries. Statements from a speech by Samuel Insull at Purdue University in 1924 indicated that Minneapolis, St. Paul, St. Louis, Louisville, and Cincinnati soon would be interconnected and that extension of these systems to Pittsburgh and across the Allegheny Mountains to the Atlantic Seaboard could be easily conceivable in a few years. This was but one example across the United States. To quote Insull, "It makes electric service available in new places for new purposes, so that aggregate demand for service is spread over more hours of day and night, and thus opens the way to utmost economy in production and distribution."

The interconnected power system grew and the load demand increased, over time, at varying rates and reached an average growth of some 10 percent, or a doubling every ten years. Utilities financed, constructed, and controlled the vast majority of the generation and transmission facilities in what was classified a "monopolistic society." In this environment, while the processes were complicated, the data necessary to plan these extensive projects were readily available.

One example of the complexity of the process was in load forecasting technology. By the 1960s and 1970s, utilities had developed processes to replicate the extensive electric power system for study purposes to include generation resources, transmission networks and individual points of load service to customers. The most complicated of these were the load forecast and load distribution techniques. No matter how fast technology developed, certain parameters of forecasting created significant challenges. Companies forecasted load as a composite for their territory, primarily using historical trends. Often, these forecasts were economically based, thus having boundaries driven by non-technical issues. Load growth varied within areas of individual companies. New and increased industrial loads, which made up significant portions of the forecasts, diluted the forecast and often appeared on the system later (or earlier) than planned.

In the 1980s, as the complexity of forecasting became more challenging, utilities chose to augment their forecasting methodologies through the use of consultants who possessed sophisticated databases that included topography, diversity, trends, population growth and other intricate variables and impactors. These forecasts enabled a finer tuning of the process and in many cases a more efficient and economical forecast.

No matter how accurate the forecast, other planning techniques, while state of the art, introduced trending, estimating, and engineering judgment in the facility planning cycle. One such issue involved the time and expense to plan a project, budget the project, and place the project in service when required. Because of constraints, most planning studies considered only the estimated annual peak load for a company. This peak could occur in the summer or winter, depending on the region of the country. Obviously other variations on the annual load curve occurred, such as off-peaks (spring and fall seasons) and shoulder peaks (related to time of day). Normally these variations were trended or estimated or ignored in the interest of time, and an ultimate plan was developed. In the very early years of planning, utility planners assumed that planning for the peak load covered all other times of the year. History recorded it to be a valid theory.

In later years, studies involving variations of the peak load were initiated to answer specific questions. Generation maintenance schedules, which were initially developed to accommodate only peak load periods, were revised to protect extended off-peak periods of large energy usage. The components of generation and load include real power for load service and reactive power for voltage support. The effects of various load levels to voltage profiles were tested to plan economic generation reactive parameters and adequate capacitor levels, and to define optimum capacitor locations. In addition, off-peak voltage profiles were sensitive to generator outages and maintenance. Power transfer capabilities between entities had been routinely studied at peak load. As peak periods began to lengthen and as generation locations were more widely distributed, it became necessary to consider the impacts to transfers at load levels less than peak. The transfer capability issues have become of more specific importance with the move to greater open access and competition.

Transmission parameters for existing systems were translated into appropriate input data for calculation systems available at the time. During early periods of the twentieth century, calculations for expansion of the transmission network or grid were actually made by laborious longhand manual monotonous iterations. Fortunately, the areas of study were small or abbreviated sections not affecting other areas. In the 1940s, network analog calculators were introduced for use in modeling the power system. These calculators were large "black boxes" composed of resistors, capacitors, meters, controls, dials, plotting boards, etc., and often covered an entire room, usually at an area university.

The analog calculator employed a quantitative methodology. The engineer calculated the impedance (parameters) of a transmission line and physically dialed in the valve between points A and B. The calculator also included metering windows where various information could be readily observed. With this method an experienced planner could actually sense a "feel" of the system as the dials and data were manipulated. Likewise, the analog computer became a significant training tool for utility engineers.

By the late 1960s, digital computers began to replace the "black boxes." Despite forebodings of utility planning engineers regarding training of engineers and planning, adaptations to the new technology became commonplace.

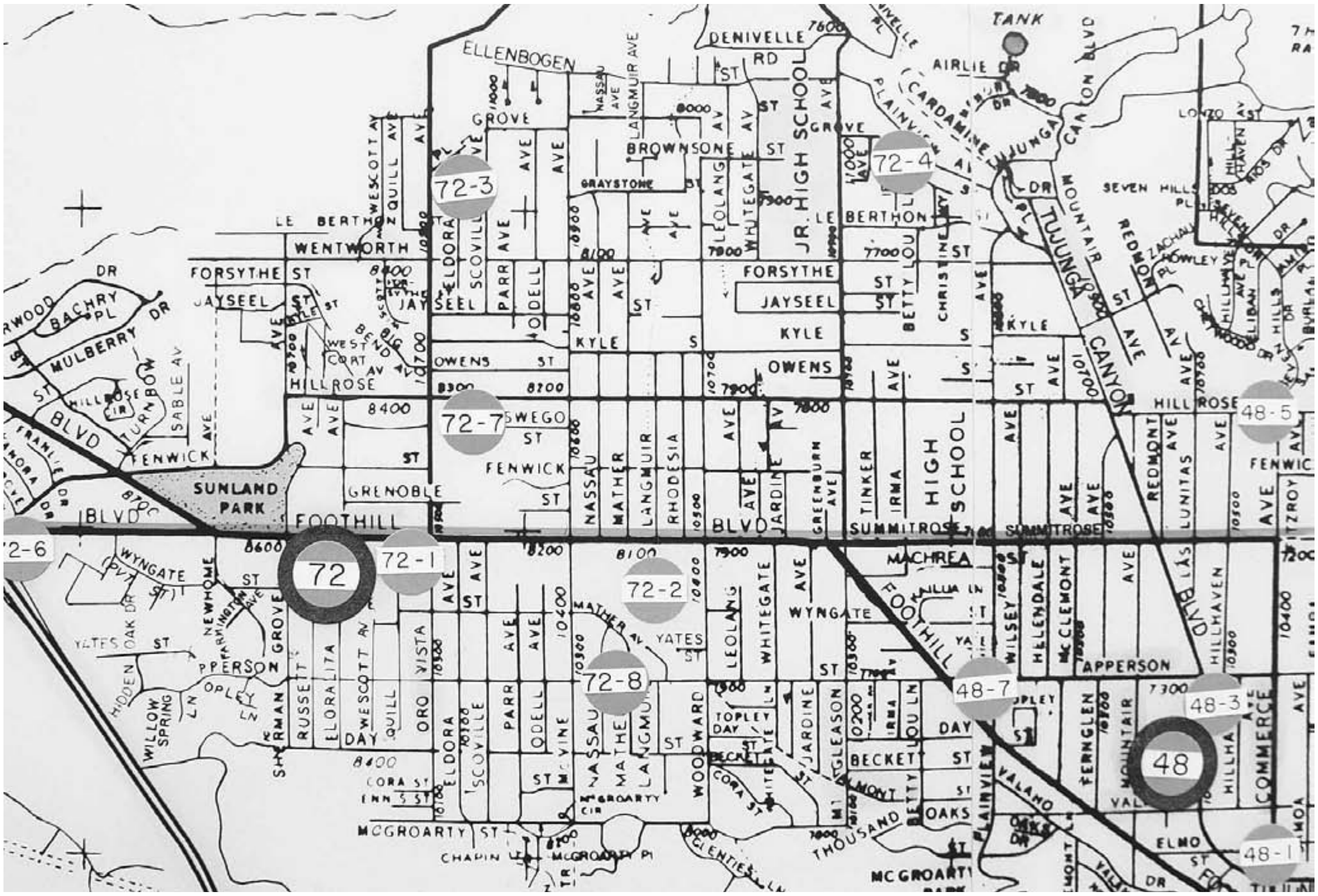
The digital computer began as a mammoth blue or black giant mechanism hidden in some corporate information resource office. The planner's view of data processing was keypunching a "deck" of cards and dropping them into an open "hopper" to mysteriously disappear into the computer. Similarly, once the magic of the calculations was accomplished, white punched computer paper arrived from an output device with guaranteed solutions to difficult questions.

The process improved. Computers became smaller, and today it is rare for a planning engineer to be void of computer power available at his desk many times greater than the early corporate computer could muster. Computer tools of the future can only be envisioned as technology advances.

Generation data to complement the planning study were plentiful. As long as the utilities owned the physical facilities and were in control of future facility expansion, data were generally a nonissue. The parameters of generating plants were dialed into the calculations or computers, similar to transmission. Obviously, alternate locations for new plants could be evaluated, and oftentimes studies were finalized some ten years ahead of construction, with only limited analysis required as the in-service date approached.

Planning studies to identify the magnitude of generation required on a particular system for a specific time period were based on parameters such as loss of load probability (LOLP) or loss of load energy (LOLE). These studies modeled the system load curve throughout the year for projected years into the future and included existing generation resources. The purpose of the studies was to identify the probability statistics where the system could not serve its load requirement or to identify an amount of unserved energy. These data would be compared to specific criteria, and future generation requirements would be developed. The ability to obtain emergency or specific contracted generation resources from neighboring systems would normally be factored into the data. In addition, in many cases generation reserve sharing agreements were factored into the final determination of capacity required. Reserve sharing agreements were contractual arrangements, usually within power pools (groups of systems) or areas where each system involved factored in the pro rata share of neighboring generation available to that system. Both concepts of utilizing off-system resources as an integral part of a system's total resources involved risk, which had to be coordinated with economic advantages and reliability levels.

Given that there were numerous alternatives for the resources, including demand-side management, interruptible loads, and off-system purchases, many case studies were necessary. Each plan for service to future loads had to be tested for conformance with system and regional reliability criteria. The plans were further examined for cost, including losses, and examined for flexibility. Any acceptable plans had to



A map at Los Angeles's energy troubleshooting facility in 1987. Circuit 72–4, the source of an earlier blackout, is marked in the upper right. (Corbis-Bettmann)

be considered for ease of modification, should future load patterns differ from present projections.

Generation was normally modeled in system studies using economic dispatch schedules developed from individual unit fuel cost and heat-rate data. To serve a particular load level, the units would be "stacked" (added to the system) in order of priority based on cost and performance. Additional capacity options were available from off-system purchases or reserve sharing arrangements with neighboring systems. A system's ability to import power was a strong indicator of its territorial reserve requirements.

Prior to the 1990s, generation lead time was generally the critical factor in project planning. As smaller units, low run-time peaking plants, environmental issues, and more diversified fuel availability appeared, the lead time for transmission construction began to be of more concern than the generation construction period. In addition, line citing, environmental impact, and condemnation issues contributed to longer lead times for transmission projects.

The planning processes described herein were traditional in nature, with little variation from year to year. Growth was steady, equipment technology advancements were available for the conditions expected, and while utility facilities were capital-intensive, a regulated guaranteed return on investment resulted in adequate financing capital. The utility planner had a vision of the future with acceptable accuracy.

TRANSITION

Many events signaled the beginning of a lengthy transition period for the utility industry in general, and more specifically, for the facility planning activity. The timeline in many instances was blurred, with some of the blurring attributable to a reluctance to change.

Perhaps the Public Utilities Regulatory Policies Act (PURPA) of 1978 was the initial formal driving force for change. The introduction of nonutility generators (NUGS) with qualified facilities and almost guaranteed sales opportunities was the first major nonutility resource activity. Whether the cause was the actual construction of these facilities or the resource atmosphere it stirred, transition was on the way. Independent power producers (IPPs) soon followed. These projects were introduced into the competitive resource market on an equal footing with utility-sponsored plants and could not be ignored by utility management.

In parallel with the new approaches to generating capacity additions, the utilities, with encouragement from regulators, introduced incentives during the 1980s for reducing load demand. Since the system peak hour load provided the inertia for capacity requirement definition, “shaving” of the peak became the focus of these incentives.

Enticements to involve the retail customer base, as well as the industrial sector, in the solutions desired became popular. Interruption of large commercial load, a provision by contract, had long been used to offset capacity shortages. Interest in this method mounted. In addition, retail customers were provided enticements to “cut load,” either by manual participation or by automatic devices. These enticements will like continue into the future.

Further involvement by the typical industrial or commercial utility customer, both large and small, was stimulated by time-of-day price incentives. Encouragement was provided in the form of reduced rates if use of electricity was shifted from peak periods of the day to off-peak or shoulder-peak periods. Even the residential customer was invited to participate in load shifting with price incentives or rewards. A popular example of the day was encouraging the household laundry activity to be moved to late-night hours. This suggestion was met with varying enthusiasm.

Initially there were attempts to quantify and project magnitudes of power shifted and relief provided to the capacity resource requirements. Ultimately the shifting of load became embedded in the subsequent forecast and became harder to identify.

The impact of these tactics to augment capacity resources was successful to some degree. Currently, the values identified are marginal; however, much of the impact is unidentified and embedded in the load growth.

Industrial customers, early a driving force in the industry, began to react not only to local utility incentives but also to more competitive pricing opportunities in other systems and exerted significant pressure on regulators. Forces in Washington, D.C., following the industry trends, finally acted in 1992 with open access of the transmission system to the wholesale market, and more extensive competition resulted. Possibly this one act, more than any other, provided the stimulus for a major shift in the industry through deregulation, restructure, and more specifically, in the planning process of the bulk electric system.

With the opening of the transmission network to all resource suppliers, many marketing entities entered the game in the mid-1990s. Many of these marketers are subsidiaries of already established gas suppliers. Some have been created solely for the electric industry. Still others have been formed from the utilities themselves. All of these entities are competition-motivated. Facility planning and reliability issues, while important to their business, are left to other organizations.

The planning and construction of generation and transmission facilities by utilities came to a virtual halt. With no guaranteed market, no control over the use of their transmission networks, no assurance of stranded investment recovery, and no assurance of federal remedial treatment, the economic structure of utility planning and construction of generation facilities basically stopped in place.

The transition for utility planners has been difficult. Generation and transmission margins have deteriorated. Shortages have been noted. The number of transactions within and across systems has escalated to levels that not only test the reliability of a transmission network not designed for this level of activity but that also have challenged the ability to operate such a system with existing technology. Prices have reacted. Emergency disturbances have occurred. Basic questions regarding authority, planning, responsibility, and finances have been raised.

Several alternatives have been offered by the Federal Energy Regulatory Commission (FERC). Regional transmission groups (RTGs) were suggested. While the basic premise of these wide-area coordinating and planning organizations was sound, many of the questions involving economics and responsibility were left unanswered.

Systems of independent system operators (ISOs) were introduced as another alternative. Some have been implemented and many have been evaluated. Adoption of these systems has been slow, the same basic reasons as those affecting RTGs. The regional transmission organization (RTO) is the most current alternative being proposed and initial plans will be provided to FERC during 2000. All of these organizational arrangements are devised to segregate ownership, planning, construction, and operation in an appropriate manner to produce a nondiscriminatory, totally competitive marketplace for electricity. All of this activity will likely result in the complete deregulation of the industry, wholesale and retail. Restructuring of the electric business on all fronts is required to make deregulation complete.

The North American Electric Reliability Council (NERC), established in 1968, and its associated ten regional councils have been reliability monitors of the electric bulk power system. Their emphasis on compliance with reliability standards through peer pressure was effective in the past. NERC and the councils, with oversight from FERC, are engaged in a complete restructuring of the reliability organizations. This move will ultimately involve a move from peer pressure to mandatory compliance.

In the meantime, the utility planner struggles with how to approach the planning process in this changing environment. Perhaps these issues are more critical in the transition, since so many of the traditional parameters of planning have disappeared, and few new parameters have taken their place.

Future-generation location, timing, and ultimate customer designations are generally unknown. Transmission construction has become more of a patchwork activity. Power transactions across the various systems are at an all-time high, a condition the network was not designed to handle. System operators struggle with daily and weekly operational planning because of the volume of transactions. Many of these transactions are not identified beyond buyer and seller. The marketing of power continues to follow the traditional contractual-path power flow route from seller (source) to buyer (load), as opposed to the actual system flow path. The contract path is a predetermined contractual route from the source system to the load system, through any interconnected intermediate system or systems. The contract assumes that all or a majority of the power will flow along this path. Only systems outlined in the contract are involved or compensated. In reality the power flows over the path of least resistance, as determined by the parameters (resistance) of the transmission lines involved. While there are economic considerations involved between the two alternatives, this is an issue of significant magnitude to the system planners and operators, who must know the path of actual power flow to properly address facility planning and operating security. In addition, many operating problems arise from the lack of automated software, a requirement for the volume of daily business being handled.

One alternative to planning in this environment of unknowns is called “scenario” planning. Utility planners in 2000 continue to estimate resource requirements for five to ten years into the future. These resources could be constructed indigenous to their systems or external to their systems, or purchased from off-system. By considering multiple alternatives for generation sources, the planner can simulate power transfers from within and outside each system. The results of these scenario analyses can be used to estimate where critical transmission might be constructed to be most effective for wide-area power transfer. Similarly, analyzing multiple transfers across a system can provide further justification for a new transmission path.

Other new planning processes are being considered to aid the transition. These include variations of probability analysis, optimum planning tools, and short-lead-time projects. None of these addresses all of the constraints discussed above. This should not, however, be construed as an impossible task.

As new generation is announced from all market segments, and as physical locations are determined, the planning picture will slowly evolve as well. It is also assumed that as retail access is introduced during the first decade of the twenty-first century, it will be controlled to some extent by contract terms long enough to provide practical development of demand forecasts. It is further assumed that future legislation and regulation also will assist in defining many aspects of the planning process. While the transition may be lengthy, the utility planner may be immersed in a full competitive environment without realizing that the transition has been completed.

FUTURE

The future of utility planning is uncertain. Good engineering judgment and technological advancements, however, will prevail as future system requirements are defined.

Major wholesale load shifts, unidentified transactions, dynamic scheduling (services provided by systems remote from the load served), unknown generation locations and timing, and retail wheeling will contribute to difficulty in planning for the future. The issues of planning, however, are likely to be easier to solve than the political issues that have a major impact on the health of the electrical infrastructure. The threats are many. An environment created by new deregulation legislation that fails to consider or impacts critical electrical phenomena and proper division of responsibility between the state and federal domains will be difficult to plan for. Environmental restrictions that curtail or impact generation resources or transmission availability will be costly to overcome and may lead to undesirable shortages. Retail customer access to systems remote from the host system will present future projection issues and change obligation-to-serve regulations. While stranded investment recovery will continue to be an economic issue, it may impact future planning in the restriction of alternative choices. Finally, mandatory compliance with reliability standards, while necessary to maintain a reliable system in a competitive marketplace, can become a constraint to good planning practices if the program becomes more bureaucratic than functional.

Based on history and the inherent unknowns related to planning, it is likely that full retail access and customer choice will present the utility planner with the most difficulty in future planning of the system. Since load demand is the principal driver of the facility developmental process, it has been necessary in the past to have a high degree of probability in the forecast. Without significant improvement in planning techniques for more accurate forecasts, planning for a changing (fluctuating) load demand in a given service area will be a significant challenge.

The various restructures of the industry are all planned to address these major issues. Wide-area planning, while undefined, may solve certain issues resulting from the unknown parameters. Divestiture of the industry into generation, transmission, and distribution companies will strengthen emphasis on the facilities involved. Many of these new structures will move the industry toward full competition; however, coordination of their activities could become more difficult.

Entry into a true competitive market suggests that the market will resolve all issues created by this move. Supply and demand will likely prevail. Full retail wheeling will introduce issues into utility planning never before addressed. Planning strategy for the future is difficult to envision based on history and current transition difficulties. It is assumed, however, that the need will create the solutions.

New technology, in both long-range planning and operational planning, will aid the entrance into a full competitive market. New tools and ideas will offset increased business and downsizing of manpower created by a more competitive environment. Each significant transition in the past was met with creative solutions. Utility planning expertise and experience will be major factors in the twenty-first-century electric industry. The goal, as always, is to continue the planning, construction, and operation of an economic and reliable electric power system.

James N. Maughn

See also: Capital Investment Decisions; Economically Efficient Energy Choices; Economic Growth and Energy Consumption; Electric Power, System Protection, Control, and Monitoring of; Energy Management Control Systems; Government Intervention in Energy Markets; Regulation and Rates for Electricity; Risk Assessment and Management; Subsidies and Energy Costs; Supply and Demand and Energy Prices.

BIBLIOGRAPHY

Arrillaga, J.; Arnold, C. P.; and Harker, B. J. (1983). *Computer Modelling of Electrical Power Systems*. New York: John Wiley & Sons.

Edison Electric Institute. (1991). *Transmission Access and Wheeling*. Washington, DC: Author.

Fox-Penner, P. S. (1990). *Electric Power Transmission and Wheeling: A Technical Primer*. Washington, DC: Edison Electric Institute.

Knable, A. H. (1982). *Electrical Power Systems Engineering: Problems and Solutions*. Malabar, FL: Robert E. Krieger.

McDonald, F. (1962). *Insull*. Chicago: University of Chicago Press.

Rosenthal, B. (1991). *Transmission Access and Wheeling*. Washington, DC: Edison Electric Institute, Power Supply Division.

Spiewak, S. A., and Weiss, L. (1997). *Cogeneration and Small Power Production Manual*, 5th ed. Lilburn, GA: Fairmont Press.

Stevenson, W. D., Jr. (1982). *Elements of Power System Analysis*, 4th ed. New York: McGraw-Hill.

Wang, X., and McDonald, J. R. (1994). *Modern Power System Planning*. New York: McGraw-Hill.

VOLTA, ALESSANDRO (1745–1827)

Alessandro Volta, now known as the inventor of the electric battery and eponym of the volt, the unit of electrical potential, was a prominent figure in late eighteenth-century science. A younger son from a family of the lesser nobility, he was born in 1745 in the commercial town of Como, in Northern Italy, at that time part of Austrian Lombardy. He received an irregular education and did not attend university. As a sometime pupil of the Jesuits, however, he developed a lifelong interest in natural philosophy, combining it with a commitment to Enlightenment culture and the notion of "useful knowledge," then fashionable among the educated classes and the public administrators of Lombardy. Having chosen the science of electricity and chemistry as his fields of expertise, at twenty-four he published a treatise "on the attractive force of the electric fire." At thirty he embarked on a career as a civil servant and a teacher in the recently reformed educational institutions of Lombardy. Appointed professor of experimental physics at the University of Pavia in 1778, he held the position until he retired in 1820. He traveled extensively, sharing his enthusiasm for natural philosophy with colleagues in Switzerland, the German states, Austria, Britain, the Low Countries, and France where in 1782 he met Benjamin Franklin.

Regarded in some circles as "the Newton of electricity," by the 1780s Volta had won a European fame as an "electrician," a chemist specializing in the chemistry of airs (especially "the inflammable air found in marshes," i.e., methane, his discovery), and the brilliant inventor of intriguing machines. Volta's machines prior to the battery included the electrophorus, or the "perpetual bearer of [static] electricity," a eudiometer, the electric pistol, the "condensatore" (a device that made weak electricity detectable), and a straw electrometer. Volta's contributions to the science of electricity included the notions of tension, capacity, and actuation (an ancestor of electrostatic induction). Thanks to painstaking measurements taken throughout his life, Volta managed to combine these notions into simple, quantitative laws that offered effective guidance through the intricacies of eighteenth-century investigations into electricity. In 1794, the Royal Society of London awarded Volta the Copley Medal for his work on Galvanism. In 1801, in the wake of his discovery of the battery, Napoleon publicly rewarded him. Volta died in 1827 in Como.

The Voltaic battery conceived and built toward the end of 1799 in Como, was the first device to produce a steady flow of electricity, or electric current. The instrument enabled other natural philosophers, notably Humphry Davy, to develop electrochemistry as a new branch of science, and still others, notably Hans Christian Oersted and Georg Simon Ohm, to explore electromagnetism. Because of these later developments—showing that chemical, electrical

Alessandro Volta. (Corbis Corporation)

cal and magnetic phenomena could be converted into each other—after the early 1840s the battery was a frequent topic for reflections on what was subsequently known as energy conversion, and energy conservation.

Volta saw the battery in a different light. He had conceived it as a demonstration device to show his contact theory of electricity at work. He had developed this theory to refute Galvani's notion of a special electricity intrinsic to animals. Volta claimed that the mere contact between different conductors (especially metals) was able "to set the electric fluid in motion." He also claimed that the electric fluid—one of the several, imponderable fluids then found in physics—was the same in organic and inorganic bodies. He built the battery after reading a paper by William Nicholson, who suggested imitating the electric organs of the torpedo fish by means of an apparatus combining many electrophoruses together. Having discarded his own electrophorus as unable to perform as Nicholson expected, Volta tried instead with pairs of discs of two different metals (like silver and zinc) that he knew could give weak signs of electricity. When he managed to pile up several such pairs, always in the same order and inserting a wet cardboard disc between each metallic pair, he found that an electric current was produced at the two ends of the pile, and that the power of the current increased with the number of pairs making the pile.

In the first circulated description of the battery—two letters addressed to the Royal Society of London on March 20 and April 1, 1800—Volta mentioned no chemical phenomena associated with the new apparatus. It was William Nicholson and Anthony Carlisle who, having had access to Volta's letters prior to publication, first observed the "decomposition of water" (electrolysis) while experimenting with the battery in London in May 1800. Nicholson, in particular, emphasized the chemical phenomena accompanying the operations of the battery. After that, a struggle between Volta's contact interpretation and the chemical interpretation of the battery developed; a struggle that Wilhelm Ostwald still regarded as unsettled in 1895. The struggle had obvious if complex implications for reflections on energy conversion and conservation; the more so because people like Nicholson already perceived the battery as the herald of a new family of machines, and wondered how the "intensity of action" of these machines could be measured.

Viewing things from the perspective of his physical theory of contact electricity, Volta was intrigued by the apparently endless power of the battery to keep the electric fluid in motion without the mechanical actions needed to operate the classical, friction, electrostatic machine, and the electrophorus. He called his battery alternately the "artificial electric organ," in homage to the torpedo fish that had supplied the idea, and the "electromotive apparatus," alluding to the "perpetual motion" (his words) of the electric fluid achieved by the machine. To explain that motion Volta relied, rather than on the concepts of energy available around 1800, on his own notion of electric tension. He occasionally defined tension as "the effort each point of an electrified body makes to get rid of its electricity"; but above all he confidently and consistently measured it with the electrometer.

Giuliano Pancaldi

BIBLIOGRAPHY

Heilbron, J. L. (1981). "Volta, Alessandro." In *Dictionary of Scientific Biography,* edited by C. C. Gillispie, Vol. 14, pp. 69–82. New York: Charles Scribner's Sons.

Pancaldi, G. (1990). "Electricity and Life. Volta's Path to the Battery." *Historical Studies in the Physical and Biological Sciences*, Vol. 21. Berkeley: University of California Press.

VOLTAGE

See: Units of Energy

W–Z

WASTE HEAT RECOVERY

See: Cogeneration Technologies

WASTE-TO-ENERGY TECHNOLOGY

Solid waste disposal has been an issue since the dawn of humanity. In the earliest times, land filling, or, more simply, land application, was the most likely scenario for disposal, as anything unusable or nonessential for immediate survival was discarded and left by the trail as the hunter-gatherers moved to follow life-sustaining herd migrations. As humans became more "civilized" and established permanent residences and villages and obtained an existence through means of agriculture, trade, etc., the accumulation of wastes increased due to the localized populations and increased permanence of those societies. Waste disposal practices included both landfill and combustion, as anything of fuel value was most likely burned for heat or cooking purposes. In all likelihood, landfill accounted for a minimal fraction of the waste in those early times, until the advent of the industrial revolution opened the way for greater leisure time and the economy to create and consume greater quantities of nonessentials, which, in turn, created a greater flow of disposable waste from the thriving communities.

Solid waste disposal has always consisted of two methods, burning or discarding. Requirements of communal living conditions and a greater understanding of the health and sanitary implications of haphazard waste disposal created the need to concentrate solid wastes into a landfill and bury the material. Convenience and availability of seemingly unlimited space favored land filling as the universal means of waste disposal until midway through the twentieth century. As populations increased, however, capacities in existing landfills were rapidly used up and sites for new "garbage" dumps were being pushed farther and farther from the population centers, leading to increased hauling and operating costs for the newer facilities. Since the early 1990s, a new environmental regulations continue to increase both the complexity and the costs for new landfill facilities in the United States and the other more developed nations worldwide.

Where any pressures arose concerning the siting or operating costs of land filling, consideration has been immediate for the option of combustion of the same waste stream. The earliest systems were designed to incinerate the incoming waste. With no energy recovery capabilities and only basic wet scrubber gas cleanup technology, this practice reduced the quantity of waste ultimately routed to the landfill by seventy-five to eighty-five percent. These early incinerators were typically small, only fifty to one-hundred tons per day capacity, mass burn, starved air, multistaged incinerators. During operation, the waste would be introduced into the furnace onto the first of a series of ram-activated cascading grates. In this primary zone, the fresh fuel contacted some combustion air and a stream of recycled gases from the exhaust header. As the waste began to volatilize and burn, the ram feeders on the grate pushed the pile out on the grate where it fell to the next step. In the process, the material was stirred and agitated to present a fresh combustion surface to the air stream. Likewise, the material from this second step was ram-fed inward until it "cascaded" onto the next grate. This process continued until the remaining material, primarily unburnable material plus some uncombusted carbon, fell into the water-cooled ash trough and was removed from the furnace in an ash conveyor. The combustion and volatile gases driven off from the fuel pile were mixed with combustion air in the primary chamber directly above the grate. Further combustion and quench occured in the secondary and tertiary chambers where additional ambient air and recycled flue gases were mixed into the combustion exhaust stream. All of the gases from the tertiary chamber were conveyed into the wet scrubber where particulate was removed.

In the early 1970s, the surge in energy costs, spurred by the oil embargo, created a demand for energy recovery in these waste disposal facilities. At that time, while still in its infancy stage, waste-to-energy was poised to take one of two paths to implementation—mass burn (MB) or process fuel, typically called refuse derived fuel (RDF). Like the predecessors, mass-burn technology opted to burn the waste virtually as it is received, eliminating preprocessing of the material prior to burning. The mass-burn units are typically very large furnaces, most often field erected, wherein the waste is literally plucked from a receiving floor with a grappling hook and ram-fed into the furnace onto a cascading grate design. Following the success of other European designs beginning in the 1950s, these systems were typically of the waterwall furnace design, as opposed to a more "conventional" refractory furnace. One advantage of this was the direct removal of much of the furnace heat by the waterwalls, minimizing the need for significant volumes of excess combustion air or recycled flue gas to maintain acceptable operating temperatures in the furnace. These plants have also evolved into much larger capacity facilities. They were already field erected, so up-sizing the capacity did not represent any further economic disadvantage, and have been built to capacities as great as 2,500 tons per day.

Because the mass burn systems are designed to handle the fuel as it is received, implying the obvious variations in quality, sizing, and handling therein, the combustion systems must be extremely robust and conservative in design with regard to the quality of fuel being introduced. In addition, the ash removal system must be capable of handling the size and capacity of material coming out the furnace as is fed into the furnace on the front end. While this universal acceptability has proven to be one of the most important features of the mass burn success, it has also created two major liabilities in the performance of the furnace. First, extreme variations within the fuel pile on the grate can creat significant temperature variations across the grate which, in turn, can cause high temperature damage to the equipment or generate greater emission from products of incomplete combustion. Second, unpredictable fuel composition can create surges in ash quantities, resulting in slagging and fouling of the boiler surfaces and increased emission levels of unburned hydrocarbons, acid gases, and dioxins/furans in the outlet gases. In spite of these drawbacks, the mass burn technology has captured a majority of the waste-to-energy market in the United States in the last three decades of the twentieth century. Part of that success came as a direct result of the early failures of RDF in the industry.

