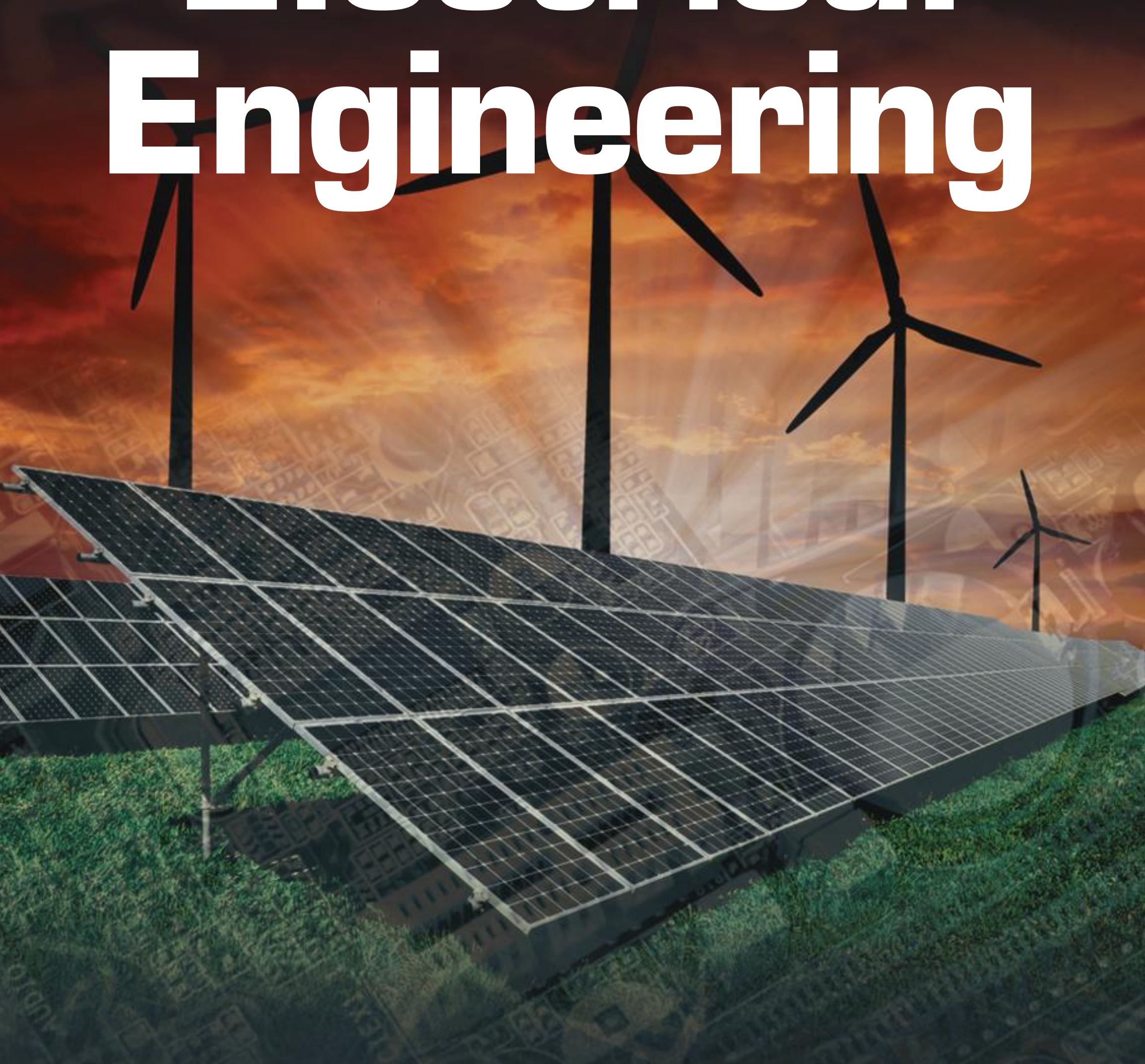
FUNDAMENTALS OF

Electrical Engineering



Charles A. Gross
Thaddeus A. Roppel

CRC Press
Taylor & Francis Group

FUNDAMENTALS OF

Electrical Engineering

Charles A. Gross
Thaddeus A. Roppel

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20120110

International Standard Book Number-13: 978-1-4398-9807-9 (eBook - PDF)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

To Dodie and Tammy

This page intentionally left blank

Contents

Preface

Contemporary engineering requires that practitioners specialize. There is simply too much to know for any one individual to deal with all of the technical aspects of most engineering projects. Still, real-world engineering problems are almost never neatly divided into mechanical, electrical, chemical, civil, and other categories. Thus, there is a need for all engineers to have at least a basic knowledge of the full spectrum of specialties throughout the engineering profession, not to mention consideration of economics and environmental, political, and social issues. It is paradoxical that while we need an ever-increasing depth of knowledge in our particular field, we also need an ever-broadening general education.

This book presumes to present the fundamentals of electrical engineering. It is recommended to those who seek an understanding of the fundamentals of electrical engineering, separated from important, but nonessential, digressions into detailed technical analyses that are essential to in-depth specialized studies. This is not to say that the material is written in the "EE for dummies" style, which is so popular in contemporary writing. The fundamentals are treated with the same scientific and mathematical rigor as one would rightfully expect in any engineering textbook. It is just that many topics which are necessary for one to learn in order to claim the title of "expert" are intentionally omitted. The book is intended as the beginning, and not the end, of EE study.

Of necessity, the content is colored by the judgment of the authors as to just what these "fundamentals" are, based on their personal experience with a combined 73 years of teaching and practice. Still, a conscious effort has been made to include the many suggestions and constructive criticisms of colleagues within and outside of electrical engineering.

In the context of contemporary engineering curricula, this book is recommended as a textbook for two types of courses:

- A Fundamentals of Electrical Engineering course which would be taken by EE majors and offered at institutions which prefer to give their students an overview of EE before in-depth EE courses are required. One benefit of such a course is that much of the time spent on the fundamentals in specialty courses could be then reduced.
- A course which would be taken by non-EE majors to provide an understanding of EE fundamentals and a general appreciation of just what EE is all about.

Outside of academia, the book is recommended to anyone, but especially practicing engineers of all specialties, who seek an understanding of the fundamentals of electrical engineering. The book assumes an engineering background on the part of readers, with appropriate undergraduate competencies in science and mathematics. On the other hand, an understanding of advanced mathematics and science is not required.

There are two fundamental applications of electromagnetic phenomena:

- The processing and transmission of energy
- The processing and transmission of information

Thus, the book is divided into two parts:

- Fundamentals of electrical energy processing (Chapters 1 through 5)
- Fundamentals of electrical information processing (Chapters 6 through 9)

These two parts match the authors' spheres of expertise (Professor Gross the former, and Professor Roppel the latter). The case for the study of EE is strong. The necessity to control energy is essential to modern civilization, and energy in the electrical form is an important part of that. The capabilities and applications of electronic information processing surpass even the imaginations of most science fiction writers.

A proposed catalog description of a course for which this book would serve as an appropriate textbook might be as follows:

> **Electrical Engineering Fundamentals** Prerequisites: engineering physics and mathematics. A presentation of the fundamentals of electrical engineering.

Part I, fundamentals of electrical energy processing, includes a comprehensive discussion of electrical energy, generators, motors, transformers, and loads, with several contemporary applications such as electrical vehicles and wind machines. Part II, fundamentals of electrical information processing, includes a discussion of electrical components such as transistors, diodes, and operational amplifiers, with contemporary applications such as robotics.

The table of contents serves as a recommended first draft of a course syllabus, and is essentially self-explanatory. Instructors should be able to tailor this outline to their individual needs. Appendix A provides a convenient reference for SI units and selected conversion factors. Appendix B provides data for the SN7400 series TTL NOR Gate.

Experienced engineering educators know that students need to solve relevant problems to solidify their understanding of technical material. End-of-chapter problems are provided for that purpose. Each problem has its own unique set of educational objectives, and is intended to reinforce and augment the material of that chapter, as well as provide continuity to earlier material in previous chapters. Therefore, it is recommended that all problems be assigned. If it is desired to change the answers from term to term to prevent students from copying solutions from files, it is a simple matter to change certain key data elements to generate a different set of answers to every problem.

We agree that the order of authorship in no way ranks the relative contributions by each author. This was a mutual effort, with both authors contributing equally to the work. Furthermore, no book is entirely the product of two individuals, as certainly is the case with this work. Dozens of individuals have contributed in many ways, including, but not limited to, Chris Roberts, Tom Burch, P. K. Raju, Bruce Tatarchuk, Ram Gupta, and Sandy Johnson. The cooperation and support of many companies and institutions, including General Electric, the Alabama Power Company, the Southern Company, Reliance Electric, Rockwell, ABB, and Siemens, are gratefully acknowledged. Finally, the sacrifices made by our families, but particularly our wives Dodie Gross and Tammy Roppel, are gratefully noted.

This page intentionally left blank

The Authors

Charles A. Gross is presently Professor Emeritus of Electrical and Computer Engineering at Auburn University where he continues to teach. His technical areas of interest include electrical power and control systems; energy conversion processes; and power system protection. Dr. Gross has over forty years teaching experience, including two years as Visiting Professor at the United States Military Academy, at West Point. He also has significant industrial experience, including work with the US Corps of Engineers; the Tennessee Valley Authority; City of Los Angeles Water and Power; Southern Co. Services; Alabama Electric Cooperative; General Electric Company, EPRI, and ABB Corporation. Dr. Gross is a former USAF officer, a Life Fellow of IEEE, and a registered professional engineer.

Professor Gross has numerous teaching awards, including the national IEEE-PES Outstanding Power Engineering Educator Award for 2001. The undergraduate electrical engineering teaching laboratories at Auburn were named after Dr. Gross in 2011. He has contributed to over a hundred technical publications, including his textbooks *Power System Analysis, 2nd Ed.* (Wiley, 1986) and *Electric Machines* (CRC Press, 2007).

Thaddeus Roppel is Associate Professor in the Electrical and Computer Engineering Department at Auburn University. He earned the PhD degree in Electrical Engineering from Michigan State University. He is a member of the Institute of Electronic and Electrical Engineers, the American Society for Engineering Education, and the Electrical Engineering Honor Society, Eta Kappa Nu. While at Auburn, he has earned numerous awards for undergraduate teaching and mentoring, including the Pumphrey Award for outstanding teacher in the College of Engineering and the Spencer Creative Mentoring Award. He has served as faculty advisor to the Electrical Engineering Honor Society and the Student Projects Incubator, and has directed numerous senior design projects and honors theses. He has served on the ECE Department's curriculum committee for many years. His research has covered topics ranging from low-temperature semiconductor fabrication to artificial neural networks and fuzzy logic. He currently leads the cooperative robotics research laboratory, which has the goal to develop teams of biomimetic robots using vision as a primary means of navigation.

Chapter 1

What Is Engineering?

Joe finishes the last of his soft drink and drops the empty can into a recycling bin. As he turns to leave, he pauses and retrieves the can. Examining it closely, he reflects on the object. Here is a nearly perfect aluminum container, decorated and with clear informative lettering, hermetically sealed until the owner opens a cleverly designed hatch. An amazing design, balanced between form and function, that would have been an object of great interest and admiration two hundred years ago. In an epiphany, Joe realizes that this is but one of the hundreds of manufactured items Joe will encounter in a typical day: some with elegant simplicity, others with mind-boggling complexity. How have we managed to surround ourselves with such a vast array of objects, without which human existence would be extremely different or, in some cases, impossible?

The readers of this book are either engineering students or practicing engineers, or at least have an interest in engineering. But just what is this area of human endeavor we call “engineering”? Thomas Tregold offered a definition in 1828:

> Engineering is the art of directing the great sources of power in nature for the use and convenience of man.

The American Engineers’ Council for Professional Development (ECPD) has defined engineering as follows:

> The creative application of scientific principles to design or develop structures, machines, apparatus, or manufacturing processes, or works utilizing them singly or in combination; or to construct or operate the same with full cognizance of their design; or to forecast their behavior under specific operating conditions; all as respects an intended function, economics of operation and safety to life and property.

The Preamble to Engineers Code of Ethics[1] captures the essence of the profession:

Preamble

> Engineering is an important and learned profession. As members of this profession, engineers are expected to exhibit the highest standards of honesty and integrity. Engineering has a direct and vital impact on the quality of life for all people.

[1] From the National Society of Professional Engineers (NSPE) Web site. See http://www.nspe.org/Ethics/CodeofEthics/index.html for the full text of the code.

> Accordingly, the services provided by engineers require honesty, impartiality, fairness, and equity, and must be dedicated to the protection of the public health, safety, and welfare. Engineers must perform under a standard of professional behavior that requires adherence to the highest principles of ethical conduct.

We offer a more concise statement. Engineering is...

> the application of science, mathematics, and technology to the design, fabrication, and use of practical devices and systems.

Still, no definition completely captures the full nature of engineering. Perhaps Samuel Florman said it best:

> These, then, are what I take to be the main elements of the engineering view: a commitment to science, and to the main elements of science, and to the values that science demands—independence and originality, dissent and freedom and tolerance; a comfortable familiarity with the forces that prevail in the physical universe; a belief in hard work, not for its own sake, but for the quest for knowledge and understanding and in the pursuit of excellence; a willingness to forego perfection, recognizing that we have to get real and useful products "out the door"; a willingness to accept responsibility and risk failure, a resolve to be dependable; a commitment to social order, along with a strong affinity for democracy; a seriousness that we hope will not become glumness; a passion for creativity; a compulsion to tinker; and a zest for change.[2]

In a word, engineers are problem solvers. Engineers use science, but are not scientists; use mathematics, but are not mathematicians; and use technology, but are not technicians. They solve problems related to the betterment of the human condition and the larger ecosphere. Practicality and safety are critical requirements. A properly engineered solution must be economical in the use of material, financial, and temporal resources.

Although engineering is truly a global profession, the scope of this book confines our discussion primarily to its development in the United States. The goal of this chapter is to acquaint the reader with some of the more important historical milestones in engineering, with particular emphasis on electrical engineering issues.

To know where we're going, and why things are as they are, we must have some understanding of where we came from, and who got us here. For example,

[2] Samuel Florman, *The Civilized Engineer* (New York: St. Martins Press, 1987), 76–77.

the standard spacing of 4 feet 10 inches between rails on railroads came to pass because this was the spacing between wagon wheels and, still earlier, the spacing between chariot wheels used throughout the Roman Empire. Like science, engineering is by its nature evolutionary. Hence, it is appropriate that we begin with a brief history of engineering.

1.1 A Brief History of Engineering

Since feeding itself has always been a primal concern of humanity, it is not surprising that this activity was at the beginnings of engineering. Tool making was one of the earliest technological developments, including knives, scrapers, spears, and clubs for hunting and gathering. As rudimentary farming began, plows and other tools were developed. Around 3500 BCE,[3] cities began to form. An early center of urbanization was the Tigris–Euphrates Valley, which included the cultures of the Assyrians and the Babylonians. As people moved from the farm to the city, it became necessary to develop more efficient farming techniques, including irrigation and transportation systems. Cities require public structures, and water delivery and sanitation systems, which provided the beginnings of civil engineering.

Flood control in the Tigris–Euphrates Valley was also of great concern. The biblical account of the great flood may have been linked to this problem, inspiring the development of dikes, levees, canals, and storage basins. Such early engineering projects are documented by records on ancient clay tablets.

In the Nile Valley, Egyptian civilization began to rise as early as 4000 BCE. The great pyramids, constructed around 2600 BCE, stand in mute testimony to Egyptian prowess in structural engineering. It is clear that the Egyptians were adept at quarrying sandstone, limestone, and granite. Later, around 600 BCE, the use of metals, including iron, nickel, and bronze, was developed.

One can argue that the construction of the Great Wall of China was one of the greatest engineering feats of all time. Beginning around 300 BCE, completed in the 1600s CE, and extending over 1500 miles, it is said to be the only manmade structure to be visible from outer space. The Great Wall is as much as 50 feet high, 26 feet wide at the base, and 16 feet wide at the top. It has over 40,000 watchtowers along its length. A volume of material in excess of

[3] "CE" ("common era") has replaced the older notation "AD" ("anno Domini"), as has "BCE" ("before the common era") replaced the notation "BC" ("before Christ"). The meanings are the same: 100 CE = 100 AD and 100 BCE = 100 BC.

100 great Egyptian pyramids was used in its construction, and a titanic investment of manpower was required.