Simplified in process terms, RDF involves processing the incoming municipal solid waste (MSW) stream to remove a substantial portion of the non-combustible components, namely aluminum, ferrous, glass and dirt. Various sources list these components in the range as follows:

Aluminum	2%
Ferrous	6-11%
Glass	11-12%
Dirt/ grit	2-20%

A review of this composition would indicate the noncombustibles in the raw MSW range from twenty to forty percent. By removing as much of these fractions as possible from the fuel stream, the quality of fuel presented to the combustor is improved as well as the contaminants from the combustor being reduced.

Unfortunately for the RDF industry, the first attempts at implementing an RDF processing system met with disappointment and failure. With no European technology to draw from, RDF processing evolved from experience and inspiration gained from the U.S. applications. In the earliest processes, the design called for all of the incoming waste to be shred-

The Bridgeport RESCO power plant mixes garbage to facilitate the drying process, which improves the efficiency of energy generation by incineration. (Corbis Corporation)

ded as it entered the process. Conceptually, this was a good idea, providing a means to get all of the material sized to a maximum particle size, thereby enabling further separation based upon particle sizing and density; however, the actual results were less favorable. First, shredding all material as it was introduced into the process included numerous explosive items such as propane bottles, gas cans, etc. The damage caused to the equipment in these early processes was significant but the loss of life and limb from some of the early accidents was even more devastating. Aside from the catastrophic failures due to explosions, this initial shredding also contaminated the waste mixture and made further separation difficult. Shards of glass and shredded metal became imbedded in the paper or biomass fraction of the material in the shredder and were carried through the process in that fraction, thereby reducing the removal efficiencies of those noncombustible fractions. Although many of the earliest RDF processes are still in operation, numerous modifications have been made to improve their performance.

With such a rocky beginning, the RDF option soon fell out of favor and yielded the market to the mass burn technology. The complexity, added costs of material handling, and the poor operating history of the RDF processes proved to be sufficient negative factors for any significant consideration of RDF fired waste-to-energy facilities for the next ten to fifteen years. That situation did not turn around until the push toward more waste recycling activities in the late 1980s and early 1990s. With that new impetus, plus many years of experience in operating RDF processing systems, the new generation of automated waste processing plants gained favor from a political/social base as well as from a more proven technical design basis. In numerous instances, RDF systems were promoted strictly to comply with mandated recycling directives. These material recovery facilities, or MRFs, accomplished essentially the same function as the earlier RDF processes in separating combustible from noncombustible materials, but were now done in the name of "recycling" with the recovery of ferrous and aluminum metals, glass, newsprint and corrugated paper, and plastics being the primary objective. Although oftentimes reduced yields of RDF were achieved because of the higher removal of recyclable materials, the quality of the fuel stream was strongly enhanced by this approach. With a national mandate in the United States on recycling, the cost and need for preprocessing of the waste became a burden to be borne by all new waste management plans. The advantages of mass burn technology had just been eliminated via the legislated mandate for higher recycling.

With more proven methods for RDF processing being demonstrated, the increase in RDF combustion technology has followed. Some of the facilities burning RDF utilize similar grate and furnace technology as the mass burn, but others, most notably fluidized bed or circulating fluidized bed combustion, offer a new and enhanced means of combusting the waste. Fluidized bed technology refers to the concept of burning a fuel in a combustion zone comprised of small particles of sand suspended in a stream of upward flowing air. The air velocity provides a buoyancy force to lift and suspend the individual sand particles such that they are able to float throughout the mixture and move freely within the furnace. The sand bed displays the characteristics of a pot of boiling water, hence the term, fluid bed. In a circulating fluid bed, the air velocities are actually increased to the terminal velocity of the sand particle, thereby carrying the sand up and out of the furnace. The sand is mechanically separated from the air/gas stream at the outlet of the furnace, and the sand is recirculated back into the furnace at the bottom of the bed. Hence, the term circulating, or recirculated, fluid bed.

The advantages of fluid bed combustion over the more traditional technology arise from the increased turbulence provided by the bed particle action. This fluidization increases the interaction of the fuel particles with the combustion air and creates a very accelerated combustion environment for the incoming fuel. Additionally, the sand, initially heated to an ignition temperature for the incoming fuel, provides a means of heating and drying the new fuel introduced into the furnace and igniting that fuel as it heats up. The thermal stability provided by the hot sand in the bed plus the turbulence of the fluidization causes the fuel particles to be consumed in a very short period, typically a few minutes or less. Complete combustion is achieved by the uniform temperature zone within the furnace plus the enhanced intermixing of fuel and combustion air. Historically, fluidized bed combustion systems have been able to achieve better operating flexibility, lower emission levels, and improved combustion and boiler efficiencies.

All methods of combustion of the waste pose certain "problems." The fuel, or waste, stream is nonhomogeneous, so the likelihood of variances in operation and performance are typical. With mass burn systems, the inventory of fuel within the furnace is maintained fairly high, for up to thirty minutes, and the "management" of the fuel feeding system is responsible for selecting a feed blend that maintains some measure of constancy throughout the process. With RDF systems, the processing of the fuel and sizing/ shredding has enhanced the fuel homogeneity significantly, but variations in quality, content, etc., can still be expected. The fuel is known to contain measurable levels of sulfur, chlorine and other contaminants that can generate acid gases in the combustion products. These acids can, and do, attack the boiler surfaces and other equipment in the gas train, causing corrosion and necessitating continuous maintenance. Most boilers establish steam conditions of 650 psi and 750°F (398.9°C) as the maximum design condition in order to minimize the corrosion losses in the boiler. Recent improvements in metallurgy and furnace designs have enabled these design limits to be pushed to 850 psi and 825°F (440.6°C) or higher.

As a fuel, municipal solid waste (MSW) does not compare too favorably with more traditional solid fuels, such as coal. MSW averages somewhere around 4500 Btu/lb, versus coal at 10,500–13,000 Btu/lb. However, given the current U.S. population of 250 million and the annual generation of waste per person of fifteen hundred pounds, the potential energy content in the annual waste generated in the U.S. alone is comparable to nearly seventy million tons of coal and has the potential to generate over 13,000 MW of electrical power. As of a published report in 1993, 128 facilities were actually in operation, with an additional forty-two planned or under construction. Of the existing plants, ninety-two produced electricity to some degree. The remaining thirty-six produced only steam or hot water. The number of plants in the various sizes range from twenty plants less than one-hundred tons per day to twenty plants greater than two-thousand tons per day. Roughly forty plants are in the capacity of one-hundred to five-hundred tons per day and about twenty-five plants are sized for each of the five-hundred to one-thousand tons per day and the one-thousand to two-thousand tons per day capacity. Most of the smaller plants were installed in the early period of waste to energy development. Most of the facilities installed in the last ten years represent sizes from five-hundred to one-thousand tons per day. As of 1993 the capacity of waste consumed in these facilities was approximately 103,000 tons per day, representing approximately 18 percent of the projected annual waste generation in the United States.

Although waste-to-energy probably will never supply more than one percent of U.S. electricity (U.S. electricity generation in 1997 was 3,533 billion kilowatts), it is still a very useful renewable energy source. A state-of-the-art RDF cogeneration plant (hot water and electricity) burns much cleaner than the majority of existing coal-fired plants. At the same time, it is an attractive option in communities that desire inexpensive electricity and predictable waste disposal costs ten to fifteen years into the future, and that lack the real estate for additional landfill space. The major obstacle facing new waste-to-energy facilities is the considerable resistance contingency of the community who voice the "not-in-my-backyard" objection to the smell, noise and emissions. New facilities are likely to be located in rural areas, or as supplemental energy suppliers for industrial complexes, or on or near existing landfill operations.

The potential economic benefit exists for many regions to elect to become "host" facilities for disposal of wastes shipped in from greater metropolitan areas, such as New York City, which is closing its Freshkill landfill and already ships thousands of tons to Ohio, Georgia, and elsewhere for disposal, mostly in landfills. As the cost for energy (oil and natural gas) increases, the energy value of the waste will escalate. Its value as an alternate fuel will compensate for some of the underdesirable attributes and create a greater demand for future waste-to-energy facilities.

Michael L. Murphy

See also: Efficiency of Energy Use; Environmental Economics; and Environmental Problems and Energy Use

BIBLIOGRAPHY

Electric Power Research Institute. (1991). *Proceedings: 1989 Conference on Municipal Solid Waste as a Utility Fuel*, ed. E. Hughes. Palo Alto, CA: Author.

International Conference on Municipal Waste Combustion. (1989). *Vol. 1, Conference Proceedings*. Hollywood, FL: Author.

North American Waste-to-Energy Conference. (1997). *Proceedings of Fifth Annual North American Waste-to-Energy Conference*. Research Triangle Park, NC: Author.

Power Magazine Conference and Waste-to-Energy Report. (November 1986). *Energy from Wastes: Defining Public-and-Private Sector Markets*. Washington, DC: Author.

Solid Waste and Power (May/June 1993). *"WTE in North America: Managing 110,000 Tons Per Day."* Kansas City, MO: HCL Publications.

Meridian Corporation. (1986). *Waste-to-Energy: A Primer for Utility Decision-Makers*. Golden, CO: Western Area Power Administration.

WATER HEATING

Heating water consumes over 3.5 quadrillion Btus per year of primary energy which makes it the third largest energy end use for the residential and commercial sector in the United States, after space conditioning and lighting. Although indoor domestic hot water is nearly universal in the United States, this is a relatively recent historical development.

One of the earliest recorded uses of heated water was the Roman public baths. Starting in the first century C.E., many Roman cities had public baths with pools of cold, warm and hot water. The water was heated with wood or charcoal, although some of the baths incorporated passive solar features for space heating.

Prior to World War II, only about half of households in the United States had complete indoor plumbing facilities, including hot and cold water. Before indoor plumbing became widespread, water was heated on stoves in pots and pans, a difficult process that meant a lot less hot water was used.

The first water heaters used heat from a wood or coal stove to heat water in an uninsulated storage tank. Eventually, separate heating sources were

Water Heater	Installed Cost	Energy Factor	Annual Energy Cost	Life/ Reliabilty	Market Share 1995	Best Application
Gas						
Standard	\$200-\$600	.56	~\$150	9±6 Years	47%	If Gas Available
High Efficiency	\$1600+	.74	~\$110	Data Unavailable	<1%	Heavy Hot Water Use Space Heating
Instantaneous	\$700+	.64	~\$125	Additional Maintenance	<1%	Limited Space, Low Hot Water Use
Electric						
Standard	\$200-\$600	.86	~\$320	12±7 Years	53%	Cheap Electricity, Gas Not Available
Heat Pump	\$1500+	1.7+	~\$175	Poor Record To Date	<0.1%	High Hot Water Use Costly Electricity Gas Not Available
Instantaneous	\$200-\$400	.98		Data Unavailable		Small Intermittent Uses

Table 1.
Summary of Typical Water Heater Costs and Efficiencies

added. Either coal or wood heaters provided by a side-arm coil outside the tank. The hot water moved into the tank by convection.

The first instantaneous water heaters which started appearing in the 1890s when gas or liquid fuels started becoming available, were un-pressurized. The first models had no automatic controls and very limited safety features. Some early models were more efficient than standard modern gas-fired water heaters.

Early in the twentieth century over 150 different manufacturers were making storage tank, instantaneous, and solar water heaters. Then as indoor plumbing became more common and as people started using showers more and baths less, storage type water heaters became more popular. Accurate hot water temperatures were not possible in early instantaneous water heaters. This was not much of a problem with baths where the water temperature is adjusted before getting in the water, but it is a much bigger concern when using showers.

Solar water heating became commercially available in Southern California in the 1890s. Early models consisted of four large cylindrical tanks of heavy galvanized iron mounted horizontally in a wooden box under a glass cover.

In 1909 one company began selling solar water heaters with separate collectors and insulated storage tanks. The collectors were made of copper tubing soldered to a copper plate in a glass covered box. The hot water was transferred to the storage tank by thermosyphon action, so the insulated tank had to be installed above the collector, typically in the attic or on the roof. It was often connected to an auxiliary heater on the stove or furnace to supplement solar heat. Well insulated storage tanks meant hot water was available the next morning.

Solar water heating declined in Southern California in the 1920s due to the development of natural gas, but it continued in Florida where natural gas was very expensive. In 1941 more than half Miami's population had solar water heaters, and more than 80 percent of new homes built then were equipped with solar water heaters. By the end of the 1950s in Florida, solar water heating was displaced by electricity as the price dropped and the storage tanks of solar water heaters failed because of galvanic corrosion from connecting steel tanks to copper collectors.

Corrosion of the tank is the major factor limiting water heater lifetimes. The A. O. Smith company invented the glass-lined tank in 1939, where an enamel coating is baked onto the inside surface of the tank at high temperatures. The technology has subsequently been adopted by other manufacturers, although other linings are used such as cement or plastic.

Total sales of water heaters have climbed steadily over the past several decades with sharp increases corresponding to times of high home construction rates.

In response to the oil crisis in 1970s, and temporary curtailment of new gas connections, the percent of sales of electric water heaters grew to almost match gas-fired water heater sales. Since 1987, however, the percent of water heaters sales that are gas-fired has exceeded electric water heater sales by 5 to 10 percent.

Tax credits were provided for solar water heating systems in response to the energy crisis. After tax credits expired, the solar water heater business declined drastically. The only segment of solar water heater market that has continued to have growth is the pool heater.

Outside of the United States and Canada, storage water heaters are common in Australia, Southern Africa, Latin America, and Britain. Instantaneous water heaters are much more common particularly in Asia and Europe. In countries that were once part of the British Empire, unpressurized storage tanks that relied on gravity were common, however most have changed over to pressurized storage tanks.

STANDARD STORAGE TANK WATER HEATER CONSTRUCTION

A typical storage water heater consists of a steel tank that is lined with ceramic coating fused to the inside of the tank at high temperature during manufacturing. The tank is typically about sixteen inches in diameter and about four to five feet tall. The top of the tank is domed upward and the bottom of the tank is also domed upward in a concave manner. The outside of the tank is insulated with a polyurethane foam insulation that is squirted into the gap between the tank and a thinner sheet metal jacket. The polyurethane is made of two different components that react and harden when mixed. Included in the mixture is a blowing agent that causes the polyurethane to expand in a foam-like manner. Prior to about 1980, water heaters were insulated with fiberglass insulation. The foam insulation process was developed to allow automation and increased manufacturing speed and reduced costs. A side benefit was improved insulating ability leading to a slight increase in efficiency.

In response to concerns about ozone depletion, the original blowing agent, CFC-22 was phased out. In 1994 and 1995 the manufacturers switched to HCFC-141b, a substance with lower ozone depletion potential. HCFC-141b is much more difficult to use, and manufacturers had some difficulty learning to use the new blowing agent effectively. New formulations of polyurethane and much tighter tolerances on the manufacturing processes, such as temperatures and pressures, were required.

Electric water heaters typically use two 4,500 watt heating elements. One element is located in the lower part of the tank and provides the bulk of the energy. The other element is located near the top of the tank and is used to quickly heat a small amount of hot water after a large draw empties the tank of hot water. The elements are each controlled by separate thermostats and are interlocked so only one can come on at a time. The thermostats on electric water heaters are snap type devices that are installed directly on the outside of the tank, but inside the jacket. They are located a few inches above the element which they control.

Gas-fired water heaters use the same general method of construction, except that the elements are replaced with a burner beneath the tank. The combustion products from the burner are vented through a flue made out of the same thickness steel as the tank, that goes up through the center of the tank. To increase heat transfer from the hot flue gases to the inner wall of the flue, a baffle is inserted down the flue. This baffle is a twisted strip of sheet metal with folds and tabs on it. The folds and tabs are designed to increase the turbulence of the hot combustion products as they pass through the tank. The burner is normally a circular burner made out of sheet metal or occasionally cast iron. Nearly all gas-fired water heaters use a standing pilot for ignition purposes. The pilot is a safety pilot: If it goes out, the small amount of electricity generated by a thermocouple in the pilot flame will no longer hold open an solenoid valve, and gas to the main burner will be turned off. The thermostat on a gas-fired water heater is an iron rod within a copper tube that is inserted into the water near the bottom of the tank. As the copper tube is heated it expands and allows a click disc to close a gas valve and shut off the flow of gas to the burner. The pilot continues to burn even when the burner is not needed.

In the North American market, water heaters are almost always made with the cold water inlet and hot water outlet lines coming out of the top of the tank. The hot water outlet opens right into the top of the tank and so draws off the hottest water. The hot water has risen to the top of the tank because of its lower density. The cold water on the inlet side is directed to the bottom of the tank by a plastic dip-tube. In some models the dip-tube is curved or bent at the end to increase the turbulence at the bottom of the tank. This is to keep any sediment from settling on the bottom of the tank. As sediment— usually calcium carbonate or lime—precipitated out of the water by the increased temperature builds up, it will increase the thermal stress on the bottom of a gas-fired water heater and increase the likelihood of tank failure. On electric water heaters the sediment builds up on the surface of the elements, especially if the elements are high-density elements. Low-density elements spread the same amount of power over a larger surface of the element so the temperatures are not as high and lime doesn't build up as quickly. If the lower elements get completely buried in the sediment, the element will likely overheat and burn out.

Another required safety feature on water heaters is a temperature and pressure relief valve. This is mounted near or on top of the tank. The temperature and pressure valves are designed to vent water if the temperature or pressure in the tank becomes too high. If there is no relief valve and the controls fail to limit water temperature, there is a danger of a powerful explosion from superheated, pressurized water flashing to steam as the water heater bursts.

All glass-lined steel tanks have at least one anode. This is a long metal rod made of magnesium, aluminum or zinc, metals that are more reactive than steel. The glass lining is never perfect, and any pinholes or defects around welds would allow water to touch steel. The anode protects the steel in the tank from corrosion by "sacrificing" itself and corroding first. This prolongs the life of the tank and protects against galvanic corrosion when two dissimilar metals are in contact in water.

Storage type water heater tanks also have a drain valve at the bottom so the tank can be drained for maintenance, or in case it ever needs to be removed. These drain valves are often plastic, although higher quality models use a brass drain valve.

EFFICIENCY STANDARDS

Under the National Appliance Energy Conservation Act of 1987, residential water heaters in the United States are required to meet a minimum energy efficiency. This is measured as energy factor. The minimum allowed energy factor depends on the fuel type and decreases with increasing rated volume. The minimum energy factor for fifty gallon electric water heaters is 0.86, for forty gallon gas-fired water heaters it is 0.54.

Average efficiency of new gas-fired water heaters has increased from an estimated 47 percent in the mid-1970s to about 56 percent in 1999. Over the same period the efficiency of electric water heaters has risen from about 75 percent to 86 percent. Revised efficiency standards were expected to be adopted during 2000.

Energy factor is a measure of average service efficiency at a specified condition and hot water draw pattern. It includes the effects of both standby losses and recovery efficiency of the water heater. Currently, water heaters are shipped from the manufacturers with thermostats set at 120°F (48.9°C) to reduce the risk of scalding. This is a drop from values of 140°F (60°C) that were reported from the early 1970s.

Efficient models of water heaters have thicker insulation, up to three inches thick, on some of the most efficient electric water heaters. Another means to increase efficiency is installing heat traps, or anti–convection devices, on the inlet and outlet pipes. Standard heat traps consist of short pipe nipple containing a small plastic ball. On the inlet side the ball is lighter than water and floats up to seal the inlet pipe. On the outlet side the ball is heavier than water and sinks against the seal. This prevents the heated water, which is slightly buoyant, from setting up convection loops within the pipes.

Gas-fired water heaters are also made more efficient by a variety of designs that increase the recovery efficiency. These can be better flue baffles; multiple, smaller-diameter flues; submerged combustion chambers; and improved combustion chamber geometry. All of these methods increase the heat transfer from the flame and flue gases to the water in the tank. Because natural draft systems rely on the buoyancy of combustion products, there is a limit to the recovery efficiency. If too much heat is removed from the flue gases, the water heater won't vent properly. Another problem, if the flue gases are too cool, is that the water vapor in the combustion products will condense in the venting system. This will lead to corrosion in the chimney and possible safety problems.

One design that is used for installation convenience, but can also allow higher efficiency, is an induced draft fan that pulls in extra air to cool the flue products to temperatures that can be safely vented through plastic pipe. Because of the fan, long horizontal venting runs are possible with this system.

Flue dampers that block the flue when the burner is not firing increase the efficiency of gas-fired water heaters. These can operate electrically or thermally. Because gas-fired water heaters lose so much heat up the flue during standby periods, this can provide significant savings. These are readily available on larger water heaters used in commercial settings but haven't been applied in the residential market because of their cost.

The most effecient gas-fired water heaters are condensing models. These are designed to be so efficient that the water vapor in the combustion gases condenses out and must be drained off. The recovery efficiencies of these models can be as high as 94 percent. As of 1999, four of the major manufacturers offered condensing water heaters in the small commercial sizes (over 100 kBtu/hr rated input). To reach this efficiency level, most used a pre-mix power burner with a central flue that starts up the center of the tank, but instead of going out the top as in standard water heaters, the flue (or flues) turn and spiral down inside the tank and exit out the side at the bottom. Because of the corrosive nature of the condensate, the flues are specially protected, either by glass coating on the gas side as well as the water side or by being made out of stainless steel alloys. These models can cost as much as $1,000 to $1,500 more than standard water heaters. The best applications are where heavy hot water loads can justify the extra cost.

On standard electric resistance water heaters, not much can be done to improve efficiency other than adding more insulation and heat traps. Heat pump water heaters extract heat from air and add it to the water in the tank. Energy factors as high as 2.60 have been achieved, meaning about two-thirds less electricity is used compared to a typical electric resistance water heater. This technology is mature in space heating and household refrigerator applications (where heat is pumped out of the refrigerator into the surrounding room); however, despite several attempts by different manufacturers, heat pump water heaters have never gotten established in the market. The main problem seems to be much higher equipment costs than simple electric resistance water heaters and a poor reliability record on the part of some models.

A small fraction of homes use the same appliance to provide both space heating and water heating. This has been done for over a century with tankless coils in cast-iron boilers. Water heating can be provided by modern boilers supplying heat to an indirect water heater that uses the water from the hydronic system to heat water for domestic uses in a storage tank. Since the early 1980s a technology has been developed that uses hot water from the water heater to heat a forced air system via fan coils. Ground-source heat pumps that provide space conditioning while extracting or storing heat from the ground are often also equipped with the capability to provide water heating.

Perhaps the ultimate combined appliances, which are currently under development, will be microturbines or fuel cell power generators used as small-scale co-generation systems. These would supply not only electricity, but space and water heating as well.

James D. Lutz

See also: Heat and Heating.

BIBLIOGRAPHY

American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc. (1995). "Service Water Heating." In *1995 Handbook - HVAC Applications (I-P)*. Atlanta, GA: Author.

Barbour, C. E.; Dieckmann, J. T.; and Nowicki, B. J. (1996). *Market Disposition of High-Efficiency Water Heating Equipment, Final Report*, edited by U. S. D.o.E. Office of Buiding Equipment. Washington DC: Arthur D. Little, Inc.

Butti, K., and Perlin, J. (1980). *A Golden Thread: 2500 Years of Solar Architecture and Technology*. New York: Van Nostrand Reinhold Company.

Energy Information Administration. (1998). *Annual Energy Outlook 1999 with Projections to 2020*. Washington, DC: U.S. Department of Energy.

Gas Appliance Manufacturers Association. (1998). *Consumers' Directory of Certified Efficiency Ratings for Residential Heating and Water Heating Equipment*. Arlington, VA: GAMA.

Gas Appliance Manufacturers Association. (various years). *GAMA Statistical Release*. Arlington, VA: Author.

Greenberg, J.; Reeder, B.; and Silberstein, S. (1985). *A Review of Energy Use Factors for Selected Household Appliances*, prepared for Office of Building Energy Research and Development, Building Equipment Division, U.S. Department of Energy. Gaithersburg, MD: National Bureau of Standards.

Smith, K. (1993). *The Gas Fitter's Guide to Domestic Hot Water*. Lewiston, NY: KLS Training Corporation.

U.S. Census Bureau. (1999). "Census of Housing, Housing: Then And Now, 50 Years of Decennial Censuses." <http://www.census.gov/hhes/www/housing/census/histcensushsg.html>.

U.S. Department of Energy. (1998.) "Energy Conservation Program for Consumer Products." In *Title 10, Code of Federal Regulations, part 430*. Washington, DC: U.S. Government Printing Office.

Weingarten, L., and Weingarten, S. (1992). *The Water Heater Handbook: A Hands-On Guide to Water Heaters*. Monterey, CA: Elemental Enterprises.

Weingarten, L., and Weingarten, S. (1998). "The History of Domestic Water Heating." *PM Engineer* 4(8):74–77.

WATT

See: Units of Energy

WATT, JAMES (1736–1819)

Born fourth of five children to a merchant family in coastal Greenock, Scotland, on January 19, 1736, Watt was an ingenious modelmaker as a youth, being exposed to all the ship chandler's tackle and instruments, and was fascinated with natural philosophy, as science was called. His grandfather Watt was a very successful professor of mathematics in this remote locale. His mother's death when he was seventeen set his determination to pursue instrument making.

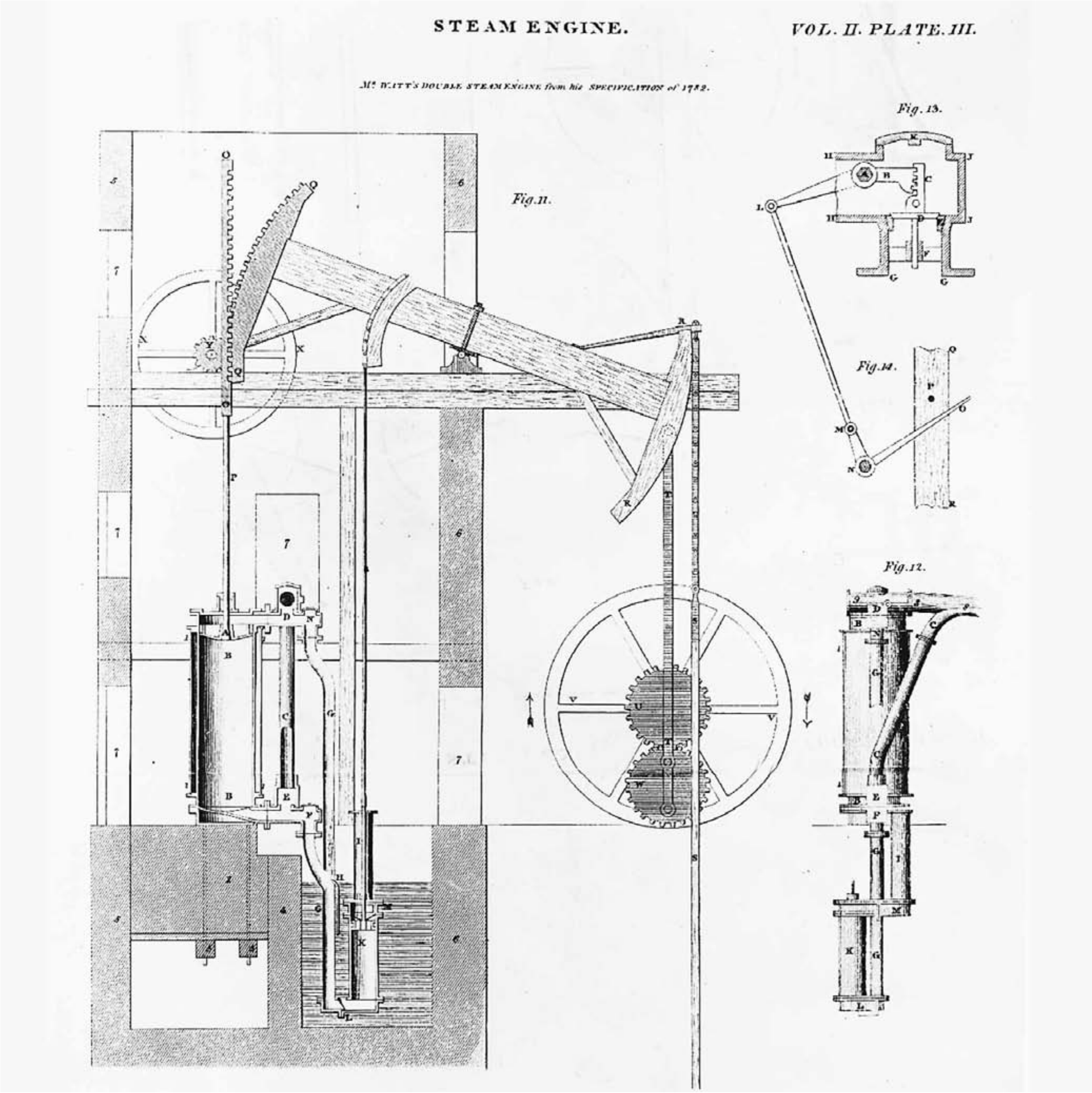
After nine years of some difficulty but considerable learning opportunity in London and at the University of Glasgow, while he was repairing a poorly built Newcomen engine for the university, he had the flash of inventive intuition that a separate condenser would greatly improve its efficiency. For this patented improvement, and its eventual technical and economic success, he is popularly credited with the invention of the steam engine. In 1765, while at the University of Glasgow, he married his cousin Margaret Miller.

At the age of thirty-three, Watt wrote, "Of all things in life, there is nothing more foolish than inventing." Just widowed with six children, he had put away his love of science, engineering and fine technical instrumentation to feed his family by contracting to survey for the Caledonian canal. Bankruptcy of his financial partner in steam development, John Roebuck, had derailed his efforts to improve the steam engine.

For some years previous, the successful Birmingham manufacturer Matthew Boulton had enjoyed an increasing correspondence in natural philosophy with such notables as Benjamin Franklin, Erasmus Darwin and Watt's partner, Roebuck, regarding practical uses of fuel and steam. Boulton's discussions with Watt in 1768 led him to discontinue his own developments, since Watt was clearly ahead technically and had obtained basic patents. With the bankruptcy of Roebuck, Boulton bought out Roebuck's interest, and Boulton's investment continued to fund Watt's demonstrations and developments. By then however, Watt's patent had nearly expired. With the leverage of Boulton's contacts in London, a Bill of Parliament extended the patent 24 years until 1800. The success of Watt's engine applications in the next few years brought Boulton and Watt together as partners, Boulton having two-thirds interest and managing the business, and Watt having one-third interest and managing the engineering and erection of engines.

While traveling on business, Watt met a Miss MacGregor who was connected with the University of Glasgow, and they married in 1776. Their two children did not live into adulthood.

The business acumen of Boulton provided profound industrial leverage to the technical creativity of Watt. The success of the steam engine for the rest of



Plan of a low-pressure steam engine designed by James Watt. (Library of Congress)

the century was substantially controlled by this partnership. The partners initially obtained annual royalties based on the horsepower of their engines, then set the royalty at one-third the cost of fuel saved by the use of their engines over the Newcomen engine.

The engines were initially used in pumping and winding in the mines, then were extended to iron-rolling, spinning and weaving mills, and virtually any factory use of rotative power.

Retiring from their business just as the patents began to expire in 1800, the firm of Boulton & Watt was turned over to the junior Boulton and Watt as its industrial preeminence slipped away and royalties dwindled. Watt's friends and family aged and passed, and he settled into twilight development of three-dimensional replicating equipment for sculpture. His workshop from his home at Heathfield Hall, near Birmingham, is preserved at the Science Museum in South Kensington. The most famous engineer of his time, Watt died at home on August 19, 1819 at 82, and was buried alongside his partner Boulton in his country churchyard. A monument was erected in his memory in Westminster Abbey. Contemporary eulogies and biographies show Watt as a fine person with a brilliant mind, little business sense, and a tendency toward melancholy. It is clear that he valued Boulton as a friend as much as a business partner.

James Watt recognized the steam engine to be a heat engine, not a pressure engine, demanding the conservation and complete use of heat energy. His long acquaintance with scientists at the University of Glasgow, and his study of new information on heat and energy, gave him the intellectual tools to make these great steps, which he backed up with technical analysis. While many of the principles in Watt's 1769 patent had been previously described academically or constructed in other mechanical examples, Watt combined these principles and patented them just at the time when they could be technically applied to advantage. Coincidently, the quality of heavy machine work necessary for their success was at last becoming practical on a commercial basis. Previously, fine machine work was confined to technical, medical and other instruments.

To paraphrase the claims of Watt's 1769 patent: Keep the engine cylinder hot; condense the steam separately and keep the condenser cold; air in the steam system is defective to performance and must be removed by pumping as necessary; use the expansive energy in high-pressure steam without using a condenser; re-use the steam by compressing and reheating to avoid loss of the latent heat of evaporation. Additional patent claims were the concept of a rotary expansive engine and using materials other than cold water to seal and lubricate the engine to avoid loss of heat energy.

To more fully exploit the key patent claim for the separate condenser, the firm avoided the use of high-pressure steam since it did not absolutely require a condenser. The patent claim for expansive use of steam without a condenser was used only as ammunition to restrict development in that field, although high-pressure steam replaced low-pressure steam during Watt's lifetime, and noncondensing systems became the norm for railway locomotives and steam road vehicles.

Many of Watt's further innovations for the steam engine added to its utility, remaining in use long after patent protection expired, while some were merely stopgaps to avoid using the patents of other inventors. These include the double-acting engine in which the piston is powered in both directions with the rod passing through a seal, a parallel motion using links to guide the piston rod through a relatively straight path (Watt linkage), and the sun-and-planet gear method of driving a rotary shaft from a connecting rod, avoiding another's patent for the crank.

Watt also applied the flyball governor to the steam engine and invented the steam engine indicator, which gives a graphical representation of cylinder pressure versus displacement. In other fields, Watt invented the copying press for hand-written letters, which was in universal use for over a century, and a cloth drying machine using copper cylinders internally heated by steam.

Karl A. Petersen

See also: Trevithick, Richard.

BIBLIOGRAPHY

Dickinson, H. W. (1936). *James Watt, Craftsman & Engineer*. Cambridge: The University Press.

Robinson, E., and Musson, A. E. (1969). *James Watt and the Steam Revolution: A Documentary History*. London: Adams & Dart.

Thurston, R. H. (1878). *A History of the Growth of the Steam-Engine*. New York: D. Appleton and Company.

WAVES

Waves which are involved in many aspects of life, are disturbances that propagate through a medium with a definite speed. A wave of light conveys information to the eyes, a wave of sound brings music to the ears, a water wave rolling onto a beach can topple the swimmer, and electromagnetic waves cook food (microwave oven), and carry television reception. It takes energy to create the wave disturbance, and how the wave travels through the medium (elastic or damped) is quite variable.

There are two important categories of waves, mechanical and electromagnetic. Mechanical waves require a material medium. Water waves in a pond require water, sound waves in a room require air, a stadium "wave" must have people, and a wave produced by plucking a guitar string needs a string. Mechanical waves are produced when the medium is disturbed. A stone dropped in a pond disturbs the water, a spoken word results from a pressure disturbance, the guitar player disturbs the string by plucking it, and the sports fan disturbs the audience by standing up. The disturbance moves through the medium with a speed that is determined by the properties of the medium. In a stringed musical instrument, for example, the mass of the string and how tightly it is stretched determine the speed. That is why there are different size strings and a mechanism for tightening them.

A stone dropped in a pond pushes the water downward, which is countered by elastic forces in the water that tend to restore the water to its initial condition. The movement of the water is up and down, but the crest of the wave produced moves along the surface of the water. This type of wave is said to be transverse because the displacement of the water is perpendicular to the direction the wave moves. When the oscillations of the wave die out, there has been no net movement of water; the pond is just as it was before the stone was dropped. Yet the wave has energy associated with it. A person has only to get in the path of a water wave crashing onto a beach to know that energy is involved. The stadium wave is a transverse wave, as is a wave in a guitar string.

When air in a room is disturbed by a person speaking the molecules of the air have movements that are along the path of the wave. If you were to draw a line from the speaker's mouth to your ear, the movement of the molecules would be along this line. This type of wave, called an acoustical wave, is said to be longitudinal. The pleasant sounds of music are produced by acoustical waves. On the other hand, destruction by a bomb blast also is caused by acoustical waves. Instead of oscillating up and down, molecules in the acoustical (or compression) wave bunch together as the wave passes. It is not a transverse wave.

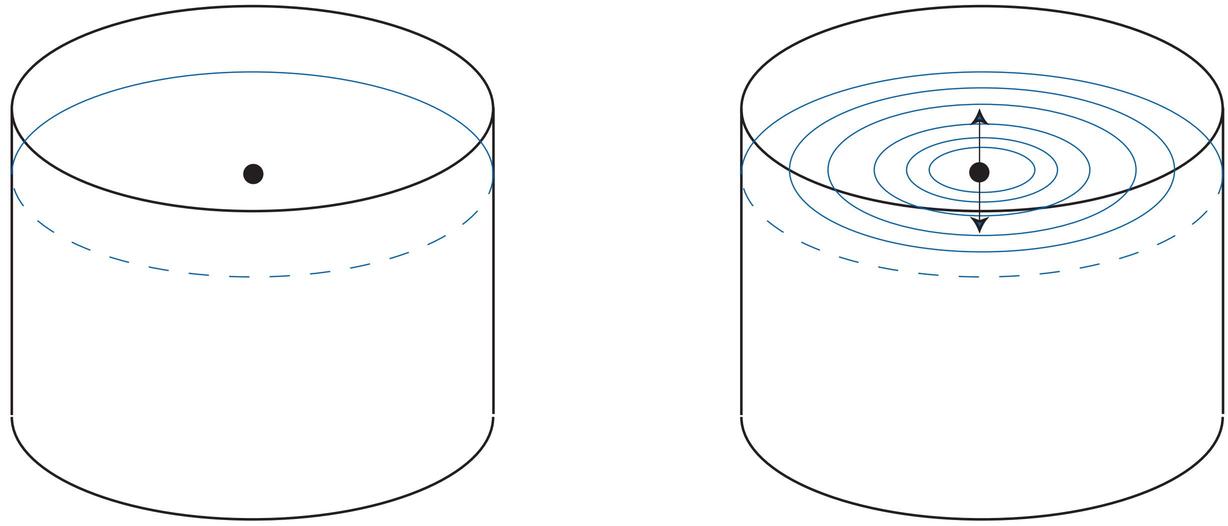
Imagine something floating in a pan of water with both the float and the water motionless (Figure 1A). If you were to continually bob the float up and down, you would see a continuous train of crests and troughs moving away from the float (Figure 1B). The separation of adjacent crests is called the wavelength, and given the lower case Greek lambda (λ) for a symbol. The time it takes one of the crests to travel a distance equal to the wavelength is called the period, and given the symbol T. The speed that the crest moves is just the distance traveled divided by the time so that $v = 1/T$. The period measured in seconds is just the time for one cycle of the oscillation. The reciprocal of the period is the number of cycles per second and is called the frequency, symbol f. The speed of the wave can also be written $v = f\lambda$.This simple relationship is a very important feature of all types of waves.

When waves encounter something in their path, they may bounce off (reflect) somewhat like a ball bouncing off a wall. A sound echo is produced by sound waves reflecting from something, a building, for example. We see our image in a mirror because of reflection of light. Reflection as well as other phenomena involving light led to the notion that light is a wave. It is decidedly different from a sound wave because there is no material medium required to sustain its motion. Light traveling to Earth from a distant star encounters virtually no matter until it reaches Earth's atmosphere. Modeling light as a wave has many virtues and uses.

Light is an example of electromagnetic waves. Electromagnetic waves propagate electric and magnetic fields. All electromagnetic waves travel with the same speed in a vacuum. This speed is given the symbol c and has a value 3.00×10^8 m/s (300,000,000 m/s). The general equation $v = f\lambda$ becomes $c = f\lambda$ for electromagnetic waves. Because c is the same for all electromagnetic waves in vacuum, $f\lambda$ is constant. Therefore, when the frequency of an electromagnetic wave increases, the wavelength must decrease. The range of wavelengths of electromagnetic waves is staggering: about 10^{-14} m (roughly the diameter of the nucleus of an atom) to about 1,000,000 m (approximately the distance from New York to Chicago). Because $f\lambda$ is constant, the frequency varies from a high of about 10^{22} Hz to a low of about 10^3 Hz.

Electromagnetic radiation is classified according to ranges of wavelength. Although the classifications are not precise, they are useful. For example, visible light to which our eyes are sensitive has wavelengths between about 4×10^{-7} m and 7×10^{-7} m. Microwaves have much longer wavelengths, about 0.001 m to 0.3 m.

Energy streams to Earth from the sun by electromagnetic waves. Photosynthesis depends on solar energy, and humans and animals rely on photosynthesis for food. In this sense, solar energy is not a luxury, it is essential.



A small piece of wood floats motionless in a pan of water

A.

Bobbing the piece up and down produces circular waves centered on the piece

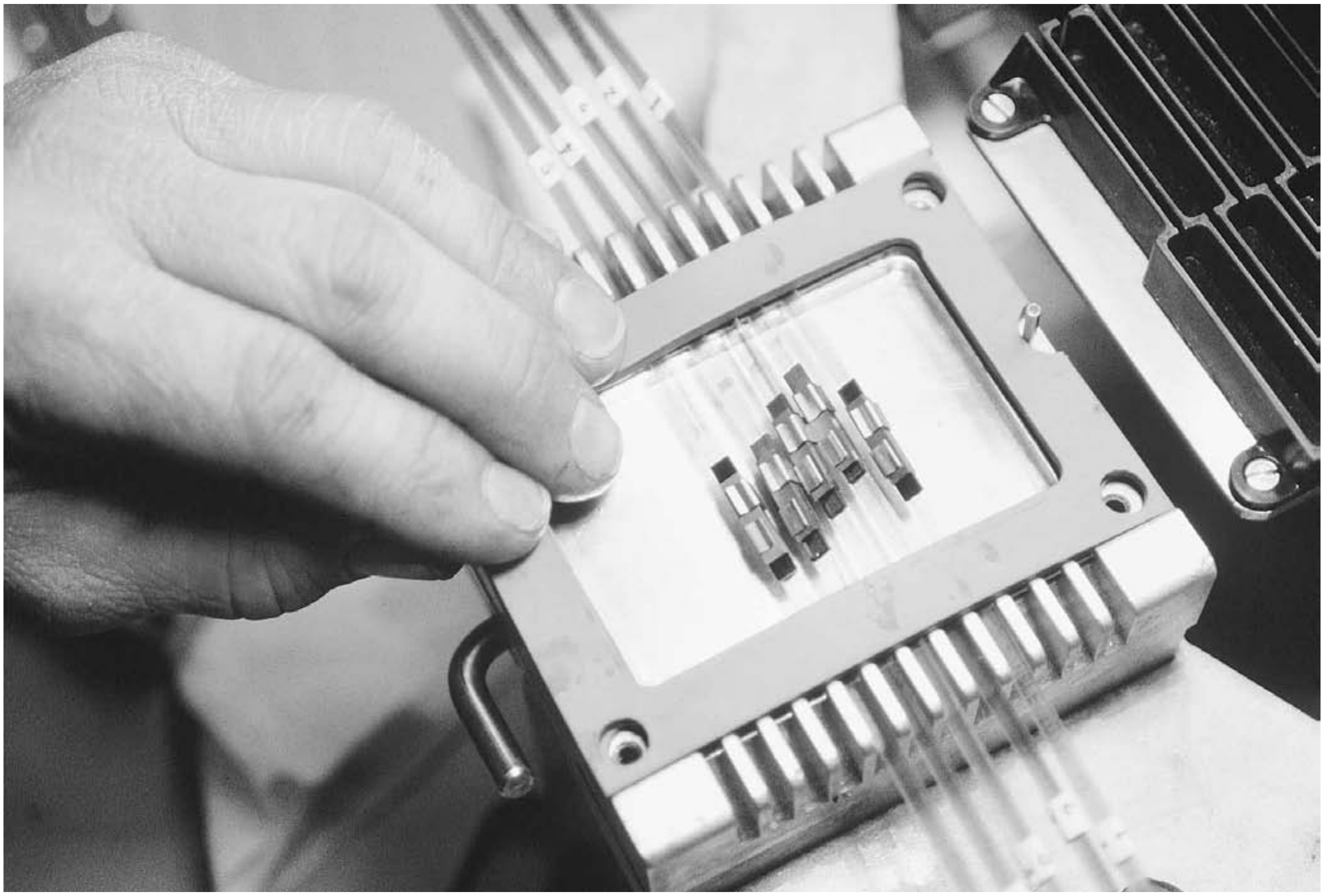
B.

Figure 1.
Illustration of wave action inside a cylinder.

All objects, including the sun and incandescent light bulbs, emit electromagnetic radiation. The radiant energy and the type of radiation depend on the temperature. When an electric toaster is switched on, the heating element begins to glow and one can feel the radiant energy increase as the temperature increases. Visible radiation is produced by both the sun and a light bulb because the temperature of both is roughly 6,000 K. When the bulb cools to room temperature, about 300 K, the visible radiation disappears for all practical purposes. Nevertheless, the bulb at 300 K is still radiating, but the radiation is predominantly infrared that our eyes do not sense.

Radiation from the sun includes significant ultraviolet and infrared radiation in addition to visible radiation. Contributions of each type to the radiation that reaches Earth's surface are reduced significantly through absorption in Earth's atmosphere. The rate at which solar radiation falls on a square meter of Earth's solid surface is called the solar insolation. Solar insolation depends on the time of day, the day of the year, and where the square meter is located on Earth's surface. During an eight-hour period, the solar insolation in midwestern United States is roughly 600 W/m^2. If the 600 watts striking a 1 m^2 surface could be collected and converted to electricity, there would be enough electricity to operate six typical 100-watt light bulbs.

Solar energy can be used in many ways: heating buildings and houses, providing heat for producing steam for a turbogenerator in an electric power plant, producing electricity with photovoltaic cells, and so on. But since solar energy is not highly concentrated, it takes a considerable area of collection to even heat



Fiber optic cables, such as this one being spliced, allow for more information to be passed in smaller "packages." (Corbis Corporation)

a small home, which makes it uneconomical and impractical in many situations.

One of the most interesting and fascinating methods of electronic communication uses electromagnetic waves of light. The light is literally piped around in tiny bundles of transparent strings called optical fibers. Because the frequency of the light is thousands of times greater than the frequency of radio and television waves, a much higher concentration of information is permitted. The telephone industry makes great use of transmission of information via optical fibers. Using this technology allows information to be transmitted between North America and Europe via optical fiber cables strung beneath the surface of the Atlantic Ocean.

The back-and-forth motion of a child's swing is an example of an oscillating system, oscillator, for short. Set into motion, these systems oscillate at a frequency determined by properties of the oscillator. For example, the length of a child's swing and the acceleration due to gravity (g) determines the frequency of the back-and-forth motion of a child's swing. Periodically feeding energy into such a system at the same rate as the natural frequency gradually increases the energy of the oscillator. This is the phenomenon of resonance. Children achieve resonance with a swing by "pumping" at the resonant frequency of the swing. Molecules are microscopic oscillators that can be "pumped" by electromagnetic waves. If the frequencies match, the molecules absorb energy. Water molecules have a resonant frequency that falls in the microwave region of the electromagnetic spectrum. Water-based materials exposed to microwaves of the proper frequency absorb energy and heat up. This principle is exploited in a microwave cooker. When food on a plastic plate is placed in a microwave oven. the food warms because of its water content, and the plate does not warm because it is essentially devoid of water.

Ozone is a molecule formed from binding together three oxygen atoms. Gaseous ozone readily absorbs selected frequencies of ultraviolet radiation because ozone molecules have a resonant frequency in the ultraviolet region of the electromagnetic spectrum. A natural concentration of ozone exists twenty to thirty kilometers above Earth's surface. Ozone molecules absorb nearly all ultraviolet radiation from the sun having a wavelength less than about 0.3 micrometers. Accordingly, the natural ozone layer protects humans from harmful effects such as skin cancer. Human-made products migrating into the ozone layer can react with ozone molecules and convert them to other forms that do not absorb ultraviolet radiation. Chlorinated fluorocarbons, CFCs for short, once commonly used in refrigerators and aerosol spray cans, were found to be depleting the ozone, and in 1978 were banned from use in aerosol sprays. Banning the use of CFCs has not totally solved the ozone depletion problem.

Carbon dioxide is a molecule formed from binding together two oxygen atoms and one carbon atom. Gaseous carbon dioxide readily absorbs selected frequencies of infrared radiation because carbon dioxide molecules have a resonant frequency in the infrared region of the electromagnetic spectrum. Coal, natural gas, and oil all have carbon as a component, and carbon dioxide is produced when they are burned. It is a fact that the concentration of carbon dioxide in the atmosphere has been increasing since the Industrial Revolution. The carbon dioxide has little effect on solar radiation penetrating Earth's atmosphere. Earth absorbs the solar energy and, like any object, emits electromagnetic radiation. The temperature of Earth is more than 5,000 K lower than the sun, and Earth's radiation is mostly infrared. Carbon dioxide can absorb energy from the infrared radiation, resulting in a temperature increase of Earth. This is the so-called "greenhouse effect."

Coping with the greenhouse effect is a very difficult sociopolitical problem. A greenhouse effect existed on Earth long before the Industrial Revolution. Had it not, Earth's surface would be much colder than it is now. The introduction of gases absorbing infrared radiation only enhances the greenhouse effect. Carbon dioxide is not the only gas of importance; water vapor and methane, for example, are also of concern. The industrial world is dependent on coal, oil, and natural gas, so it is not easy to quit burning them to stop producing carbon dioxide.

The radiation one sees from a neon advertising sign or a laser is not produced by accelerated electric charges or by thermal effects. It comes from atoms or molecules that have absorbed energy and given this energy up as photons. A photon is a unit of electromagnetic energy having energy equal to Planck's constant (6.63×10^{-34} joule·seconds) times the frequency of the radiation ($E = hf$). The frequency of the radiation is determined by the magnitude of energy associated with atoms and molecules. Radiation from atoms and molecules is usually in the infrared, visible, and ultraviolet regions of the electromagnetic spectrum. Electromagnetic radiation also results from energy transitions in the nucleus of an atom. However, the nuclear energy scale is roughly a million times larger, making the frequency of the radiation correspondingly larger. This radiation is called gamma radiation.

All electromagnetic radiation has energy to some extent. Absorption of electromagnetic energy can be put to good use, as in photosynthesis or a microwave oven. On the other hand, electromagnetic energy can do damage. Skin cancer can result from absorption of ultraviolet radiation from the sun. This is why young and old are encouraged to stay out of the summer sun between the hours of 10 A.M. and 2 P.M. Gamma radiation is more energetic and more penetrating than ultraviolet, and the damage it can do is not confined to the surface of the skin. This is why some are concerned about the gamma radiation produced by radioactive nuclei in the spent fuel of a nuclear reactor. Ironically, gamma radiation is used to treat certain types of cancer.

Joseph Priest

See also: Atmosphere; Climatic Effects; Communications and Energy.

BIBLIOGRAPHY

Hobson, A. (1995). *Physics: Concepts and Connections.* Englewood Cliffs, NJ: Prentice-Hall.

Serway, R. A. (1998). *Principles of Physics*, 2nd ed. Fort Worth, TX: Saunders College Publishing.

WHEATSTONE, CHARLES (1802–1875)

Charles Wheatstone was born on February 6, 1802, at the Manor House, Barnwood, near Gloucester, England, and died in Paris, France, on October 19, 1875. His father, William, was in business as a shoe maker in Gloucester but he also taught music and several members of the family made musical instruments. On the death of an uncle, Charles and his older brother inherited the uncle's musical instrument manufacturing business in London. Although a competent businessman, he much preferred to give his attention to understanding the basic physics of music and musical instruments. He enjoyed the delicate work of designing and manufacturing musical instruments, and he devised new ones, in particular the highly successful concertina, patented in 1827.

Wheatstone's research into the physics of music addressed two practical questions. What is the difference between the sound vibrations of two notes having the same pitch but different timbre such as the same note played on a flute and a violin? How is sound transmitted through the bridge-piece from the string to the sound board of a violin or piano, and how far can sound be transmitted? These studies brought him into contact with Faraday, who gave several lectures at the Royal Institution in London about his discoveries. They also led to his appointment as Professor of Experimental Philosophy in the new King's College London in 1834.

Before 1831, the usual way of producing an electric current was by chemical means in the electric battery. Each cell of a battery had two different metals, or one metal and one carbon, separated by an acidic liquid. All electrical research in the first third of the nineteenth century made use of such batteries, and many combinations of materials were expired.

In 1831 Michael Faraday showed that an electromotive force is produced when a wire is moved through the lines of force of a magnet. If the wire is part of a complete circuit, a current flows. In the following years, several inventors made magneto-electric generators ("magnetos") in which coils of wire were rotated close to the poles of a fixed magnet, or a magnet was rotated close to a coil of wire. Most such machines were turned by hand and generated electricity without the consumption of expensive chemicals. They offered the prospect of larger machines driven by steam power that could generate electricity in large quantities.

A fundamental limit to the current that could be generated was imposed by the strength of the available permanent magnets. Some inventors sought to overcome this by using pairs of generators, the first being a magneto as just described, and the second being a similar machine except that electromagnets were used in place of the permanent magnets of the magneto. Current produced by the first machine was passed through the coils of the second, which then generated a larger current than could be produced by a magneto alone. Wheatstone was not primarily interested in producing large currents, but he used magnetos in some of his telegraphs.

In February 1837 Wheatstone was introduced to William Fothergill Cooke, an enthusiastic but unscientific inventor and entrepreneur who saw the commercial possibilities of a telegraph, but whose own instruments would not operate through long lengths of wire. The two went into partnership and established the first few working telegraphs in England. Their very first connected Paddington and West Drayton stations on the new Great Western Railway, using Wheatstone's five-needle system. The system also required five wires, which made it costly. Cooke preferred a telegraph with a single wire and a single needle, using a code such as Morse. Sadly, the partners quarrelled and parted. Cooke, with his Electric Telegraph Company, concentrated on the main commercial telegraph business. Wheatstone concentrated on instrument design and on user-friendly telegraphs for private use where letters were indicated directly on a dial, rather than being sent in code.

Wheatstone thought it might be possible to develop a communication system in which the vibrations of speech were transmitted many miles through solid rods or stretched wires, but he could not find a suitable material. Having become interested in the idea of communication at a distance, he took up the idea of the electric telegraph, and soon had an arrangement working in the College with which he could send signals through several miles of wire and deflect the needle of a galvanometer at the far end. He designed an arrangement in which any two of five galvanometer needles in a line could be deflected to point out a letter on a board.

The direct-reading instruments required a series of pulses to be transmitted along the line, each pulse causing the needle to advance one position. Some of Wheatstone's telegraphs used batteries and a switching system to produce the drive pulses, but magnetos naturally produced a series of pulses, and Wheatstone wanted a magneto-electric machine as his telegraph transmitter. In the course of this work he studied the action of the magneto and made some of the earliest measurements of the waveform of the current it produced. A machine that was not dependent on permanent magnets was undoubtedly an attraction for his purposes. He described his self-excited generator in a paper to the Royal Society on February 14, 1867. At that time there was widespread interest in the possibility of more powerful generators. Wheatstone's demonstration machine was driven by two large handles, possibly turned by two of his students. William Siemens described a similar machine at the same meeting. Siemens' title was "On the Conversion of Dynamical into Electrical Force without the aid of Permanent Magnets." Wheatstone approached the matter differently. His title was "On the Augmentation of the Power of a Magnet by the reaction thereon of Currents induced by the Magnet itself."

Wheatstone's self-excited generator was an important step in the development of the electric generator. Although it proved to be the last of Wheatstone's inventions in that field, the remainder of his life was devoted to further improvements in the telegraph.

Brian Bowers

BIBLIOGRAPHY

Bowers, B. (1975). *Sir Charles Wheatstone FRS.* London: HMSO for the Science Museum.

Wheatstone, C. (1867). "On the Augmentation of the Power of a Magnet by the reaction thereon of Currents induced by the Magnet itself." *Proceedings of the Royal Society,* February.

WIND ENERGY

See: Atmosphere; Turbines, Wind

WINDOWS

HISTORICAL BACKGROUND

In primitive homes, the only essential opening was a door for entry and exit. The first window was probably a hole to vent smoke from cooking and heating fires. That first opening offered the same advantages and disadvantages of windows that please and plague building occupants today: it allowed indoor pollutants to escape and daylight to enter; at the same time, it allowed heat to escape in cold weather and to enter in warm weather, creating discomfort for occupants.

Shutters were eventually added. When open, they permitted light and air to enter, along with undesirable visitors, such as intruders, rain, insects, and dust. Closed shutters provided darkness as well as protection. The next addition was translucent materials such as oiled paper or animal skin, framed into window openings.

Transparent glass was first used in Roman times, allowing daylight to enter unaccompanied by unwelcome elements, and giving building occupants a view of the outdoors. The largest surviving piece of Roman glass, three feet by four feet, came from a bath in Pompeii.

Venice became the world's glassmaking center in medieval times. Small, flat panes were produced by cutting and rolling flat hot, blown glass. This technique was used for the buildings constructed by early European settlers in America.

The next advancement in glassmaking was called sheet glass. In this method, a long sheet of glass was developed by dipping a long rod into molten glass and drawing it upward. The glass from this vertical process was then fire-polished.

During the 1600s, a plate glass casting method developed in France produced larger, higher-quality pieces of glass than previously possible. It was the first horizontal process, where glass was flattened through rollers, then ground to polish. The first uses of glass on a grand scale resulted, including the Hall of Mirrors at Versailles and the architectural feature still known as the French door.

Nineteenth-century innovations made larger, stronger, higher-quality pieces of glass more widely available. During the 1950s the float glass technique

Marble walls and crystal lamps flank windows that let light into the Hall of Mirrors in the Chateau de Versailles. (Corbis Corporation)

was developed, in which glass is produced floating on a tank of molten tin. This method produces extremely flat, uniform surfaces with few visual distortions. The desired thickness of glass can be achieved by varying the flow of molten glass from the tin bath. This process is still commonly used for residential windows in the United States.

Prior to 1965, single-pane (single-glazing) windows were standard in U.S. buildings. In cold weather, second panes of glass—storm windows—could be installed for added insulation.

WINDOWS FROM 1970 TO THE PRESENT

The energy crisis of 1973 drew attention to windows' impact on the heating and air conditioning of buildings. Windows were responsible for approximately 5 percent of total energy loss in the United States at the time, or the equivalent of all the energy provided by the Alaska pipeline. Research priorities, which had previously focused on sizing heating, ventilation, and air conditioning (HVAC) systems to meet the heating and cooling energy loads created by windows, turned to improving window energy efficiency. Double-glazed windows—two panes of glass sealed in a frame with air space in between—had been developed prior to the 1960s, but the seals on which their insulating value depended were not always reliable, and they were not widely used. As the quality of sealants improved, the market penetration of those double-pane products increased.

Window Energy Basics

To understand the energy efficiency of windows, it is necessary to define some solar, thermal, and optical properties. All heat transfer through a window is performed through conduction, convection, or radiation. Conduction is transfer of heat through solids by molecular interaction; an example of thermal heat transfer by conduction is heat being lost from a warm room, through the frame of a window. Convection is transfer of heat by fluid—in this case, air—that moves as a result of temperature differences; convection currents involving room air that comes in contact with cold glass are commonly called drafts. Drafts also can be caused by infiltration, described below. Radiation is the physical phenomenon of emitting heat; the Sun and all other objects radiate heat. Radiation can occur in winter when a fire radiates its heat to the objects and surfaces of a room to warm it. Radiation is also at work in the summer when the Sun's energy penetrates a window and heats up the room inside.

To measure the efficiency of a whole window, special testing takes into account all heat transfer from conduction, convection, and radiation. Certain values are used to represent the thermal and solar efficiency of high-performance windows by measuring reduced thermal heat loss (measured by the U-factor), solar heat loss and gain (measured by solar heat gain coefficient, or SHGC), and air infiltration (measured by air leakage).

U-factor. This is the total heat-transfer coefficient of a window system (window plus frame). It represents the amount of heat that moves through the window glass and frame. U-factors are expressed in watts per square meter per degree of temperature difference (Centigrade) or BTUs per hour per square foot per degree of temperature difference (Fahrenheit). In climates with heating requirements, low U-factors are recommended. U-factor is generally expressed as a fraction between zero and one, but in cases of very inefficient products the number can be greater than one.

Solar heat gain coefficient (SHGC). This is a measure of how much of the Sun's energy that is transmitted through the window becomes heat. It includes both the heat that is transmitted through the window and the heat that is absorbed by the window and then reradiated into a room. SHGC is also expressed as a fraction between zero and one. The SHGC can be determined for any angle at which sunlight hits the glass, but the most commonly used reference is for "normal incidence" —sunlight striking the glass at a ninety-degree angle. In climates with cooling requirements, low SHGC numbers are recommended. Because low SHGC numbers also improve the thermal performance of a window (i.e., U-factor), they are also recommended in heating climates, depending on the architectural design of the building.

Visible transmittance (VT). This is a window system's transmittance across the visible portion of the solar spectrum; it is a measure of the visible light that comes through window glass without being reflected or absorbed. VT is expressed as a fraction between zero and one. It is desirable to preserve visible transmittance—building occupants' access to daylight and views of the outdoors—while reducing U-factors, SHGC, and transmission of UV radiation. In all climates, a high VT is recommended.

Infiltration rate. This is the rate at which air moves through a window system as a result of pressure differences between the inside and the outside. Infiltration rates are expressed as airflow rates—cubic meters per hour or cubic feet per minute per square foot of window. An infiltration rate below 0.3 cfm/sq ft is recommended in all applications.

Technological Advances in Window Energy Efficiency

Reducing unwanted heat loss. Technological advances in window performance have come in the components of windows (glass, frames, spacers, sealants, warm edge design) and in the way those components fit together to make the whole window more efficient. A major goal in window energy efficiency starting in the late 1970s has been to design products with lower U-factors to achieve the benefits described above. Efforts to produce highly insulating window products in the 1970s and early 1980s focused on existing technologies: double-glazed windows led to triple- and quadruple-glazed windows, which lowered U-factors by increasing the number of still air spaces in the window system. Each layer reduced U-factors by approximately 1 hr-ft^2-F/Btu but also increased the window's weight and cost of fabrication as well as significantly lowering visible transmittance, which diminished the quality of the light and view.

In the mid-1970s it became clear that most of the heat transfer through a typical double-glazed window was through thermal radiation exchange from one layer of the window to another. Although glass blocks direct transmittance of long-wave infrared radiation in sunlight, it absorbs the radiation and reemits it at a high rate. To solve this problem, the first transparent, optically clear, low-emissivity (low-e) coating was developed in 1983 to reduce radiative heat transfer without compromising visible transmittance.

Radiant heat transfer had historically been the biggest heat transfer mechanism for windows. Low-e materials were developed and have historically been used to control for heat transfer. An example of the popularity of using metals to reflect heat to control for radiant heat transfer is the thermos bottle. Applying that new technology to windows, and getting materials that normally would affect transparency of the product to remain visually neutral, was a huge advance to the industry.

Low-e coatings are microscopically thin, virtually invisible metal or metallic oxide layers deposited on a window or glazing surface. Low-e coatings are virtually transparent to visible light. At the time of its invention, technology had not been developed to apply this product, a thin polyester film with a silver-based coating, directly to large pieces of glass, so it was applied to a substrate that was then suspended

An estate of houses installed with solar panels and other energy-saving devices at Shenley Lodge in Milton Keynes, England. (Corbis Corporation)

between two glazing layers. This technique had the advantage of creating another insulating air space in the double-glazed window unit, cutting window heat loss rates by up to 50 percent, but it also increased manufacturing costs.