Greek technology developed as early as 4000 BCE, and later overlapped that of the Romans. Athough the Greeks were primarily noted for their contributions in philosophy, science, and mathematics, their engineering achievements were not insignificant, including inventions such as gears, screws, a water clock, a water mill, catapults, and primitive steam engines. Their architectural achievements are astounding, including such magnificent structures as the Parthenon (447 BCE) and the Temple of Apollo (540 BCE). One particular noteworthy Greek invention, the Antikythera computing device, is said to be the most complex instrument of antiquity. It contained thirty-two gears, which correlated solar days to motions of the moon to an accuracy of one part in 86,000. It also tracked the position of the earth within the zodiac, and is thought to have located the celestial position of the known planets as well.

But no ancient society demonstrated the full range of engineering achievements better than the Romans. Although many Roman inventions were improvements on those created by Greek, Egyptian, and other cultures, the Romans improved and deployed them far beyond earlier work. Their road-building skills are legendary, and some highways still exist today. Structures such as the Colosseum, the Circus Maximus, and the Pantheon are testimony to their construction expertise. "Modern" conveniences such as running water, sanitary sewer systems, and centrally heated buildings were in existence in Rome at the time of Christ.

Roman bridges were spectacular for the period, many setting new records in bridge building. One Roman bridge, Trajan's Bridge, was longest in the world for over 1000 years. Roman bridges were usually made of stone, featuring elegant arches and use of concrete.

Roman aqueducts, canals, holding basins, and dams were built throughout the empire to meet the water needs of communities. The aqueducts in particular were the crown jewels of Roman engineering. Sextus Frontius, water commissioner of Rome in the first century CE, proudly wrote, "Will anybody compare the idle pyramids, or those other useless, though renowned, works of the Greeks with these aqueducts, these indispensible structures?" The water available to the citizens of Rome compares favorably to that available to the citizens of present large cities.

The Romans also used advanced technology in mining. Gold, copper, and iron ore were mined, as well as tin, silver, zinc, mercury, arsenic, and lead.

Water and hydraulic mining techniques were extensively used, again using aqueducts to bring water to the mining site. It is highly likely that they also developed water-powered crushing mills to crush the ore.

Military engineering has always been a major application. In Roman times, this included the development of catapults and siege machines, and the mass production of swords, spears, shields, bows and arrows, chariots, and wagons. The Romans were adept at the construction of fortified camps, roads, and bridges.

After the fall of Rome, engineering progress in the West slowed, but it did not cease altogether. There was some engineering progress in structural design and in the design of labor-saving machines and devices. The medieval Gothic cathedrals, with their stained glass windows, high central arches, and thin walls supported by flying buttresses, demonstrate sophisticated engineering and construction expertise. Fortified homes of wealthy landowners (or "castles") were constructed, exhibiting features such as moats, drawbridges, towers, and crenellated walls. Wind- and water-powered mills were developed and improved. Advances were made in ship building, including the hinged rudder. Other mechanical advances that appeared in Europe during medieval times include the spinning wheel and a hinged rudder for ships.

The word "engineer" seems to have been first used during the Middle Ages (circa 1000–1200 CE), derived from the Latin term "in generare," meaning "to create." Thus, the person who created or designed engines came to be known as the "ingeniator" or "engine-er."[4] Many of the medieval engineering advances were introduced in the Far East, including the invention of gunpowder, paper making, and textile manufacturing.

As the so-called Dark Ages, as the Middle Ages are sometimes called, drew to a close, and the "Renaissance," or rebirth, began from the fourteenth century onward, science and technology were stimulated by the work of several of the finest minds in all of recorded history. Perhaps the greatest engineer of all time, Leonardo da Vinci (1452–1519), provided a stream of inventions and machines, many of which were far ahead of their time. Da Vinci was a genius in several areas, including science, art, architecture, and medicine. Earlier, the invention of the printing press by Johannes Gutenberg (c. 1398–1468) in 1450 had a major impact on scientific progress.

[4] See http://uconngalileo.engr.uconn.edu/engineering/engineering.php?subtitle=1&part=5 from Paul Wright's *Introduction to Engineering*.

Major advances in science were contributed by Nicolaus Copernicus (1473–1543), Robert Boyle (1627–1691), and Robert Hooke (1635–1703).

This period of scientific genius was capped by the work Galileo (1564–1642), the first physicist to link mathematics to physics in a rigorous way. It is interesting to note that Sir Isaac Newton (1642–1727) was born the year Galileo died. Some argue that the genius and spirit of Galileo were passed on to Newton. Newton is generally recognized to be among the, if not *the*, greatest physicist of all time. His work remains the foundation of statics and dynamics even today and stood as our best understanding of physical mechanics until refined by Einstein's relativistic concepts in the twentieth century.

Engineering continued to evolve, typified by people like Thomas Newcomen, who built a practical steam engine in 1712, some seventy-five years before it was replaced by James Watts's device. Progress continued in manufacturing, mining, and transportation. Experimentation with using engines to drive boats were made, culminating in Robert Fulton's *Clermont* in 1807. Less than twenty years later, George Stephenson demonstrated the feasibility of a steam-powered railroad locomotive.

River and lake systems have been used as natural arteries of transportation since ancient times. However, rivers do not always go along optimum routes. Extensive canal systems were developed to augment natural waterways throughout the nineteenth century, with perhaps the most famous American example being the Erie Canal. At the time the Erie Canal was said to have been the greatest (American) engineering achievement of the nineteenth century. The Erie Canal was completed in 1825, running from Buffalo to Albany, New York, and linking up with the Hudson River to provide a water route to New York City. The canal was some 363 miles long, 40 feet wide, and 4 feet deep. Overhead aqueducts were used to allow streams to cross the canal. The canal had eighty-five locks to manage a change in elevation of 500 feet between Buffalo and Albany.

The project was funded by the State of New York at a cost of $7 million. The state was reimbursed within nine years from tolls along the canal. The Erie Canal had a dramatic effect on the region's economy. For example, shipping rates from Buffalo to New York dropped from $100 per ton to $10 per ton. With the canal completion, the population of western New York grew dramatically. In the following decades, many improvements and repairs were made. By 1862, the canal had been widened to 70 feet and deepened to 7 feet, and minor changes were made in its route.

In addition to canals, great progress was made in roads and railways throughout the nineteenth century. Railroad mileage increased from 35,000 miles at the end of the Civil War to nearly 200,000 miles in 1900. In road building, the most famous road builder of this era was John Macadam (1756–1836) from Scotland, who developed a method of road construction by compacting layers of broken stone (known as a "macadamized" road).

And then advancement began on an entirely new front. Since antiquity, electrical and magnetic phenomena had been observed in nature. As early as 650 BCE, the Greek philosopher Thales of Miletus had observed "natural" magnetism in certain rocks ("lodestones"). Lightning was (and is) spectacular and well known, if not well understood. Thales had also discovered that amber rubbed with animal fur attracted light objects. Chinese generals during the Han Dynasty (206 BCE–220 CE) were known to have used lodestones as compasses. Starting in the late eighteenth and throughout the nineteenth centuries, scientific progress in electricity and magnetism accelerated, including such notable events as the following:

- 1752: Benjamin Franklin's kite experiment demonstrating that lightning was a form of static electricity.
- 1820: Hans Oersted's work in electromagnetism.
- 1821: Michael Faraday's demonstration of a primitive electric motor, and the process of electromagnetic induction in 1831.
- 1826: André-Marie Ampère's work connecting magnetic fields to electric current.
- 1826: Georg Simon Ohm presents his famous law.
- 1827: Alessandra Volta devised the first electric battery.
- 1830: Sir Humphrey Davy discovers electromagnetism. Davy had also demonstrated a working electric arc light in 1809.
- 1834: Thomas Davenport invents the first dc electric motor.
- 1880: Thomas A. Edison invented a practical incandescent bulb and discovered that lamps could be connected in parallel, permitting one or more to be turned off without disconnecting the whole system.
- 1882: Edison's Pearl Street electricity-generating station was placed in operation in New York City.
- 1888: Nikola Tesla secures patents for an induction motor and for a new polyphase alternating current system.

- 1888: After organizing the Westinghouse Electric Company in 1886, George Westinghouse was granted a contract to provide generators for the Niagara hydroelectric project, the first such project in history.

In addition to these, dozens of others made significant contributions to the development of various types of motors and electric light sources throughout the nineteenth century.

It seems appropriate at this point to define "electrical engineering." Recall our definition of engineering:

> Engineering is the application of science, mathematics, and technology to the design, fabrication, and use of practical devices and systems.

Continuing, we define electrical engineering (EE) as follows:

> Electrical engineering is that branch of engineering that primarily uses electromagnetic science and technology to achieve the engineering mission.

In contemporary academic parlance, "EE" has been largely replaced by "ECE" (electrical and computer engineering), but this is mainly for commercial reasons. Academic areas of specialization tend to become territorial, and EE's want to stake out computers as being on their turf. One can argue that the electronic computer is an extremely important, if not the most important, product of EE. Still, it easily fits within the scope of classical EE, and doesn't really warrant a name change of the discipline, at least not on intellectual grounds. Also, the term "wireless" has come into common usage, referring mainly to broadband communication. Again, wireless issues have traditionally been included in classical EE. Bear in mind that when the authors use the term "EE," it is inclusive of all things electromagnetic.

Starting in the latter half of the nineteenth century, practitioners of engineering began to recognize the need to organize themselves as a formal profession. One definition of a profession is an occupation requiring specialized education in a field that necessitates a measure of public trust. Actually, this need was recognized much earlier, resulting in trade organizations, or guilds, requiring some certification of competency for participants. It is important that those entering engineering be somewhat familiar with its professional attributes.

1.2 Engineering as a Profession

Since most engineering activities impact directly on public safety and well-being, and require skills and knowledge well beyond the abilities of laypeople, there is a clear need to restrict its practice to individuals with the necessary competencies and accountabilities. To deal with the situation, it was decided to form professional engineering societies.

One of the earliest professional societies was the American Society of Civil Engineers (ASCE) founded in 1852, followed by the founding of the American Society of Mechanical Engineers (ASME) in 1883 and the American Institute of Electrical Engineers (AIEE) in 1884. A little more than a decade later, in 1908, the American Institute of Chemical Engineers (AICE) came into being.

A partial listing of contemporary American engineering societies includes the following:

- American Academy of Environmental Engineers
- Audio Engineering Society
- American Institute of Aeronautics and Astronautics
- American Institute of Chemical Engineers
- American Railway Engineering Association
- American Society for Engineering Education
- American Society of Agricultural and Biological Engineers
- American Society of Civil Engineers
- American Society of Heating, Refrigerating and Air-Conditioning Engineers
- American Society of Mechanical Engineers
- American Society of Naval Engineers
- American Society of Plumbing Engineers
- American Society of Test Engineers
- ASM International: The Materials Information Society
- Engineering Society of Buffalo
- Association for Computing Machinery
- Institute of Biological Engineering
- Institute of Electrical and Electronics Engineers
- Institute of Industrial Engineers
- Institute of Transportation Engineers

- National Academy of Forensic Engineers
- National Society of Professional Engineers
- Order of the Engineer
- Society of American Military Engineers
- Society of Automotive Engineers
- Society of Broadcast Engineers
- Society of Fire Protection Engineers
- Society of Hispanic Professional Engineers
- Society of Manufacturing Engineers
- Society of Marine Port Engineers
- Society of Naval Architects and Marine Engineers
- Society of Petroleum Engineers
- Tau Beta Pi Engineering Honor Society

By 1912, many engineers within the AIEE felt that the institute was too specialized on electric power engineering, and felt a need to focus on the increasing development of electrical communications. A new organization, the Institute of Radio Engineers (IRE), was formed to meet that need and coexisted with the AIEE for the next fifty years. Because of many overlapping interests, and the fact that many electrical engineers were forced to hold joint memberships, the organizations decided to merge into the Institute for Electrical and Electronic Engineers (the IEEE). The IEEE has since become the largest professional society in the world, with a current membership of over 300,000. IEEE Societies include the following:

Aerospace and Electronics
Antennas and Propagation
Broadcast Technology
Circuits and Systems
Communications
Components, Packaging and Manufacturing Technology
Computational Intelligence
Computer
Consumer Electronics
Control Systems
Dielectrics and Electrical Insulation
Education
Electromagnetic Compatibility

Electronic Devices
Engineering in Medicine and Biology
Geoscience and Remote Sensing
Industrial Electronics
Industry Applications
Information Theory
Instrumentation and Measurement
Intelligent Transportation Systems
Lasers and Electro-Optics
Magnetics Society
Microwave Theory and Techniques
Nuclear and Plasma Sciences
Oceanic Engineering
Power Electronics
Power Engineering
Product Safety Engineering
Professional Communications
Reliability
Robotics and Automation
Signal Processing
Social Implications of Technology
Solid-State Circuits
Systems, Man and Cybernetics
Technology Management
Ultrasonic, Ferroelectrics and Frequency Control
Vehicular Technology

In addition to the societies that focus on technical specialties, there are those which have universal appeal to all engineers. The National Society of Professional Engineers (NSPE) and the American Engineers' Council for Professional Development (ECPD) are two such organizations. The ECPD was founded in 1932 to supply guidance to engineering students, to assist engineers in professional development, to promote public recognition of engineering achievements, and to accredit engineering curricula. As the accreditation role proved to be the most important, the ECPD was renamed the Accreditation Board for Engineering and Technology (ABET) in 1980. In 2005, the Accreditation Board for Engineering and Technology formally changed its name to ABET, Inc. Today, ABET is recognized as the primary accrediting agency for all engineering educational programs throughout North America.

Since the relationship between the public and practitioners of any profession requires a degree of trust, professions recognize the need to articulate a code of ethics. All of the engineering societies have such codes, and all express similar convictions. Typical of these is the IEEE Code of Ethics:[5]

> We, the members of the IEEE, in recognition of the importance of our technologies in affecting the quality of life throughout the world, and in accepting a personal obligation to our profession, its members and the communities we serve, do hereby commit ourselves to the highest ethical and professional conduct and agree:
>
> 1. to accept responsibility in making decisions consistent with the safety, health and welfare of the public, and to disclose promptly factors that might endanger the public or the environment;
> 2. to avoid real or perceived conflicts of interest whenever possible, and to disclose them to affected parties when they do exist;
> 3. to be honest and realistic in stating claims or estimates based on available data;
> 4. to reject bribery in all its forms;
> 5. to improve the understanding of technology, its appropriate application, and potential consequences;
> 6. to maintain and improve our technical competence and to undertake technological tasks for others only if qualified by training or experience, or after full disclosure of pertinent limitations;
> 7. to seek, accept, and offer honest criticism of technical work, to acknowledge and correct errors, and to credit properly the contributions of others;
> 8. to treat fairly all persons regardless of such factors as race, religion, gender, disability, age, or national origin;
> 9. to avoid injuring others, their property, reputation, or employment by false or malicious action;
> 10. to assist colleagues and co-workers in their professional development and to support them in following this code of ethics.

It is clear then that the engineering profession ranks with medicine, law, and the arts as essential to a civilized society.

---

[5] Approved by the IEEE Board of Directors, February 2006, and accessible from the IEEE Web site, http://www.ieee.org.

1.3 Engineering Education

Specialized education in the engineering arts has been in existence since ancient times. However, institutions exclusively dedicated to engineering are of relatively recent origin. One of the first "engineering" schools in the West was the École Polytechnique in Paris, established in 1794. The École Polytechnique became a military school under Napoleon in 1804, and its faculty featured some of the finest mathematicians and engineers of the time, including such notables as Jean-Baptiste Biot and Jean Baptiste Fourier.

Engineering education spread to the United States with the founding of the U.S. Military Academy at West Point, New York, in 1802. Sylvanus Thayer, an honor student at Dartmouth, accepted an appointment to West Point by President Thomas Jefferson and graduated in 1808. After Thayer had served a short but distinguished military career, President James Monroe appointed Thayer to serve as the USMA's superintendent in 1817. Thayer proved to be an outstanding choice, and served with distinction until 1833. Thayer firmly established USMA's reputation as a prestigious engineering school, and became known as "the Father of West Point."

From West Point, engineering spread to a number of schools, mainly in the Northeast. In 1867, Sylvanus Thayer had left an endowment to his alma mater to establish the Thayer School of Civil Engineering. The *Dartmouth Engineer* (the magazine of the Thayer School of Engineering), in its spring 2006 issue, recounts the interesting history of the college. A young Robert Fletcher was selected by Thayer and Dartmouth President Asa Smith to serve as its first director (and only professor) at age twenty-three. Fletcher and Thayer designed the first curriculum, consisting of some fourteen courses, all of which Fletcher had to become proficient in. In the fall of 1871, the school opened. After Thayer's death in 1872, Fletcher shouldered the full responsibility for upholding Thayer's goals, which intended that Dartmouth was to become the foremost engineering school in the country.

Fletcher revised the curriculum in 1879, and except for the inclusion of electrical engineering courses, and the updating of courses, it was basically intact until 1918. Early on, Fletcher typically taught four courses per term, as well as administering the program, for a modest annual salary of $2,500. He served for the next thirty-nine years without a raise. Even after his retirement at age seventy, Fletcher continued to serve the program until his death at age eighty-eight in 1936, ending a truly remarkable career!

Civil engineering was offered at Rensselaer Polytechnic Institute as early as 1828 and the University of Virginia by 1833. Norwich University was probably in the business even earlier. But it was the Morrill Act, signed into law by President Abraham Lincoln on July 2, 1862, that established the notion of the land grant college. The purpose of the land grant colleges was as follows:

> without excluding other scientific and classical studies and including military tactic, to teach such branches of learning as are related to agriculture and the mechanic arts, in such manner as the legislatures of the States may respectively prescribe, in order to promote the liberal and practical education of the industrial classes in the several pursuits and professions in life.[6]

At first excluded, Confederate states were included after they were readmitted into the Union after the Civil War. The provisions of the Morrill Act essentially guaranteed that land grant schools would have colleges of engineering. The first land grant college was Iowa State University at Ames, Iowa.

At most institutions, the study of electricity was considered to be a subfield of physics until in the latter part of the nineteenth century, when independent programs of study in electrical engineering began to form. Cornell University claims to have offered the first course of study in electrical engineering in the world in 1883, and the University of Missouri subsequently established the first department of electrical engineering in the United States in 1886. Today, the number has grown to well over two hundred.

The American Society for Engineering Education (ASEE), founded in 1893, is an organization dedicated to promoting and improving engineering education. It provides an excellent resource for up-to-date information on engineering education at American institutions of higher learning. Through the years, engineering educational programs in the United States have earned a reputation as among the finest in the world.

1.4 Standards and Codes

Anyone who has used metric and traditional socket sets appreciates the need for standards. It is critical that those who make the nuts and those who make the sockets agree on standard sizes. A standard is a rule or measure (either minimum or optimum) designed to insure quality or level of performance.

[6] U.S. Code, Title 7.

Standards are produced and/or adopted by national standards organizations, professional associations or societies, international standards organizations, and/or private companies. In addition to the professional societies, some of the more important organizations that have direct impact on engineering standards include the following:

The International Organization for Standardization (ISO) is a network of the national standards institutes that includes most of the world's countries. Although the ISO is a private sector organization, many of its standards are mandated by some member governments.

The American National Standards Institute (ANSI) has served as a private sector voluntary standardization system for more than ninety years. Founded in 1918 by five engineering societies and three government agencies, the institute remains a private, nonprofit membership organization supported by a diverse constituency of private and public sector organizations.

The National Institute of Standards and Technology (NIST) is an agency of the U.S. Department of Commerce. NIST was founded in 1901 as the nation's first federal physical science research laboratory.

ASTM International, originally known as the American Society for Testing and Materials (ASTM) and founded in 1898, is one of the largest voluntary standards development organizations in the world and a source for technical standards for materials, products, systems, and services.

The National Fire Protection Association (NFPA) is a nonprofit organization which provides consensus codes and standards, research, training, and education directed at reducing the impact of fire and other hazards on the quality of life. It is recognized as the world's leading advocate of fire prevention and is an authoritative source on public safety.

One of the most important NFPA publications is the National Electrical Code (NEC), or NFPA 70. The NEC is a U.S. standard for the safe installation of electrical wiring and equipment, and is commonly mandated by state or local law, as well as in many jurisdictions outside of the United States.

The National Electrical Manufacturers Association (NEMA) is the leading trade association in the United States representing the interests of industrial manufacturers of products used in the generation, distribution, and utilization of electrical energy.

The reader is advised that there may be other agencies that impose codes and standards on engineering in a particular area of specialization. Also, one's employer may impose certain engineering design practices. So, while engineering is a very creative pursuit, the engineer must also be aware that he or she is always constrained by relevant codes, standards, and laws.

1.5 Use of Numbers in Engineering

Mathematical analysis is an essential part of engineering and, as such, requires the use of numbers. In the base 10 number system, ten characters, or digits, are used to represent all numbers, and are defined by enumeration (counting) from zero to nine. The table below lists the digit set for three number systems.

Base	Description	Digit Set
2	binary	0 1
10	decimal	0 1 2 3 4 5 6 7 8 9
16	hexadecimal	0 1 2 3 4 5 6 7 8 9 A B C D E F

In base 10 systems, continuing the count past nine uses positional notation, such that "163" means one hundred (1) plus six tens (6) plus three ones (3). This scheme is continued indefinitely to form the set of positive integers (the "whole numbers"). Data in integer form are typically exact (e.g., a four-pole motor); when it isn't, engineers typically provide some statement clarifying the accuracy (e.g., we estimate that we will sell 7950 units ± 10%).

Rational numbers are those which can be expressed as the ratio of two integers. Examples are as follows:

$$\frac{3}{2} = 1.5$$

$$\frac{3}{4} = 0.75$$

$$\frac{2}{3} = 0.66666\ldots = 0.\overline{6}$$

$$\frac{1}{9} = 0.0101010\ldots = 0.\overline{0}\overline{1}$$

If a rational number consists of an infinitely repeating series of digits, the repeating string may be represented by an overscore.

Decimal numbers are sometimes written as follows:

1,234.56: U.S. standard decimal notation
1.234,56: European standard decimal notation

Note that the meanings of the comma and period are reversed when comparing U.S. and European standard decimal notation. U.S. notation is used throughout this book.

Engineering data are rarely exact, and the issue of accuracy is always important. Accuracy is the measure of how close a number is to its true value. In engineering, accuracy is implied by the number of significant digits. For example, consider the number 3.21. It is understood that the leading digits (3,2) are exact, but the last digit (1) is an estimate. That is, the number is approximately known to be one part in 1000. Such accuracy is routinely described as "three significant figures" or "three significant digits." Leading zeros don't count as significant digits, but trailing zeros do (or at least should; sometimes this rule is violated when using trailing zeros as a decimal place marker).

Many times, numbers have been rounded up (or down) to a specified number of significant digits. Consider the following numbers to be rounded to three significant digits:

324.587	rounds to 325
324.483	rounds to 324
324.500	rounds to 324
325.500	rounds to 326

At three significant digits, the number to be rounded is the third significant digit (4).

- In the first case, the digits to the right are 0.587, which is closer to 1 than to 0. Therefore, we replace 0.587 with 1, which when added to 324 produces 325.
- In the second case, the digits to the right are 0.483, which is closer to 0 than to 1. Therefore, we replace 0.483 with 0, which when added to 324 produces 324.
- In the third case, the digits to the right are 0.500, which is exactly halfway between 0 and 1. Now what? Some folks have a preference for even numbers. If so, round to the nearest even number. In this case, 324 is even and 325 is odd. So we keep 324.

- In the fourth case, the digits to the right are 0.500, which again is halfway between 0 and 1. Since 326 is even and 325 odd, we round up to 326.

There are some numbers which cannot be expressed as the ratio of integers. These are called "irrational" numbers. Three famous examples are the following:

$$\pi = 3.14159265...$$

$$\varepsilon = 2.718281828459....$$

$$\sqrt{2} = 1.4142136...$$

The first is pi (p), the ratio of the circumference of a circle to its radius; the second is the natural logarithm base (e); and the third is the square root of two.

Such numbers are of special interest, and people interested in such matters have computed these to incredible accuracies. For example, Alexander J. Yee and Shigeru Kondo announced on August 2, 2010, that they have successfully computed pi to 5 *trillion* digits![7]

Very large, or very small, numbers are generally expressed in scientific notation. Also, the choice of SI prefixes can be used to great advantage in making numbers more meaningful. For example:

The speed of light in a vacuum is approximately

300,000,000 m/s
or 3×10^8 m/s (scientific notation)
or 0.3 m/ns (use of SI prefixes)

The wavelength of visible light ranges from

0.000000380 to 0.000000750 m
or 380×10^{-9} to 750×10^{-9} m (scientific notation)
or 380 to 750 nm (use of SI prefixes)

The standard SI prefixes are supplied in Table 1.1.

It is possible to plot all numbers discussed to this point on a "number line," as shown in Figure 1.1.

[7] For the same reason that people climb Mount Everest, we presume.

Table 1.1 SI Prefixes

yatto	Y	10^{24}		yocto	z	10^{-24}
zetta	Z	10^{21}		zepto	z	10^{-21}
exa	E	10^{18}		atto	a	10^{-18}
peta	P	10^{15}		femto	f	10^{-15}
tera	T	10^{12}		pico	p	10^{-12}
giga	G	10^{9}		nano	n	10^{-9}
mega	M	10^{6}		micro	μ	10^{-6}
kilo	k	10^{3}		milli	m	10^{-3}
hecto	h	10^{2}		centi	c	10^{-2}
deka	da	10^{1}		deci	d	10^{-1}

Because all the numbers have a unique position on the number line, we can argue that such a number system is one-dimensional. But now consider the following equation:

$$z^2 - 6z + 25 = 0$$

$$z = \frac{6 \pm \sqrt{36 - 100}}{2} = 3 \pm 4\sqrt{-1}$$

Where can we locate these two numbers on the number line? Nowhere! There is no place that is consistent with the set of integers, rational numbers, and irrational numbers! This must be a "new kind" of number, which requires a broader geometrical perspective. The specific problem is the square root of minus one. Mathematicians were aware of such numbers dating back to the early Greeks. René Descartes coined the term "imaginary" for the square root of minus one in 1637, with the intention to ridicule it as useless, writing z as

$$z = 3 + 4i$$

$$\text{where } i = \sqrt{-1}$$

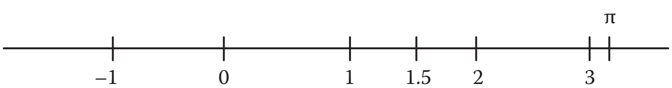


FIGURE 1.1 The number line.

and referring to "4*i*" as an "imaginary" number. It was logical, then, to refer to "3" as a "real" number. "3 + 4*i*" became known as a "complex" number, with real and imaginary parts. Actually, the choice of this terminology was unfortunate in that "imaginary" numbers are no more "imaginary" or "make-believe" than any other number. All numbers are abstract. What counts for engineers is their usefulness, and complex numbers have proved to be extremely important in that regard. Many mathematicians have made contributions to the issue, but it was the work of the great mathematician Leonhard Euler (1707–1783) that earned widespread acceptance of complex numbers.

The complex plane (z-plane) is a Cartesian plot of complex numbers, with the real part plotted on the horizontal (x) axis, and the imaginary part plotted on the vertical (y) axis. Such a plot is sometimes called an "Argand diagram," in honor of Jean-Robert Argand (1768–1822), although like many topics in mathematics, this was understood and used by many others. So, our "number **line**" expands to a number **plane** (i.e., an Argand diagram), as shown in Figure 1.2a, in order to include complex numbers.

It happens that complex numbers are particularly important in electrical engineering, as we shall see. The symbol "*i*" for the square root of minus one in EE is unfortunate because "*i*" is also commonly used for "electrical current." Hence, we adopt the common practice of using "*j*" instead, and choose to write "*j*" in the lead position:

$$z = 3 + 4i \quad \text{(standard mathematical notation)}$$

$$z = 3 + j4 \quad \text{(standard EE notation)}$$

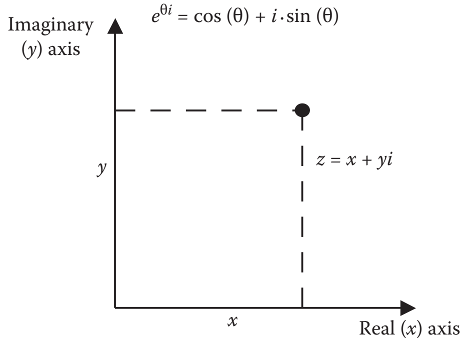
Notation and symbology in engineering are of course arbitrary. One can define a symbol to mean whatever one wishes. However, there are two concerns. Bad choices can lead to awkward constructions and impediments to understanding, and, if one strays too far from traditional and historic choices, confusion can result. The choice of symbols is not a major issue, but neither is it trivial. We define the following general notation for complex numbers:

$$\bar{Z} = X + jY$$

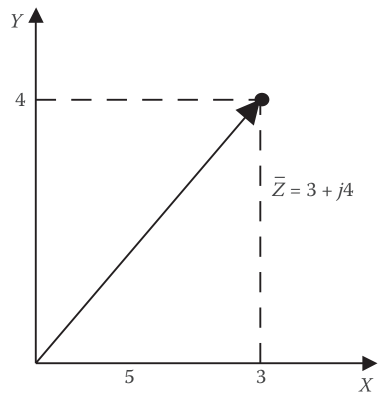
$$\bar{Z} = \text{complex number}$$

$$X = \text{real part of } \bar{Z}$$

$$Y = \text{imaginary part of } \bar{Z}$$



(a)



(b)

FIGURE 1.2 **The complex number plane. a. Complex plane (Argand diagram). b. Argand diagram for Example 1.1.**

Symbols for complex variables shall be overscored. Euler's work with complex numbers leads to one of the most famous equations in all of mathematics (i.e., "Euler's identity"):

$$e^{j\theta} = \cos(\theta) + j\sin(\theta)$$

Applied to complex numbers in our notation, this produces the following:

$$\overline{Z} = Z \cdot e^{j\theta} = Z \cdot \cos(\theta) + jZ \cdot \sin(\theta)$$

$$Z = |\overline{Z}| = \text{magnitude (modulus) of } \overline{Z}$$

$$\theta = \arg(\overline{Z}) = \text{angle (argument) of } \overline{Z}$$

This leads to two forms for complex numbers:

$$\bar{Z} = X + jY \quad \text{rectangular form}$$

$$\bar{Z} = Z \cdot e^{j\theta} \quad \text{exponential (polar) form}$$

Engineers generally find the formal exponential (polar) notation awkward and commonly use an alternate polar notation:

$$\bar{Z} = Z \cdot e^{j\theta} = Z\underline{/\theta} \quad \text{(engineering) polar form}$$

In formal exponential notation, θ *must* be in radians. In engineering polar notation, θ is normally expressed in degrees. Conversion between forms is as follows:

$$\text{rectangular to polar form:} \quad Z = \sqrt{X^2 + Y^2} \quad \theta = \mathrm{Tan}^{-1}\left(\frac{Y}{X}\right)$$

$$\text{polar to rectangular form:} \quad X = Z \cdot \cos(\theta) \quad Y = Z \cdot \sin(\theta)$$

An example will be instructive.

Example 1.1

a. Given $\bar{Z} = 3 + j4$. Convert $\bar{Z}$ into polar and exponential forms.

$$Z = \sqrt{X^2 + Y^2} = \sqrt{3^2 + 4^2} = 5$$

$$\theta = \mathrm{Tan}^{-1}\left(\frac{Y}{X}\right) = \mathrm{Tan}^{-1}\left(\frac{4}{3}\right) = 53.1° = 0.9273 \text{ rad}$$

$$\bar{Z} = 5\underline{/53.1°} = 5 \cdot e^{j0.9273}$$

b. Given $\bar{Z} = 5\angle 53.1°$. Convert $\bar{Z}$ into rectangular form.

$$X = Z \cdot \cos(\theta) = 5 \cdot \cos(53.1°) = 3$$

$$Y = Z \cdot \sin(\theta) = 5 \cdot \sin(53.1°) = 4$$

$$\bar{Z} = 3 + j4$$

See Figure 1.2b for the corresponding Argand diagram.

The arithmetic of complex numbers is important. Addition and subtraction are best done in rectangular form, and multiplication and division in polar form.