Two processes, referred to as sputtered and pyrolitic, were developed to produce large volumes of quality, low-e coated glass. Pyrolitic coatings are incorporated into float glass production and tend to be more durable. Sputtered systems use a stand-alone vacuum deposition process to produce coatings that are have lower emissivities but that are softer and need more protection than pyrolitic coatings.

Reducing heat transfer with gas fill. Conduction and convection cause heat transfer across the air spaces in multilayer windows. Although air is a relatively good insulator, other gases that have lower thermal conductivity can be sealed into the cavities between panes. These gases include argon, carbon dioxide, and krypton. Developments in sealants and quality control during the 1980s reduced gas leakage in some window units to less than 1 percent per year.

By the late 1980s, many manufacturers offered double-glazed units with low-e coatings and inexpensive gas fills such as argon. Even greater efficiency could be achieved with extra-low-conductivity gases such as krypton filling small gaps between panes. Three-layer units with two gas-filled gaps and low-e coatings to reduce radiative heat transfer were designed at Lawrence Berkeley National Laboratory (LBNL), a leader in U.S. energy-efficient windows research. Products using these concepts appeared on the market in 1991.

Reducing heat transfer through warm-edge design. As new efficient glass technologies emerged, the relative fraction of heat lost through window frames decreased. During the 1990s, existing metal-framed windows were increasingly replaced by lower-conductivity vinyl. Pultruded fiberglass also became an energy-efficient framing option. The high-conductivity aluminum spacers typically used to hold panes in place in multilayer windows were replaced with more energy-efficient materials: metal-reinforced butyl, stainless steel, thermally broken aluminum or steel, and silicone foam.

Reducing infiltration. Traditional "brute force" approaches to weather stripping were redesigned to reduce infiltration with less material that contacted a larger surface area. New materials such as thermoplastic elastomers offered additional potential to control for infiltration. But perhaps the most important way to control infiltration is proper installation.

Controlling unwanted solar heat gain. The first generation of low-e coatings, described above, was designed for colder climates, where controlling heat loss is important. The next generation, introduced during the 1980s, selectively blocked the UV and shortwave infrared or "heat" radiation parts of the solar spectrum while admitting as much daylight as possible. These new spectrally selective coatings were useful in warmer areas, where reducing solar heat gains in summer and interior heat loss in winter is important. Where building energy codes are restrictive about the percentage of glazing allowed or the total energy performance of the unit, use of low-solar gain coatings can permit architects to increase total window area compared to what would be allowed in a building design with less efficient glazing.

TODAY'S WINDOWS

Today, nearly 90 percent of all residential windows sold in the United States have two or more layers of glazing. Approximately one-third of all windows sold for residential buildings and 20 percent of windows sold for commercial buildings in 1996 had low-emissivity coatings and insulating frames appropriate for the local climate. Besides the energy savings, energy-efficient windows increase thermal comfort, increase daylight and view, decrease condensation, decrease utility bills, and offer protection from UV fading.

Consumers buy homes with windows for the aesthetics, daylighting, and improved view of windows, yet they don't want to sacrifice comfort, that condition of mind that expresses satisfaction with the environment. Energy efficient windows transfer less energy, so less indoor heat is lost in cold weather and less unwanted heat from the outside enters in warm weather. Heating and cooling energy costs are therefore reduced, and occupants are more comfortable because the window's surface temperature is closer to the ambient indoor temperature. Windows without energy-efficiency features have very warm surfaces in warm weather and cold surfaces in cold weather, so occupants near a window feel uncomfortable.

During the heating season, high performance windows create warmer interior glass surfaces, reducing frost and condensation. The development of low-conductance spacers to reduce heat transfer near the edge of insulated glazing, and improved seals have significantly reduced condensation under all conditions.

Fading is probably one of the least understood consequences of visible and ultraviolet (UV) light transmitted through windows. Fabrics fade at different rates; UV radiation can fade organic materials such as carpet, artwork, paints, and wood differently than it fades synthetic materials. Energy efficient windows that reduce transmission of the UV spectrum not blocked by ordinary glass can reduce the potential for fading damage. Although most fading damage occurs from UV, visible light also has the potential to cause some damage.

People prefer windows in buildings to add light, open up space, and create aesthetically pleasing environments in which to live and work. Windows are also used for natural ventilation in some climates. Energy efficient windows allow increased daylight to enter a building without compromising the comfort of a room by allowing in unwanted heat or cold, particularly infiltration that can create drafts. Several studies also suggest daylighting can increase productivity, health and welfare.

Homeowners can realize substantial reductions in energy bills by selecting energy efficient windows. It is possible to save up to 30 percent on the heating and cooling portion of one's utility bill by selecting high performance windows. Moreover, building owners can rely on smaller heating and cooling equipment to meet smaller loads, and power suppliers that can meet customer demand with less generating capacity.

Research has shown that if all new windows sold in warmer U.S. regions had low-solar gain coatings, cooling energy use could be reduced by 25 percent in 2010; if residential windows in all parts of the United States had these coatings, heating energy use would

Buildings in downtown Detroit with sunlight reflecting off the myriad windowpanes. (Field Mark Publications)

decline by 19 percent. The total projected energy savings would be $1.2 billion per year. However, the market share for windows purchased with low-e coatings in the late 1990s is less than 20 percent in the southern United States, compared to 50 percent or more in the coldest states.

Choosing Efficient Windows

Window recommendations for a particular installation depend on local climate variations and the intent of the occupant and designer. If reducing energy costs is the primary goal, local utility rates must be taken into account along with which space-conditioning loads (heating or cooling) dominate. If comfort is the primary goal, occupant preferences must be addressed along with local site characteristics such as nearby shade trees that may help keep the building cool.

To achieve maximum energy performance, an energy-efficient window's characteristics must compensate for local climate conditions. In residential buildings in cold climates, the majority of building conditioning energy goes to heating, so efficient windows must keep heat indoors. In some cases, the solar gains that enter through a window may exceed the energy lost through the window, capturing passive solar gains and retaining captured heat. By contrast, solar heat gain is a liability in a residential building in a warm climate where cooling accounts for the majority of space conditioning energy use. Passive solar design in cooling climates necessarily employs different strategies: maximizing shading, using light roof colors and, in some cases, allowing natural ventilation.

Commercial buildings, such as offices with large windows, present a different challenge from residences. In almost all climates, commercial buildings suffer from excess solar heat gains, which produce large cooling loads. These buildings are best served by windows that filter out unnecessary solar radiation (and complementary technologies that redirect daylight to spaces away from windows to offset electrical lighting needs, see below).

A number of efforts assist U.S. consumers and homeowners in choosing windows. The National Fenestration Rating Council (NFRC), a non-profit organization created in 1989 by members of the window industry, code officials, energy officials, utilities and architects to devise a technically sound, impartial way to compare windows, has established a rating, labeling, and certification system. All NFRC ratings are for the whole window, that is, window plus frame. In the past, ratings were based on glass performance only, which sometimes over-rated the performance of a real installed window production. Laboratory tests and computer simulations determine a window's basic thermal and optical properties, which appear in NFRC certification documents for most windows sold in the U.S. Air infiltration and ratings of annual heating and cooling energy performance may be added sometime in the early 2000s. Work is under way to rate condensation resistance and long-term energy performance.

The U.S. Department of Energy (DOE) and the Environmental Protection Agency (EPA), have established the Energy Star program for energy-efficient windows to certify which products meet minimum energy performance criteria. Because window efficiency depends on the local climate, Energy Star window certifications are based on three U.S. climate zones: northern (heating energy dominated), southern (cooling energy dominated), and central (roughly equal heating and cooling).

The Efficient Windows Collaborative (EWC), managed by the Alliance to Save Energy, is a coalition of window and window component manufacturers, research organizations, non-profits, and federal and state government agencies interested in expanding the market for high-efficiency window products. Established in 1997, the EWC educates builders, remodeling contractors, home energy raters, utilities, and consumers about efficient window products.

WINDOWS OF THE FUTURE

Future energy-efficient window technologies will likely emphasize further control of heat loss and solar gain and integration with other building components. New technologies also will likely find ways to harness the power of thermal and solar heat transfer rather than just controlling it. Harnessing the power of the Sun, wind, and elements to make windows power-generation sources rather than energy losers will occur in the future. Designers, planners, and window industry visionaries foresee windows as dynamic, integrated, controllable "appliances in the walls." Windows of the future may very well be automated home systems that serve many purposes. The technologies listed below are only a few in a list of hundreds that may improve efficiency, performance, and visibility.

Future Window Technologies for Controlling Heat Loss

The gas fills that reduce the U-factors of low-e, insulated windows require very low pressure—that is, a vacuum or evacuated window. Research is focusing on improving seals, which are critical in vacuum windows, and methods of trapping the gas fill, which would otherwise diffuse through the seals. Spacers that keep the glass layers from collapsing under the required low pressure are significant sources of heat transfer; reducing their thermal bridging effect is under study.

Another promising means to reduce conductive, convective, and radiative heat transfer through windows is to fill the space between panes with a transparent insulating material. Silica aerogel is under development for this purpose. Because the silica particles are much smaller than the wavelengths of visible light, aerogel is transparent to the human eye. Its thermal conductivity is lower than that of air, and it is opaque to most longwave infrared radiation, so radiative heat losses through an aerogel-insulated window are minimal. It is much easier to vacuum-seal a window with aerogel insulation than with gas insulation. Aerogel is also strong enough to balance external atmospheric pressure on the window, so no spacers are needed. Testing and evaluation to confirm the performance of silica aerogel is under way.

Future Window Technologies for Controlling Solar Heat Gain

Solar heat gain through windows can add to or detract from energy efficiency, depending on a building's type and use, the local climate, the season, and even the time of day. For some applications, SHGCs should be maximized; for others, SHGCs should be minimized.

For residential buildings in climates where heating energy use dominates, it is important to maximize solar gains through windows. Solar heat transmission rates can be decreased by minimizing impurities in the glass (e.g., using low-iron glass). Low-e coatings may also be designed to be "smart" about maximizing or minimizing gains as a result of inside environmental conditions.

Controlling solar heat gains is equally important because

1. Residences are increasingly being built with air conditioners; in northern climates, cooling costs may be similar to heating costs even though the annual demand is heating-dominated, because electricity used for cooling is generally much more expensive than gas for heating.
2. Almost all commercial buildings are cooling-dominated, and solar gains through windows are major contributors to cooling loads.
3. Cooling loads are most significant on hot summer afternoons when electric utilities typically have annual peak loads. Lower solar heat gains lead to reduced peaks, which reduce peak demand charges and power plant capacity requirements.

Solar heat gain needs to be controlled in three ways: by intensity, by variations in time, and by spatial distribution

Controlling solar heat gain intensity. As described above, low-solar gain low-e glazing blocks out parts of the solar spectrum. Approximately half of incident solar energy is in the solar infrared; this solar radiation will not light interior spaces, but it is responsible for most of the heat gain, so it should be the first to be rejected by a selective glazing.

Controlling solar heat gain variations in time. Optical coatings that control solar heat gain on a time-dependent basis may be advantageous in applications such as commercial buildings where the benefits of solar gain and daylight vary daily or hourly, and in residential buildings where the desirability of solar gains varies seasonally. Glazings with variable properties, typically called chromogenic (switchable) glazings or "smart" windows, can be more effective and efficient than traditional shading devices.

There are several different types of switchable glazings. All are activated by a physical phenomenon (light, heat, or electricity), and, once activated, they switch, either incrementally or completely, to a different state. The state in which they permit the highest solar transmittance (typically the inactivated state) is called the bleached state; the state with the lowest solar transmittance is called the colored state.

Three switching mechanisms have been identified: photochromic (light-sensitive), thermochromic (heat-sensitive), and electrochromic (electrically activated). Photochromic materials alter their optical properties when exposed to light, typically UV, and revert to their original state in the absence of light. Probably the most familiar and widespread consumer use of photochromics is in sunglasses. Photochromic glass cannot yet be produced using the float glass process, so it is prohibitively expensive for windows. Current research is on developing plastic photochromic layers that are durable enough for use in windows. Photochromic glazing may be appropriate to control daylight transmission but is insensitive to solar heat gain.

Thermochromic materials change optical properties when subjected to temperature change. Their primary commercial uses to date are in inks, paints, security devices, and temperature indicators. Research in window thermochromics is focusing on durability, specification of switching or activation temperatures and ranges, and achievement of sufficient optical clarity in the colored state. Thermochromics can respond well to thermal effects but may not allow desirable daylight transmittance and may not distinguish between high ambient temperatures and incident solar radiation.

Electrochromic glazings are complex, multilayer coatings whose optical properties vary continuously from a highly transmissive (bleached) state to a low-transmitting (colored) state. A small electrical potential difference, typically one to five volts, triggers the change in state. The longer the potential difference is applied, the more the electrochromic layer will switch until it reaches its maximum or minimum. Electrochromics can be tied into building energy management control systems that optimize trade-offs between lighting energy and cooling consumption. Electrochromics appear to have greater potential for energy savings and user comfort than photochromics or thermochromics.

Controlling solar heat gain distribution. Commercial buildings often have a surplus of daylight near exterior windows and a lack of daylight deep in their interiors. Skylights can help on the top floor; strategies are being explored to redistribute daylight from side windows to the center of the building on lower floors.

Various technologies are being studied to maximize daylight, direct daylight and reject heat gain, and avoid visual glare. Two types of technology are being studied: holographic and angular selective glazing. Light shelves and complex light pipes also can redirect daylight. Light shelves often are simple horizontal elements on a building's exterior. Light pipes are rectangular "pipes" extending into a building from the exterior. Variations on these kinds of architectural features are being studied.

Transparent or translucent insulating materials (TIMs) can provide light or solar gains without view. TIMs typically have thermal properties similar to conventional opaque insulation and are thicker than conventional insulating glass units, providing significant resistance to heat transfer.

TIMS often have the same thermal conductance as a wall, so they are appropriate to admit daylight in areas where insulated walls would otherwise be installed. In Europe, where there are many uninsulated masonry buildings, TIMs with solar control mechanisms could be excellent retrofits. TIMs' optical properties are different from those of conventional insulating glass units; solar radiation often is scattered as is passes through a TIM. As a result, measuring the optical properties of TIMs can be challenging. Laminated and safety glazings were originally designed to help structurally support the building. They also may help to control daylight and reduce solar heat gain.

Integration of Windows with Other Building Components

Windows often are thought of as independent components of a building. However, to maximize windows' energy efficiency, they must be integrated with the rest of a building system.

Projects around the country are beginning to demonstrate that the energy consumption of new houses can be reduced by as much as 50 percent with no behavioral changes and little or no impact on the cost of construction. This is possible because cost savings in one component can then be reinvested to improve energy performance and product quality. New techniques for improving the building envelope, such as upgrading to high-performance windows, often enable builders to install smaller, less expensive heating and cooling systems than would be needed with traditional windows. Equipment cost savings in heating, ventilation, and air conditioning (HVAC) offset the added envelope improvement costs.

At a demonstration project in Las Vegas, low-solar gain low-e windows were a key technology used to reduce measured cooling energy by 20 percent and heating energy by 50 percent in new homes. After improvements to the construction, including energy-efficient windows with an SHGC of 0.4 and a U-factor of 0.35, the HVAC was sized at three tons, with a resulting cost savings of $750 to the builder. Energy-efficient windows accounted for half a ton to one ton of those savings.

Specific integrated residential applications. An Integrated Window System (IWS) is a window-wall panel with a number of energy-efficient features:

- movable interior or exterior insulating night shades recessed into the wall cavity;
- seasonally deployed solar shading, that can be interior, exterior, or in-between;
- optional overhangs;
- insulated box beam headers;
- larger glazed area;
- elimination of structural window form elements.

IWSs would most likely be factory-built and integrated into conventional construction, replacing likely framing components. IWS prototypes are being designed.

Commercial applications. Exterior building design, glazing, lighting systems, and heating, ventilation, and air-conditioning equipment are typically specified and installed separately in commercial buildings. However, to realize the energy-saving potential of advanced products such as electrochromics or light shelves, it will require concurrent advances in sensor, software, and hardware of energy management control systems.

Other Integrated Energy Issues

In addition to the examples cited above, many manufacturers and researchers in the industry are trying to identify methods of reducing embodied energy costs. Glassmaking is an energy-intensive industry; vinyl, aluminum, and fiberglass extrusion also is energy-intensive. Making wood windows out of an ever-shrinking supply of forestry is becoming difficult. Researchers hope to identify whether there is some advantage to recycling whole window products at the end of their useful life and whether that process is more energy- and resource-efficient than beginning again with virgin materials.

Alicia Ward

See also: Air Conditioning; Building Design, Commercial; Building Design, Energy Codes and Building Design, Residential; Climatic Effects; Cool Communities; Efficiency of Energy Use; Energy Management Control Systems; Insulation.

BIBLIOGRAPHY

Arasteh, D. (1996). "Advances in Window Technology: 1973–1993." In *Advances in Solar Energy*, ed. K. W. Boöer. Boulder, CO: American Solar Energy Society.

Arasteh D. and Selkowitz, S. (1989). "A Superwindow Field Demonstration Program in Northwest Montana." *ASHRAE Transactions* 95(1).

Arasteh, D.; Selkowitz, S.; and Hartmann, J. (1985). "Detailed Thermal Performance Data on Conventional and Highly Insulating Window Systems." *ASHRAE Transactions* 91(1).

ASHRAE. (1997). *ASHRAE Handbook: Fundamentals*. Atlanta: Author.

ASHRAE. (1981). *Thermal Environmental Conditions for Human Occupancy*. Atlanta: Author.

ASTM. (1988). *Standard Test Method for Solar Absorptance, Reflectance, and Transmittance of Materials Using Integrating Spheres*. Philadelphia: Author.

ASTM. (1992). *Standard Test Method for Determining the Rate of Air Leakage Through Exterior Windows, Curtain Walls, and Doors Under Specified Pressure Differences Across the Specimen*. Philadelphia: Author.

Beck, F. A., and Arasteh, D. (1992). "Improving the Thermal Performance of Vinyl-Framed Windows." Proceedings of the ASHRAE/DOE/BTECC Conference on the Thermal Performance of the Exterior Envelopes of Buildings V, Clearwater Beach, FL.

Carmody, J.; Selkowitz, S.; and Heschong, L. (1996). *Residential Windows*. New York: W. W. Norton.

Carpenter, S., and McGowan, A. (1993). "Effect of Framing Systems on the Thermal Performance of Windows." *ASHRAE Transactions* 99(1).

Chahroudi, D. (1991). *Weather Panel Architecture*. Albuquerque NM: Suntek.

Collins, R. E.; Fischer-Cripps, A. C.; and Tang, J. Z. (1992). "Transparent Evacuated Insulation." *Solar Energy* 49(5): 333–350.

Ducker Research Company. (1997). *Study to Quantify and Profile the U.S. Market for Residential and Light Commercial Windows and the Technology for High-performance Windows*. Berkeley, CA: Lawrence Berkeley National Laboratory.

Efficient Windows Collaborative. <http\\:www.efficientwindows.org>.

Elsherbiny, S. M.; Raithby, G. D.; and Hollands, K. B. T. (1982). "Heat Transfer by Natural Convection Across Vertical and Inclined Air Layers." *ASME Transactions* 204:96.

Finlayson, E. U.; Arasteh, D.; Huizenga, C.; Rubin, M. D.; and Reilly, M. S. (1993). WINDOWS 4.0: *Documentation of Calculation Procedures*. Berkeley, CA: Lawrence Berkeley National Laboratory.

Frost, K.; Arasteh, D.; and Eto, J. (1993). *Savings from Energy-Efficient Windows: Current and Future Savings from New Fenestration Technologies in the Residential Market*. Berkeley, CA: Lawrence Berkeley National Laboratory.

Frost, K.; Eto, J.; Arasteh, D.; and Yasdanian, M. (1996). *The National Energy Requirements of Residential Windows in the U.S.: Today and Tomorrow*. Berkeley, CA: Lawrence Berkeley National Laboratory.

Geller, H., and Thorne, J. (1999). *U.S. Department of Energy's Office of Building Technologies: Successful Initiative of the 1990s*. Washington, DC: American Council for an Energy-Efficient Economy.

Heschong, L. (1999). *Daylighting in Schools: An Investigation into the Relationship between Daylighting and Human Performance*. Fair Oak, CA: Heschong Mahone Group.

Hunt, A. J. (1992). "Optical and Thermal Properties of Silica Aerogel." In *Chemical Processing of Advanced Materials*, ed. L. I. Hench and J. K. West. New York: John Wiley & Sons.

Johnson, R.; Connell, D.; Selkowitz, S.; and Arasteh, D. (1985). "Advanced Optical Materials for Daylighting in Office Buildings." Tenth National Passive Solar Conference Proceedings, American Solar Energy Society, Inc.

Kim, J. J.; Papamichael, K.; Selkowitz, S.; Spitzglas, M.; and Modest, M. (1986). "Determining Daylight Illuminance in Rooms Having Complex Fenestration Systems." *International Daylighting Conference Proceedings* 2:219.

Klems, J. (1983). "Methods of Estimating Air Infiltration through Window." *Energy and Buildings* 5:243–252.

Lampert, C. M., and Ma, Y. (1992). *Fenestration 2000: Advanced Glazings Materials Study*. St. Helens, UK: Pilkington Glass.

Lee, E. S., and Selkowitz, S. E. (1993). "Integrated Design Yields Smart Envelope and Lighting Systems." *Consulting-Specifying Engineer* 64.

Papamichael, K.; Beltraán, L.; Furler, R.; Lee, E. S.; Selkowitz, S.; and Rubin, M. (1993). *The Energy Performance of Prototype Holographic Glazings*. Berkeley, CA: Lawrence Berkeley National Laboratory.

Papamichael, K.; Klems, J.; and Selkowitz, S. (1988). "Determination and Application of Bidirectional Solar-Optical Properties of Fenestration Systems." Proceedings of the 13th National Passive Solar Conference, Massachusetts Institute of Technology.

Reilly, M. S.; Arasteh, D.; and Rubin, M. (1990). "The Effects of Infrared Absorbing Gasses on Window Heat

Transfer: A Comparison of Theory and Experiment." *Solar Energy Materials* 20:277–288.

Reilly, S.; Arasteh, D.; and Selkowitz, S. (1991). "Thermal and Optical Analysis of Switchable Window Glazings." *Solar Energy Materials* 22:1–14.

Selkowitz, S. E. (1981). "Thermal Performance of Insulating Window Systems." *ASHRAE Transactions* 85(2).

Selkowitz, S. E., and Lampert, C. M. (1989). "Application of Large-Area Chromogenics to Architectural Glazings." In *Large Area Chromogenics: Materials and Devices for Transmittance Control*, ed. C. M. Lampert and C. G. Granqvist. Bellingham, WA: Engineering Press.

Shakerin, S. (1987). "Wind-Related Heat Transfer Coefficient for Flat-Plate Solar Collectors." *Journal of Solar Energy Engineering* 109:108–110.

Sullivan, R.; Chin, B.; Arasteh, D.; and Selkowitz, S. (1991). "RESFEN: A Residential Fenestration Performance Design Tool." *ASHRAE Transactions* 98(1).

Sullivan, R., and Selkowitz, S. (1987). "Residential Heating and Cooling Energy Cost Implications Associated with Window Type." *ASHRAE Transactions* 93(1):1525–1539.

Warner, J.; Reilly, S.; Selkowitz, S.; and Arasteh, D. (1992). "Utility and Economic Benefits of Electrochromic Smart Windows." Proceedings of the ACEEE 1992 Summer Study on Energy Efficiency, Asilomar Conference Center, Pacific Grove, CA.

Weidt, J. L.; Weidt, J.; and Selkowitz, S. (1981). "Field Air Leakage of Newly Installed Residential Windows." *ASHRAE Transactions* 85(1):149–159.

Windows and Daylighting Group. (1992). *WINDOWS 4.0, A PC Program for Analyzing Window Thermal Performance*. Berkeley, CA: Lawrence Berkeley Laboratory.

Wright, J. (1993). *VISION 3*. Waterloo, Ontario: University of Waterloo.

Yazdanian, M., and Klems, J. (1994). "Measurement of the Exterior Convective Film Coefficient for Windows in Low-Rise Buildings." *ASHRAE Transactions* 100.

WOOD

See: Biofuels

WORK AND ENERGY

Work, in general, is an activity in which one exerts strength or faculties to do or perform something. In physics, work is defined as a force on an object multiplied by the displacement of the object. Measuring force in newtons (N) and displacement in meters (m) yields newtons times meters (Nm) as the units of work. To honor James Prescott Joule, who did pioneering work in the science of thermodynamics, a newton-meter is called a joule (J). A person pushing on a box with a force of 200 N (about 50 pounds) and moving the box 1 m would do 200 J of work. Energy is capacity, or ability, for doing work. Work and energy are both measured in joules.

After being outside on a cold day, a person's hand warms when shaking hands with someone who has spent time in a warm room. Heat flows from the warm hand to the cool hand to produce the warming. Rather than shaking hands, the person could briskly rub his hands together. In this case, warming is a result of doing work, not due to a flow of heat. Heat is energy in transit and, accordingly, is measured in joules. It takes 4,200 joules of heat to raise the temperature of 1 kilogram of water 1 degree Celsius. The same temperature rise would occur if 4,200 joules of mechanical work were done on the water using a food mixer, for example.

The concepts of heat and work are very important for understanding the roles of automobiles, airplanes, trucks, trains, and electric power plants that are essential for a modern industrial society. Examining the technology, one always sees the involvement of an engine of some sort. Most automobiles employ a gasoline engine, trucks and trains use diesel engines, and an electric power plant uses a steam turbine to drive an electric generator. There are significant differences in the size and complexity of the engines involved, but they all convert heat to useful work.

The word kinetic implies motion. An object in motion has a capacity for doing work and has kinetic energy. Forces acting on an object do work when the object moves and the kinetic energy changes. When pulling on a sled to start it moving, force and displacement are in the same direction. Work done by this force is taken to be positive. While the sled is moving, frictional forces on the runners pull backwards. The force and displacement are in opposite directions and the work is taken to be negative. The net amount of work is the algebraic sum of the work due to all forces acting on an object. For example, if pulling on the sled produced 100 J of work and friction produced –50 J the net amount of work is 100 J – 50 J = +50 J.

A very important principle called the work-energy principle reads as follows: net amount of work is equal to the change in kinetic energy. Emphasis is placed on net and change. If the net work is positive, this means the net force and displacement have the same direction, and the kinetic energy increases. Increasing the kinetic energy of a car requires positive work. Conversely, decreasing the kinetic energy requires negative work. When a braking system brings a car to rest, forces must act in a direction opposite to the direction the car is moving. If a car travels down a highway at constant speed, its kinetic energy does not change. This means the net work on the car is zero. Air always pushes against the car and does negative work. To keep the speed from changing, forces that propel the car forward must do positive work of equal magnitude. The faster the car travels, the greater the negative work done by the air. As a result, forces driving the car forward must do more work. This means more consumption of gasoline. This is the basic reason why the fuel economy of a car is sensitive to speed.

Joseph Priest

See also: Joule, James Prescott.

BIBLIOGRAPHY

Hobson, A. (1995). *Physics: Concepts and Connections.* Englewood Cliffs, NJ: Prentice-Hall.

Serway, R. A. (1998). *Principles of Physics*, 2nd ed. Fort Worth, TX: Saunders College Publishing.

YUCCA MOUNTAIN

See: Nuclear Waste

ZERO-WATER DISCHARGE

See: Emission Controls, Power Plant

Z-PINCH

See: Nuclear Fission

ENERGY TIMELINE

ENERGY SOURCE, PROCESSING, AND STORAGE EVENTS

B.C.E.

2600	Construction of Lake Moeris in Egypt, a reservoir created by a dam 27 miles long.
2589	Construction of Great Pyramid of Khufu begins and lasts until 2566 B.C.E.; innovative use of available power for transportation and construction.
1500	Water clocks being used by Egyptians.
850	Natural gas utilized in China.
400	An oil well is completed on an island in the Ionian Sea and the oil used in lamps.
180	Revolving mill invented; it is powered by slaves and asses.
65	Earliest known reference to the use of a windmill is made by Antipater of Thessalonica.

C.E.

600	Arabs develop a windmill in which the paddlewheel revolves in the vertical plane.
1000	Burmese successfully drill oil wells.
1003	Wells are drilled in China for natural gas, which flows through bamboo pipes to be used perhaps in porcelain manufacture.
1200	Coal being mined in Europe.
1269	Frenchman Pèlerin de Maricourt writes a treatise on magnetism that includes the earliest description of the compass in the Western world.
1430	Turret windmill invented.
1600	England suffers from timber and fuel wood scarcity.
1619	Coke first used instead of charcoal in a blast furnace.
1635	John Winthrop, Jr., opens America's first chemical plant in Boston. They produce saltpeter (used in gunpowder) and alum (used in tanning).
1640	Oil well completed in Italy; kerosine from the oil later used for lighting
1670	First distillation of gas from coal.
1678	Dutch mathematician Christiaan Huygens first states his wave theory of light, published in *Traité de la lumière* in 1690.
1682	Law of gravitation announced by English mathematician and physicist Isaac Newton; five years later his *Philosophiae Naturalis Principia Mathematica* is published, setting forth the laws of motion as well as gravitation.
1694	Oil produced in England by retorting oil shale and cannel coal.
1793	British physicist Benjamin Thompson (Count Rumford) shows that work is convertible into heat and vice versa.
1799	Production of coal gas.
1800	Italian physicist Alessandro Volta demonstrates the galvanic cell, also known as the voltaic cell.
1802	Gas lighting.
1819	English chemist and physician John Kidd obtains naphthalene from coal tar, pointing the way toward the use of coal as a source of many important chemicals.
1830	American physicist Joseph Henry discovers the principle of electromagnetic induction. English chemist and physicist Michael Faraday independently discovered the same principle a year later but is the first to publish his findings.

1850s	The first petroleum refinery consisting of a one-barrel still is built in Pittsburgh by Samuel Kier.
1850	Scottish chemist James Young starts to produce “coal oil” (kerosine) from coal.
1853	Kerosene is extracted from petroleum.
1854	The Pennsylvania Rock Oil Company becomes the first oil company in the United States.
1855	Chemist Benjamin Silliman, of New Haven, Connecticut, obtains valuable products by distilling petroleum. They include tar, gasoline, and various solvents.
1856	The first synthetic dye is developed by William H. Perkin (English); he accidentally creates a mauve dye from the impure aniline in coal tar.
1857	Oil is discovered in Romania and Ontario.
1859	The first commercially successful U.S. oil well is drilled by E. L. Drake near Titusville, Pennsylvania. This 70-foot well launches the petroleum industry.
1867	Swedish philanthropist Alfred Nobel patents dynamite.
1886	There are forty to fifty water-powered electric plants reported to be online or under construction.
1888	The first dry cell battery, consisting of a moistened cathode and a swollen starch or plaster of paris separator, is invented.
1895	The Edison dry cell nickel/cadmium battery and the Jungers nickel/iron cell are developed; work on these endeavors lasts until 1905.
1900	American chemist Charles Palmer makes a breakthrough in devising a thermal process to produce gasoline from crude petroleum.
1900	First off-shore wells, fixed to piers, are drilled in the Caspian Sea.
1901	Oil drilling begins in Persia.
1905	German-American physicist Albert Einstein formulates the Law of Mass-Energy Equivalence ($E=mc^2$) and the Photon Theory of Light.
1907	Atlantic Refining Company introduces the tower still refinery, in which petroleum is separated in a continuous process rather than in batches.
1913	The world’s first geothermal power station begins operation in Italy.
1918	Shell introduces the first drill with diamond tooth edges.
1921	Federal Energy Regulatory Commission statistics show hydropower generation at 3,700 MW (cf. 1992 figure of 91,600 MW).
1918	Ethyl gasoline is developed by General Motors Laboratories in the United States.
1926	Du Pont and Commercial Solvents begin synthetic methanol production in the United States.
1931	Oil drilling becomes more accurate because of the gyroscopic clinograph that stabilizes the drill.
1932	After American physicist Ernest Orlando Lawrence invented the cyclotron three years earlier, the E. O. Lawrence Cyclotron becomes operational. It helps scientists discover what an atom is composed of, how it behaves, and how its energy can be tapped.
1933	Construction of Grand Coulee Dam begins. Originally built to meet irrigation needs, it has more electric generating capacity than any dam in North America by 1975.
1935	From 1920 to 1935, 6.5 million windmills have been erected in the United States to pump water, run sawmills, or generate electricity.
1936	The Hoover Dam is completed.
1936	The Houdry Process is used in the catalytic cracking of petroleum.
1937	Westinghouse constructs its “Atom Smasher” in Forest Hills, Pennsylvania. The five million volt van de Graaff generator represents the first large-scale program in nuclear physics established in industry, makes possible precise measurements of nuclear reactions, and provides valuable research experience for the company’s pioneering work in nuclear power.
1939	Enrico Fermi (Italian-American), Otto Hahn (German), F. Strassman, Lise Meitner (Austrian), and Otto Frisch (Austrian) discover and describe nuclear fission.
1940	Standard Oil Co. (Indiana) develops catalytic reforming to produce higher octane gasoline.