Given that

$$\bar{A} = a + jb = A\angle\alpha$$

$$\bar{B} = c + jd = B\angle\beta$$

Addition

$$\bar{A} + \bar{B} = (a + jb) + (c + jd)$$
$$= (a + c) + j(b + d)$$

Subtraction

$$\bar{A} - \bar{B} = (a + jb) - (c + jd)$$
$$= (a - c) + j(b - d)$$

Multiplication

$$\bar{A} \cdot \bar{B} = (A\angle\alpha) \cdot (B\angle\beta)$$

$$\bar{A} \cdot \bar{B} = A \cdot B\angle(\alpha+\beta)$$

Division

$$\frac{\bar{A}}{\bar{B}} = \frac{A\angle\alpha}{B\angle\beta} = \left(\frac{A}{B}\right)\angle(\alpha - \beta)$$

Example 1.2

Given $\bar{A} = 3 + j4 = 5\angle 53.1^\circ$ $\bar{B} = 5 - j12 = 13\angle -67.4^\circ$

Compute $\bar{C} = \bar{A} + \bar{B}$; $\bar{C} = \bar{A} - \bar{B}$; $\bar{C} = \bar{A} \cdot \bar{B}$; and $\bar{C} = \frac{\bar{A}}{\bar{B}}$

$$\bar{C} = \bar{A} + \bar{B} = (3 + j4) + (5 - j12) = 8 - j8$$

$$\bar{C} = \bar{A} - \bar{B} = (3 + j4) - (5 - j12) = -2 + j16$$

$$\bar{C} = \bar{A} \cdot \bar{B} = (5\angle 53.1^\circ) \cdot (13\angle -67.4^\circ) = 65\angle -14.3^\circ$$

$$\bar{C} = \frac{\bar{A}}{\bar{B}} = \left(\frac{5\angle 53.1^\circ}{13\angle -67.4^\circ}\right) = 0.3846\angle 120.5^\circ$$

The Conjugate Operation

The conjugate of a complex number is formed by reversing the sign of the imaginary part in rectangular form, or, equivalently, reversing the sign of the angle in polar form:

$$(\bar{Z})^* = \text{conjugate of } \bar{Z} = (X + jY)^* \triangleq X - jY$$

$$(\bar{Z})^* = (Z \cdot e^{j\theta})^* \triangleq Z \cdot e^{-j\theta} = Z\angle -\theta$$

Example 1.3

$$\text{Given} \quad \bar{A} = 3 + j4 = 5\angle 53.1^\circ$$

$$\text{Find} \quad (\bar{A})^*$$

$$(\bar{A})^* = (3 + j4)^* = 3 - j4$$

$$(\bar{A})^* = (5\angle 53.1^\circ)^* = 5\angle -53.1^\circ$$

All standard mathematical functions (sine, cosine, exponentials, square root, etc.) work for complex numbers, as the commercial product MATLAB© confirms. MATLAB© treats all numbers as complex. In fact, any real number **is** complex (the imaginary part just happens to be zero).

In engineering, numbers usually represent physical quantities, and as such, they carry physical units. In EE, in most cases, SI units are used. See Appendix A for a tabulation of the most important variables in EE and their SI units.

1.6 Summary

Engineering is the application of science, mathematics, and technology to the design, fabrication, and use of practical devices and systems. Its history parallels the history of human civilization. Electrical engineering (EE) is that branch of engineering that primarily uses electromagnetic science and technology to achieve the engineering mission. The purpose of this book is to address those EE topics which are most important to non-EE specialists.

Practitioners of engineering have created professional organizations to articulate a code of ethics, to offer continuing educational opportunities, to enforce professional standards, to provide an authoritative advisor on technical matters to the general public, and to provide a fellowship of those with mutual scientific interests. Standards organizations are also necessary for EE: the most important of these include the IEEE, ABET, NEMA, NFPA, ISO, and ANSI.

Since engineers routinely deal with questions of "how strong," "how heavy," "how large," and so on, they routinely work with numbers. It happens that complex numbers are particularly important to EE, and competency in complex number analysis is essential.

Electrical engineering is concerned with two broad areas of application:

- Processing of energy
- Processing of information

This book is divided into two parts, the first dealing with energy processing and the second with information processing, and is designed to serve as a textbook for a course by the same name.

There are two basic ways of explaining electromagnetic phenomena, the first using the circuits model, and the second using the fields model. We begin our study by considering the EE circuits model.

References

Dorf, Richard C. *Handbook of Engineering Tables*. Boca Raton, FL: CRC Press, 2004.

Florman, Samuel C. *The Civilized Engineer*. New York: St. Martin's Press, 1987.

Linden, David. *Handbook of Batteries and Fuel Cells*. New York: McGraw-Hill, 1984.

Kirby, Richard S., Sydney Withington, Arthur B. Darling, and Frederick G. Kilgour. *Engineering in History*. New York: McGraw-Hill, 1956.

Reece, Holtzapple. *Foundations of Engineering*. New York: McGraw-Hill, 2003.

Wright, Paul H. *Introduction to Engineering*, 3rd ed. New York: John Wiley & Sons, 2003.

Problems

1.1. a. Define engineering.
b. Define electrical engineering.

1.2. Identify the following professional societies: IEEE, ASME, ASCE, AICE, and ASEE.

1.3. Identify the following standards organizations: ISO, ANSI, NIST, ASTM, NFPA, NEC, and NEMA.

1.4. What is the oldest engineering institution in the United States?

1.5. The distance to the moon varies from a perigee of 363,104 km to an apogee of 405,696 km. Express the perigee in SI units to four significant digits.

1.6. The sodium EM spectrum is dominated by the bright yellow doublet at 588.9950 and 589.5924 nanometers, the brighter line at the shorter wavelength. Express the bright-line wavelength in SI units to four significant digits.

1.7. Complete the table of complex numbers (to four significant digits).

Rectangular	Polar	Exponential
$+3 + j4$		
$-3 + j4$		
$-3 - j4$		
$+3 - j4$		
	$13\angle 22.62^0$	
	$13\angle -22.62^0$	
	$-13\angle 22.62^0$	
	$-13\angle -22.62^0$	
$0 + j4$		
$0 - j4$		
$4 + j0$		
$-4 + j0$		
		$10 \cdot e^{j1.0472}$

1.8. Draw the Argand diagram for each complex number in Problem 1.7.

1.9. Given that

$$\overline{A} = 3 + j4 \quad \overline{B} = 5 - j12 \quad \overline{C} = -8 - j8$$

Evaluate

a. $\overline{A} + \overline{B} \cdot \overline{C}$

b. $\overline{A} + \dfrac{\overline{B}}{\overline{C}}$

c. $\dfrac{\overline{A} + \overline{B}}{\overline{C}}$

Chapter 2

Electrical Circuit Concepts

It is astonishing to realize that we have never written an *exactly correct* equation on any physical system! And, furthermore, we never will! What engineers do is to create idealized *models* of physical systems, models to which we can apply the laws of physics in mathematical form, in order to predict ideal system performance. If we have created "good" models, actual system performance closely matches that of the ideal. This modeling approach has proved to be remarkably effective, permitting the development of all sorts of products from large modern aircraft to the tiny cell phone and millions of products in between. Still, always bear in mind that there are no perfect models.

In the modeling process, we normally create models which focus on those phenomena in which we have an immediate interest. For example, consider an electric motor. Engineer A may be interested in issues like winding voltages and currents; Engineer B, the associated magnetic fields; Engineer C, its thermodynamics, winding temperatures, and heating; Engineer D, its torques, speed, and acceleration; and so on. Whereas there are interactions between all these phenomena, there are models which exclusively deal with each.

There are two basic ways of dealing with electrical phenomena:

The electromagnetic (EM) fields perspective
The "circuits" perspective

Of the two, the circuits approach has proved to be the most useful and, fortunately, the easiest. For these reasons, we dedicate this chapter to electric circuits.

Four fundamental forces have been identified in contemporary physics:

- The *strong nuclear force* is very intense, but very short ranged.
- The *weak nuclear force* is responsible for radioactive decay. It has a very short range and is much weaker than the strong force.
- The *gravitational force* is weak, but very long ranged. It is attractive, and acts between any two masses.
- The fourth force is that which is related to the so-called electromagnetic properties of matter, and gives rise to the subject of this book.

2.1 Charge, Current, Voltage, Power, and Energy

After gravity is accounted for, when an electron is placed near an elementary atomic particle "*X*," one of the following situations is observed:

Case 1. A force of attraction is observed between X and the electron.

Case 2. A force of repulsion is observed between X and the electron.
Case 3. No force is observed.

This so-called electrostatic force is accounted for by assigning X and the electron a property called "charge," such that X is said to be

"positively charged" (Case 1),
"negatively charged" (Case 2), or
"uncharged" (Case 3).

If X is also an electron, Case 2 is observed, and the charge (Q_e) can be quantified from physical experimentation.[1] All atomic particles are observed to have integer multiples of Q_e.[2] The unit selected for charge is the coulomb (C).

The electron, and therefore charge, can freely move through certain materials, particularly metals. "Electric current," or simply "current," is defined as the flow rate of charge through a given path or

$$\text{current} = i = \frac{dq}{dt}, \text{ in coulomb per second (C/s), or ampere (A)}$$

Current is a signed quantity, and has a defined positive direction[3] (an arrow) in a diagram. A "negative" current means that flow is in the direction opposite to that defined as positive.

The energy per unit charge required to move a unit charge from point a to point b in a circuit is defined as potential difference from a to b or voltage (between a and b). We say there is a voltage "drop" from *a* to *b* if *a* is more positive than *b*; we say there is a voltage "rise" from *a* to *b* if *b* is more positive than *a*.

$$\text{voltage drop from a to b} = v_{ab} = \frac{dw}{dq} \text{ in joules per coulomb (J/C), or volts (V)}$$

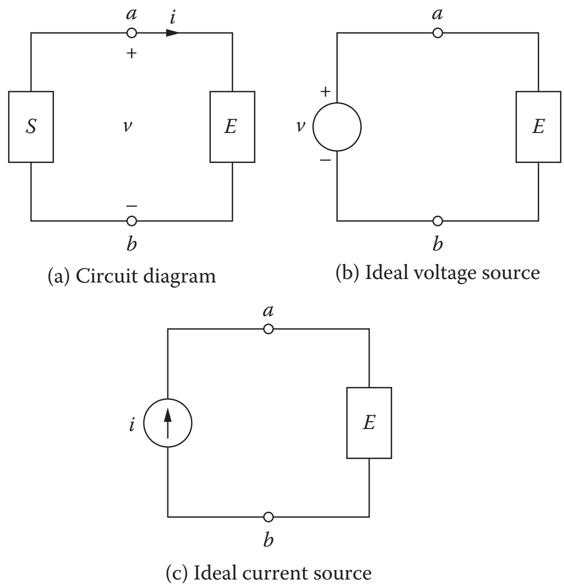
Consider Figure 2.1. Any interconnection of elements is called a "circuit," and the diagram that shows the interconnection details is called a "circuit diagram."

---

1 Q_e = charge on the electron = 0.1602 aC.

2 This is not strictly true in quantum mechanics. Subjects addressed in this book require only classical physics, without considering either quantum mechanics or relativity.

3 Positive *physical* current direction is defined as the direction that a unit *positive* charge would move through a circuit, which is opposite to the direction electrons move, since the electronic charge is negative.



(a) Circuit diagram (b) Ideal voltage source

(c) Ideal current source

FIGURE 2.1 **Circuit fundamentals.**

Note that

$$v \cdot i = \left(\frac{dw}{dq}\right) \cdot \left(\frac{dq}{dt}\right) = \frac{dw}{dt} = p$$

Here, p represents the flow rate of energy, or instantaneous power, into element E, or, equivalently, the power out of element S. We say that p is the power flow *into* E, or p is the power flow *out of* S. Since v and i are signed variables, so is p. An example would be useful.

Example 2.1

Consider the circuit of Figure 2.1. Given that $v = 10$ V and $i = 2$ A, what are the following?

A.	The voltage drop from a to b?	+10 V
B.	The voltage rise from a to b?	−10 V
C.	The voltage drop from b to a?	−10 V

D.	The voltage rise from b to a?	+10 V
E.	The current from a to b through E?	+2 A
F.	The current from b to a through E?	–2 A
G.	The current from a to b through S?	–2 A
H.	The current from b to a through S?	+2 A.
I.	The power absorbed by E?	+20 W.
J.	The power delivered by E?	–20 W.
K.	The power absorbed by S?	–20 W.
L.	The power delivered by S?	+20 W.

2.2 Kirchhoff's Laws

Gustav Robert Kirchhoff (1824–1887) presented two fundamental properties of electric circuits in his 1845 work "The Laws of Closed Electric Circuits." His work can be summarized in two famous general principles, which have become known as "Kirchhoff's laws." Kirchhoff's laws serve as the foundation of circuit analysis.

The physics of the electric circuit dictate that two fundamental conservation principles must be obeyed:

- Conservation of charge
- Conservation of energy

Kirchhoff's current law (KCL) is based on conservation of charge; Kirchhoff's voltage law (KVL) is based on conservation of energy.

Kirchhoff's Current Law (KCL)

The current law requires that the net current flowing into any node in a circuit must be equal to the net current flowing out. Consider a node terminated in n branches, where branch k has a current i_k defined flowing into the node. We write,

$$\sum_{k=1}^{n} i_k = 0$$

An example will clarify matters.

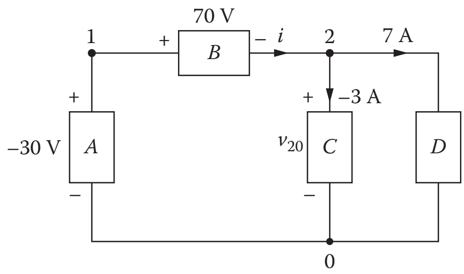


FIGURE 2.2 Circuit illustrating KCL and KVL.

Example 2.2

Consider the circuit of Figure 2.2. Using KCL at node 2, find *i*.

There are three ways of thinking about KCL:

$$\text{“}\sum \text{in's} = 0\text{”} \qquad i - (-3) - (7) = 0; \quad i = 4 \text{ A}$$

$$\text{“}\sum \text{out's} = 0\text{”} \qquad -i + (-3) + (7) = 0; \quad i = 4 \text{ A}$$

$$\text{“}\sum \text{in's “=”} \sum \text{out's”} \qquad i = +(-3) + (7); \quad i = 4 \text{ A}$$

Considering Example 2.2, we note that all three approaches result in the same equation, and hence the same answer. Therefore it makes no difference which way one thinks: it's mostly a matter of personal preference. The authors have a slight preference for equating the "ins" to the "outs," because it seems the most "natural" (i.e., "what goes in must come out"). Try all three approaches and use the one you prefer. Also, be tolerant of those who don't think like you do.

Kirchhoff's Voltage Law (KVL)

Conservation of energy requires that whatever energy gained by a charge *dq* moving around any closed path in a circuit must also be lost, or given back, to the circuit elements. Dividing by *dq* converts energy to voltage, producing a general principle commonly called "Kirchhoff's voltage law." KVL requires that the net voltage drop must be equal to the net voltage rise around any closed path traced out in a given circuit. A drop in voltage is a transition from plus to minus as we cross over an element; a rise is a transition from minus to plus.

Consider a closed path through a circuit, crossing n branches, where branch k is terminated at nodes i and j, and v_{ij} is the voltage **drop** from node i to node j. We write

$$\sum_{k=1}^{n} v_{ij} = 0$$

An example will clarify matters.

Example 2.3

Consider the circuit of Figure 2.2. Using KVL around the path containing elements A, B, and C, find v_{20}.

There are three ways of thinking about KVL:

$$\text{“}\sum \text{rises”} = \text{“}\sum \text{drops”}: \qquad (-30) = 70 + \boldsymbol{v}_{20}; \quad \boldsymbol{v}_{20} = -100\ \boldsymbol{V}$$

$$\text{“}\sum \text{rises} = 0\text{”}: \qquad (-30) - (70) - \boldsymbol{v}_{20} = 0; \quad \boldsymbol{v}_{20} = -100\ \boldsymbol{V}$$

$$\text{“}\sum \text{drops} = 0\text{”}: \qquad -(-30) + (70) + \boldsymbol{v}_{20} = 0; \quad \boldsymbol{v}_{20} = -100\ \boldsymbol{V}$$

Considering Example 2.3, we note that all three approaches result in the same equation, and hence the same answer. Therefore it makes no difference which way one thinks. The authors have a slight preference for equating “rises” to “drops,” because it seems the most “natural” (i.e., “What dq gains in energy must be returned to the circuit”). Again, try all three approaches and use the one you prefer.

As we progress in our study of circuit analysis, you will appreciate the power and generality of Kirchhoff’s laws: KCL applies at *any* node in *every* circuit, with no exceptions. Also, KVL applies around *any* closed path in *every* circuit, with no exceptions. Observe that in Examples 2.2

and 2.3, we solved the problems *without knowing the nature of elements A, B, C, and/or D!*

2.3 The Ideal Circuit Elements

We return to the situation of Figure 2.1. If energy flows out element *S* and into element *E*, we say that element *S* is a source and may be one of two types:

- The ideal voltage source. Maintains a specified voltage *v* at its terminals, independent of its termination (i.e., element *E*).
- The ideal current source. Maintains a specified current *i* at its terminals, independent of its termination (i.e., element *E*).

The symbols for the ideal sources are shown in Figure 2.1b and Figure 2.1c, respectively.

Observe that immediately we encounter a problem. In Figure 2.1a, suppose *S* is an ideal 100 V voltage source, and element *E* is an ideal 50 V voltage source! *How can both elements simultaneously conform to their definitions and satisfy the voltage law?* The answer is that they can't, and such a circuit is referred to as indeterminate. Likewise, in Figure 2.1a, suppose *S* is an ideal 10 A current source, and element *E* is an ideal 5 A current source! *How can both elements simultaneously conform to their definitions and satisfy the current law?* Again, this is impossible, and again we have an indeterminate situation.

Fortunately, these are never encountered in physical systems. We note them only in passing to realize that it is possible to create indeterminate abstract circuit structures using ideal components. Should some perverse individual ask us to analyze such a circuit, we write "indeterminate" across the diagram, and move on to the next problem. We pledge to never create such an atrocity.

If element *E* is not a source, it is said to be "passive." We turn our attention to element *E*, for the case where *E* is passive. First consider two very simple but important special cases.

Suppose that *S* is an ideal non-zero-voltage source *v* and that the current (i) is zero. *E* is said to be an "open circuit," abbreviated as an "open," or OC. Now suppose that *S* is an ideal non-zero-current source *i* and that the voltage *v* is zero. *E* is said to be a "short circuit," abbreviated as a "short," or SC.

Given that element E is not a source, an open, or a short, consider the electrical energy flowing into element E:

$$w = \int p \cdot dt = \int v \cdot i \cdot dt$$

In general, passive elements have three fundamental circuit properties related to how they deal with this energy.

- The energy is dissipated (i.e., nonrecoverable) internally. This is called the "resistive," or R, property.
- The energy is stored (i.e., recoverable) in the internal magnetic field. This is called the "inductive," or L, property.
- The energy is stored (i.e., recoverable) in the internal electric field. This is called the "capacitive," or C, property.

We define the following:

R = resistance, ohm (Ω)
L = inductance, henry (H)
C = capacitance, farad (F)

All physical passive elements have all three of these properties. However, it is possible to design a component which maximizes one property while minimizing the other two. Hence we speak of

- a *resistor*, a component which maximizes R, with negligible L and C;
- an *inductor*, a component which maximizes L, with negligible R and C; and
- a *capacitor*, a component which maximizes C, with negligible R and L.

That said, it is natural to define three *ideal* elements, according to

- an *ideal resistor*, a component which has exclusively resistance;
- an *ideal inductor*, a component which has exclusively inductance; and
- an *ideal capacitor*, a component which has exclusively capacitance.

The ideal passive elements, and their v–i relations, are summarized in Figure 2.3. The v–i relation for the resistor is commonly referred to as

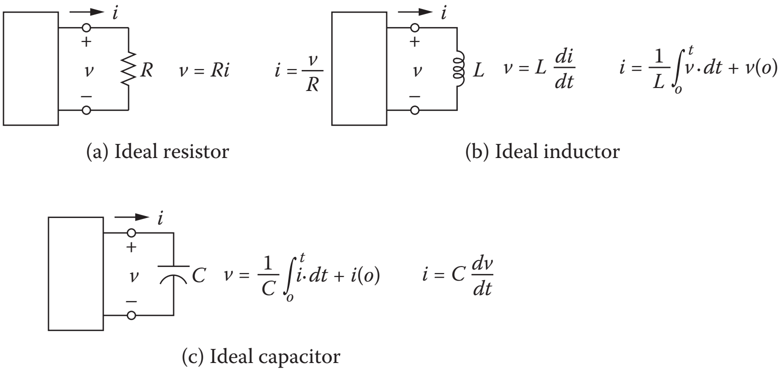


FIGURE 2.3 **The ideal passive elements.**

"Ohm's law" (ΩL), after Georg Simon Ohm, as presented in his famous 1827 book:

$$\Omega L: \quad v = R \cdot i$$

Example 2.4

Consider the circuits of Figure 2.3, each terminated in a voltage source, where

$$v = 10 \cdot e^{-3t}; \quad t \geq 0; \quad i(0) = 0$$

A. In circuit (a), for $R = 2\ \Omega$ find $i(t)$.

$$i = \frac{v}{2} = \frac{10 \cdot e^{-3t}}{2} = 5 \cdot e^{-3t}$$

B. In circuit (a), for $L = 2$ H find $i(t)$.

$$i = \frac{1}{L}\int_{-\infty}^{t} v(t) \cdot dt = \frac{1}{L}\int_{-\infty}^{0} v(t) \cdot dt + \frac{1}{L}\int_{0}^{t} v(t) \cdot dt = i(0) + \frac{1}{L}\int_{0}^{t} v(t) \cdot dt$$

$$i = 0 + \frac{1}{2}\int_{0}^{t} (10 \cdot e^{-3t}) \cdot dt \quad i = \frac{5}{3}(1 - e^{-3t})$$

C. In circuit (c), for $C = 2$ F find i(t).

$$i = C \cdot \frac{dv}{dt} = 2\frac{d}{dt}(10 \cdot e^{-3t}) = -60 \cdot e^{-3t}$$

We now have a total of five ideal elements (voltage and current sources, plus R, L, and C). We can now define the general problem of circuit analysis:[4]

> For a given interconnection of any combination of known elements, solve for all voltages, currents, and powers throughout the circuit.

The problem is solved by systematically applying KCL and KVL, and the v–i relations for each element, which results in the formulation in a set of linear differential equations, which are subsequently solved.[5]

To apply circuit analysis to engineering problems, it is necessary to determine "an equivalent circuit" for each physical device we encounter. This task occupies a large part of the remainder of this book. But for now, we focus on circuit analysis.

The topological concepts of "node" and "branch" are defined as follows:

Node: the nontrivial connection point of two or more circuit elements.
Branch: a path for current flow between two nodes; an element.[6]

In Figure 2.4 there are six nodes (a, b, c, d, e, and f). Points b' and o' are not nodes because they can be merged with nodes b and o, respectively. The circuit has nine branches.

[4] There are a few more ideal elements, some of which we will introduce later when we have need of them. However, they will introduce no substantive changes in the process that we normally think of as circuit analysis.

[5] Linearity is extremely important, but we resist the temptation to digress into this topic due to space limitations. If the reader doesn't understand linearity and its implications, he or she is encouraged to consult any of the many excellent references on the subject.

[6] If interior nodes are of no interest, a branch may contain more than one element.

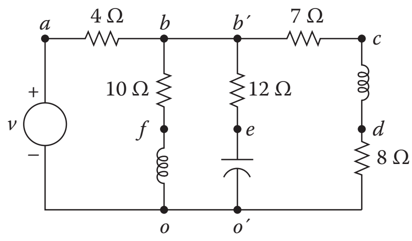


FIGURE 2.4 An example circuit.

Circuits can be said to operate in one of four possible modes,[7] three of which are stable steady-state conditions. These are, from the simplest to the most complex:

- dc
- ac
- periodic
- transient

Figure 2.5 illustrates the nature of voltage (and/or current) in circuits operating in steady-state modes.

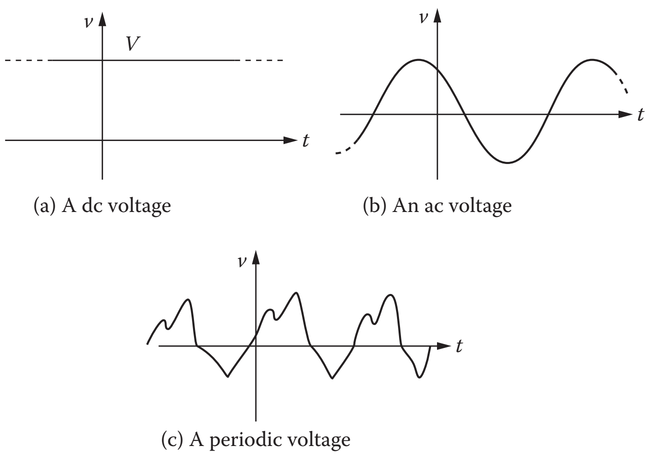


FIGURE 2.5 Steady-state voltages.

[7] There is a fifth possibility, where the sources are time-varying and not dc, ac, or periodic, nor have they ever been or will they ever be. This mode is of limited practical interest, and beyond the scope of our study.

2.4 The dc Mode

The term "dc" historically means "direct current," but it means "time invariant" in contemporary usage. A "dc circuit" is a circuit operating in the dc mode. Observe that, for dc:

$$v = L \cdot \frac{di}{dt} = 0$$

$$i = C \cdot \frac{dv}{dt} = 0$$

This means that inductors and capacitors become shorts, and opens, respectively, and hence are trivialized. When a general circuit is redrawn for dc operation, the *L*'s and *C*'s "disappear," the "*L*'s" are replaced by shorts, and the "*C*'s" are replaced by opens! Thus, the basic set of ideal circuit elements is reduced from five (v-source, i-source, R, L, and C) to three (v-source, i-source, and R) for the dc circuit.

There is a conventional notation that is commonly used in electrical engineering, and one which we will use as well (except when otherwise noted).

$$v \text{ or } i = v(t) \text{ or } i(t) = \text{ instantaneous voltage or current}$$

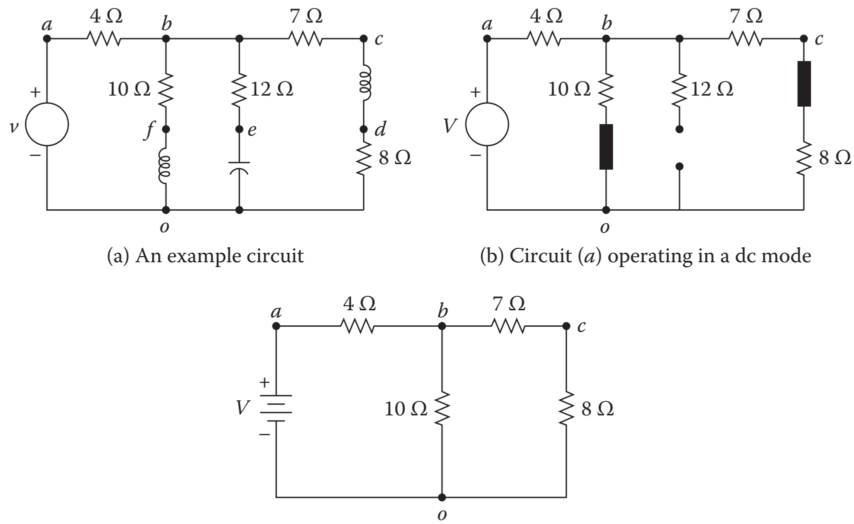
$$V \text{ or } I = \text{ constant (dc) voltage or current}$$

For the dc circuit, the v, i relation for the resistor is *Ohm's law* (ΩL).

Example 2.5

Consider the circuit of Figure 2.6a, for which $v(t) = 100$ V for $-\infty \le t \le +\infty$, putting the circuit in the dc mode. Draw the corresponding dc circuit, replacing *L*'s with shorts and *C*'s with opens.

Solution: See Figure 2.6b and 2.6c. Note the alternate symbol for a dc voltage source. Also note that nodes *d*, *e*, and *f* are eliminated (*d* merges with *c*, *e* is trivial since it terminates in an open, and *f* merges with *o*).



(a) An example circuit

(b) Circuit (*a*) operating in a dc mode

(c) Circuit (*b*) Redrawn

FIGURE 2.6 An example circuit.

THE SERIES CONNECTION

There are usually some simplifications that one can make in a given circuit which will simplify its analysis. One such simplification is series reduction. Consider "*n*" resistors connected as shown in Figure 2.7a, such that the same current flows through each. We write,

$$V_{ab} = \sum_{i=1}^{n} R_i \cdot I_{ab} = \left(\sum_{i=1}^{n} R_i \right) \cdot I_{ab} = R_{ab} \cdot I_{ab}$$

$$\text{where } R_{ab} = \sum_{i=1}^{n} R_i$$

The resistors are said to be "in series," and can be replaced with R_{ab}, as shown in Figure 2.7b. The circuit is said to be "simplified" since we've replaced "*n*" elements with one. We have also eliminated "*n*–1" nodes (the connection points between the resistors), which certainly is simpler, although not necessarily desirable.

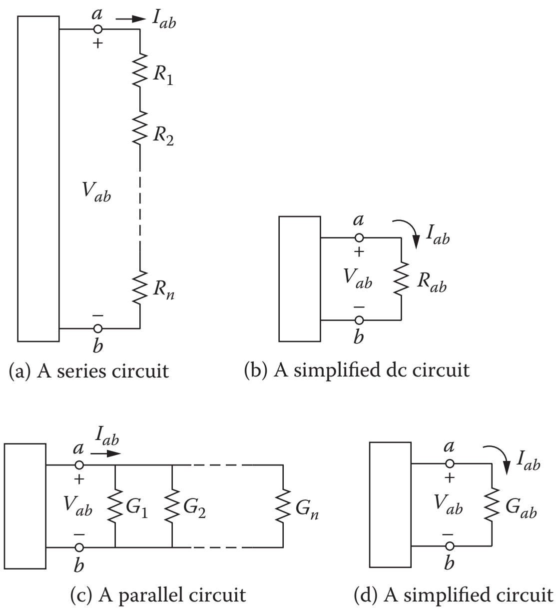


FIGURE 2.7 Series/parallel concepts.

THE PARALLEL CONNECTION

There is an alternate way of representing resistance, for example by specifying its "conductance."

$$G = \frac{1}{R} = \frac{I}{V} = \text{ conductance in siemens (S).}$$

Consider "n" resistors, with specified conductances, connected as shown in Figure 2.7c, such that the same voltage is across each. We write,

$$I_{ab} = \sum_{i=1}^{n} G_i \cdot V_{ab} = \left(\sum_{i=1}^{n} G_i \right) \cdot V_{ab} = G_{ab} \cdot V_{ab}$$

$$\text{where } \mathrm{G}_{ab} = \sum_{i=1}^{n} G_i$$

The resistors are said to be "in parallel," and can be replaced with G_{ab}, as shown in Figure 2.7d. The circuit is simplified since we've replaced "n" elements with one. We have also eliminated "$n-1$" branches (the current paths between the top and bottom nodes), which may not be desirable.

An example will demonstrate the usefulness of series and parallel simplifications.

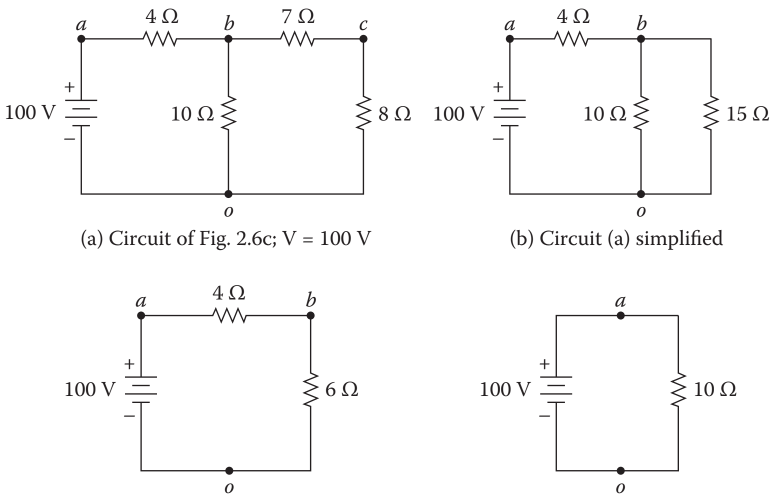
Example 2.6

Consider the circuit of Figure 2.8a (which is the circuit of Figure 2.6c).

a. Using series reduction, draw the simplified circuit.

$$R_{ab} = 7 + 8 = 15\ \Omega$$

See the circuit of Figure 2.8b.