1941 American nuclear chemist Glenn Seaborg's team of experimenters isolates plutonium, which proves to be a better fuel for nuclear reactors than uranium because of its greater energy yield.

1942 The Manhattan Project at the University of Chicago Laboratory, headed by Enrico Fermi (Italian-American), creates the first self-sustaining nuclear chain reaction.

1942 Natural gas liquified for first time in Cleveland, Ohio.

1944 Seismic profiling for oil deposits begins in the Gulf of Mexico.

1947 Offshore oil wells drilled off coast of Louisiana.

1951 An electricity-producing nuclear breeder reactor commissioned by the U.S. Atomic Energy Commission.

1952 The first hydrogen (fusion) bomb to be tested by the United States is exploded at Bikini Atoll.

1954 First submersible drilling unit, "Mr. Charlie," is used.

1954 First solar cell developed by Bell Telephone Laboratories researchers.

1959 Liquified natural gas is shipped via cryogenic tanker from Lake Charles, Louisiana, to London.

1960s First large-scale U.S. nuclear power plants go on line.

1960 Geysers near San Francisco begin supplying geothermal electric power.

1976 Clinch River Breeder Reactor Project, the first large-scale demonstration breeder reactor is constructed near Oak Ridge, Tennessee. The project died for lack of support, however.

1982 The Tokamak Fusion Test Reactor at the Princeton Plasma Physics Laboratory produces fusion.

1992 Federal Energy Regulatory Commission statistics show hydropower generation at 91,600 MW (cf. 1921 figure of 3,700 MW).

TRANSPORTATION AND AGRICULTURE

B.C.E.

8000 Settled agriculture occurs in the Near East and other centers of human habitation.

7500 Dugout canoes are used in northwestern Europe. Reed boats are developed in Mesopotamia and Egypt.

7000 Agriculture starts in Mexico.

3500 The earliest illustration of a sail dates from this period and was found near Luxor, Egypt. The sail is fixed to a single mast and there is a shelter aft.

3200 The first wheeled vehicles are believed to have appeared in Sumer (now Iraq).

2700 Spoked wheel appears; traction plow already developed.

2400 First canal for ships built at Elephante in Egypt.

2000 Horse-drawn vehicles are used.

312 Roman road building prowess is exemplified in the construction of the Appian Way.

C.E.

400 Paddlewheel propulsion invented in China.

500 Invention of the modern horse collar in China helps produce agriculture surpluses.

700 Lateen sailing vessels are established in the Mediterranean Sea, increasing directional sailing ability.

730 Stern post rudder invented for sailing vessels.

1701 English agriculturist Jethro Tull invents a seed drill that sows seed in neat rows, saving seed and making it easier to minimize weeds.

1705 English inventor Thomas Newcomen builds his atmospheric steam engine.

1769 French engineer Nicholas-Joseph Cugnot builds his steam road-carriage.

1786 Scottish inventor Andrew Meikle builds his first thresher.

1783 First manned flight via balloon.

1787 American John Fitch launches the first U.S. steamboat.

1804 English inventor Richard Trevithick's steam railway locomotive.

1805 Twin-screw propeller developed by American John Stephens.

1814 Railway locomotive by Englishman George Stephenson.

1816 The forerunner of the bicycle is patented by German inventor Karl D. Sauerbronn.

1819 Atlantic first crossed by a steam-powered vessel, the *Savannah*.

1825 Stockton and Darlington Railway completed, first public steam powered railway.

1827 Steam automobile invented by Hancock.

1829 First horsedrawn carriage “omnibus,” which carries eighteen passengers, introduced by G. Shillibear in London.

1830 Liverpool and Manchester railroad.

1831 American inventor Cyrus McCormick develops the harvester.

1834 Electric streetcar invented by Thomas Davenport (American).

1844 American inventor Charles Goodyear patents “vulcanizing” of rubber.

1847 German-English electrical engineer William Siemens creates the regenerative steam engine.

1852 American inventor Elisha G. Otis develops the first “safety” elevator; it incorporates a brake that prevents elevators from falling even if the main cable is completely cut.

1852 The first nonrigid, powered, manned airship is flown by its builder, French engineer Henri Giffard; this marks the beginning of the practical airship.

1860 French inventor Jean-Joseph-Étienne Lenoir builds the first practical internal-combustion engine, fueled by illuminating gas.

1865 Samuel Calthrop (American) creates a streamlined locomotive.

1867 Gas engine built by German engineers Nikolaus August Otto and Eugen Langen.

1868 American engineer George Westinghouse introduces the air brake. The new power braking system uses compressed air as the operating medium.

1869 The Transcontinental Railroad is completed as the Golden Spike is driven in at Promontory Point, Utah.

1875 Internal combustion engine invented by Siegfried Marcus (Austrian).

1877 German aeronautical engineer Otto Lilienthal invents his first glider.

1877 American meatpacker Gustavus Franklin Swift invents the refrigerated railcar.

1883 Elevated electric railroad opens in Chicago.

1884 English engineer Charles Algernon Parsons invents his compound steam turbine.

1885 The gasoline automobile is developed by German engineer Karl Friedrich Benz. Before this, gasoline was an unwanted fraction of petroleum that caused many house fires because of its tendency to explode when placed in kerosene lamps.

1886 The first modern oil tanker, *Gluckauf,* is built for Germany by England.

1887 Scottish inventor John Boyd Dunlop creates an air-inflated rubber tire.

1888 Richmond Union Passenger Railway electric street railway system designed by American electrical engineer Frank Julian Sprague.

1889 Carl Gustaf de Laval (Swedish) improves the steam engine by devising a small, high-speed turbine in which jets of steam hit a single set of blades set on a rim of a wheel.

1889 Petroleum-fueled agricultural tractor developed in the United States.

1900 Worldwide railways total 470,000 miles of track.

1901 First merchant vessel, *King Edward,* driven by steam turbines in Scotland.

1902 Steam superheaters dramatically improve the performance of railway engines.

1903 American inventors Orville and Wilbur Wright fly the first powered aircraft at Kitty Hawk, North Carolina.

1903 The Ford Motor Company is founded.

1905 Long Island Railroad is the first to abandon steam completely in favor of electrification.

1911 American industrialist and electrical engineer Elmer A. Sperry creates a gyrocompass.

1924 Constant speed propeller brings better efficiency to aircraft.

1926 American physicist R. H. Goddard builds and launches the first liquid-fuel rocket.

1927 First electrically powered, automatically controlled pipeline is built in California.

1930 English inventor Frank Whittle patents the basic design of the turbojet engine.

1939 Russian-American aeronautical engineer Igor Sikorsky makes the first successful tethered helicopter flight.

1941 The Whittle jet engine is flown from Britain to the United States and provides the model for the first practical American jet engines that will be built by General Electric.

1947 Diesel-electric locomotive is built.

1952 First regular jet air passenger service begins; it goes from London, England, to Johannesburg, South Africa.

1954 U.S. Navy launchs the first nuclear-powered submarine, *U.S.S. Nautilus,* the first use of nuclear propulsion. It could cruise 62,500 miles before refueling.

1956 German engineer Felix Wankel develops the prototype for his Wankel engine, a rotary-piston engine.

1956 English engineer Christopher Cockerell designs the hovercraft.

1959 *NS Savannah* launches; it is the first nuclear-powered merchant ship.

1961 *USS Enterprise,* a nuclear powered aircraft carrier and the world's largest ship, is launched. It operates at speeds of 20 knots for distances of up to 400,000 miles.

1966 Electronic fuel injection is developed; it eventually replaces the carburetor.

1968 Supersonic transport plane.

1972 Ford invents the sodium sulfur battery.

1977 Hydrogen gas used to power vehicles in two California experiments.

LIGHT, HEAT, ELECTRICITY, AND COMMUNICATION

B.C.E.

c. 500000 Peking man uses fire for heat, light, and cooking.

c. 17000 Stone oil lamps enable Paleolithic artists to paint and engrave the walls of caves.

c. 2400 Greek civilization designs buildings to take advantage of solar heating.

c. 8000 Oil lamps are used in Mesopotamia.

c. 2500 Glass-making occurs in Egypt and Mesopotamia.

c. 600 Greek philosopher Thales observes that amber, a fossilized type of tree sap, attracts bits of paper and certain materials, like straw, when rubbed. This is the first mention of static electricity.

285 The Lighthouse at Alexandria is constructed; a mirror projects the light of fire for thirty miles.

1 Roman engineers develop a vertical watermill known as a Vitruvian mill.

C.E.

551 Jerome Cardan, an Italian mathematician, determines that while amber attracts light objects, a magnetic black stone attracts only iron. This is the first in a series of discoveries that links electricity to magnetism.

1340 Blast furnaces are developed in Europe.

1600 English physician William Gilbert discovers that materials like glass, sulfur, and diamonds behave just like amber. He calls these materials electrics, which means amber in Latin.

1646 Walter Charlton coins the word "electricity" to explain the attraction between these substances.

1672 German physicist Otto von Guericke molds a large sphere out of sulfur. Holding a piece of wool against this spinning sphere produces a large spark. This is the first generator to use friction to create electricity.

1729 Englishman Steven Gray first discovers that metals are conductors and non-metals are non-conductors.

1745 Dutch mathematician and physicist Pieter van Musschenbroek invents the Leyden Jar, which stores an electric charge.

1747 American statesman and inventor Benjamin Franklin deduces the existence of positive and negative electric charges.

1752 Benjamin Franklin flies a kite during a thunderstorm with a key dangling on the end of a wire. A silk string collects a charge from the thunderclouds which is conducted into a Leyden Jar. Thus, he makes the connection between lightning and electricity. This experiment leads to his invention of the lightning rod.

1762 Oil street lamps used in New York City.

1767 The city of Philadelphia lights streets with whale-oil lamps.

1790 First steam-heated factory.

1792 Coal-gas lighting invented by William Murdock in Cornwall.

1794	Italian physicist Allessandro Volta creates the first continuous electrical current by making a battery out of silver and zinc strips placed in salty water. Prior to this discovery all man-made electrical sources came from static.
1795	British physicist Benjamin Thompson (Count Rumford) invents the Rumford stove, a close-toped range that economizes fuel.
1800	The first commercial battery is manufactured. Scientists realize that if chemical changes can create electricity, then electricity can create chemical change.
1803	First factory illuminated by coal-gas lighting, in James Watt's foundry (Scotland).
1803	Italian cities of Genoa and Parma are lighted by kerosine from an oil well in Modena.
1813	London Bridge is lighted by gas.
1819	Danish physicist Hans Christian Oersted creates a magnet with electrical current, establishing the connection between electricity and magnetism.
1826	German physicist George Simon Ohm publishes *Die galvanische kette, mathematisch bearbeitet,* in which he describes his discovery that the voltage across an electrical conductor is proportional to the electrical current, and that the current is inversely proportional to the resistance of the conductor. His formulation becomes known as Ohm's Law.
1828	English chemist and physicist Michael Faraday discovers electromagnetic induction.
1829	American physicist Joseph Henry develops a coil magnet that grows stronger as more wire is wound around an iron core. He succeeds in lifting more than a ton of metal.
1830	Thermostat is invented.
1831	English chemist and physicist Michael Faraday creates the first electrical generator by using a magnet and a spinning copper plate to produce a current. Using a steam engine to keep the copper plate spinning within the magnetic field, electrical current is produced.
1831	American physicist Joseph Henry, by reversing Faraday's discovery, passes an electrical current through a magnetic field to turn a copper wheel, creating the first electric motor. For the first time in history, electrical energy can be used to power machines to do work that was formerly done by humans and animals.
1834	The first practical liquid refrigerating machine is patented by American inventor Jacob Perkins.
1835	The electric automobile is created by American inventor Thomas Davenport.
1838	American inventors Samuel F. B. Morse and Alfred Vail first demonstrate practical telegraphy.
1844	Samuel F. B. Morse builds the first electric telegraph. By transmitting short or long signals along a wire, messages can be sent anywhere. The Morse code makes it possible to send messages long distances at the speed of light.
1845	Safety matches are developed.
1849	American engineer James B. Francis invents the hydraulic turbine.
1851	American inventor John Gorrie patents an expansion cycle refrigerating machine.
1852	Heat pump is invented.
1868	The Siemens brothers (German) design the regenerative gas furnace.
1864	Scottish physicist James Clerk Maxwell publishes his theory of light and electricity.
1866	The first transatlantic cable is laid, creating a permanent electrical communications link between the old world and the new.
1868	French chemist Georges Leclanche invents the zinc-carbon battery, a precursor of the dry cell and the modern portable battery.
1875	A building in France is iluminated by electricity.
1876	American inventor and educator Alexander Graham Bell develops the telephone by converting electrical impulses into sound.
1878	American inventor Charles F. Brush invents the arc lamp.
1879	The streets of Cleveland are lit by carbon-arc lamps.
1879	Thomas Edison invents the incandescent lamp.

1879	English physicist William Crookes develops the cathode ray tube.
1882	October. The pioneering firm of United States Electric Illuminating Company starts up South Carolina’s first central station for incandescent electric lighting one month after Thomas Edison opened his central station on New York City’s Pearl Street. In the following years, U.S. Electric becomes one of Edison’s main competitors.
1885	American electrical engineer William Stanley invents the transformer.
1886	Austrian chemist Carl Auer (Freiherr von Welsbach) invents the Welsbach mantle, tripling the output of kerosene lamps and gas burners.
1884	American engineer Lester A. Pelton develops the hydraulic turbine.
1884	First large-scale use of natural gas in Pittsburgh.
1887	Italian-American physicist Nikola Tesla invents a motor that produces alternating current. This discovery changes the way electricity is transmitted over long distances.
1889	The first commercial, long-distance transmission of electricity takes place when a direct-current line provides power from Willamette Falls for street lights in Portland, Oregon.
1893	German physicists Julius Elster and Hans F. Geitel invent the first photoelectric cell as a result of studying the photoelectric effect.
1895	The first hydroelectric generator at Niagara Falls, New York, produces alternating current from a Nikola Tesla design.
1900	A hydroelectric plant is built at Niagara Falls.
1901	The first reception of transatlantic radio signals.
1904	Gas is used for the first time for heating and hot water in London, England.
1904	English electrical engineer John Ambrose Fleming patents the first electron tube, which he calls a diode vacuum tube.
1904	American electrical engineer Ernst Alexanderson creates a high-frequency alternator; it allows reliable transoceanic radiotelegraph communication.
1909	Shoshone Transmission Line, power generated by the Shoshone Hydroelectric Generating Station to Denver.
1910	French chemist and physicist Georges Claude invents the first neon light.
1911	English engineer Charles Algernon Parsons improves the turbo-alternator for generating electricity in power stations.
1913	Austrian inventor Victor Kaplan patents his turbine; the invention enables hydroelectric power stations to be more consistently efficient.
1920	Martin Hochstadter introduces a three-core power cable that does not become deformed or burn by the high voltage electricity.
1930	First domestic gas water heater to work efficiently is developed.
1934	Power from the Boulder (Colorado) dam is transmitted 270 miles to Los Angeles, California.
1937	First commercial convective heater equipped with an electric fan.
1937	The five million-volt van de Graaff generator, Westinghouse “Atom Smasher, 1937,” represents the first large-scale program in nuclear physics established by industry.
1939	The Massachusetts Institute of Technology builds a solar house.
1939	American physicist John Vincent Atanasoff collaborates with Clifford E. Berry to design the first digital electronic computer.
1940	The first gas-powered turbine to generate electricity is developed in Switzerland.
1943	Electric cables are filled with pressurized gas for insulation in England.
1945	Perry L. Spencer (United States) patents the first microwave oven.
1948	Bell Telephone Laboratories first demonstrates the transistor, a non-vacuum device that will eventually replace the conventional electron tubes.
1952	Four thousand die in London, England, from smog during air-inversion event.
1953	Four hundred die in New York City from smog event.
1954	Americans D. M. Chapin, C. S. Fuller, and G. L Pearson develop the silicon photovoltaic cell.

1955 Narinder Kapany invents optical fiber in Germany.

1958 American Gordon Gould develops the laser.

1959 Jack Kilby and Robert Noyce invent the integrated circuit.

1962 American Nick Holonyak, Jr., invents the light-emitting diode.

1963 Solar furnace capable of generating temperatures greater than 9,300°F becomes operational in Japan for scientific research.

1964 Generating electricity without turbines via magnetohydrodynamics (MHD) could theoretically double the output from nuclear power plants.

1964 George Heilmeier develops the liquid-crystal display.

1966 First superconducting motor.

1971 Ted Hoff invents the microprocessor.

1986 April. A severe nuclear accident occurs at Chernobyl in the former Soviet Union.

1986 High Temperature Superconductors developed by J. Georg Bednorz and Karl A. Muller.

MECHANICAL ENGINEERING

B.C.E.

1550 Levers used for well sweep in Egypt as well as India.

1500 Simple pulley employed in Egypt.

900 Rotary bucket pump invented.

700 First mention of the pulley; chain of pots used to raise water.

600 Compound pulley crane first used in Greece.

300 Tooth wheels and gears developed.

180 Quern or revolving mill is invented; turned by slaves or asses.

150 Force pump appears.

100 Undershot water-wheel first designed.

27 Vitruvian waterwheel created; first known instance of the transmission of power through gearing.

C.E.

1–100 Aeolipile, earliest recognized steam-powered mechanism, built by Hero of Alexandria.

200–300 Barbagal water-mills developed in Provence, France.

600–700 Windmills appear on Iranian plateau.

1240 Water-powered saw and jack invented by Honnecourt in France.

1335 Mechanical clock erected in the tower of Milan.

1472 In the next fifty years, Leonardo Da Vinci constructs the following devices: centrifugal pump, dredge for canal building, breech-loading cannon, rifled firearms, universal joint, rope-and-belt drive, link chains, bevel gears, spiral gears, parachute, and propeller.

1530 Foot-driven spinning wheel invented.

1698 Steam pump created by Thomas Savery (English).

1698 Denys Papin’s steam engine, France.

1705 Reciprocating steam engine developed by Thomas Newcomen (English).

1720s Thomas Newcomen’s steam engine comes into general use.

1769 James Watt receives main patent on condensing steam engine.

1769 Richard Arkwright invents his spinning water frame to spin yarn and silk; later to be used in North America as an industrial mill.

1772 Ball-bearings developed.

1772 James Watt’s double-acting steam engine first used in Britain.

1782 Double-acting steam engine patented by James Watt.

1785 Screw propeller invented invented by Joseph Bramah.

1801 Richard Trevithick demonstrates first successful steam-powered road vehicle in Cornwall, England.

1816 Robert Stirling patents a forerunner of the Stirling engine, tauted as “A New Type of Hot Air Engine with Economiser.”

1827 Benoit Fourneyron develops a water turbine in France.

1877 Invention of a hydraulic elevator with jugger mechanism.

1884 First practical steam turbine patented by Charles Algernon Parsons.

WARFARE AND SPACE

B.C.E.

30000 Bow and arrow first used.

1000	Assyrians use battering rams mounted on wheeled fighting towers.
1000	Crossbow invented in China.
400	Catapult and mechanical crossbow (ballista) invented at Syracuse and used against Carthage.
400	Catapult first used in China.
215	Archimedes' catapult used against the Romans.

C.E.

950	Gunpowder invented in China.
1000	The trebuchet, a missile thrower of great force operated by a hundred or more men, appears in China.
1118	Cannon used by Moors.
1232	Chinese use rockets against the Tartars and invent hot-air balloons.
1259	First cannons used in China.
1405	Portable firearms appear.
1490s	First portable siege guns used in a military campaign in France.
1515	Wheel-lock gun invented in Germany.
1525	Rifled musket designed.
1835	American Samuel Colt invents the revolver.
1840s	Guncotton (nitrocellulose) and nitroglycerine developed.
1850s	First ocean-going steam-powered naval vessels.
1860s	Railroads provide important support for armies in the U.S. Civil War and the Wars of German Unification.
1864	Self-propelled torpedo invented by Robert Whitehead (England).
1867	Alfred Nobel invents dynamite.
1880s	Electric power introduced on warships, which allows significant improvement in their capabilities.
1904–1905	Japan defeats the Russian army; first use of telephone and telegraph to coordinate troops and supplies.
1914	Tank invented by E. D. Swinton (England).
1926	Liquid-fuel rocket developed by R. H. Goddard (United States).
1939	August 2. Albert Einstein writes to U.S. President Franklin Roosevelt to warn him of the possibility of Nazi Germany developing atomic weapons.
1944	Ballistic missile developed by Wernher von Braun (Germany).
1945	July 16. First atomic bomb test at Alamogordo, New Mexico, under the code name "Manhattan Project."
1945	August. United States drops atomic bombs on Hiroshima and Nagasaki, Japan, during World War II.
1949	The Soviet Union tests its first atomic bomb, which launches the "arms race" with the United States.
1952	United States tests the first hydrogen bomb.
1955	World's first nuclear submarine, *Nautilus,* is tested (United States).
1957	First artificial satellite, Sputnik, in orbit (Soviet Union).
1958	Neutron bomb developed in the United States
1965	First nuclear reactor in space is launched.
1983	U.S. President Ronald Reagan announces the Strategic Defense Initiative, intended to shield the United States from nuclear attack.
1991	January 15. United States and allied countries launch Operation Desert Storm against Iraq, a military operation characterized by some as an "energy war."
1992	May. The nuclear arms race ends—for the first time since 1945, the United States builds no nuclear weapons.

BUSINESS, GOVERNMENT, AND CULTURE

C.E.

529	St. Benedict founds Monte Cassino; the Benedictine Order, which is to become a very powerful force in Western Christianity, adopts manual labor as a virtuous action.
716	St. Boniface, an English Benedictine, visits Germany and establishes abbeys and country estates as centers of industry and material progress.
1066	With the Norman conquest of England, industry moves from the abbeys and country estates to the towns because of energy—water mills and water transportation.
1095	The First Crusade begins (ending in 1099); the Crusades, of which there were ten—if one counts the tragic Children's Crusade—reflected a growing energy surplus (as did

	the universities) in western Europe, not yet channeled into national military establishments.
1150	Cistercians introduce use of city garbage and sewer water as fertilizer near Milan (Italy); both Benedictines and Cistercians drain swamps and lakes in Germany, France, the Low Countries, and Italy.
1273	Coal smoke in London, England, provokes complaints from the gentry.
1306	King Edward I of England makes it a capital offense to burn coal in London.
1345	Division of hours and minutes into sixties (Germany); without linear time, the Industrial Revolution would not have been possible.
1798	T. R. Malthus publishes *An Essay on the Principle of Population,* as pessimistic in its conclusions as the Marquis de Condorcet's work was optimistic
1863	The British government passes the "Alkali Works Act" in an attempt to control environmentally harmful emissions.
1870	John D. Rockefeller founds the Standard Oil Company.
1882	Standard Oil Trust is formed and buys controlling interest in a number of oil companies.
1901	Oil found at Spindletop, Texas, which leads to the formation of Gulf, Texaco, and Sun oil companies.
1907	First drive-in gas station opens in St. Louis, Missouri.
1908	Oil is discovered in Persia; Anglo-Persian (later British Petroleum) is formed.
1911	U.S. Supreme Court breaks Standard Oil monopoly into thirty-four companies.
1920	The U.S. Congress passes the Federal Water Power Act of 1920, which authorizes the first Federal Power Commission (later Federal Energy Regulatory Commission). It has authority to issue licenses for hydroelectric projects that are best adapted to the comprehensive development of a waterway.
1930	Huge oil discovery in East Texas.
1935	Federal Water Power Act becomes part of the Federal Power Act to regulate interstate commerce in electricity.
1937	The Bonneville Project Act creates the Bonneville Power Administration (BPA), which markets electricity generated at Federal hydro projects to the northwestern United States; it also owns the nation's largest network of long-distance, high-voltage transmission lines needed to bring hydropower to market.
1938	Oil discovered in Kuwait and Saudi Arabia.
1941	The United States, Britain, and the Netherlands put an oil embargo on Japan after it takes over Indonesia.
1952	Smog identified for the first time in Los Angeles, California, from the combination of a large number of automobiles, bright sunlight, and frequently stagnant air.
1953	December 8. U.S. President Dwight Eisenhower delivers his "Atoms for Peace" speech before the United Nations.
1954	The Atomic Energy Act of 1954 permits and encourages the participation of private industry in the development and use of nuclear energy.
1955	Atomic Energy Commission announces the Power Demonstration Reactor Program, a cooperative effort with private industry in constructing experimental power reactors.
1956	Celilo Village, a traditional Indian tribal fishing ground, is flooded by the Dalles Dam. The sovereign rights of tribes lead to several court cases and agreements that affect use of rivers and hydropower in the United States.
1957	The United Nations establishes the International Atomic Energy Agency in Vienna, Austria, to promote the peaceful use of nuclear energy.
1960	The Organization of Petroleum Exporting Countries (OPEC) founded in Baghdad, Iraq.
1961	The United States and Canada sign the Columbia River Treaty. Under the treaty, Canada builds two storage dams and one dam for generation, resulting in greater power and flood control benefits at U.S. facilities downstream.
1964	The Pacific Northwest Coordination Agreement is signed; it seeks to meet the region's electricity needs most efficiently by

operating the diverse generating resources as a coordinated system, as if they were owned by a single utility.

1965 First major blackout occurs in northeast United States and Canada.

1966 The Public Power Council is formed to give a voice to publicly owned utilities in the Northwest. PPC represents and advocates the common legal and technical interests of the Northwest's consumer owned utilities.

1967 The Pacific Northwest-Pacific Southwest Intertie creates the only direct way to move electricity between the Northwest and California. Billions of dollars are saved by the Northwest trading some spring and summer surplus power for fall and winter power from California.

1967 The Air Quality Act becomes law in the United States.

1968 Oil is discovered on Alaska's North Slope.

1969 Santa Barbara, California, oil spill.

1970 Production of crude oil in the United States reaches an all-time peak.

1970 Clean Air Act goes into effect in the United States.

1973 October 6. The Yom Kippur War breaks out in the Middle East.

1973 October 17. The Organization of Arab Petroleum Exporting Countries declares an oil embargo; prices rise to nearly $12 a barrel from $3.

1975 December. U.S. President Gerald Ford signs the Energy Policy and Conservation Act, extending price controls into 1979, mandating automobile fuel economy standards, and authorizing creation of a strategic petroleum reserve.

1977 June 3. The U.S. Department of Energy is created by the consolidation of the Federal Energy Administration, the Energy Research and Development Administration, and the Atomic Energy Commission.

1978 U.S. Congress passes Public Utility Regulatory Policies Act (PURPA). This law requires utilities to purchase electricity from qualified independent power producers. Portions of the act helped stimulate growth of small scale hydro plants as a means of meeting the nation's energy needs.

1978 November. U.S. President Jimmy Carter signs the National Energy Conservation and Policy Act, which promotes conservation activities, requires development of standard measures of energy efficiency and its reporting to the public.

1978 Last new nuclear power plant ordered in the United States.

1979 March. An accident occurs at the Three Mile Island nuclear power plant in New York.

1979 June. U.S. President Jimmy Carter announces program to increase the nation's use of solar energy.

1979 July. U.S President Jimmy Carter proposes $88 billion effort to enhance production of synthetic fuels from coal and shale oil reserves.

1980 U.S. Congress passes the Pacific Northwest Electric Power Planning and Conservation Act. The Northwest Power Planning Council is formed. The Council is charged with developing a plan to meet Northwest energy needs. The act also called for the Council to develop a fish and wildlife mitigation and enhancement plan.

1980 June. U.S. President Jimmy Carter signs the Energy Security Act.

1981 July. U.S. President Ronald Reagan signs Executive Order 12287, effectively decontrolling crude oil and refined petroleum products.

1979–1981 Panic caused by Iran's revolution and the Iran-Iraq war sends oil prices as high as $34 a barrel from $13.

1983 January. U.S. President Ronald Reagan signs the Nuclear Waste Policy Act of 1982, the first comprehensive nuclear waste legislation.

1986 Congress amends the Federal Power Act, increasing environmental review of hydropower projects.

1986 Oil price collapses to $12 a barrel.

1987 December. U.S. Congress approves Yucca Mountain, Nevada, as the only repository site for high-level nuclear waste.

1988 The Northwest Power Planning Council designates 44,000 miles of Northwest streams as "protected areas" because of their importance as critical fish and wildlife habitat.

1989 New York Governor Mario Cuomo and the Long Island Power Authority announce that the already built Shoreham Nuclear Power Plant will never open.

1989 *Exxon Valdez* runs aground off the Alaska coast.

1990 August. Iraq invades Kuwait, triggering an international crisis.

1991 January 15. United States and allied countries launch Operation Desert Storm against Iraq to end its invasion of Kuwait.

1991–1995 By some estimates, fish and wildlife protection measures reduced firm electric generation by about 850 megawatts annually.

1992 June. Representatives from many nations convene at the Earth Summit in Rio de Janeiro, Brazil.

1993 April. U.S. President Bill Clinton announces that the United States will stabilize greenhouse gas emissions at 1990 levels by the year 2000.

1994 The U.S. Supreme Court rules that states have the authority under the Clean Water Act to establish minimum streamflows at hydro projects. The ruling gives states more authority in hydro licensing and relicensing decisions.

1997 December 11. The Kyoto Protocol is adopted by the United Nations Framework Convention on Climate Change.

1998 United States oil and utility industry companies spend over $100 million to influence federal government energy policy.

1998 Oil prices fall sharply; collapses in Asian economies severely curtail demand.

2000 Oil prices surge to highest levels since the mid-1970s.

INDEX

Key to codes: b=box, f=figure, t=table. Bold page numbers indicate main articles. Italic numbers indicate photos.

A

- Acceleration
  - automobile performance, 98–99
- Acid deposition, 1–6
  - aquatic effects of, 4–5
  - Clean Air Act of 1970, 3
  - controlling effects of, 5
  - in eastern U.S., 2f
  - emission trends in U.S., 2–3, 3f
  - formation of, 2–3
  - sources of, 3–4
  - Terrestrial effects, 5–6
- Acid rain, **1–6.** *See also* Acid deposition
  - first use of term by Angus Smith, 1
  - historical references, 1
- Adenosine triphosphate (ATP), 167–178
  - proteins and, 172–173
  - structure and formation of, 168f
- Aerodynamic drag, 9–14
  - Automobile performance, 99–100
  - energy consumption, 13f, 13–14
  - induced drag, 12
  - skin friction drag, 11–12
  - streamlining, 12f, 13
  - vehicle drag reduction, 14f
- Aerodynamic forces
  - drag and friction, 9–10
  - pressure and shear stress, 7f
  - resolution of, 8f
  - shear force, 7
  - skin friction, 9–10
  - sources of, 7–8
- Aerodynamics, **7–15**
  - boundary layer, 9f, 9–10
  - energy efficiency, 13f, 13–14
  - laminar flow, 9–10
  - lift and drag, 8f
  - turbulent flow, 9–10
  - Wright Flyer, 34–36
- Aeronautics
  - golden age (1930s), 37–40
- Agricultural production
  - commercial energy use, 19–21
  - energy efficiency, 20–21
  - fossil fuel use in, 16
  - growth of, 18–19
  - role of energy in, 19
  - tractors use in 1905, *17*
- Agriculture, **15–22**
  - classification of, 16–19
  - commercial agriculture, 18–19
  - commercial energy use in, 19–21
  - early societies use of, 73–74
  - energy use in, 15–16, 20t
  - irrigation energy requirements, *21,* 74–75
  - renewable energy use of, 19–20
  - tools and technology, 74–75
  - traditional agriculture, *15,* 16–17
  - transitional agriculture, 18
- Air conditioning, **22–31**
  - active heating and cooling, 1055–1056, 1057f
  - applications circa 1900–1920, 25–26
  - automotive, *27*
  - in buildings, 191
  - central air conditioning systems, 26f, 26–28
  - definition of, 22
  - fossil fuel use in, 26
  - future of, 30
  - history of, 25–28
  - household use of, 27–28
  - invention of, 25
  - passive heating and cooling, 1054f, 1054–1055
  - residential buildings, 206–207
  - solar energy and, 1053–1056
  - trends in, 28–30
- Air conditioning systems
  - absorption systems, 24f
  - CFC refrigerants phase out, 28
  - chlorofluorocarbon refrigerants, 26–28
  - energy efficiency, 29
  - hydronic heating systems, 23
  - rating systems for, 24
  - types of, 22–24
  - vapor compression systems, 23f
- Aircraft, **31–47**
  - aerodromes, 33–34
  - aviation fuel, 39, **109–113**
  - Boeing 707, 43f
  - Concorde supersonic transport, 46f
  - engine performance, 37
  - gliders, 33
  - high-lift devices, 39f, 39–40
  - jet aircraft, 40–45
  - Lockheed F-104, 42f
  - Lockheed Vega, 38f
  - North American F-86H, 41f
  - origins of, 32–34
  - payload capacity comparison, 44t
  - streamlining, 40
  - Wright Flyer, *31–32,* 34–36
- Aircraft cabins
  - commercial jet transport, 44–45
  - first pressurized aircraft, 40
- Aircraft engines
  - air-cooled design, 37–38
  - cowlings, 37–38
  - gas turbines, 40–41
  - pressurization of, 40
  - superchargers, 40
- Aircraft performance
  - high-lift devices, 39f, 39–40
  - World War I, 36–37
- Air filtration
  - indoor air quality, 55, 58
- Airfoils
  - laminar flow, 10, 11f
- Airline industry, 60–64
  - aircraft efficiency, 61–62
  - deregulation, 61–62
  - market research, 63
  - U.S. government subsidies, 61
- Air pollutants
  - carbon monoxide, 50
  - identifying damaging, 48
  - indoor, 54–58, 56
  - lead, 50–51

Air pollutants (cont'd)
  location of non attainment areas, *48*
  nitrogen dioxide, 51
  ozone, 49
  particulate matter, 49–50
  sulfur oxides, 51
  volatile organic carbon compounds (VOC), 49t
Air pollution, **47–53**
  composition of, 85–86
  constituents of, 47–48
  effects of emission control, 444–445
  energy use and, 48–51
  environmental activities in, 475
  history of, 47
  industry as leading cause of, *51*
  Kyoto protocol to reduce greenhouse gases, 249t, 249–250
  Los Angeles view of, *47*
  National Ambient Air Quality Standards, 50t, 51–52, 52t
  regulations, 51–52
Air pollution control
  electric vehicles, 437–439
Air quality, indoor, **53–60.** *See also* Indoor air quality
Air traffic control, 62–63
Air travel, **60–64**
  airport congestion, 62–63
  air traffic control, 62–63
  growth in, 1159
  jet aircraft for, 61
  lifestyle choice of, 135
Alexanderson, Ernest Frederik Werner, **64–65**
Alkaline manganese cells
  alkaline rechargeable batteries, 120–121
Alkaline rechargeable batteries, 120–121
  alkaline manganese cells, 120–121
  lithium-ion cells, 120
Alternative fuels
  competition with gasoline, 555
  derived from natural gas, 833f
  economic barriers, 68–69
  estimated costs of, 67f
  ethanol, 67–68, 160–162
  fuel costs, 68
  history, 66
  legislation, 67
Alternative fuels and vehicles, **66–69** *See also* Alternative fuels
  market barriers, 67–69
American Society of Heating, Refrigeration and Air-conditioning Engineers (ASHRAE)
  building standards, 199–200, 200t
Ames Laboratory
  description of, 814
Ampère, André-Marie, **69–71**
Animal and human energy, **71–75**
Animals
  endothermic, 183–184
  energy effects on, 183–186
  habitat adaptations, 185–186
  metabolism of, 183
  mobility and energy in, 184–185
Antimatter
  properties of, 778–780
Appliance labeling rule
  appliance energy regulations, 380–382, 381f
Appliances, **75–83**
  appliance labeling rule, 79, 380–382, 381f
  efficiency changes in, 81f
  energy efficiency standards, 78–81
  energy policy, 75–77
  energy savings in, 371–372
  history of energy efficiency in, 77–81
  market trends, 78f
  refrigerators and energy use, 77f
Archaeology
  use of tools and energy, 71–73
Archimedes, **83–84**
Argonne National Laboratory
  description of, 814–815
Atmosphere, **84–94**
  air pollution effects on, 85–86
  composition of, 85–86, *87*
  hydrological cycle, 86, 88–89
  outer space view of, *88*
  solar radiation effects, 86
  weather, 89–92
  wind speed averages, *93*
Atomic Energy Act of 1954
  nuclear power industry, 854
Atomic Energy Commission
  description of, 813–814
  nuclear energy development, 853–854
Auditing of energy use, **95–97,** *96*
  financing of, 96–97
  trends in, 97
  types of, 95
Automobile design
  fuel economy effects on, 105f
Automobile engines
  advances in energy efficiency, 374–375
  fuel economy, 103–105
  performance, 101–105
  reserve power, 102f, 102–103
Automobile performance, **97–108**
  acceleration, 99, 103f
  aerodynamic drag, 99–100
  engine performance, 101–102
  fuel economy, 103–105
  fuel economy trade-off, 106f, 107f
  history of, 105–108
  performance metrics, 98
  power requirements, 99–100
  tires and rolling resistance, 99
Automobile production
  Ford Model T, 1158
Automobiles
  electric hybrid cars, 375
  public transit *vs.,* 133–134
  standard U.S. driving schedules, 104f
Automotive
  air conditioning, *27*
Aviation fuel, **109–113**
  additives, 111–112
  composition of, 112
  environmental issues, 112–113
  fuel types and specifications, 109–110
  octane levels, 39
  production of, 112
  properties of, 110t, 110–112

B

Ball bearings, *125,* 126–127
Batteries, **115–124**
  advanced batteries, 123–124
  alkaline rechargeable batteries, 120–121
  bad batteries and energy efficiency, 117b
  basic principles of, 115–117
  button primary cells, 121
  commercially developed, 233t
  development of, 234–235
  development of storage, 396
  for electric vehicles, 122–123
  electrochemistry and, 231–234
  lead-acid secondary cells, 121–122
  nickel alloy batteries, 123
  smart battery technology, 121b
  types of, 117–122
  Zinc-manganese dry cells, 118–120, 119f
Battery discharge
  hindered by polarization, 395
Bearings, **124–127**
  applications of, 127
  ball bearings, *125,* 126–127
  components of, *125*
  roller bearings, 127
  type comparison of, 126t
Becquerel, Alexandre-Edmond, **127–129**
  photoelectric effect observation, 1058
Behavior, **129–141**
Bell X-1 aircraft
  first flight, 43
Bernoulli, Daniel, **141–142**
  steam propulsion with screw propellers, 957
Bethe, Hans Albrecht, **142–144**

- Bicycles
  - Biopace chainwheel, 151f
  - chainwheel design, 150–151
  - dynamics of, 147–149
  - efficiency of, 144–145
  - history of, 145–147
  - as transportation, 152–154
- Bicycling, **144–154**
  - air and sea pedal-craft, 150b
  - commuting with, 153–154
  - human energy and technique, 147–151
  - lifestyles and culture, *310*
  - mass transit and, 768–769
  - in New York City, *153*
  - pedal cadence, 149–150
- Big bang theory, **154–157**
  - Albert Einstein and relativity theory, 154
  - concept illustration, *155*
  - expanding universe, 154–156
  - fate of the universe, 157
  - galaxy formation and, *156*
- Biochemistry
  - Adenosine triphosphate (ATP), 168–169
  - enzymes effect on chemical reactions, 169–170, 170f
  - proteins and ATP, 172–173
- Biodiesel
  - production of, 162t, 162–163
- Biofuels, **157–166**
  - biodiesel, 162t, 162–163
  - biogas production, 160
  - corn as source for ethanol, *161*
  - current use of, 158
  - energy conversion efficiency, 163
  - ethanol, 160–162
  - ethanol powered snowplow, *165*
  - future use of, 163–166
  - gasification of, 159–160
  - greenhouse gases and, 163–164
  - pryolysis, 160
  - soil nutrient loss, 164
  - transportation fuels, 160–163
- Biogas
  - production of, 160
- Biological energy
  - adenosine triphosphate (ATP), 167–178, 168f
  - animal adaptations, 183–186
  - animal mobility and, 184–185
  - bioenergetics, 808
  - cells capturing energy from catabolism, 169–171
  - cellular processes, 166–178
  - coupled reactions releasing, 167–169
  - effects on animals, 183–186
  - energy storage provided by body, 176–178, 177f
  - energy transformation, 179–183
  - enzymes and activation energy, 172f
  - metabolism and food, 171–172
  - muscles and glycogen, 176–178
  - muscles and myosin molecules, 173f, 174f, 175
  - myosin crossbridge releasing, 173f, 174f, 175
  - photosynthesis of plants as, 180–181
  - Sun as source of, 180–181
  - thermodynamics laws and, 166–167
- Biological energy use, cellular processes of, **166–178** *See also* Biological energy
- Biological energy use, ecosystem functioning of, **178–188** *See also* Biological energy
- Biomass. *See also* Biofuels
  - combustion of, 158–159
  - electricity generation from, 158
- Black, Joseph, **188–190,** *189*
- Blanchard, Jean-Pierre
  - propellers on balloons, 958
- Boeing 707 aircraft, 43f, 45
- Boeing 747 aircraft, 45
- Bohr, Niels
  - model of atom, 804
- Boundary layer (aerodynamics), 9f, 9–12
  - laminar flows, 10–12
  - transition region, 10–11
  - turbulent flows, 10–12
  - velocity profiles, 10f
- British Association for the Advancement of Science
  - energy terminology defined, 1034
- Brookhaven National Laboratory
  - description of, 815
- Building construction
  - energy savings in, 371
  - regulations for, 198–199
- Building design, commercial, **191–197**
  - air conditioning effects on, 191
  - building envelope, 193, 195t
  - energy crisis effects on, 192–193
  - future of, 196–197
  - history of, 191–192
  - lighting, 195–196 *See also* Lighting
  - office equipment energy use in, 196
  - thermal comfort in, 194–195
- Building design, energy codes and, **197–203**
  - adoption and enforcement, 200–201
  - energy code evolution, 201t
  - Energy Policy Act of 1992, 202–203
  - future of energy codes, 203
  - International Code Council (ICC), 202–203
  - legislative history, 199t, 199–200
  - regulations for, 198–199
- Building design, residential, **203–211**
  - air sealing and insulation, 204–206, 205f, 206t
  - appliances, 207 *See also* Appliances
  - energy efficiency in, 204–208, 209t
  - energy savings in, 371
  - heating and cooling, 206–207, 207f
  - history of, 204
  - lighting, 207 *See also* Lighting
  - trends in, 209–210
  - ventilation, 207–208
  - weather protection, 208–209
- Buildings
  - air conditioning in, 191
  - energy use in, 192–193
  - energy use in commercial, 196f
  - heat flow in modern, 194f
  - HVAC systems for comfort in, 194
  - lighting systems, 195–196
  - office equipment energy use in, 196
- Building standards
  - energy regulations for, 199t, 199–200
  - history of, 201t
- Busernann, Adolf
  - swept wing theory, 42
- Button primary cells
  - alkaline batteries, 121

C

- Capacitance
  - energy storage, 213f
- Capacitors and ultracapacitors, **213–216**
  - circuits of, 214f
  - electrochemical, 214f
  - electrolytic, 215
  - energy storage, 213f
  - pseudo capacitance, 215
  - structure of, 214f
- Capital investment decisions, **216–219**
  - conventional *vs.* energy conserving design, 218t
  - life-cycle cost analysis, 216–218
  - steps in, 216–217
- Carbon dioxide
  - climatic effects of, 240f
  - control in oil and gas drilling, 915
- Carbon monoxide
  - air pollutants, 50
- Carnot, Nicholas Léonard Sadi, **219–221,** *220*
  - Carnot cycle and Kelvin scale, 1032
  - high pressure steam engines, 1031–1032
- Carson, Rachel, **221–223,** *222*
- Catabolism
  - glucose, 169–171
- Catalysis. *See also* Catalysts
  - commercial processes, 227
  - kinetics of, 225–226
  - transport effects, 226–227
- Catalysts, **223–227.** *See also* Catalysis
  - heterogeneous reaction mechanisms, 225

- Catalysts (cont'd)
  - history of, 223–225
  - homogeneous reaction mechanisms, 225
  - kinetics of, 226
  - multifunction catalysts, 227
- Cayley, Sir George
  - aircraft concepts, 32–33
- Charcoal, **228–230** *See also* Coal headings
  - furnaces, 228f, 229f
  - history of, 228
  - processing and production of, 229
  - properties of, 228t
  - use of, 228–229
- Chemical energy
  - animal muscles and, 230
  - battery development, 233–235
  - chemistry of combustion, 273–274
  - commercial batteries, 233t
  - fuel cells, 234–235
  - galvanic cells, 230–231
  - history of, 230–234
- Chemical energy: historical evolution of use, **230–236.** *See* Chemical energy
- Chlorofluorocarbon refrigerants (CFCs)
  - air conditioning systems, 28
  - atmospheric effects, 86
  - Climatic effects of, 242f
  - ozone depletion, 28
- Clausius, Rudolf Julius Emanuel, **236–238,** *237*
- Clean Air Act of 1970
  - acid deposition, 3
  - power plant emission control, 444–445
- Climate change. *See also* Climate effects
  - electric power system reliability, 426–427
  - greenhouse effect theory, 239–240, *239*
- Climatic effects, **238–251** *See also* Climate change
  - atmospheric elements, 242–243
  - carbon dioxide, 240f
  - Chlorofluorocarbon refrigerants (CFCs), 242f
  - climate change, 238–240, 248
  - Intergovermental Panel on Climate Change (IPPC), 240
  - Kyoto protocol (United Nations), 249t
  - methane, 241f
  - observed changes, 243–246
  - reducing greenhouse gases, 249–250
  - research needs in, 246–247
  - temperature trends, 244–245
  - water and snow trends, 244–246
- Coal, transportation and storage, **262–265**
  - barge transportation, 263
  - coal slurry pipelines, 264
  - ocean transportation, 263
  - rail transportation, 263–264
  - storage and oxidation, 264–265
  - truck transportation, 264
- Coal combustion
  - health effects of, 252–253
- Coal consumption, **251–257**
  - coal emissions reduction, 253, 255–257
  - history, 251–252
  - Kyoto protocol (United Nations), 255–257
  - power plant usage of, *253*
  - production and, 505
  - quantitative evolution, 254–255
  - train engine coal use, *252*
  - U.S. use, 254t
  - world use, 255t
- Coal production, **257–262.** *See also* Charcoal
  - clean coal technology, 445–446, 1179
  - environmental regulations for, 262
  - geology of coal, 257
  - mining machines, *258, 259*
  - mining methods, 258–262, 260f, 261f
  - types of coal, 257–258
  - world coal reserves and, 258
- Cockroft, John
  - first particle accelerator, 936
- Cogeneration, **265–266**
  - energy conversion, 265–266
  - energy efficiency, 265–266
  - thermal energy, 265
- Cogeneration technologies, **266–270**
  - electric power cogeneration site, *267*
  - gas turbines, 269
  - history of, 266–269
  - internal combustion engine, 268–269
  - nuclear power plant cooling tower, *269*
  - smoke jacks, 266
  - steam engines, 267–268
- Color rendering index (CRI)
  - lighting measurement, 713
- Combustion, **270–276**
  - bunsen burner, 271f
  - chemistry of, 273–274
  - diffusion flame, 271f
  - environmental effects, 273–274
  - flame types, 271f, 272f
  - fuels for, 273
  - laminar diffusion flame jet, 271f
  - research in, 274–276, *275*
  - science of, 271–273
- Commercial agriculture
  - crop production rates, 19–21
  - energy use in, 20t
  - fertilizer and pesticide use, 22
  - history and development of, 18–19
  - irrigation use, 21
  - land use, 21–22
  - machinery use, 22
- Commercial energy
  - agriculture use of, 19–21
- Commercial transport aircraft
  - aircraft efficiency, 61
  - airline industry origins, 60
  - Boeing 707, 43f, 45
  - Boeing 747 aircraft, 45
  - first flight of, 44–45
  - jet aircraft, 60–61
  - supersonic aircraft, 45–46, 46f
- Communications and energy, **276–280**
  - basic signaling, 276–277
  - computer development, 277–278
  - electric telegraph, 277
  - satellite-based systems, 279
  - television evolution, 278–279
- Communications satellites, 278–279
- Compressed natural gas
  - storage and consumption, 830–831, 831t
- Computer technology
  - communications and energy in, 277–278
  - Control systems, 299–301
- Concorde
  - supersonic transport aircraft, 45–46, 46f
- Conservation of energy, **280–287** *See also* Energy conservation
  - chemical and nuclear reactions, 285
  - conservative forces, 280
  - correlation of forces and, 1029–1030
  - first law of thermodynamics, 282
  - heat engines, 282–283
  - internal combustion engine cycle, 283f
  - mechanical energy, 280–281, 281f
  - model for, 286
  - refrigerators, 284–285
  - spring mass oscillator, 281–282
  - thermodynamics and, 1032
- Conservation supply curves, **287–291**
  - cost of conserved energy, 287–288
  - macro supply curves, 289–290
  - supply curves of U.S. residential sector, 290f
- Consumption, **291–296**
  - by economic sectors, 293–295
  - efficiency dual role, 291–292
  - sources of energy, 292f, 293f
  - United States energy, 291
- Control systems
  - electromechanical controls, 297–298
  - electronic loop controls, 298–299
  - energy and load control, 302–303
  - history of, **296–303**

  - HVAC controls, 302–303
  - mechanical controls, 297
  - microcomputer systems, 299–301
  - sensor types, 301–302
  - steam engine pressure control, 298f
  - switches and relays, 297–298
  - types of controls, 297–301
- Cool communities, **304–308**
  - cool pavements, 306–307
  - cool roofs, 304–305
  - Los Angeles average summer temperatures, 305f
  - smog predictor graph, 303f
  - temperatures in urban areas, 304, 305f, 307f
  - urban trees, 305–306
- Corporate Average Fuel Economy (CAFE)
  - standards, 450
- Correlated color temperature (CCT)
  - lighting measurement, 713
- Crude oil
  - production and refineries, 988
  - products and markets, 942–943
  - types of, 978–980
- Culture and energy usage, **309–316**
  - bicycling used as transportation, *311*
  - cultural evolution and, 309–311
  - energy policy, 311–312
  - lifestyles, *311*
  - public values and energy, 314–315
  - social organization, 312–314
  - social scientists and, 309–311
- Curie, Marie Sklodowska, **316–317,** *316*
- Curie, Pierre, 316–317
  - radium heat flux measured, 1036
- Cyclotron, *936*, 936–937

D

- Dams
  - history of buttress, 649f
  - history of hydroelectric, 648–651
- Darrieus, D. G. M.
  - vertical-axis wind turbines, 1195
- Davy, Sir Humphry
  - electrochemical experiments, 230
- Debye, Peter, 928
- Decision-making theory
  - energy usage and, 132–133
- de Lavel, Carl G. P.
  - impulse turbine, 1085
- Demand-side management (DSM), **319–324**
  - future of, 323–324
  - growth in, 320–322, 322f
  - history of, 320–322
  - market transformation and, *759*, 759–760
  - National Energy Conservation Policy Act, 321
  - program types, 317–318
- Department of Agriculture (DOA)
  - description of, 590
- Department of Commerce (DOC)
  - description of, 589
- Department of Defense (DOD)
  - description of, 590–591
- Department of Energy (DOE)
  - administration policy and, 586–589
  - description of, 585f, 586–589
- Department of the Interior
  - description of, 590–591
- Department of Transportation (DOT)
  - description of, 590
- Deregulation
  - airline industry, 61–62
- Diesel, Rudolph, **324–326,** *325*
- Diesel cycle engines, **326–336**
  - applications of, 326–327
  - combustion chamber designs, 333f
  - combustion process, 334
  - combustion types, 332
  - constant pressure cycle, 328–329
  - cylinder combustion cycle, 330f
  - designed by principles of thermodynamics, 1033
  - exhaust emissions, 334–335
  - four stroke engine, 327–329, 329f
  - gasoline engines comparison, 335–336
  - pressure volume curve, 329f, 329–331
  - technological advances, 336
  - thermal efficiency, 332–334
  - two stroke engine, 331–332, 332f
- Diesel fuel, **336–342**
  - characteristics of, 338–339
  - crude oil refining system, 338f
  - end-products from crude oil production, 337f
  - grades of, 339
  - production of, 337–338
  - properties of, 339–341
  - synthetic, 341
- District heating and cooling, **342–345**
  - development and status of, 343–345
  - history of, 342–344
- DOE. *See* Department of Energy (DOE)
- Domestic energy use, **345–349**
  - coal use, 347
  - electricity, 348
  - fireplaces, 347
  - gas for cooking, 347–348
  - history of, 345–348
- Double layer capacitors, 214f, 215
- Drag reduction
  - aircraft cowlings, 37–38
- Drive train, **349–356**
  - differential, 355
  - fluid coupling, 353f
  - torque converter, 352–354, 354f
  - transmissions, 350–355
  - universal joint, 355–356
- Dunlop, John Boyd
  - tire inventor, *1139*
- Durand, William F.
  - propeller design coefficients, 958

E

- Economically efficient energy choices, **357–361** *See* Economics
- Economic development
  - emerging fossil fueled economies, 624–626
  - energy consumption, 569–571, 569t
- Economic externalities, **361–363**
  - energy pollution, 362–363
  - environmental issues, 361
  - market equilibrium, 362f
- Economic growth and energy consumption, **363–367**
  - Economic maturity effects on, 364–365
  - economy tied to energy usage, 363–364
  - energy intensity ratio, 364t
  - environmental issues of, 367
  - world's largest energy users, 366f
- Economics
  - economists' views, 360–361
  - energy prices, 358–359
  - prices and external costs, 359–360
  - selection of sources, 357–358
  - timing of costs and benefits, 358–359
- Ecosystems
  - animal population control, 181–183, 186f
  - food chain, 179f, 181–183
  - habitat adaptations, 185–186
- Edison, Thomas Alva, **367–369,** *368*
- Efficiency of energy, economic concerns, **377–380** *See also* Energy efficiency, economic concerns
- Efficiency of energy use, **369–377** *See also* Energy efficiency
- Efficiency of energy use, labeling of, **380–382**
  - appliance labeling rule, 380–382, 381f
- Einstein, Albert, **382–385,** *383*
  - Big bang theory, 154–156
  - general relativity and gravity, 1037
  - letter to FDR regarding nuclear research, *850*
  - mass energy equation, 1036
- Elastic energy, **385–387**
  - materials with, 386f–387
  - springs and, 385–386
- Electrical resistance heating, 603–605
  - welding machine using, *604*

Electric generators, 391–392, 393f
Electricity, **387–394**
  atomic structure, 389f
  circuits, 387–390f, 1099f
  current flow, 387–390
  electrical power and energy, 391–393
  electric charges and force, 387
  generators, 391–392
  history of, 394–399
  magnetic fields and currents, 391–392, 392f
  magnets and electromagnets, 391, 392f
  power delivery, 392–393
  reliability and usage, 136
  static electricity experiments, 388f
  transformers, 392–393
Electricity, history, **394–399**
  battery discharge hindered, 395
  electron research, 397–399
  frictional machine, 394
  generators and dynamos, 396
  Leyden jar, 394–395, *395*
  lighting systems, 396
  storage battery development, 396
  transistors, 399
  vacuum tubes, 398
Electric motor systems, **400–404**
  alternating current, 402
  direct current, 401f–402
  history of, 400
  motor types, 400–402
  performance and efficiency advances, 402–403
  reliability advances, 403–404
Electric power generation, **404–415**
  combined cycle plants, 413
  demand growth of electricity, 408f
  economic regulation of, 409–411
  electricity prices, 407f
  from geothermal energy, 576t, 576–577
  planning and operation, 405–409
  projected changes in total capacity, 414f
  restructuring of industry, 411–412
  types of power, 404–405
  using residual fuels, 1016
Electric power industry
  combined cycle plants effects on, 413
  competitive marketplace, 411–412
  deregulation, 1003–1004
  distributed generation, 413
  economic regulation of, 409–411
  Energy Policy Act of 1992, 409
  Federal Energy Regulatory Commission (FERC), 409
  global production, 569f
  growth of, 234
  National Energy Act of 1978, 408
  natural-gas technologies and, 413
  North American Electric Reliability Council, 423–424
  ownership percentage of U.S. capacity, 410f
  Public Utility Commissions (PUCs), 409
  rates and restructuring, 1004–1005
  regulation and rates, 1003–1005
  restructuring, 1003–1005
  restructuring of industry, 411–412
  time-of-use rates, 1004
Electric power substations, **428–433**
  bus types in, 429t
  circuit breakers, 430
  classes of, 428
  control box in, *431*
  interconnection of, 428–430
  monitoring and protection of, 430–432
  voltage control, 432
Electric power system protection
  control and monitoring of, **415–422**
  magnitude comparison relaying, 417f
  monitoring, 421–422
  nature of protection, 415–417
  protection schemes, 421
  relay design, 419–421
  relay operating principals, 417f, 417–419
  three-zone step relaying, 419f
  zones of protection, 416f
Electric power system reliability, **422–427**
  climate change, 426–427
  customer service, 426
  North American Electric Reliability Council, 423–424
  Northeast blackout of 1965, 424–425
  reliability indices, 422–423
Electric power transmission and distribution systems, **431–436**
  environmental impact of, 437–438
  history of, 433
  system interconnections, 433–435
  transmission line characteristics, 435–437
  transmission line types, 437
Electric transformers
  power delivery and, 392–393
Electric utilities
  Demand-side management, 319–324
  deregulation of power utilities, 1202–1203
  sulfur oxide emission reduction at, 2–3, 3f
Electric vehicles, **438–442**
  air pollution control, 439–441
  basics of, 438–439
  batteries for, 122–123
  battery power, 235
  development of, 439
  internal combustion engine *vs.*, 440–441
  recharging challenge, 441–442
  trains, 973–974
Electrochemical capacitors, 214f, 215
Electrochemistry
  animal muscles and, 230
  batteries and, 233–234
  electric vehicle power, 235
  electrochemical theory, 231–233
  fuel cells, 234–235
  history of, 230–234
  Volta's elements of, 232f
Electrolytic cell
  comparison with voltaic/galvanic cell, 231f
Electromagnetism
  early experiments, 394–395
  invention of electromagnet, 232
Emission control, power plant, **443–449**
  Clean Air Act, 444–445
  clean coal technology, 445–446
  environmental issues, 444–445
  nitrogen oxide control technologies, 447
  particulate control technologies, 447–448
  pollutant regulations, 443–444
  sulfur dioxide control technologies, *446*, 446–447
  technologies and, 445–446
  utility boiler pollutants, 443–444
Emission control, vehicle, **449–458**
  air pollution effects of, 449
  catalytic converters, 451–456, 452f
  cleaner fuels, 454
  engine start emissions, 451–452
  enrichment emissions, 454–455
  fuel economy and, 457
  heavy-duty vehicles, 454
  inspection and maintenance programs, 456
  in-use vehicles, 453–454
  light-duty trucks and SUVs, 453–454
  regulations and standards, *450*, 450–453
  tailpipe standards, 453f
  technology and, 450–453
Energy audits. *See* Auditing of energy use
Energy conservation
  corporate policies, 139–140
  cost of conserved energy, 287–288
  government policies, 138–139

  indoor air quality, 58–59
  interconvertability of forces, 1030
  model for, 286
  National Energy Conservation Policy Act, 321
Energy conservation (in physics). *See* Conservation of energy.
Energy consumption
  aerodynamic drag, 13f, 13–14
  appliances, 77–81
  economic development, 569–571, 571t
  economic maturity effects on, 364–365
  energy intensity, 365f, 366f
  future energy needs, 376–377
  geographic composition, 568f, 568–569
  global electricity production, 569f
  lifestyle decisions, 129–130, 131t
  mass transit impact on, 763t, 763–764
  mass transit use of fossil fuels, 765t
  new technologies, 765–766
  petroleum, 940–948
  railway passenger service, 972f
  U.S. energy per dollar GDP, 462t
  for U.S. furnaces and boilers, 541
Energy conversion
  cogeneration, 263–264
Energy crisis
  effects on building design, 192–193
Energy economics, **458–461**
  costs, subsidies and fuel cycle, 1169–1170
  efficiency and supply technologies, 1179
  energy production steps, 1168f
  external and internal costs, 1167–1178
  integrated fast refiner, *1169*
  key policy issues, 1170–1171
  mineral depletion, 459–460
  pollution as external cost, 1167
  research in, 459
  specialties in, 459
  subsidies, 1102–1106, 1168–1169
  taxation of energy, 1119–1122
  true cost of energy example, 1170
Energy efficiency
  aerodynamics, 13f, 13–14
  agricultural production, 20–21
  air conditioning systems, 29
  annual gain in, 375–376
  appliances and, 75–82, 371–372
  audits, 95–97
  automobile engine advances, 371–375
  buildings and, 371–372
  coefficient of performance (COP), 607–609
  cogeneration, 265–267
  commercial sector, 133
  economic concerns, 377–380
  effects on building design, 193–196
  energy efficient investments, 378–379, 379t
  energy intensity trends, 370–371
  factory productivity and, 671–673, *673*
  future energy needs, 376–377
  gasoline engines, 562–566
  housing choices, 135–136
  industrial sector, 133, 372–374
  iron and steel industry, 751–752
  manufacturing, 750–754
  manufacturing energy use, 373f
  office productivity and, 670
  origins of, 370
  pulp and paper industry, 752–753
  railway passenger service, 970–973, 971f
  residential buildings, 204–208
  residential sector, 135–136
  return on investment for, 377–379
  stirling engines, 1093
  transportation sector, 133–135
  transportation sector savings, 374f
Energy industry
  open access restructuring for green pricing, 598–599
Energy intensity trends, **461–463**
  Corporate Average Fuel Economy (CAFE), 461–462
  energy consumption in U.S., 462t
  history 1972–1986, 461–462
  history 1986–1997, 462–463
  manufacturing trends in, 749f, 750f
  trends in, 751f
Energy management control systems, **463–470**
  characteristics of, 464–466, 465f
  data monitoring, 467f, 468f
  evolution of, 463–464
  function integration, 466–467
  real-time pricing, 467
  trends in, 468–470
Energy marketplace
  Canadian hydroelectric market potential, 665
  energy trading, 662f, 662–663
  government intervention and, 593–596
Energy Policy Act of 1992
  building codes, 202–203
  Electric power industry, 409
Energy production
  materials in, 769–770
Energy requirements
  irrigation pumps effects on, *21*
Energy research and history, 1028–1037
  Carnot cycle and Kelvin scale, 1032
  correlation of forces and conservation of energy, 1029–1030
  doctrine of energy, 1030
  electric and magnetic forces = transverse force, 1029–1030
  energy as quantized packages, 1035
  energy research, 1105–1106
  energy terminology defined, 1033–1034
  equipartition theorem, 1034–1035
  era of correlation (1800–1840s), 1029–1030
  era of energy (1840s–current), 1029–1030
  general relativity and gravity, 1036–1037
  gravitation and space-time curvature, 1037
  Heisenberg’s uncertainty principle, 1035
  high pressure steam engines, 1031–1032
  horsepower, first measurement, 1031
  interconvertability of forces, 1030
  isothermal heat transfer, 1032
  linguistic definition of, 1028
  Newcomen’s atmospheric engine, 1031
  origin of term, 1028–1029
  physical forces and kinetic energy, 1029
  quantum mechanics development, 1035
  reversibility paradox, 1033
  scientific definition of, 1028–1029
  steam engine history, 1030–1031
  theory of gravity, 1037
  thermal and thermodynamic efficiencies, 1032
  thermodynamic engines, 1034
  thermodynamics, 1032
Energy standards
  Corporate Average Fuel Economy (CAFE), 461–462
Energy Star program
  office equipment, 901–902, 902t
Energy storage, **1095–1097**
  battery storage, 1101
  capacitance, 213f
  capacitors, 213f
  capacitors and ultracapacitors, 213–216
  coal reserves, 1096
  compressed air energy storage, 1097
  cool thermal, 1101–1102
  electric power plants, 1097
  electric utilities peak demand, 1098
  evaluating systems by energy density, 1098–1099
  future growth, 1102