(a) Circuit of Fig. 2.6c; V = 100 V

(b) Circuit (a) simplified

(c) Circuit (b) simplified

(d) Circuit (c) simplified

FIGURE 2.8 Using series/parallel simplifications.

b. Using a parallel simplification, draw the simplified circuit.

$$G_{ab} = \frac{1}{10} + \frac{1}{15} = \frac{1}{6}\ S$$

$$R_{ab} = \frac{1}{G_{ab}} = 6\ \Omega$$

See the circuit of Figure 2.8c.

c. Using series simplification, draw the simplified circuit.

$$R_{ab} = 4 + 6 = 10\ \Omega$$

See the circuit of Figure 2.8d.

Example 2.7b shows that it is awkward to combine two resistors in parallel when resistor values are given in terms of *R* rather than *G*). In two steps, we computed,

$$G_{ab} = \frac{1}{R_{10}} + \frac{1}{R_{15}} = \frac{R_{15} + R_{10}}{R_{15} \cdot R_{10}}$$

$$R_{ab} = \frac{1}{G_{ab}}$$

We could have used a more direct one-step approach:

$$R_{ab} = \frac{R_{15} \cdot R_{10}}{R_{15} + R_{10}}$$

That is, to combine two resistors in parallel, divide their product by their sum. Most analysts find this direct approach to be more convenient.

Observe that dc circuit analysis results in a set of linear algebraic equations. We now have the tools to pursue a comprehensive analysis of the dc circuit. The methodology will be demonstrated in Example 2.7.

Example 2.7

Perform a comprehensive analysis of the dc circuit of Figure 2.8a, which is also the circuit of Figure 2.6c, with $V = 100$ V.

Combining the 7 Ω and 8 Ω resistors in series produces the circuit of Figure 2.8b. Combining the 10 Ω and 15 Ω resistors in parallel produces the circuit of Figure 2.8c. Combining the 6 Ω and 4 Ω resistors in series produces the circuit of Figure 2.8d. Using the circuit of Figure 2.8d:

$$I_{ab} = \frac{100}{10} = 10\ A$$

Using the circuit of Figure 2.8c:

$$\Omega L: \quad V_{ab} = (4)I_4 = 40\ V$$

$$KVL: \quad V_{b0} = 100 - V_{ab} = 60\ V$$

$$\Omega L: \quad I_{b0} = \frac{V_{b0}}{10} = \frac{60}{10} = 6\ A$$

$$KCL: \quad I_{bc} = I_{ab} - I_{b0} = 10 - 6 = 4\ A$$

$$\Omega L: \quad V_{bc} = (7)I_{bc} = 28\ V$$

$$\Omega L: \quad V_{c0} = (8)I_{bc} = 32\ V$$

$$KCL: \quad I_{a0} = -I_{ab} = -10\ A$$

Summarizing results:

$$V_{ab} = 40\ V; \quad I_{ab} = 10\ A;$$

$$V_{b0} = 60\ V; \quad I_{b0} = 6\ A;$$

$$V_{bc} = 28\ V; \quad I_{bc} = 4\ A;$$

$$V_{c0} = 32\ V; \quad I_{c0} = 4\ A;$$

$$V_{a0} = 100\ V; \quad I_{a0} = -10\ A;$$

An analysis of the methodology used in Example 2.6 has general implications. We performed a comprehensive analysis of a dc circuit using only ΩL, KCL, and KVL, enhanced by series and parallel reductions. We refer to this approach as the "direct method." It is possible to perform a comprehensive analysis of *any* dc circuit using the direct method.[8]

We now state an extremely important network theorem.

Tellegen's theorem: Given a dc circuit with b branches, and P_k = the power absorbed by branch k:

$$\sum_{k=1}^{b} P_k = 0$$

Stated another way, in any dc circuit, the total power delivered by all the sources equals the total power absorbed by all the resistive branches.

Example 2.8

Show that the dc circuit of Figure 2.8a satisfies Tellegen's theorem, using results of Example 2.7.

$$V_{ab} = 40\ V; \quad I_{ab} = 10\ A; \quad P_{ab} = 400\ W$$

$$V_{b0} = 60\ V; \quad I_{b0} = 6\ A; \quad P_{b0} = 360\ W$$

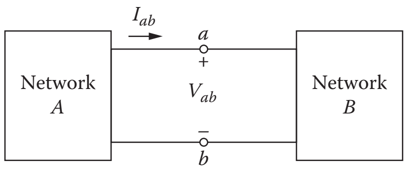
$$V_{bc} = 28\ V; \quad I_{bc} = 4\ A; \quad P_{bc} = 112\ W;$$

$$V_{c0} = 32\ V; \quad I_{c0} = 4\ A; \quad P_{c0} = 128\ W;$$

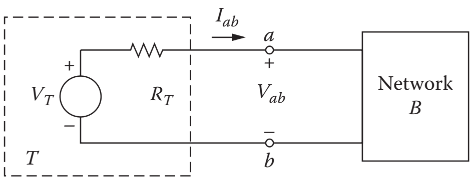
$$V_{a0} = 100\ V; \quad I_{a0} = -10\ A; \quad P_{a0} = -1000\ W;$$

$$\sum_{k=1}^{b} P_k = 400 + 360 + 112 + 128 - 1000 = 0$$

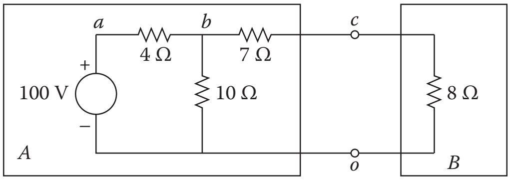
[8] The direct method of formulating the network equations can get complicated for medium and large networks, at which point we deploy more sophisticated methods, such as mesh current or node voltage methods. However, all advanced methods are simply organized ways of applying KCL, KVL, and ΩL, and aren't really necessary for our purposes, since the circuits we shall use are relatively simple.



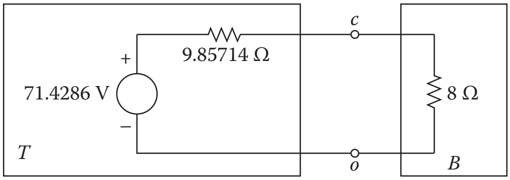
(a) Situation for Thevenin's theorem



(b) Replacing A with T



(c) Example Circuit



(d) Thevenin equivalent

FIGURE 2.9 **Thevenin equivalent circuit concepts.**

There is another important network theorem that merits our attention. *Thevenin's theorem* may be stated as follows:

> Given the situation illustrated in Figure 2.9a, Network A may be replaced with Network T, as shown in Figure 2.9b, such that all branch voltages and currents in Network B remain unchanged. Hence, Network T is said to be the "Thevenin equivalent (circuit)" of Network A at port a–b.

The obvious question that comes to mind is how do we determine the Thevenin elements V_T and R_T, given Networks A and B? There are several options, but the methodology we will use is as follows. First, to find the Thevenin voltage V_T :

- Disconnect B, producing an open circuit at port a–b for A. Compute the OC voltage at port a–b for A, calling it "V_{OC}." Recognize that $V_T = V_{OC}$.

We consider two methods for determining R_T . First the "look-in method."

- Disconnect B and turn off all voltage sources in A (replace them with shorts). Look into A at port a–b, and reduce the network to a single resistor R_{ab} (using series/parallel combinations). Recognize that $R_T = R_{ab}$.

Now the short-circuit method.

- Disconnect B and short A at port a–b. Compute the SC current at port a–b in A, calling it "I_{SC}." Recognize that $R_T = V_{OC}/I_{SC}$.

As always, an example will demonstrate.

Example 2.9

a. Calculate the port c–0 voltage and current using Network A in Figure 2.9c.
b. Determine the Thevenin equivalent at port c–0 for Network A, as shown in Figure 2.9c.
c. Calculate the port c–0 voltage and current using Network T. Comment.

Solution:

a. Observe that the circuits of Figure 2.8 and Figure 2.9 are identical. From Example 2.6:

$$I_{c0} = 4\ A$$

$$V_{c0} = 32\ V$$

b. First, we find the OC voltage in A. Disconnect B. The current I_{b0} is now zero. Therefore, the current I_{ab} is

$$I_{ab} = I_{b0} = \frac{100}{4+10} = 7.14286\ A$$

$$V_{b0} = 10 I_{b0} = 10(7.14286) = 71.4286\ V$$

$$V_{c0} = V_{b0} - (7) I_{bc} = 71.4286 - 0 = 71.4286\ V$$

$$V_T = V_{c0} = 71.4286\ V$$

We use the "look-in" method to find R_T. We see 7Ω in series with the parallel combination of 4Ω and 10Ω.

$$R_{c0} = 7 + \frac{10(4)}{4+10} = 9.85714\ \Omega$$

$$R_T = R_{c0} = 9.85714\ \Omega$$

Now use the "short circuit" method to find R_T. With port c–0 shorted, we see 4Ω in series with the parallel combination of 7Ω and 10Ω (looking right from the source terminals a–0):

$$R_{a0} = 4 + \frac{10(7)}{7+10} = 8.11765\ \Omega$$

$$\Omega L: \quad I_{ab} = \frac{V_{a0}}{R_{a0}} = \frac{100}{8.11765} = 12.3188\ A$$

$$KVL: \quad V_{b0} = V_{a0} - 4 \cdot I_{ab} = 100 - 4(12.3188) = 50.7246\ V$$

$$\Omega L: \quad I_{bc} = \frac{V_{b0}}{7} = \frac{50.7246}{7} = 7.24638\ A$$

Finally:

$$R_T = \frac{V_{OC}}{I_{SC}} = \frac{71.43}{7.246} = 9.85714\ \Omega$$

The "look-in" and SC methods check with each other. The Thevenin equivalent appears in Figure 2.9d.

C. Using Network T to calculate the port c–0 voltage and current,

$$I_{c0} = \frac{V_T}{R_T + 8} = \frac{71.4286}{8 + 9.85714} = 4\ A$$

$$V_{c0} = 8 \cdot I_{c0} = 32\ V$$

which checks the results in part a. In fact, if we changed Network B (the 8Ω resistor) to any other element (or combination of elements), the results would have checked! Of course, one example doesn't "prove" Thevenin's theorem. However, it should raise one's confidence in its validity.

For large dc networks, as we have noted, there are more sophisticated and efficient analytical methods of circuit analysis. However, these advanced methods are also based on systematic application of ΩL, KCL, and KVL. Since a comprehensive analysis of *all* dc circuits is possible using the direct method, we limit our study to this approach. The reader is admonished that the proficiency with which one can solve circuit problems is proportional to how many problems one attempts to solve. The problems provided at the end of Chapter 2 are sufficient to enable the student to reach a level of proficiency adequate to deal with the remainder of material in this book.

2.5 The ac Mode

The term "ac" historically stands for "alternating current," but it means "sinusoidal steady state" in contemporary usage. An "ac circuit" is a circuit operating in the ac mode. In ac circuits, all voltages (and currents) are of the following form:

$v(t) = V_{\max} \cos(\omega t + \alpha)$

where:

$V_{\max}$ = maximum (peak) voltage, V

ω = radian frequency, rad/s

t = time, s.

α = phase angle, formally radians, but frequently given in degrees.

Also:

$f = \omega/2\pi$ = cyclic frequency, hertz (Hz)

$T = 1/f$ = period, seconds (s).

Example 2.10

Given $v(t) = 169.7 \cos(377t + 50°)$, find $V_{\max}$; ω; α; f; and T

$V_{\max} = 169.7$ volts

$\omega = 377$ rad/s

$\alpha = 50°$

$f = \omega/2\pi = 377 / 2\pi = 60$ hertz (Hz)

$T = 1/f = 1/60 = 16.67$ ms

The RMS ("root mean square" or "effective") value for any periodic function $v(t)$ with period T is

$$V_{RMS} = \sqrt{\frac{1}{T}\int_{t_0}^{t_0+T} v(t)^2 \cdot dt}$$

Let us compute the RMS value of a sinusoidal function.

$$V_{RMS} = \sqrt{\frac{1}{T}\int_{t_0}^{t_0+T} \left\{V_{\max}\cos(\omega t + \alpha)\right\}^2 \cdot dt} = \frac{V_{\max}}{\sqrt{2}}$$

Example 2.11[9]

Given $v(t) = 169.7 \cos(377t + 50°)$, find V_{RMS}:

$$V_{RMS} = \frac{V_{\max}}{\sqrt{2}} = \frac{169.7}{\sqrt{2}} = 120\ V$$

When we specify an ac voltage (or current) as a single number we shall always default to the RMS value. For example, when we say that "the (ac) current is 10 amperes," we understand that the RMS value is 10 and:

$$i(t) = 10\sqrt{2} \cdot \cos(\omega t + \beta) = 14.14\ \cos(\omega t + \beta)$$

We will need more information to evaluate the frequency and phase.

[9] The details of the integration is left as a student exercise.

It happens that ac circuit problems can easily be solved using a technique called "phasor algebra." However, this requires an understanding of, and competency in, the arithmetic and algebra of complex numbers. It is highly recommended that readers review this topic before proceeding (see Section 1.5).

We now define a complex number called a "phasor," which applies to functions that vary sinusoidally with time, specifically voltages and currents.

Given $v(t) = \sqrt{2} \cdot V\cos(\omega t + \alpha)$, the corresponding phasor is:

$$\bar{V} = V\angle\alpha$$

Note that we shall use the RMS value for the magnitude.[10]

Example 2.12

Convert $v(t) = 169.7\cos(377t + 50°)$ to its phasor representation.
Convert $i(t) = 14.14\cos(377t - 30°)$ to its phasor representation.
Convert $\bar{I} = 15\angle 45°$ to its time domain representation.

Solution:

a. $\bar{V} = 120\angle 50°$

b. $\bar{I} = 10\angle -30°$

c. $i(t) = 21.21\cos(\omega t + 45°)$

Impedance

Next, we define a concept called "impedance." For the situation in Figure 2.10, the impedance of Network E is

$$\text{Impedance at port ab } = \bar{Z}_{ab} = \frac{\bar{V}_{ab}}{\bar{I}_{ab}}$$

[10] Some authors use the *maximum* value for the phasor magnitude.

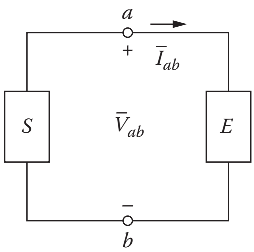


FIGURE 2.10 Defining $\bar{Z}$ and $\bar{Y}$.

Impedance of a Resistor

Suppose network E is a resistor R and

$$i_{ab}(t) = \sqrt{2} \cdot I \cos(\omega t + \beta)$$

We compute

$$v_{ab}(t) = R \cdot i_{ab}(t) = \sqrt{2} \cdot I \cdot R \cdot \cos(\omega t + \beta)$$

Converting to phasors:

$$\bar{V}_{ab} = I \cdot R \angle \beta \qquad \bar{I}_{ab} = I \angle \beta$$

Hence the impedance of a resistor is

$$\bar{Z}_R = \frac{\bar{V}_{ab}}{\bar{I}_{ab}} = \frac{I \cdot R \angle \beta}{I \angle \beta} = R + j0$$

Impedance of an Inductor

Suppose network E is an inductor and

$$i_{ab}(t) = \sqrt{2} I \cdot \cos(\omega t + \beta)$$

$$v_{ab}(t) = L \cdot \frac{di_{ab}}{dt} = \sqrt{2} IL \cdot \frac{d(\cos(\omega t + \beta))}{dt}$$

$$= -\sqrt{2} I \omega L \cdot \sin(\omega t + \beta) = \sqrt{2} I \omega L \cdot \cos(\omega t + \beta + 90°)$$

Converting to phasors:

$$\bar{V}_{ab} = I \cdot L \cdot \omega \angle \beta + 90^\circ$$

$$\bar{I}_{ab} = I \angle \beta$$

Hence, the impedance of an inductor is

$$\bar{Z}_L = \frac{\bar{V}_{ab}}{\bar{I}_{ab}} = \frac{I \cdot \omega L \angle (\beta + 90^\circ)}{I \angle \beta} = 0 + j\omega L$$

Impedance of a Capacitor

Suppose network E is a capacitor C and

$$v_{ab}(t) = \sqrt{2}V \cdot \cos(\omega t + \alpha)$$

$$i_{ab}(t) = C \cdot \frac{dv_{ab}}{dt} = \sqrt{2}VC \cdot \frac{d(\cos(\omega t + \alpha))}{dt}$$

$$i_{ab}(t) = -\sqrt{2}V\omega C \cdot \sin(\omega t + \alpha)$$

$$i_{ab}(t) = \sqrt{2}V\omega C \cdot \cos(\omega t + \alpha + 90^\circ)$$

Converting to phasors:

$$\bar{V}_{ab} = V \angle \alpha$$

$$\bar{I}_{ab} = V\omega C \angle \alpha + 90^\circ$$

Hence the impedance of a capacitor is

$$\bar{Z}_C = \frac{\bar{V}_{ab}}{\bar{I}_{ab}} = \frac{V \angle \alpha}{V\omega C \angle \alpha + 90^\circ} = \frac{1}{j\omega C} = 0 - j\left(\frac{1}{\omega C}\right)$$

There is a related concept called "admittance." For the situation in Figure 2.10, the admittance of Network E is:

$$\text{Admittance at port ab} = \bar{Y}_{ab} = \frac{\bar{I}_{ab}}{\bar{V}_{ab}} = \frac{1}{\bar{Z}_{ab}}$$

Impedances and admittances for the three passive elements are summarized in Table 2.1.