Energy storage (cont'd)
  hybrid electric vehicles, 634–636
  nuclear power plants, 1097
  petroleum reserves, 1096
  solar energy, 1097
  solar energy and photosynthesis, 1096
  superconductor magnetic, 1099f, 1099–1100
Energy transformation
  biological energy, 179–183
Energy transmission
  mechanical. *See* Mechanical transmission of energy
Energy use
  in agriculture, 20t
  air pollution and, 48–51
  audits, 95–97
  in buildings, 192–193
  in commercial buildings, 196f
  cost and effects of human, 186–188
  decision-making theory and, 132–133
  energy units and equivalents, 1198
  in global manufacturing, 748–749
  joule definition, 1197
  measurement units, 1197–1198, 1198t
  for processing materials, 769–776
  in railway passenger service, 971–975
Engines, **470–473**
  combustion engine, 470–471, 472f
  diesel engine, 473
  internal combustion engine, 471–472, 472f
  internal combustion engines and economy, 625
  steam engines and economy, 625
Environmental activities
  air pollution, 478
  Earth Day 1970, 477
  nuclear power, *477,* 481
  water pollution, 479–481
Environmental economics, **473–475**
  property rights, 473–475
Environmental effects
  combustion, 273–274
  of geothermal energy, 577–578
Environmental impact
  of electric power transmission lines, 435–436
Environmental issues
  aviation fuel, 112–113
  economic externalities, 361–363
  economic growth and energy consumption, 367
  oil refineries, 987–988
  petroleum production, 949–950
  power plant emission control, 444–448
  in production of materials, 775–776
  reductions using railways, 975–976
Environmental problems and energy use, **475–483**
  energy-environment link, 475–477
  environmental action, 477–482
  future of, 482f
  population growth and, 476
  power plant pollution model, *476*
Environmental Protection Agency (EPA)
  air pollution, 51
  description of, 588–589
  fuel economy standard, 103–104
Environmental regulations
  for coal production, 260–262
EPA. *See* Environmental Protection Agency (EPA)
Ericsson, John, **483–485**
  modern marine propeller, 957
Ethanol
  production of, 160–162
Ethical and moral aspects of energy use, **486–493**
  energy policy and individual choice, 486–488
  social ethics and policy (future), 491–493
  social ethics and policy (present), 488–491
Europe
  nuclear waste handling, 886
Explosives and propellants, **493–495**
  gunpowder, 493–495
  propellants, 493–495
  safety in manufacturing, 493
  types of explosives, 494t
Exxon Valdez, 949

F

Faraday, Michael, **497–498**, *498*
  electric generators, 391–392, 393f
  electromagnetism, 231, 395, *397,* 1156
  era of correlation (1800–1840s), 1029–1030
Federal Energy Regulatory Commission (FERC)
  Electric power industry, 409
Fermi, Enrico, **499–501,** *500*
Fermi National Accelerator Laboratory
  collider detector, *938*
  description of, 815–816
  proton accelerator, *935*
  world's largest particle accelerator, *816*
Fludd, Robert, 939
Flywheels, **501–504**
  applications of, 503t
  future advances of, 504
  history of, 501–502, *502*
  technology of, 502–504
Forests
  acid deposition effects on, 5–6
  nitrogen and nitrate effects on, 6
Fossil fuels, **504–508**
  agriculture use of, 16
  coal resources, 504
  consumption by mass transit, 765t
  effects of, 187–188
  Natural Gas Policy Act of 1978, 506
  oil and gas production, 505f, 506
  Organization of Petroleum Exporting Countries (OPEC), 506
Fourier, Jean Baptiste Joseph, **508–510,** *509*
Four-stroke piston engine, 931
  history of, 557–558, 558f
Fourth law of thermodynamics, 928
Franklin, Benjamin
  electrical experiments, 394
  Pennsylvania fire place, 346–347
Freight movement, **510–521**
  cargo quilt cooling, 520b
  containerization, 515–516
  cost comparison of, 516t
  energy requirements for, 512t
  future of, 517–521
  intermodalism in, 514t, *515*, 514–521
  loading/unloading standards, 515–516
  nature of energy and, 510t, 511t, 510–513
  railroads and waterways, 511–513
  railroads *vs.* trucking, 511–513
  railroad technology, 516–517
  trucking, 513–514
Fuel additives
  aviation fuel, 111–112
Fuel cells, **521–531**
  development of, 234–235
  electricity production by, 523–527
  fuel cell stack, 523–525, 524f
  future challenges, 529–530, 530f
  history of, 522f, 522–523
  hydrogen economy shift, 530f
  spacecraft power supply, 1076
  system diagram, 526f
  types of, 527–529, 527t
Fuel cell vehicles, **531–534**
  commercialization challenges, 531–532
  hydrogen fuel, 532–533
  naphtha fuel, 533
  natural gas fuel, 531–532
Fuel economy
  automobile design effects on, 105f
  automobile performance trade-off, 106f, 107f
  emission control and, 457
  EPA standard measurement of, 103

  - of railway passenger service, 970–973, 971f
  - reformulated fuel production and, 987
- Fuels. *See also* Alternative fuels, Rocket fuels, Gasoline and special types (hydrogen fuel, natural gas fuel, etc.)
  - cleaner automotive, 456–457
  - combustion properties, 830t
- Fuller, R. Buckminister, Jr., **534–537,** *535*
  - solar-powered geodesic dome, *1066*
- Fulton, Robert, **537–538**
- Furnaces and boilers, **539–543**
  - efficiency of, 541–542, 541f
  - gas and oil fuel for, 539f, 539–541
  - heat loss from pipes and ducts, 542f
  - history of, 539
  - typical operation of, 540f
  - U.S. energy consumption for, 541
- Futures, **543–546**
  - history of, 543
  - market features, 544
  - New York Mercantile Exchange (NYMEX), 545–546
  - options and, 544
  - overview of, 543
  - trading procedures, 545–546

G

- Gasoline and additives, **547–555**
  - alternative fuel competition, 555
  - demand growth history, 547–548
  - environment and, 551–555
  - gas additives and knocking, 549–551
  - gas additives octane level, 549f
  - manufacturing process, 553f
  - new and emerging additives, 554–555
  - octane rating, 548–549
  - reformulated gas and MTBE, 552–554
  - sulfur and, 552
  - tetraethylead gas additive, 549–550
- Gasoline engines, **556–566**
  - air-fuel ratio, 564–565
  - compression ratio, 563–564
  - diesel cycle engines comparison, 335–336
  - engine efficiency, 562–566
  - four-stroke piston engine, 557–559, 558f
  - history of, 557–561
  - ignition timing, 565
  - multicylinder engines, 561–562, 562f
  - overhead cam engine, 557f
  - parasitic losses, 565–566
  - push rod engine, 556f, 556–557
  - rotary engine, 560–561, 561f
  - two-stroke piston engine, 559–560, 560f
- Gas turbines, **1171–1182**
  - advanced, 1181
  - alternative thermal cycles, 1174–1178
  - applications of, 1180–1181
  - basics of, 1171–1173, 1172f
  - biofuels for, 1179–1180
  - clean coal technology and gasification, 1179
  - cogeneration with, 269
  - component development, 1173–1174
  - compressor intercooling, 1176
  - fuel cells, 1177–1178
  - gasification combined cycle (GCC), 1178–1179
  - humid air turbine (HAT) cycle flow, 1177f
  - hydrogen fueled, 1180
  - origins of, 1172–1173, 1173f
  - other fuels for, 1178–1180
  - partial oxidation, 1177
  - recuperative cycles, 1176
  - role of advanced turbine technology, 1181
  - Stolze power plant, 1173f
  - thermochemical recuperation, 1176
  - turbine cycle options, 1175f
  - turbine inlet temperature, 1173–1174, 1174f
- Geochemistry
  - in oil and gas exploration, 916–917
- Geography and energy use, **567–572**
  - aggregate energy use, 567–569
  - energy consumption, 567–569, 568f
  - global electricity production, 569f
  - global energy consumption, 569f, 569–571
  - quality of life and, 571
  - self-sufficiency, 567
- Geothermal energy, **572–579**
  - direct use of, 576–577, 578t
  - electricity generation, 575–576, 576t
  - environmental effects of, 577–578
  - future of, 578–579
  - power plant, *573*
  - source of, 572–574, *575*
  - thermal gradient map, *573*
  - types of, 572–574
- Gibbs, Josiah Willard, **579–581,** *580*
- Gliders, 33
- Glushko, Valentin P.
  - thermal electric engine inventor, 1076
- Goddard, Robert H.
  - first liquid propellant rocket, *1021*
- Gorrie, John
  - inventor of air conditioning, 25
- Government agencies, **581–592**
  - state and local agencies, 591
  - United States agencies, 585–591
  - world agencies, 582–585
- Government and the energy marketplace, **592–596**
  - information, 595
  - intervention in energy crisis, 596
  - intervention types and reasons, 592–596
  - regulations, 593–595
  - research and development, 595–596
  - subsidies, 593
  - taxes, 593
- Gravitational energy, **596–598**
  - energy storage as, 597
  - hydroelectric power plants, 597–598
- Gravity
  - light and, 778
  - Theory of relativity, 778
- Green energy, **598–601**
  - consumer energy preferences, 599–601
  - defining energy choices, 599–600
  - government codependance, 599
  - open access restructuring, 598
  - power subsidies for, 598
  - pricing challenges, 598–600
  - tracking and labeling issues, 600–601
- Greenhouse gases
  - biofuels and, 163–164
  - climate change theory, 239–240, *239*
  - cost and impact of reducing, 250
  - Kyoto protocol (United Nations), 248–250, 249t
- Gunpowder, 493

H

- Heat and heating, **603–606**
  - active heating and cooling, *1055*, 1055–1056
  - cooling towers and low-grade heat, 606
  - electrical resistance heating, 603–605
  - electric power transmission, 605–606
  - forced-air heating, 603
  - hot water heating, 603
  - passive heating and cooling, 1054f, 1054–1055
  - production of unwanted heat, 605
  - solar energy and, 1053–1056
- Heating systems
  - fireplaces, 346–347
  - history of, 345–347
- Heat pumps, **607–611**
  - air-to-air pumps, 609
  - basics of, 607–609

- Heat pumps (cont'd)
  - coefficient of performance (COP), 607–609
  - cooling cycle, 608f
  - energy efficiency of, 607–609
  - future and current practice, 611
  - heating cycle, 608f
  - history, 609–610
  - refrigeration systems working with, 607
  - thermoelectric, 607
- Heat transfer, **612–616**
  - conduction, 612
  - convection, 612–613
  - radiative, 613–614
  - R value coefficient for heat flow, 614–616
  - R value, wall, 614f, 614–615
  - R value, window, 615f, 615–616
- Heaviside, Oliver, **616–617**
- Heavy-duty vehicles
  - emission controls, 454
- Heisenberg's uncertainty principle
  - quantum mechanics and, 1035
- Helmholtz, Hermann Von, **618–619**
- Hertz, Heinrich Rudolf, **619–621,** *620*
- Historical perspectives and social consequences, **621–629** *See also* Human societies
- Houdry, Eugene Jules, **630–632**
  - catalytic cracking, 991–992
- Houses
  - electricity, 136
  - energy efficient choices, 135–136
- Hubble Space Telescope
  - use of solar energy, 1063
- Human behavior, 129–141
  - American policies effect on, 129–130
  - energy consumption and, 129–130
  - energy efficient, 130, 132
  - food and health, 136
  - housing choices, 135–136
  - lifestyle decisions, 130–132, 131t
- Human societies
  - early agricultural societies, 73–74
  - emerging fossil fueled economies, 624–626
  - energy and tools in, 71–75
  - energy conversion in traditional, 623–624
  - energy use in, 72f
  - future energy challenges, 629
  - high energy civilization, 627–629
  - hunting-gathering societies, 71–73, 623–624
  - prehistoric and ancient cultures, 622–623
- HVAC systems
  - in building design, 194–195
  - indoor air quality, 53–54
- Hybrid vehicles, **633–644**
  - component technologies, 634–638
  - control strategies, 638–640
  - driveline schematics, 634f, 638f
  - electrical energy storage, 634–636
  - emissions of, 642–643
  - engines and auxiliary power units, 635t, 636–637
  - fuel cells, 637–638
  - fuel economy, 640–641, 641t, 642t
  - marketing prospects for, 643
  - mechanical components, 637
  - performance of, 640–643
  - testing of, 643
  - vehicle design options, 633–634
- Hydroelectric energy, **644–652**
  - basics of, 644–648
  - Canadian market potential in, 667
  - conventional hydroelectric system, 647f
  - future of, 651
  - history of hydroelectric dams, 648–650
  - hydrologic cycle, 646f
  - Itaipú - world's largest hydroelectric plant, 650
  - types of buttress dams, 649f
- Hydroelectric power plants
  - gravitational energy storage in, 597–598
  - gravitational energy use in, 597–598
  - Itaipú - world's largest hydroelectric plant, 650
- Hydrogen, **652–660**
  - developing energy system scenarios, 658–659
  - distribution of, 654
  - economics of production and use of, 658t
  - environmental and safety considerations, 657–658, 658t
  - fuel cell, 655–656, 656f
  - fuel cell vehicles, 531–533
  - history of, 653
  - production of, 653–654, 654f, 655f
  - properties of, 653t
  - storage, 654–655
- Hydrological cycle
  - atmosphere, 86, 88–89

I

- Idaho National Engineering Laboratory
  - description of, 816
- Import/export market for energy, **661–665** *See also* Energy marketplace
  - Canadian hydroelectric market potential, 665
  - energy independence, 663–664
  - energy trading, 662f, 662–663
  - friendly country imports, 664–665
  - history of U.S. energy imports, 663
  - import/export of energy, 661–665
- Indoor air quality
  - air cleaning and filtration, 55, 58
  - energy conservation, 58–59
  - HVAC systems, 53–54
  - pollutants, 54–58, 56t
  - source control, 54–55
  - ventilation, 55
- Industry and business
  - energy measurements, 666–668
  - energy utilization chains, 668–669
  - factory productivity and energy efficiency, 671–673, *673*
  - office productivity and energy efficiency, 670
  - solar powered car energy metrics, *667*
  - workplace environment control, 670–671
- Industry and business, energy as a factor of production in, **666–669** *See* Industry and business
- Industry and business, productivity and energy efficiency in, **669–674** *See* Industry and business
- Industry sectors
  - energy efficiency in, 133–135
- Insulation, **674–678**
  - applications for, 676
  - historical development, 674
  - materials for, 675–677
  - metrics and measurements of, 675t
  - thermal resistivity or R value, 675–677, 677t
  - trends in, 677
- Intergoverment Panel on Climate Change (IPPC)
  - climatic effects, 240, 246–248
- Intermodal freight transport, 514–517, *515*
- Internal combustion engine
  - electric vehicles *vs.,* 440–441
  - military use of, 800–801
- Internal combustion engines
  - economy development, 625
- International Atomic Energy Agency (IAEA)
  - description of, 584
- International Energy Agency (IEA)
  - description of, 582–583
- Investment strategy
  - energy efficient investments, 377–379, 379f
  - return on investment on private markets, 377–379
- Ipatieff, Vladimir Nikolaevitch, **678–681**
- Iron and steel industry
  - energy efficiency in, 751–752
- Irreversible transport processes, 928, 930

J

- Jet aircraft, 40–45
  - commercial transport, 60–61
  - delta wing, 44
  - gas turbine engines, 40–41
  - North American F-86H jet fighter, 41f, 43
  - swept wing, 41–44
- Joule, James Prescott, **683–685,** *684*
  - doctrine of energy, 1030
- Junkers, Hugo
  - first all-metal airplane, 37

K

- Kamerlingh Onnes, Heike, **687–*688***
- Karcher, John Clarence, 920
- Keely, John W., *940*
- Kerosene, **689–691**
  - equipment, 691
  - history of, 689
  - properties of, 689–691, 690t
  - range burner cross-section, 690f
  - vaporizing pot burner cross-section, 690f
- Kinetic energy, **691–692**
  - in ancient world (to 500 C.E.), 692–693
  - basics of, 691–692
  - in early modern era (1500–1880), *695*, 695–697
  - medieval world (500–1500 C.E.), 694–695
  - in modern era (1880–2000), 697–699
  - Phoenician ships, *693*
  - water wheels, 695–696
  - wind mills, *695, 696, 698*
- Kinetic energy, historical evolution of the use of, **692–699** *See* Kinetic energy
- Kyoto protocol (United Nations)
  - coal consumption, 255–257
  - treaty to limit greenhouse gases, 249–250, 250t

L

- Lakes and streams
  - acid deposition effects, 4–5
  - lake acidity, 5
- Laminar flow airfoil, 10, 11f
- Langen, Eugen, 931
- Langley, Samuel P.
  - aerodynamic experiments, 33–34
- Laplace, Pierre Simon, **701–703,** *702*
- Lasers, **703–706**
  - applications of, 705–706
  - basics of, 703–704
  - development of, *704*, 704–705
  - invention of, *704*
  - types of, 705–706
- Lawrence, Ernest
  - first circular particle accelerator, *936,* 937b
- Lawrence Berkeley National Laboratory
  - description of, 816–817
- Lawrence Livermore National Laboratory
  - description of, 817
  - National Ignition Facility for nuclear fusion, 876f
- Lead
  - air pollutants, 50–51
- Lenoir engine, 930
- Lewis, Warren K., **707–708**
- Life-cycle cost analysis
  - calculations for, 217–219
  - for capital investment decisions, 216–218
  - gasoline price sensitivity to, 218f
  - steps in, 216–217
- Light-emitting diodes (LEDs)
  - lighting applications of, 718
- Lighting, **709–719**
  - basics of, 709–710
  - colorimetry, 713
  - color rendering index (CRI), 713
  - color temperature and rendering for electric, 712t
  - correlated color temperature (CCT), 713
  - fluorescent, 716–717
  - future trends, 718–719
  - human eye cross-section, 710f
  - incandescent, 716
  - lamp types, 714f
  - light and color measurement, 710–713
  - light-emitting diodes (LEDs), 718
  - luminaire types, 715f
  - luminous efficiency functions, 711f
  - types of, 713–718
- Lilienthal, Otto, *32*, 33
- Lime
  - controlling effects of acid deposition, 5
- Lippisch, Dr. Alexander
  - delta wings, 44
- Liquefied petroleum gas, **719–722**
  - commercial properties, 721–722
  - history of, 720
  - physical properties, 720t, 720–721
  - production and delivery, 722
  - propane production process, 721f
- Lithium-ion cells
  - alkaline rechargeable batteries, 120
- Lockheed F-104 Starfighter, 42f, 43–44
- Lockheed Vega aircraft, 38f
- Locomotive technology, **723–731**
  - emissions, 730–731
  - future of, 731
  - history of, 723, 725t, 726t, 727t
  - history of locomotives in service, 728f
  - operational control, 729–730
  - performance comparison, 730t
  - power generation, 723–728
  - revenue class 1 railroads, 729f
  - tractive effort, 728–729
  - types and characteristics of, 723–728, 724t
- Los Alamos National Laboratory
  - description of, 817
- Lubricants and lubrication. *See* Tribology or Refineries
- Lyell, Charles, **732–733**

M

- Maglev trains. *See* Magnetic levitation
- Magnetic fields
  - currents and, 391–392, 392f
- Magnetic levitation, **735–740**
  - history of, 735
  - linear motors for, 737–738
  - maglev train, *736,* 738–739
  - pros and cons of, 739–740
  - suspension and propulsion, 736–738
- Magnetism and magnets, **740–744**
  - applications of, 742–744
  - lodestone and magnets, *741*, 741–742
  - magnetic fields, 742
- Magnetohydrodynamics (MHD), **745–747**
  - developments in, 745–747
  - history of, 745–747
  - linear MHD generator channel, 745f
  - MHD and turbine plants temperature comparison, 747f
  - MHD disc generator, 746f
  - MHD linear generator flow chart, 746f
  - power system requirements, 747
- Maiman, Theodore H.
  - invention of lasers, *704*
- Manufacturing, **748–756**
  - energy efficiency in, 750–754
  - energy intensity trends, 749t, 750f, 751f
  - global energy usage in, 748–749
  - iron and steel industry efficiency, 751–752
  - potential efficiency improvement in, 754–756, 755t
  - pulp and paper industry efficiency, 752–753
  - steam production and use, 753t, 753–754
- Market imperfections, **757–758**
  - policy responses, 758
  - types of, 757–758

- Market transformation, **759–761**
  - definition of, 759
  - demand-side management (DSM), *759*, 759–761
  - implementation of, 760–761
- Mass transit, **761–769**
  - bicycling and, 768–769
  - bus and rail comparison, 764–765
  - distribution by type of, 762t
  - energy intensity averages for, 763t, 764t
  - ferryboats, *762*, 762–763
  - forms of, 761–763
  - fossil fuel consumption by, 765t
  - impact on energy consumption, 763t, 763–764
  - land use by, 766–768
  - social and political factors, 766–768
  - subway, *767*
  - transit mode comparison, 764–765
- Materials, **769–776**
  - ceramics recycling, 774
  - electronic materials production, 775t
  - energy production and, 769–770
  - energy use for metal processing, 770t, 770–773
  - energy use for mining, 771
  - energy use for processing iron alloys, 771t
  - energy use for producing ceramics, 773–774
  - energy use for producing commercial gasses, 775t
  - environmental concerns in production of, 774–776
  - metal recycling, 773
  - metals consumption in U.S., 773t
  - product packaging, 137–138
  - vehicle weight and efficiency, 769b
- Matter and energy, **776–780**
  - chemical reactions, 780
  - composition of universe as, 777b
  - converting matter to energy, *779*, 779–780
  - gravity and, 778
  - inertia, mass and acceleration, 776
  - matter and the speed of light, 777
  - nuclear reactions, 779, 779–780
  - properties of antimatter, 778–780
  - properties of matter, 776–778
  - Theory of quantum mechanics, 777
  - Theory of relativity, 777
- Maxwell, James Clerk, *781*, **781–783**
  - equipartition theorem, 1034–1035
  - Faraday and conservation of force, 1030
- Maybach, Wilhelm, 931, 932
- Mayer, Julius Robert, *783*, **783–784**
- Mechanical energy
  - basics of, 785
  - belts and chains, 789–790
  - gears and torque, 788–789, 789f
  - hammer and wedge, 786f, 786–787
  - inclined plane, 786f, 786–788
  - levers, 785–786
  - pendulum conserving energy, 280
  - pulley, 787f, 787–788
  - screw, 788
  - wedge, 786f, 786–787
- Mechanical transmission of energy, **785–790** *See* Mechanical energy
- Meitner, Lise, **790–792,** *791*
- Metabolism
  - of animals, 183
  - basal metabolic rate, 175–176
  - catabolism, 169–171
  - food and ATP, 171–172
  - physical activity effects on, 175–176
- Methane, **792–794**
  - control in oil and gas drilling, 915
  - material conversion to, 792–794
  - molecular structure of, *792*
  - use for energy, 792–794
- Methanol, **794–796**
  - automobiles using, 794–795
  - fuel cell vehicles use of, 795–796
  - MTBE use of, 795
  - undersea hydrate reserves supply of, 795b
- Meusnier, Jean Baptiste
  - propellers on balloons, 956
- Military energy use, historical aspects, **796–802**
  - aircraft, *801*
  - ancient history, 796–798
  - catapults, 796–797, *797*
  - communications, 800
  - gunpowder, 796
  - information gathering, 801
  - internal combustion engine, 800–801
  - kinetic energy weapons, 796
  - steam engines, 798–800
  - World War II, 801–802
- Miller SQA (building), 198b
- Mintrop, Ludger, 920
- Molecular energy, **802–809**
  - atomic structure, 802–806
  - bioenergetics, 808
  - bond formation of first 10 elements, 805t, 805–806
  - chemical bonding, 803f, 804f
  - energy and chemical change, 806–807
  - oxidation and reduction, 807–808
  - periodic table of the elements, 803f
  - quantum mechanics, 804
  - randomness and spontaneity, 807
- Morse, Samuel F.
  - invention of the telegraph, 277
- Multicylinder engines
  - history of, 561–562, 562f

N

- Nanotechnologies, **811–813**
  - atomic imaging and manipulation, 812
  - commercial nanotechnology, 813
  - computational nanotechnology, 812–813
  - Fermi National Accelerator Laboratory, *816*
  - future of, 813
  - history of, 811
  - molecular energy for, 812
  - reversible fuel oxidation, 812
- National Advisory Committee for Aeronautics (NACA)
  - aircraft cowling, 37–38
- National Aeronautics and Space Administration (NASA)
  - aircraft energy efficiency program, 957
  - description of, 591
  - propfan development, 959
- National Ambient Air Quality Standards, *50,* 51–52, 52t
- National Appliance Energy Conservation Act of 1987
  - refrigerators and freezers, 998–999
  - water heaters, 1218
- National Energy Act of 1978
  - electric power industry, 408
- National Energy Conservation Policy Act (NECPA)
  - Energy conservation, 321
- National energy laboratories, **813–820**
  - Ames Laboratory, 814
  - Argonne National Laboratory, 814–815
  - Atomic Energy Commission years, 813–814
  - Brookhaven National Laboratory, 815
  - Fermi National Accelerator Laboratory, 815–816
  - future of, 819–820
  - Idaho National Engineering Laboratory, 816
  - Lawrence Berkeley National Laboratory, 816–817
  - Lawrence Livermore National Laboratory, 817
  - locations of major labs, *815*
  - Los Alamos National Laboratory, 817
  - National Renewable Energy Laboratory, 817
  - Oak Ridge National Laboratory, 817–818

  Pacific Northwest Laboratory, 818
  Princeton Plasma Physics Laboratory, 818
  Sandia National Laboratory, 818
  Stanford Linear Accelerator Laboratory, 818
National Renewable Energy Laboratory
  description of, 817
Natural gas
  distribution, 837–838
  fuel cell vehicles, 531–532
  future of, 839–840
  legislation and regulations, 838
  pipeline construction, *837*
  pricing of, 839
  restructuring of industry, 838
  storage of, 836–837
  supply and demand, 836
  transportation of, 835, 836f
Natural gas, consumption of, **820–827** *See* Natural gas consumption
Natural gas, processing and conversion of, **828–834** *See* Natural gas processing
Natural gas, transportation, distribution and storage of, **834–840** *See* Natural gas
Natural gas consumption
  alternative fuels derived from, 833f
  chemical conversions, 832–834, 833f
  commercialization of stranded gas, 826
  competition with other fuels, 826–827
  compressed natural gas, 830–832, 831t
  consumption by country, 822f, 823–826, 824f
  European Union, 823–825
  Fischer-Trobsch diesel, 833–834
  as a fuel additive, 834
  fuel combustion properties, 830f
  history of, 820–823
  interfuel substitution, 826–827
  Japan, 826, 825f
  liquefied natural gas, 831–832
  North America, 823–825
  South Korea and Taiwan, 826
  street gas lamps, *821*
  supply and import/export destinations, 829f
  trends in, 827
  United States, 825f
Natural Gas Policy Act of 1978
  natural gas deregulation, 505
Natural gas processing
  conversions, 828, 830
  supply and import/export destinations, 829f
Nernst, Walther Hermann, **840–842,** *841*
Newcomen, Thomas, **842–843**
  atmospheric engine, 1031
Newton, Isaac, *844*, **844–846**
  Philosophiae Naturalis Principia Mathematica, *845*
New York Mercantile Exchange (NYMEX)
  futures investing, 545
Nickel alloy batteries
  for electric vehicles, 123
Nitrogen cycle, **846–847**
  dioxide as air pollutants, 51
  fertilizer and, 847
  humans, environment and, 847
  oxide control technologies, 447
North American Electric Reliability Council (NERC)
  Electric power industry, 421–422
North American F-86H jet fighter, 41f, 43
NRC. *See* Nuclear Regulatory Commission (NRC)
Nuclear energy
  atom bomb, *852*
  Atomic Energy Act of 1954, 854
  Atomic Energy Commission, 853–854
  cold war, 851–853
  defense and, 857
  electricity generating station, *856*
  integrated fast refiner, *1171*
  Japan bombing, 851, *852*
  Manhattan project, 851
  military origins, 850–853
  nuclear power industry, 853–854
  opposition to, 854–857
  radioisopotic thermoelectric generators (RTG), 1077–1079
  reaction process, *779*–780
  ship and submarine development, 853–854
  spacecraft nuclear reactors, 1077–1079
  thermionic generators, 1078–1079
  USSR spaceborne nuclear reactors, 1079
Nuclear energy, basic processes of, **847–849** *See* Nuclear energy processes
Nuclear energy, historical evolution of the use of, **850–858** *See* Nuclear energy
Nuclear energy processes
  fission and fusion definitions, 848
  nuclear power plant, *848*
  nuclear reactors, 849
Nuclear fission, **858–866**
  basics of, 858–860
  breeder reactor, 865
  chain reaction, 861f
  components of, 862f
  control rods, 862
  energy production, 863
  fuel for, 861
  mixed oxide fuel, 870
  moderator fluid, 862
  neutron-induced fission process, 860f
  plutonium fuel, 869–870
  reactor control, 864–865
  reactor water steam circuit, 864f
  spontaneous fission process, 859f
  uranium conversion and enrichment, 868f, 869
  uranium fuel, 862–863, 866–869
  uranium mining and milling, 867t, 868f
  uranium production, 867t
Nuclear fission fuel, **866–871** *See* Nuclear fission
Nuclear fusion, **871–878**
  basics of, 871–874
  conditions for energy production, 874–875
  fusion reactions, 872–874
  fusion reactor, 873f
  inertial confinement fusion, 875–*876,* 877f
  magnetic confinement fusion, 877–878
  nuclear accelerator, *872*
  Tokomak reactor, *877*, 877f, 878f
Nuclear power
  environmental activities in, *477,* 481
Nuclear power industry
  Atomic Energy Act of 1954, 854
  electricity generating station, *856*
  origins of, 853–854
  world's first power plant in USSR, 854
Nuclear Regulatory Commission (NRC)
  description of, 591
Nuclear waste, **879–887**
  calcining facility, 883f
  concerns for, 879
  disposal sites, *882*
  Europe and, 886
  low-level radioactive, 884–885
  origins of radioactivity in fuel, 879–881, *880*
  reprocessing of, 884
  SNF handling, 882–883
  SNF - isolate or reprocess, 883–884
  SNF volume for storage option, 884
  Soviet Union and, 886
  storage of, *880, 881*
  storage sites for, 885–886
  transportation concerns, 886
  types of, 879
  Yucca Mountain, Nevada, 883–884