Table 2.1 Impedance and Admittance of the Three Passive Elements

Element	$\bar{Z}$	$\bar{Y}$
Resistor	$R + j0$	$\frac{1}{R} + j0$
Inductor	$0 + j\omega L$	$0 - j\left(\frac{1}{\omega L}\right)$
Capacitor	$0 - j\left(\frac{1}{\omega C}\right)$	$0 + j\omega C$

Consider an arbitrary circuit made of any number of the five basic elements (V source, I source, R, L, and C) arbitrarily interconnected. All sources must be sinusoidal, at the same frequency. To construct the ac circuit,

1. Replace all voltages and currents with their *phasor* equivalents.
2. Replace all passive elements (R, L, and C) with their impedances (or admittances).[11]

An example would be helpful.

Example 2.13

Consider the circuit of Figure 2.11a. Construct the corresponding ac circuit:

$$v(t) = 141.4\cos(377t) \quad V$$
$$\bar{V} = 100\angle 0°$$
$$\text{(radian) frequency} = \omega = 377 \ rad/s$$
$$R: \quad \bar{Z}_R = R + j0 = 8 + j0$$
$$L: \quad \bar{Z}_L = 0 + j\omega L = 0 + j(0.377)(26.53) = 0 + j10$$
$$C: \quad \bar{Z}_C = 0 + \frac{1}{j\omega C} = 0 - j\frac{1}{0.377(0.663)} = 0 - j4$$

[11] Either Z or Y can be used for each element. However, depending on the circuit topology, one form will generally be preferable to the other. Also, it is not recommended to mix the two in a given circuit. Usually one defaults to Z's unless there is a compelling reason to use Y's.

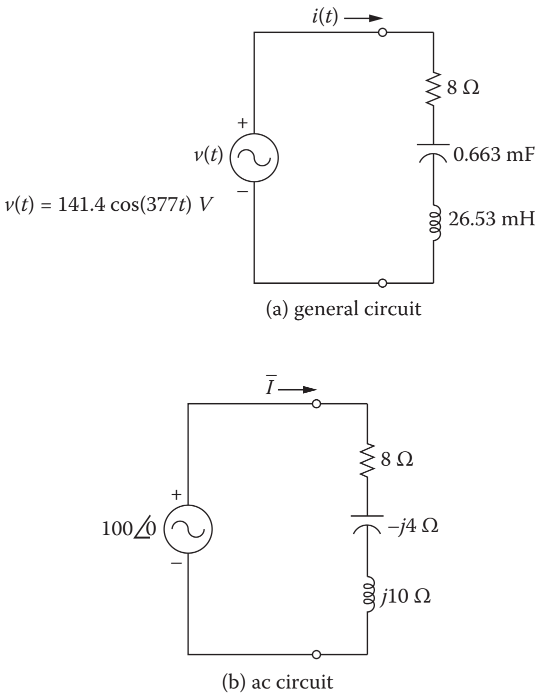


FIGURE 2.11 Circuits for Example 2.12.

The ac circuit is shown in Figure 2.11b.

We now define a more generalized version of Ohm's law for the ac case. From Figure 2.10,

$$\Omega L: \qquad \bar{V}_{ab} = \bar{Z}_{ab} \cdot \bar{I}_{ab}$$

Next, we restate *KCL* and *KVL* for the ac case. The modified current law states that "the sum of the *phasor* currents into any node in an *ac* circuit is zero."

$$KCL: \qquad \sum_{k=1}^{n} \bar{I}_{in} = 0$$

The modified voltage law states that "the sum of the *phasor* voltage drops around any closed path in an *ac* circuit is zero."

$$KVL: \qquad \sum_{k=1}^{n} \bar{V}_{ij} = 0$$

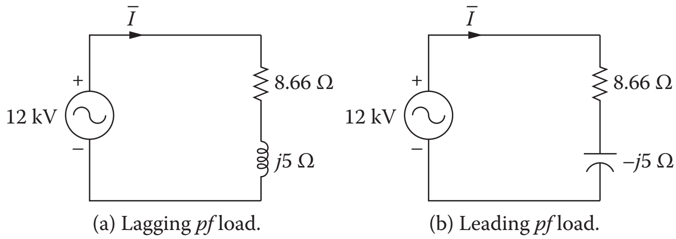


FIGURE 2.12 **Leading/lagging loads.**

We further generalize on our method of series and parallel combinations. Recall Figure 2.7a. Replace the *R*'s with *Z*'s. We can combine series *Z*'s according to

$$\overline{Z}_{ab} = \sum_{k=1}^{n} \overline{Z}_k$$

Recall Figure 2.7c. Replace the *G*'s with *Y*'s. We can combine parallel *Y*'s according to

$$\overline{Y}_{ab} = \sum_{k=1}^{n} \overline{Y}_k$$

As was in the dc case, we can combine two impedances in parallel by dividing their product by their sum:

$$\overline{Z}_{ab} = \frac{\overline{Z}_1 \cdot \overline{Z}_2}{\overline{Z}_1 + \overline{Z}_2}$$

With these changes, ac circuit analysis is nearly identical to dc analysis (except that all three passive elements are operational, and we must solve a set of *complex algebraic* equations). We now have the tools to pursue a comprehensive analysis of the ac circuit. The methodology will be demonstrated in Example 2.14.

Example 2.14

Perform a comprehensive analysis of the ac circuit of Figure 2.11b.

Using series combinations:

$$\overline{Z}_{ab} = (8 + j0) + (0 + j10) + (0 - j4) = 8 + j6 = 10\angle{+36.9^\circ}$$

Solving for the current:

$$\bar{I} = \frac{\bar{V}_{ab}}{\bar{Z}_{ab}} = \frac{100\angle 0^\circ}{10\angle +36.9^\circ} = 10\angle -36.9^\circ \ A$$

Solving for the RLC voltages:

$$\bar{V}_R = \bar{Z}_R\bar{I} = (8+j0)(10\angle -36.9^\circ) = 80\angle -36.9^\circ \ V$$

$$\bar{V}_L = \bar{Z}_L\bar{I} = (0+j10)(10\angle -36.9^\circ) = 100\angle 53.1^\circ \ V$$

$$\bar{V}_C = \bar{Z}_C\bar{I} = (0-j4)(10\angle -36.9^\circ) = 40\angle -126.9^\circ \ V$$

The voltage across, and the current through, all elements have now been determined.

Complex Power

For the ac case, power computations must be modified. We define the complex power absorbed by element k to be

$$\bar{S}_k = P_k + jQ_k = \bar{V}_k \cdot \bar{I}_k{}^* \quad \text{where}$$

$$\bar{V}_k = V_k\angle \alpha_k = \text{phasor voltage drop across element k}$$

$$\bar{I}_k = I_k\angle \beta_k = \text{phasor current through element k}$$

Continuing, we write,

$$\bar{S}_k = \bar{V}_k \cdot \bar{I}_k{}^* = (V_k\angle \alpha_k)\cdot(I_k\angle \beta_k)^* = V_k \cdot I_k\angle(\alpha_k - \beta_k)$$

$$\bar{S}_k = V_k \cdot I_k \cdot \cos(\theta_k) + jV_k \cdot I_k \cdot \sin(\theta_k) = P_k + jQ_k$$

where

$$\theta_k = \alpha_k - \beta_k$$

$$P_k = V_k \cdot I_k \cdot \cos(\theta_k) = \text{average ("real") power, watt}$$

$$Q_k = V_k \cdot I_k \cdot \sin(\theta_k) = \text{reactive ("imaginary") power, var}$$

We define "apparent power" (S) to be

$$S_k = |\overline{S}_k| = V_k \cdot I_k = \sqrt{P_k^2 + Q_k^2}$$

Now consider

$$
\begin{aligned}
P_{AVERAGE} &= \frac{1}{T}\int_{t_0}^{t_0+T} v(t)\cdot i(t)\cdot dt \\
&= \frac{1}{T}\int_{t_0}^{t_0+T} [V_{\max}\cos(\omega t+\alpha)\cdot I_{\max}\cos(\omega t+\beta)]\cdot dt \\
&= \frac{V_{\max}\cdot I_{\max}}{2}\cos(\theta) = \left(\frac{V_{\max}}{\sqrt{2}}\right)\left(\frac{I_{\max}}{\sqrt{2}}\right)\cos(\theta) = V\cdot I\cdot\cos(\theta)
\end{aligned}
$$

But this is P_k! In other words, the real part of the complex power is in fact the average power absorbed by component "k"! This gives us a physical interpretation to an abstract concept! To summarize, there are four "types" of power in ac circuits, which are

$$
\begin{aligned}
\overline{S}_k &= P_k + jQ_k = \text{complex power} \\
S_k &= \text{apparent power, VA} \\
P_k &= \text{average ("real") power, watt} \\
Q_k &= \text{reactive ("imaginary") power, var}
\end{aligned}
$$

Only one (P_k) has physical meaning. However, all four are important and useful, in the contexts of both ac circuit analysis and engineering applications. We apply these power concepts to a specific circuit in Example 2.15.

Example 2.15

Continuing the analysis of the ac circuit of Example 2.14 (Figure 2.11b), find the powers absorbed by each component.

$$\overline{S}_R = \overline{V}_R\cdot\overline{I}_R{}^* = (80\angle{-36.9^\circ})(10\angle{-36.9^\circ})^* = 800\angle 0^\circ = 800 + j0$$

$$\overline{S}_L = \overline{V}_L\cdot\overline{I}_L{}^* = (100\angle 53.1^\circ)(10\angle{-36.9^\circ})^* = 1000\angle 90^\circ = 0 + j1000$$

$$\overline{S}_C = \overline{V}_C\cdot\overline{I}_C{}^* = (400\angle{-126.9^\circ})(10\angle{-36.9^\circ})^* = 400\angle{-90^\circ} = 0 - j400$$

$$\overline{S}_{LOAD} = \overline{S}_R + \overline{S}_L + \overline{S}_C = (800 + j0) + (0 + j1000) + (0 - j400) = 800 + j600$$

The power *delivered* by the source is:

$$\overline{S}_{SOURCE} = \overline{V} \cdot \overline{I}^* = (100\angle 0^\circ)(10\angle -36.9^\circ)^* = 1000\angle 36.9^\circ = 800 + j600$$

Example 2.15 illustrates some important general principles. Observe that the resistor absorbs only real power (P). Likewise, the inductor absorbs only reactive power (Q), and the capacitor absorbs only **negative** reactive power (Q). Absorbing *negative* Q means that the capacitor is actually a Q *source*, a fact that is useful in power system design and operation. This is true in general: in every ac circuit, all R's absorb "P" only, all L's absorb "Q" only, and all C's deliver "Q" only.

Tellegen's Theorem Restated for the ac Circuit

Given an ac circuit with "b" branches, and:

$$\overline{S}_k = \overline{V}_k \cdot \overline{I}_k^* = P_k + jQ_k = \text{complex power absorbed by branch } k$$

Tellegen's theorem states that

$$\sum_{k=1}^{b} \overline{S}_k = 0 \qquad \sum_{k=1}^{b} P_k = 0 \qquad \sum_{k=1}^{b} Q_k = 0$$

Example 2.16

Show that Tellegen's theorem is satisfied in the ac circuit of Example 2.14 (Figure 2.11b).

From Example 2.15:

$$\overline{S}_R = \overline{V}_R \cdot \overline{I}_R^* = 800 + j0$$

$$\overline{S}_L = \overline{V}_L \cdot \overline{I}_L^* = 0 + j1000$$

$$\overline{S}_C = \overline{V}_C \cdot \overline{I}_C^* = 0 - j400$$

$$\overline{S}_{SOURCE} = \overline{V} \cdot \overline{I}^* = 800 + j600$$

According to Tellegen's theorem,

$$\sum_{k=1}^{b} \bar{S}_k = (800 + j0) + (0 + j1000) + (0 - j400) + (-800 - j600)$$

$$\sum_{k=1}^{b} \bar{S}_k = (800 + 0 + 0 - 800) + j(0 + 1000 - 400 - 600)$$

$$\sum_{k=1}^{b} \bar{S}_k = (0) + j(0)$$

Power Factor

The engineering applications for which complex power concepts are most useful and common are those related to the electric power utility business. The term "power factor" is commonly used in conjunction with power utility loads, and is defined as

$$\text{power factor} = pf = \frac{P_{LOAD}}{S_{LOAD}} = \frac{V \cdot I \cos(\theta)}{V \cdot I} = \cos(\theta)$$

Power factor is qualified as being either lagging or leading, the former meaning that the current *lags* the voltage (as would be the case for inductive loads) and the latter meaning that the current *leads* the voltage (as would be the case for capacitive loads), as shown in Figure 2.12.

Example 2.17

Consider the situation in Figure 2.12a.

A. The lagging (inductive) case:

$$\bar{I} = \frac{12\angle 0°}{8.66 + j5} = 1.2\angle -30° \ kA$$

$$\theta = \alpha - \beta = (0) - (-30°) = +30°$$

$$\bar{S}_{LOAD} = \bar{V} \cdot \bar{I}^* = (12\angle 0^\circ)(1.2\angle -30^\circ)^* = 14.4\angle 30^\circ = 12.47 + j7.2$$

$$S_{LOAD} = 14.40\ MVA \quad P_{LOAD} = 12.47\ MW \quad Q_{LOAD} = 7.2\ M\,\text{var}$$

$$pf = \cos(+30^\circ) = 0.866 \text{ lagging}$$

B. The leading (capacitive) case:

$$\bar{I} = \frac{12\angle 0^\circ}{8.66 - j5} = 1.2\angle +30^\circ\ kA$$

$$\theta = \alpha - \beta = (0) - (30^\circ) = -30^\circ$$

$$\bar{S}_{LOAD} = \bar{V} \cdot \bar{I}^* = (12\angle 0^\circ)(1.2\angle +30^\circ)^* = 14.4\angle -30^\circ = 12.47 - j7.2$$

$$S_{LOAD} = 14.40\ MVA \quad P_{LOAD} = 12.47\ MW \quad Q_{LOAD} = -7.2\ M\,\text{var}$$

$$pf = \cos(+30^\circ) = 0.866 \text{ leading}$$

Several points were made in Example 2.17. The term “lagging” means that the current is lagging the voltage in phase (i.e., $\theta > 0$), which happens if the load is inductive, and $Q > 0$. The term “leading” means that the current is leading the voltage in phase (i.e., $\theta < 0$), which happens if the load is capacitive, and $Q < 0$.

Thevenin’s theorem easily extends to the ac case. We need only replace the Thevenin voltage with the Thevenin *phasor* voltage and the Thevenin resistance with the Thevenin *complex impedance*. The rules for computing the Thevenin elements are the same, but using ac analysis of course.

We now know that if the sources in an electric circuit are not changing in time, all branch voltages and currents will be constant, a condition called “dc,” and the application of KCL, KVL, and ΩL results in a set of linear algebraic equations. Furthermore, if the sources in an electric circuit vary sinusoidally in time at the same frequency, all branch *phasor* voltages and currents will be constant, a condition called “ac,” and the application of KCL, KVL, and WL results in a set of linear *complex* algebraic equations. Both of these are steady-state conditions: the branch voltages and currents are constant (the dc case), or the branch *phasor* voltages and currents are constant (the ac case).

But what if a circuit is excited from a voltage or current source that is periodic *but not sinusoidal*? Now how do we solve for all branch voltages and currents? This leads to a third mode in which a circuit may operate, which we call the "periodic mode."