O

Oak Ridge National Laboratory
  description of, 817–818

- Ocean energy systems, **889–895**
  - Ocean thermal energy conversion (OTEC), 889–891
  - tidal energy, 892–895, *894*
  - wave power, 891f, 891–892
- Ocean thermal energy conversion (OTEC)
  - ocean energy systems, 889–895
  - types of power plants, 890
- Oersted, Hans Christian, **895–897,** *896*
- Office equipment, **899–903**
  - energy star program, 901–903, 902t
  - power levels, energy use and usage patterns, 901–902, 903
  - power requirements of, 901
  - types of, 899–901
- Oil and gas, drilling for, **904–915** *See* Oil and gas drilling
- Oil and gas, exploration for, **915–923** *See* Oil and gas exploration
- Oil and gas, production of, **923–927** *See* Oil and gas production
- Oil and gas drilling
  - arctic areas, 912
  - basics of, 905–909, *904*
  - carbon dioxide control in, 915
  - deep water, 911f
  - drill bits and well design, *908*, 909–910
  - drill cuttings, 909f
  - drilling rig operation, 906f, 906–907
  - drilling to total depth, 907–908
  - early methods, 905
  - fixed ocean rigs, 913f
  - future trends, 914–915
  - horizontal and directional drilling, 910f, 911
  - jacked-up rigs, 912f
  - methane control in, 915
  - modern drilling, 905–910
  - multilateral drilling, 910f, 911
  - offshore drilling, 911–915
  - semisubmersible rigs, 914f
  - shallow water, 912
  - slimhole drilling, 909
  - submersible base rigs, 911f
  - surface hole drilling, 907
  - well completion, 908–909
- Oil and gas exploration
  - basics of, 916–918
  - current methods of, 922
  - expected monetary value (EMV), 922
  - geochemistry in, 916–917
  - geological mapping, 918–919, 919f
  - geophysical methods, early history of, 919–921
  - gravity methods, 919–920
  - kerogen in source rocks, 916f, 916–917
  - oil shale rock, *917*
  - reservoir rocks, 917–918
  - refraction method, 920
  - risk evaluation, 922–923
  - Schlumberger Limited, 921
  - seismic methods, 920–921
  - source rocks, 916–917
  - traps, 918, 919f
  - vertical seismic profile (VSP), 922
- Oil and gas production, 503–504
  - basics of, 923–925
  - bubble point, 925
  - field processing, 924–925
  - infield drilling, 926
  - international, 944–945
  - market pricing, 1104–1105
  - natural gas production, 923
  - oil recovery, 959
  - oil reservoir drives, 959
  - oil well pumps, *926*
  - petroleum imports, 1013
  - pressure maintenance, 925
  - production platform, *924*
  - property rights and mineral rights, 960–961
  - reformulated fuel production, 986
  - reservoir (definition), 924
  - reservoir pressure, 924–925
  - secondary recovery, 924–925
  - for transportation, 1158–1160
  - United States, 943–944
  - unitization contracts, 961, 963
  - world crude oil production, 923
- Onsager, Lars, **927–930**, *929*
- Organization of Petroleum Exporting Countries (OPEC)
  - description of, 582
  - oil crisis of 1973, 948
  - U.S. policy and, 505f, 506
- Ostwald, William
  - work on catalysts, *224*, 224–225
- Otto, Nikolaus August, **930–932**, *931*
- “Otto silent” engine, 932

P

- Pacific Northwest Laboratory
  - description of, 818
- Parsons, Charles Algernon, **933–934**, *934*
  - epicyclodial steam engine, 933
- Particle accelerators, **935–939**
  - facts about, *936*
  - Fermilab collider detector, *938*
  - Fermilab proton accelerator, *938*
  - history of, 936
  - operation of, 936–939
  - physics of, 936
  - synchotron radiation, 938–939
- Particulate matter
  - air pollutants, 49–50
- Particulates
  - control technologies, 445
- Perkins, Jacob
  - invented vapor compression system, 996
  - inventor of refrigeration system, 25
- Perpetual motion, **939–942**
  - law of conservation of energy, 940
  - magnetic tricks for, 939–940
  - natural forces and, 940
  - pendulum, 941
  - second law of thermodynamics, 941
- Petroleum consumption, **942–950**
  - crude oil products and markets, 942–943
  - early 20th century patterns of, 945–946
  - environmental hazards, 949–950
  - international, 944–945
  - nature of, 942
  - 19th century patterns of, 945
  - 1920–1930s patterns of, 946
  - oil crisis of 1973, 948
  - pipelines, 948–950
  - postwar period, 947–948
  - production patterns, 943–945
  - regulatory environment, 949–950
  - Resource Conservation and Recovery Act (RCRA), 949
  - San Francisco gas prices (2000), *945*
  - shipping, 949
  - storage, 949
  - transport and storage, 948–949
  - United States, 945–948
  - World War II patterns of, 946–947
- Photosynthesis
  - source of biological energy, 180–181
- Photovoltaics, 1057–1061
  - applications of, 1060–1061
  - current applications, 1067–1068
  - early applications, 1065–1067
  - photoelectric effect, 1058
  - PV cell construction, 1057f
  - solar cell, module and array, 1059f
  - solar energy panels, *1058*
- Piezoelectric energy, **950-951**
  - definition of, 950
  - natural phenomenon, 951
  - quartz crystals and, 950, 951
- Planck, Max
  - radiation and Planck’s constant, 1035
- Pollutants
  - emission control regulations, 440–441
  - identifying damaging, 48
  - power plant smokestacks as, *5*
- Potential energy, **951–952**
  - atomic nuclear reactions, 952
  - definition of, 941
  - elastic potential energy, 952
  - electric forces and molecules, 950

  gravitational potential energy, 951–952
  loss in, 952
  types of, 952
Power, **952–954**
  bicyclists demonstrating torque, *953*
  definition of, 952
  force and, 954
  Joule, 952, 953
  measurement metrics, 954
  torque and, 953
Power plants
  pollution model of, *476*
  smokestacks as pollution source, *5*
Precipitation
  acid rain, 1–2, 2f
Pressure, **954–956**
  definition of, 954
  expanding air in balloon, *953*, 953–956
  expanding gas, 955
  hydraulic press, 955
  hydraulic principle, 954f
  measurement metrics, 955–956
  Pascal, Blaise, 955
  uniform surface pressure, 955
Princeton Plasma Physics Laboratory
  description of, 818
Propellants. *See also* Explosives and propellants, Rocket propellants
  burn rates, 495
  liquid propellants, *1021*, 1021–1023
  solid propellants, 1019–1020
  storable *vs.* cryogenic, 1021–1022
Propeller aircraft
  all-metal airplanes, 37
  cantilevered wing monoplane, 37
  history, 37–40
  propeller development, 38–39
Propellers, **956–960**
  aerial, 958–959
  aerial propeller noise, 957
  Archimedes, 956
  design and development, 38–39
  marine, 957–958
  NASA-developed propfan, 959
  on balloons, 958
  origin of, 956–957
  propeller design coefficients, 958
  steam propulsion with screw, 957
  turboprop engine, 959
  turboprops, 958
  variable-pitch propeller, 958–959
  Wright flyer, 958
Property rights, **960–965**
  common pool problems, 960–961
  environmental economics, 473–474
  homesteading and mineral rights, 962
  oil recovery, 960
  oil reservoir drives, 961
  oil reservoirs and production, 961
  production and mineral rights, 961
  Prudhoe Bay oil field, 963–964, *964*
  unitization and negotiation, 962–963, 964
  unitization contracts, 963
Propulsion, **965–968**
  basic physics of, 965–966
  comparison of, 967–968
  efficiency of, 966
  energy/work, 965–966
  flowing streams, 966, 968
  ion rocket engine, 966
  momentum/impulse, 965–966
  Newton's third law, 968
  rocket engines, 967
  sailboats and wind, 968
  satellites, 968
  solar sails, 968
  types of, 966
Propulsion systems, 967–968
  ion rocket engine, 968
  rocket engines, 968
  for spacecraft, 1069–1076
Prudhoe Bay oil field, *964*
Public transit
  automobiles *vs.,* 133–134
Public Utility Commissions (PUCs)
  Electric power industry, 407
Pulp and paper industry
  energy efficiency in, 752–753

Q

Quantum mechanics
  chemical bonding and, 804
  development of, 1034–1037

R

Radioisotopic thermoelectric generators (RTG)
  spacecraft energy systems, 1077–1078
Railroads
  double stack car, 515, 516t
  energy efficiency and, 511–513, 512t
  locomotive technology, 515, 721–731
Railway passenger service, **969–976**
  dynamics and design of, 971–972
  electrification of, 972–973
  energy efficiency, 970–973
  energy use by, 971–974
  environmental impact reductions, 974–975
  fuel consumption of, 971f
  fuel economy, 970–972
  land use of, 974
  transportation traffic patterns, 969–970
  types of railways, 971
  urban and city use of, 974
  world's rail traffic, 970t
Rankine, William John MacQuorn, **976–977**
  scientific definition of energy, 1028–1029
Recycling, *138*
  ceramics, 774
  energy cost, 139
  metal, 773
Refineries, **977–987** *See also* Lubricants and lubrication
  characteristics of, 979t
  conversion and integrated, 981f
  crude oil imports and types, 979–980
  desalting and distillation processes, 983
  environmental effects of, 986–987
  fluid catalytic cracking, 983–984, 992–994
  gasoline production, 981–982
  hydrocracking units, 985
  hydrotreating, 986
  Imperial oil refinery, *984*
  industry statistics, 987
  locations of, 978
  lubricants, 979
  operations of, 982–986
  products of, 982–983, 983t
  refining systems, 979–981
  reformulated fuel production, 986
  topping and hydroskimming, 980f, 980–981
  transportation fuel, 979–980
  U.S. crude oil production, 987
  vacuum distillation unit, 983
  visbreaking, 986
Refining
  catalytic cracking, 990>–*994*
  cracking process, 988
  distillation process, 987
  early technology, 988–989
  fluid catalytic cracking, 992–994
  history of, 987–995
  Houdry catalytic cracking, 990–991
  process overview, 987–988
  refinery industry statistics, 987
  thermofor cracking and air-lift systems, 991–992
  thermal cracking, *989*–990
Refining, history of, **987–995** *See* Refineries, Refining
Refrigerators and freezers, **995–1003**
  absorption refrigeration, 1001–1002
  compressors, 995–998
  energy consumption of, 998–999
  energy improvement history, 1000f
  energy standards, 1000–1001
  energy use of, 77f

- Refrigerators and freezers (cont'd)
  - evaporators, 1000
  - Frigidaire's first refrigerator, *997*
  - heat load of, 999–1000f
  - household use of, 26
  - National Appliance Conservation Act of 1987, 998–999
  - refrigerants and ozone depletion potential, 1000t
  - refrigeration systems, 26, 1000–1001, *1001*
  - thermoelectric refrigeration, 1002
  - vapor compression, 995–996, 996f
- Regulation and rates for electricity, **1003–1005** *See* Regulations
- Regulations
  - air pollution, 51–52
  - Clean Air Act of 1970, 3
  - National Appliance Energy Conservation Act of 1987, 998–999
  - National Energy Act of 1978, 408
  - National Energy Conservation Policy Act (NECPA), 321
  - Natural Gas Policy Act of 1978, 506
- Reitschel, Hermann
  - air conditioning systems, 25
- Renewable energy, **1005–1007**
  - agriculture use of, 19–20
  - biofuels, 1005–1006
  - energy production percentages in 1997, 1006f
  - production breakdown of, 1006f, 1006–1007
  - Renewable energy production incentive (REPI), 1197
  - solar and wind energy, 1006f, 1006–1007
  - types of, 1006–1007
- Reserves and resources, **1007–1014**
  - definitions of, 1008–1010
  - estimating, 1007–1008
  - estimation risks, 1010–1014
  - petroleum imports, 1013
  - probability ranking of reserves, 1012f, 1013f
  - reserves *vs.* resources chart, 1009f
  - risk evaluation, 1013–1014
  - Securities and Exchange Commission definitions of, 1008–1009
  - terminology of, 1011f
  - U.S. Geological Survey definitions of, 1008
  - World Petroleum Congress definitions of, 1009, 1010
- Residential buildings. *See also* Buildings, residential
  - air sealing and insulation, 204–206, 205f, 206f
  - Domestic energy use, 348–349
  - heating and cooling, 206–207, 207f
- Residual fuels, **1014–1015**
  - catalytic cracking, 1014
  - consumption of, 1016f
  - decrease in demand, 1016
  - electric power generation using, 1016
  - properties of, 1015f
  - thermal cracking effects on, 1015
- Reusable launch vehicles
  - Venture Star, 1074
  - X-33 technology demonstrator, 1074
- Richter, Charles
  - seismic wave scale, 1040–1041
- Risk assessment and management, **1017–1019**
  - energy price dynamics, 1017
  - financial risk management tools, 1018–1019
  - insurance programs, 1017–1018
  - managing risk, 1017–1019
  - price risk, 1017
  - seismic braces in nuclear power plant, *1018*
- Rocket engines
  - electric rocket engines, 1075–1076
  - ion engine on Deep Space 1, 1076
  - liquid propellant rocket engine, 1070f
  - propulsion efficiency, 967–968
  - space launch vehicles, 1071t
  - space propulsion systems, 1076t
  - thermal electric engine, 1075–1076
  - Venture Star aerospike engine, 1074
- Rocket propellants, **1019–1023**
  - basics of, 1019
  - energy efficiency of, 1021t
  - energy use during launch, 1070–1072
  - Goddard's first liquid propellant rocket, *1021*
  - liquid oxygen, 1022–1023
  - liquid propellants, 1021–1023
  - physical properties of cryogenic, 1022
  - physical properties of storable, 1021
  - solid propellants, 1019–1020
  - Space Shuttle Challenger, *1020*
  - specific impulse of, 1022t
  - storable *vs.* cryogenic, 1021–1022
- Rockets
  - floating launch platforms, 1074
  - Goddard's first liquid propellant rocket, *1021*
  - launch platforms, 1074
  - reusable launch vehicles, 1072–1074
  - Space Shuttle Challenger, *1020*
- Roller bearings, 127
- Rotary engine
  - history of, 560–561, 561f

S

- Sakharov, Andrei Dmitrievich, **1025–1027**, *1026*
- Sandia National Laboratory
  - description of, 818
- Satellites
  - communications, 279
  - energy systems, 1069–1079
- Savery, Thomas, **1027–1028**
- Savonius, Sigard
  - vertical-axis wind turbines, 1195
- Schrödinger, Erwin
  - quantum mechanics wave function, 804–805
- Scientific and technical understanding of energy, **1028–1037** *See* Energy research and history
- Securities and Exchange Commission (SEC)
  - definitions of reserves and resources, 1008–1009
- Seebeck, Thomas Johann, **1038–1039**
- Seismic energy, **1039–1042**
  - Earth exploration using, 1041–1042
  - earthquakes, 1040–1041
  - oil and gas exploration using, *1041*–1042
  - Richter scale, 1040–1041
  - seismograph recording, *1040*
  - types of, 1039
- Ships, **1042–1045**
  - diesel engines for, 1045
  - engine emissions, 1045
  - environmental problems, 1045
  - frictional resistance, 1042–1043
  - gas turbines for, 1045
  - nuclear power for, 1045
  - oil spills from, 1046
  - power requirements by type of, 1044t
  - propeller propulsion, 1043–1044, 1044t
  - propulsion for, 1043–1044
  - steam engines for, 1045
  - types of, 1042, 1044t
  - waterjets, 1044–1045
  - wave-making drag/resistance, 1042–1043
- Siemens, Ernst Werner Von, **1046–1048**
- Smeaton, John, **1048–1050,** *1049*
- Smith, Angus
  - acid rain, first use of term, 1
- Smith, Francis P.
  - modern marine propeller, 957
- Soils
  - biofuels - nutrient loss in, 164
  - nitrogen and nitrate effects on, 6

Solar energy, **1050–1063**
  active heating and cooling, *1055*, 1055–1056
  applications of, 1060–1062
  availability of, 1051–1052
  basics of, 1050–1051
  dish engine systems, 1056–1057
  heating and cooling, 1053–1056
  history of, 1052, **1063–1068**
  Hubble Space Telescope use of, 1063
  implementation and barriers, 1068
  ocean currents and, 1053
  parabolic collection dishes, *1067*
  passive direct gain method, 1054f
  passive heating and cooling, 1054f, 1054–1055
  photosynthesis and, 1052
  photovoltaics, 1057–1060
  photovoltaics applications, 1060–1061, 1065–1068
  potential and problems of, 1062
  power towers, 1056
  1882 printing press powered by, *1064*
  PV cell construction, 1057f
  radiative intensity, 1051f
  satellites use of, 1063
  solar arrays for spacecraft, 1077
  solar panels, *1055, 1058*
  solar thermal applications, 1065
  solar thermal power systems, 1056–1057
  spacecraft power supply, 1077
  trough systems, 1056
  water heating, 1214–1215
  wind and, 1052
Solar energy: historical evolution of the use of, **1063–1068**
Solar powered vehicles
  energy metrics of, *667*
Solar radiation
  atmospheric effects, 86
Southern, James
  engine performance optimization, 1031–1032
Soviet Union
  nuclear waste handling, 886
  USSR spaceborne nuclear reactors, 1077–1079
Spacecraft energy systems, **1069–1079**
  cryogenic engines, 1070
  electric rocket engines, 1075–1076
  fuel cells, 1076–1077
  future of, 1079
  ion engine on Deep Space 1, 1076
  launch platforms, 1074–1075
  liquid propellant rocket engines, 1070f
  nuclear electric systems, 1077–1079
  nuclear reactors, 1077–1079
  power supply systems, 1076–1079, 1078t
  propulsion systems, 1069–1076
  radioisotopic thermoelectric generators (RTG), 1077–1079
  reusable launch vehicles, 1072–1074
  Saturn V rocket, 1072
  solar arrays, 1077
  solar energy systems, 1077
  solid rocket motor, 1070f
  spaceborne nuclear reactors - USSR, 1079
  space launch vehicles, 1071t
  space propulsion systems, 1076t
  Space Shuttle, 1072–1074
  thermal electric engine, 1075–1076
  thermionic generators, 1078–1079
  Venture Star aerospike engine, 1074
Space Shuttle
  Challenger liftoff, *1021*
Space Shuttle (U.S.)
  Discovery with robot arm, *1073*
  main engines, 1072–1073
  solid rocket boosters, 1072
Sperry, Elmer Ambrose, **1080–1082**, *1080*
Sport utility vehicles
  emission controls, 453–454
Springs
  elastic energy and, 385–386
Stanford Linear Accelerator Laboratory
  description of, 818
Steam engines, **1082–1086**
  basics of, 1082–1083
  cogeneration with, 266–267
  combined cycling, 1086
  compound expansion engines, 1084
  engine performance optimization, 1031–1032
  impulse turbine, 1085
  marine turbine, 1085
  Newcomen engine and Watt cylinders, 1083–1084
  predecessors of, 1083
  reaction turbine, 1085
  reciprocating engine development, 1084
  steam turbines, 1084–1086
Steam production
  manufacturing use of, 753t, 753–754
Steam turbines, **1183–1188**
  early steam turbine, *1187*
  evolution of, 1184–1187, 1185f
  future role for power generation using, 1187–1188
  industry evolution, 1183
  nuclear power applications, 1187
  steam cycle evolution, 1186f, 1186–1187
  thermal performance, 1186f, 1186–1187
  turbine cycles, 1183–1184
  turbine generator unit, 1184f
Stephenson, George, **1086–1088**
Stirling, Robert, **1088–1090**
Stirling engines, **1090–1095**
  applications of, 1093–1094
  basics of, 1090–1093, 1091f, 1092f
  cryocoolers, 1095
  economiser, 1093
  energy efficiency of, 1093
  history of, 1090–1093
  low temperature difference engines, 1094–1095
  regeneration, 1093
Stolze, Franz
  gas turbine power plant, 1174f
Storage, **1095–1097** *See* Energy storage
Storage technology, **1098–1102** *See* Energy storage
Streamlined body
  drag reduction, 13
Subsidies and energy costs, **1102–1106,** 1170–1172
  alternative energy stimulation, 1104
  depletion allowances, 1105
  economics of subsidies, 1103
  energy research, 1105–1106
  external costs and fuel cycle, 1169–1170
  nature of subsidies, 1102–1103
  oil and gas pricing, 1104–1105
  preserving energy industries, 1103–1104
Sulfur oxides
  air pollutants, 51
  dioxide control technologies, *446*, 446–447
  element of acid deposition, 3–4
  low sulfur coal use, 4
  reduction at electric utilities, 2–3, 3f
Superconductor magnetic energy storage
  properties of, 1099f, 1099–1100
Superconductors
  ceramic, 1100
  magnetic energy storage, 1099f, 1099–1100
Supersonic aircraft
  Bell X-1 first flight, 43
  commercial transport, 45–46
  Concorde, 45–46, 46f
  Lockheed F-104 Starfighter, 42f, 43–44
Supply and demand
  change in supply and demand, 1110
  data sources, 1107
  demand elasticity, 1109
  electricity prices, 1111f
  energy demand and supply, 1108f
  energy prices, 1110
  energy supply determinants, 1107
  estimated elasticities, 1110

- Supply and demand (cont'd)
  - fossil fuel prices, 1111f, 1112
  - income elasticities, 1109–1110
  - market price, 1110, 1112
  - supply elasticity, 1107–1108
  - supply shift in response to technology, 1110f
- Supply and demand and energy prices, **1106–1112** *See* Supply and demand
- Sustainable development and energy, **1113–1114**
  - coal as a plentiful energy source, 1113
  - energy consumption and, 1113
  - nuclear energy as possible energy source, 1113
  - renewable energy as most sustainable source, 1113
- Synthetic fuels, **1114–1117**
  - basics of, 1114
  - conversion to, 1114–1116
  - direct liquification of coal, 1115–1116
  - donor solvent coal liquification process, 1116
  - economic and environmental outlook, 1116–1117
  - gas intermediates to liquid, 1115

T

- Taxation of energy, **1119–1122**
  - excise taxes, 1119–1120, 1120t
  - income taxation and energy demand, 1121–1122
  - income taxation and energy supply, 1120–1121
  - tax code and income, 1121
  - tax expenditures, 1122t
- Television
  - evolution of, 277, 278–279
- Tesla, Nikola, ***1122***–1124
- Tetraethylead
  - first gas additive, 549–550
- Theory of quantum mechanics
  - properties of light, 777
- Theory of relativity
  - gravity, 778
  - Matter and the speed of light, 777–778
- Thermal efficiency
  - diesel cycle engines, 332–334
- Thermal energy, **1124–1125**
  - basics of, 1124–1125
  - heat engines, 1125
  - watch powered by, *1125*
- Thermodynamics, **1125–1132**
  - basics of, 1125–1126
  - biological energy and, 166–167
  - Carnot cycle, 1128–1132, 1128f
  - Carnot cycle and refrigerators, 1131–1132
  - Carnot engine efficiency, 1130
  - conservation of energy and, 1032
  - diesel engine designed by principles of, 1033
  - heat engines, 1128–1129, 1129f, 1130–1132
- Thermodynamics laws
  - Clausius form of second law of, 1130–1132
  - first law, 1126–1127
  - second law, 1127–1132
- Thompson, Benjamin (Count Rumford), **1132–1134**
- Thomson, Joseph John, **1134–1136**, *1135*
  - electromagnetic experiments, 1035–1036
- Thomson, William (Lord Kelvin), **1134–1139,** *1137*
  - scientific definition of energy, 1028–1029
- Tires, **1139–1141**
  - history of, 1139–1140
  - manufacturing energy used for, 1141
  - pneumatic, 1140
  - rolling resistance and, 99, 1140–1141
  - rolling resistance coefficient, 1140t
- Townes, Charles Hard, **1141–1143,** *1142*
- Traditional agriculture, 16–17
  - energy use in, 20t
- Traffic flow management, **1144–1154**
  - carpool and ride share activities, 1146–1147
  - carpool lanes, 1149
  - congestion pricing theory, 1148
  - demand management strategies, 1144–1149
  - economic incentives, 1147–1149
  - emission impact modeling of, 1151–1152
  - employer-based trip reduction programs, 1145–1147
  - gasoline taxes, 1148
  - high occupancy toll (HOT) lanes, 1149
  - high occupancy vehicle (HOV) lanes, 1149
  - improvement strategies for, 1149–1151
  - intelligent transportation systems, 1151
  - parking pricing, 1148–1149
  - problems of, 1144
  - public information and education, 1147
  - ramp metering, 1151
  - rapid incident response programs, 1150–1151
  - regulatory mandates and trip reduction, 1145–1147
  - signal-timing optimization, 1149–1150
  - status and research in, 1152–1153
  - traffic jam, *1145*
- Transformers, **1154–1156**
  - balance of power in, 1155–1156
  - basics of, 1154–1155
  - power transformer, *1155*
  - types of, 1155–1156
- Transistors
  - invention of, *398*, 399
- Transmissions
  - automatic, 351–354
  - clutch schematic, 341f
  - continuously variable, 354–355
  - manual, 340–351
  - manual gearbox schematic, 341f
  - planetary gear set, 352f
- Transportation, **1156–1161**
  - aircraft, 1159, *1160*
  - air travel, 1160
  - automobile production, 1158
  - bicycles, 152–154, *153*
  - biofuels, 160–163
  - de Haviland Comet aircraft, 1159, *1160*
  - electric trolley cars, 1157–*1158*
  - energy use of, 145t
  - freight movement, 508–518
  - future of, 1160–1161
  - highway system, 1159
  - petroleum usage for, 1158–1160
  - railways and energy use, 1156–1157
  - traffic patterns, 969–970
  - travel growth trends, 1160
  - truck and trailer production, 1158–1159
- Transportation, evolution of energy use and, **1156–1161** *See* Transportation
- Trevithick, Richard, **1161–1163**
- Tribology, **1163–1167** *See also* Lubricants and lubrication
  - basics of, 1163–1164
  - energy and lubricant conservation, 1165
  - future of lubrication, 1166–1167
  - history of friction, 1165–1166, 1166f
  - history of lubricants, 1164–1165
  - research in, 1164
  - surface roughness, 1164f
- Trucks, light-duty
  - emission controls, 453–454
- Truck transportation *See also* Freight movement
  - coal transportation, 264
- True energy costs, **1167–1171** *See* Energy economics
- Tsiolkovsky, Konstantin
  - pioneered rocket technology, 1021
- Turbines, gas, **1171–1182** *See* Gas turbines

Turbines, steam, **1183–1198** *See* Steam turbines
Turbines, wind, **1188–1195** *See* Wind turbines
Two-stroke piston engine
    history of, 559–561, 560f

U

United Nations (UN)
    description of, 583–584
United States
    energy consumption, 289–290
    petroleum consumption, 945–948
Units of energy, **1197–1198** *See under* Energy use
Universe
    Big bang theory and expanding, 154–157, *155*
    composition of, 777b
    fate of the, 157
Uranium
    conversion and enrichment, 867–869, 868f
    fuel for nuclear fission, 862–863, 866–869
    mining and milling of, 866–868, 867t, 868f
    production of, 867t
Urban communities
    cool communities, 304–308
U.S. administration policy
    Department of Energy (DOE), 585f, 586–587
U.S. Department of Agriculture (DOA)
    description of, 590
U.S. Department of Commerce (DOC)
    description of, 589
U.S. Department of Defense (DOD)
    description of, 590–591
U.S. Department of Energy (DOE)
    administration policy and, 586–589
    description of, 585f, 586–589
U.S. Department of the Interior (DOI)
    description of, 590–591
U.S. Department of Transportation (DOT)
    description of, 590
U.S. Geological Survey
    definitions of reserves and resources, 1008
Utility planning, **1198–1204** *See also* Electric utilities
    deregulation of power utilities, 1202–1203
    future of, 1203–1204
    history of, 1199–1201, *1201*
    independant power producers, 1202
    scenario planning, 1203
    transition period, *1201*, 1201–1205

V

Vapor compression
    basics of, 995–996, 996f
    cycle of, 996f
    history of, 996f, 996–998, 998f
    invention of, 996–997
    in refrigerators and freezers, 996–998
Volatile organic carbon compounds (VOC)
    air pollutants, 49–50
    amounts by source, 49t
Volta, Alessandro, **1205–1207,** ***1206***
    battery development, 395
    galvanic cells, 230
Voltaic/galvanic cell
    comparison with electrolytic cell, 231f
von Braun, Werhner
    developed rocket technology, 1021

W

Walton, Ernest
    first particle accelerator, 936
Waste-to-heat technology, **1209–1213**
    mass burn technology, 1210–1211, *1211*
    municipal solid waste (MSW) fuel, 1210, 1212–1213
    refuse-derived fuel (RDF), 1209–1210
    solid waste disposal, 1209–1210
Water heating, **1213–1218**
    efficiency standards, 1216–1217
    heater costs and efficiencies, 1214t
    history of, 1213–1215
    National Appliance Energy Conservation Act of 1987, 1216
    solar water heaters, 1214–1215
    storage tank heater construction, 1215–1216
Water pollution
    environmental activities in, 479–481, *480*
Watt, James, **1218–1220**
    expansion steam engine, 1083–1084
    horsepower defined, 1030
    invention of the condenser, 1031
Waves, **1220–1224**
    electromagnetic, 1221–1222, 1223
    fiber optic cables, *1223*
    mechanical, 1220–1221
    radiation and environment, 1224
    radiation and solar energy, 1222–1223
    types of waves, 1220–1221
    wave action example, 1222f
Weather
    electric power system reliability, 425–427
    lightning, 91–92
    thunderstorms, 89–90
    tornadoes, 90–91
    tropical storms, 89
    winds, 92
Weather protection
    residential buildings, 208–209
Wheatstone, Charles, **1225–1226**
Whitcomb, Richard
    area rule, 44
Whittle, Frank
    jet aircraft, 40–41
Wind energy
    wind speed, 92–94, *93*
Windows, **1226–1236**
    advances in window energy efficiency, 1228–1230
    Chateau de Versailles, *1227*
    choosing efficient, 1231–1232
    commercial buildings, 1231–1232
    controlling heat loss, 1232
    controlling solar heat gain, 1232–1234
    current window design, 1230–1232
    energy basics of, 1227–1228
    energy-saving house design, *1229*
    future of, 1232–1235
    history of, 1226–1227
    infiltration rate, 1228
    integrating with other building components, 1234
    solar heat gain coefficient, 1228
    U-factor coefficient, 1228
    visible transmittance, 1228
    warm edge design, 1229–1230
Wind speed
    averages in U.S., *93*
    wind energy, 92–94, *93*
Wind turbines, **1188–1195**
    basics of, 1190f
    economics of, 1194–1195
    history of, 1188–1190
    Horizontal-axis wind turbines (HAWT), 1191–1192
    levelized cost of energy (LCOE), 1194–1195
    Renewable energy production incentive (REPI), 1194
    site selection for, 1193–1194
    small scale applications, 1193
    turbine designs, 1190–1193
    U.S. power capacity of, 1191f
    utility scale applications, 1193
    vertical-axis wind turbines (VAWT), *1192*–1193
    vertical windmill, *1189*
Work and energy, **1236–1237**
    definition of work, 1236
    heat and, 1236
    kinetic energy and, 1236–1237
World Bank
    description of, 584–585

World Petroleum Congress
  definitions of reserves and resources, 1009–1010
Wright, Orville, 31, *31*, 32
Wright, Wilbur, 31, *31*, 32
Wright Flyer, *31*
  aerodynamics, 34–35
  design, 34–36
  propellers, 958
  propulsion, 34
  structure of, 36

Y

Yeager, Capt. Charles (Chuck), 43
Yucca Mountain, Nevada
  nuclear waste storage site option, 883–884

Z

Zinc-manganese dry cells, 118–120
  cross section of Leclanché cell, 119f
Zon, Michael, 930, 931