2.6 The Periodic Mode

Consider a voltage $v(t)$ defined such that

$$v(t) = v(t \pm nT); \quad 0 \le n \le \infty$$

Such a repeating waveform is said to be "periodic with period T" and is illustrated in Figure 2.13.[12]

A related concept, frequency (f), is defined as

$$f = \frac{1}{T}$$

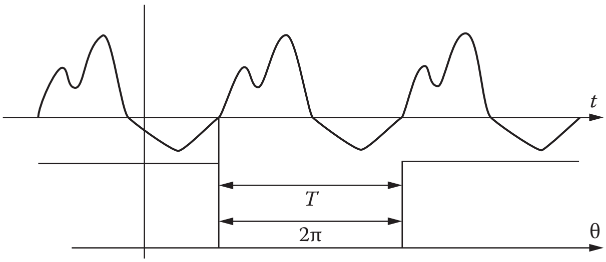


FIGURE 2.13 Periodic waveform.

[12] Actually T is the *smallest* value for which the repeating property holds. Any function periodic in T is also periodic in 2T, 3T, and so on. Also, functions can be periodic in any independent variable (distance, for example). Our interest is restricted to periodicity in time. Furthermore, the function need not be restricted to voltage. Any function that varies in time may be periodic.

We shall have occasion to convert the repeating interval in time (T) to angle measure (θ), the latter measured in radians, such that "T" seconds correspond to "2π" radians. As such, we define radian frequency (ω):

$$\omega = \frac{2\pi}{T} = 2\pi f$$

$$\theta = \omega \cdot t = \left(\frac{2\pi}{T}\right) \cdot t$$

Example 2.18

Consider the waveform $v(t)$ of Figure 2.13 such that T = 10 ms. Find f and ω.

$$f = \frac{1}{T} = \frac{1}{0.01} = 100 \text{ Hz}$$

$$\omega = 2\pi f = 628.3 \text{ rad/s}$$

Jean Baptiste Fourier (1768–1830) showed in an 1807 paper that a periodic $v(t)$ could be approximated by the summation of sinusoids of different frequencies which are integer multiples of a base, or fundamental, frequency:

$$v^*(t) = V_0 + \sum_{n=1}^{N} V_n\sqrt{2} \cdot \cos(n\omega_0 t + \phi_n)$$

where

V_0 = dc voltage

$v_n(t) = V_n\sqrt{2} \cdot \cos(n\omega_0 + \phi_0) = n^{\text{th}}$ harmonic voltage

$f_n = n \cdot f_0 = \text{n}^{\text{th}}$ harmonic frequency; $\omega_n = n \cdot (2\pi f_0)$

Observe that

$$\lim_{N\to\infty}\left\{v^*(t)\right\} = v(t)$$

This has become known as "Fourier's theorem" and is perhaps the single most important concept in signal processing. The constant term (V_0) is called the "dc" term; and each sinusoidal term is called a "harmonic." The first harmonic (n = 1) is called the "fundamental," and has the frequency of $v(t)$.

In fact, Fourier's theorem is so important that it bears repeating:

Given a periodic function $v(t)$ with period "T" and frequency f = 1/T, $v(t)$ can be approximated with function $v^*(t)$, where $v^*(t)$ is expressed as a summation of sinusoids, such that

$$v^*(t) = V_0 + \sum_{n=1}^{N} V_n\sqrt{2}\cdot\cos(n\omega_0 t + \phi_n)$$

where

$$f_0 = f; \quad \omega_0 = 2\pi f_0$$

$$\lim_{N\to\infty}\left\{v^*(t)\right\} = v(t)$$

Example 2.19

Consider the waveform v(t) of Example 2.18. Find the first, second, and third harmonic frequencies.

$$f_1 = f_0 = f = 100\text{ Hz}$$

$$f_2 = 200\text{ Hz}$$

$$f_3 = 300\text{ Hz}$$

Given a periodic $v(t)$, it is straightforward to compute the coefficients in the Fourier series from the following expressions:

$$v^*(t) = V_0 + \sum_{n=1}^{N} V_n \sqrt{2} \cdot \cos(n\omega_0 t + \phi_n)$$

where

$$V_0 = \text{dc voltage} = \frac{1}{T} \int_{-T/2}^{T/2} v(t) \cdot dt$$

$$A_n = \frac{2}{T} \int_{-T/2}^{T/2} v(t) \cdot \cos(n\omega_0 t) \cdot dt$$

$$B_n = \frac{2}{T} \int_{-T/2}^{T/2} v(t) \cdot \sin(n\omega_0 t) \cdot dt$$

$$C_n = \sqrt{A_n^2 + B_n^2}$$

$$\phi_n = Tan^{-1}\left(\frac{-B_n}{A_n}\right)$$

$$V_n = \frac{C_n}{\sqrt{2}}$$

$$f_0 = \frac{1}{T} = \text{fundamental frequency;}$$

$$f_n = n \cdot f_0 = n^{\text{th}} \text{ harmonic frequency;}$$

$$\omega_n = n \cdot (2\pi f_0) = n^{\text{th}} \text{ harmonic radian frequency.}$$

However, since our purpose is to determine the response of linear circuits to periodic excitation, we will relegate calculations of the coefficients to the mathematics department, and assume that the Fourier coefficients are known. Example 2.20 addresses the issue.

Example 2.20

Consider the waveform $v(t)$ defined by

$$v(t) = 200\ V \qquad 0 \le t \le 0.6T$$

$$v(t) = 0\ V \qquad 0.6\mathrm{T} \le t \le 1.0T$$

Determine the first twenty-one harmonics of the Fourier series of $v(t)$, plotting $v(t)$ and $v^*(t)$. Also plot the frequency spectrum of $v^*(t)$.

$v(t)$ is plotted in Figure 2.14a, as well as is $v^*(t)$, the Fourier approximation, using the first twenty-one harmonics. The Fourier coefficients were computed using a program called FSAP, and are presented in Table 2.2.

Table 2.2 Fourier Coefficients

Harm	An	Bn	Cn	Dn	Rms	phin
0	120.000	0.000	120.000	120.000	120.000	0.0
1	–37.420	115.167	121.094	60.547	85.626	–108.0
2	30.275	21.996	37.421	18.711	26.461	–36.0
3	–20.184	14.665	24.949	12.475	17.642	–144.0
4	9.357	28.797	30.279	15.140	21.411	–72.0
6	–6.239	19.203	20.191	10.096	14.277	–108.0
7	8.655	6.288	10.698	5.349	7.565	–36.0
8	–7.574	5.503	9.362	4.681	6.620	–144.0
9	4.162	12.809	13.469	6.734	9.524	–72.0
11	–3.407	10.486	11.025	5.513	7.796	–108.0
12	5.055	3.672	6.248	3.124	4.418	–36.0
13	–4.667	3.391	5.769	2.885	4.079	–144.0
14	2.679	8.247	8.671	4.336	6.131	–72.0
16	–2.346	7.221	7.593	3.796	5.369	–108.0
17	3.575	2.597	4.418	2.209	3.124	–36.0
18	–3.378	2.454	4.175	2.087	2.952	–144.0
19	1.979	6.089	6.403	3.201	4.527	–72.0
21	–1.792	5.515	5.799	2.899	4.100	–108.0

Translating from FSAP to our notation:

$$V_n = Rms \quad \phi_n = phin \quad n = Harm$$

For example, the third harmonic is

$$v_3(t) = 17.64\sqrt{2} \cdot \cos(3\omega_0 t - 144.0°)$$

The frequency spectrum is plotted in Figure 2.14b.

We recall that linear circuits have a mathematical property known as "linearity," which means that the "principle of superposition" is applicable. The principle of superposition may be stated as follows:

> The response of a linear system to a number of different inputs may be computed by applying the inputs one at a time, computing the response to each input, and finally superimposing (or adding together) the individual responses.

Suppose the input to a linear circuit is $v(t)$ and the desired response is $i(t)$. Further suppose that

$$v(t) = V_0 + \sum_{n=1}^{N} V_n\sqrt{2} \cdot \cos(n\omega_0 t + \alpha_n)$$

According to superposition, we can compute the response one frequency at a time. We apply

$$v_0(t) = V_0$$

and compute the response. This is a dc circuits problem, so the dc response is

$$i(t) = i_0(t) = I_0$$

Next, we apply the fundamental:

$$v(t) = v_1(t) = V_1\sqrt{2} \cdot \cos(\omega_0 t + \alpha_1)$$

And compute the response. This is an ac circuits problem, so:

$$i(t) = i_1(t) = I_1\sqrt{2} \cdot \cos(\omega_0 t + \beta_1)$$

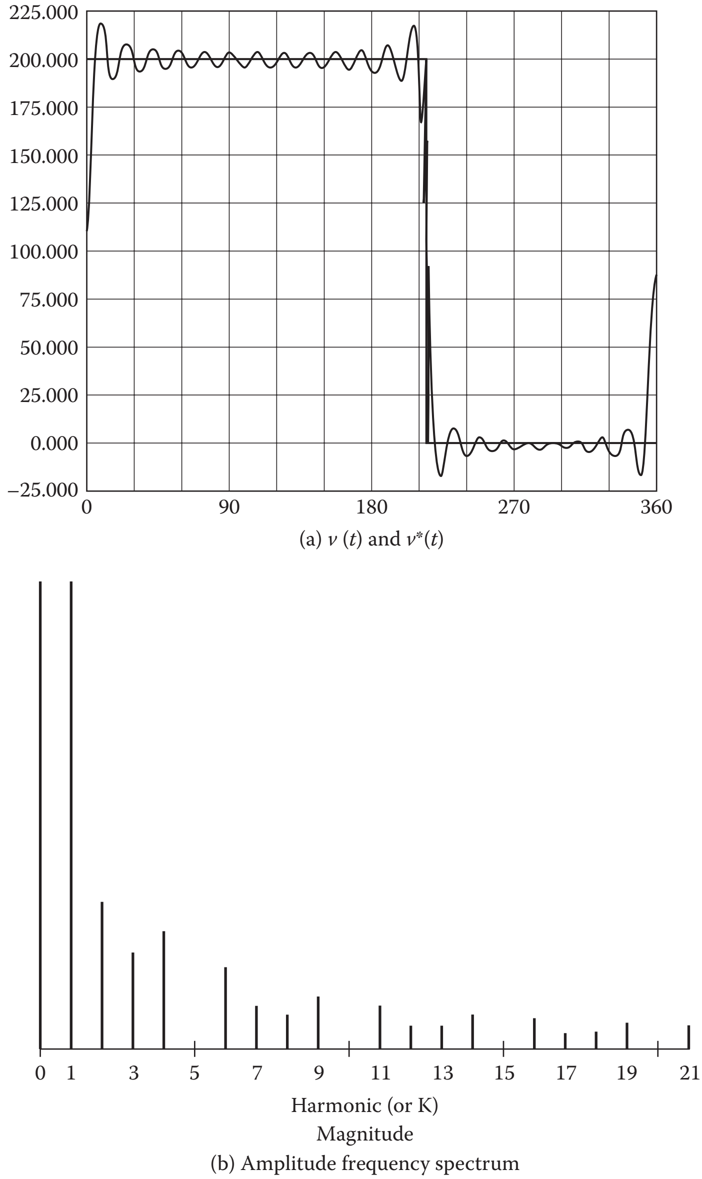


FIGURE 2.14 Plots for Example 2.20.

We continue this approach out to the N^{th} harmonic. That done, we employ superposition and add up all the harmonic responses:

$$i(t) = I_0 + \sum_{n=1}^{N} I_n \sqrt{2} \cdot \cos(n\omega_0 t + \beta_n)$$

To summarize, we work on the dc circuits problem and "N" ac circuits problems, and use superposition to add up the responses. The methodology will be demonstrated using a specific example.

Example 2.21

Continuing Example 2.20, suppose $v(t)$ has a period of T = 10 ms. Compute the fundamental, third harmonic, and twenty-first harmonic frequencies:

$$f_0 = \frac{1}{T} = 100 \text{ Hz}; \quad \omega_0 = 2\pi f_0 = 628.3 \text{ rad/s}$$

$$f_3 = 3f_0 = 300 \text{ Hz}; \quad \omega_3 = 2\pi f_3 = 1.885 \text{ krad/s}$$

$$f_{21} = 21 f_0 = 2100 \text{ Hz}; \quad \omega_{21} = 2\pi f_{21} = 13.195 \text{ krad/s}$$

Example 2.22

Given $v^*(t)$ as determined in Example 2.20 is applied to the circuit of Figure 2.15. Determine the Fourier Series for $i^*(t)$. Plot $i^*(t)$.

The dc term:

$$I_0 = \frac{V_0}{R} = \frac{120}{10} = 12\ A$$

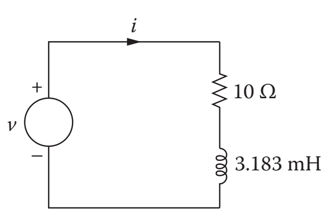


FIGURE 2.15 Circuit for Example 2.19.

The fundamental:

$$v_1(t) = 85.63\sqrt{2}\cdot\cos(\omega_0 - 108°) \quad \bar{V}_1 = 85.63\angle{-108°}$$

$$\bar{I}_1 = \frac{\bar{V}_1}{R + j\omega_0 L} = \frac{85.63\angle{-108°}}{10 + j2} = \frac{85.63\angle{-108°}}{10.2\angle{11.3°}} = 8.396\angle{-119.3°}\ A$$

$$i_1(t) = 8.396\sqrt{2}\cdot\cos(\omega_0 - 119.3°)$$

The second harmonic:

$$v_2(t) = 26.46\sqrt{2}\cdot\cos(2\omega_0 - 36°) \quad \bar{V}_2 = 26.46\angle{-36°}$$

$$\bar{I}_2 = \frac{\bar{V}_2}{R + j2\omega_0 L} = \frac{26.46\angle{-36°}}{10 + j4} = \frac{26.46\angle{-36°}}{10.77\angle{21.8°}} = 2.457\angle{-57.8°}\ A$$

$$i_2(t) = 2.457\sqrt{2}\cdot\cos(\omega_0 - 57.8°)$$

The rest of the harmonic currents appear in Table 2.3. A plot of v^* and i^* is provided in Figure 2.16.

The second harmonic:

Fundamental Frequency: f_o, w_o = 100.0 Hz; 628.3 rad/s

R = 10.000 ohm;

L = 3.183mH; XL@ fo = 2.000 ohm

The procedure is general. We can compute any branch current or voltage in any linear circuit subject to periodic excitation using the following methodology:

1. Compute the Fourier series of the excitation.
2. a. Draw the dc circuit (L's are shorts; C's are opens).
   b. Apply the dc excitation (the first term in the Fourier series).
   c. Solve for the desired branch currents and/or voltages using dc circuit analysis.

Table 2.3 Results of Circuit Analysis

Harm	Vrms @	Vang	Irms @	Iang	Zmag @	Zang	P
0	120.000	0.0	12.000	0.0	10.000	0.0	1440.00
1	85.627	−108.0	8.396	−119.3	10.198	11.3	704.99
2	26.461	−36.0	2.457	−57.8	10.770	21.8	60.36
3	17.642	−144.0	1.513	−175.0	11.662	31.0	22.89
4	21.411	−72.0	1.672	−110.7	12.806	38.7	27.95
6	14.277	−108.0	0.914	−158.2	15.620	50.2	8.35
7	7.565	−36.0	0.440	−90.5	17.204	54.5	1.93
8	6.620	−144.0	0.351	−202.0	18.868	58.0	1.23
9	9.524	−72.0	0.463	−132.9	20.591	60.9	2.14
11	7.796	−108.0	0.323	−173.6	24.165	65.6	1.04
12	4.418	−36.0	0.170	−103.4	25.999	67.4	0.29
13	4.079	−144.0	0.146	−213.0	27.856	69.0	0.21
14	6.131	−72.0	0.206	−142.3	29.731	70.3	0.43
16	5.369	−108.0	0.160	−180.6	33.525	72.6	0.26
17	3.124	−36.0	0.088	−109.6	35.439	73.6	0.08
18	2.952	−144.0	0.079	−218.5	37.362	74.5	0.06
19	4.527	−72.0	0.115	−147.3	39.293	75.3	0.13
21	4.100	−108.0	0.095	−184.6	43.173	76.6	0.09

Rms values V, I: 154.326, 15.075. True pf = 0.9768.
Powers S, P, Q, and D: 2326.40, 2272.43, 226.71, and 443.59.

Note: Series circuit analysis: fundamental frequency: fo, wo = 100.0 Hz; 628.3 rad/s; *R* = 10.000 ohm; and *L* = 3.183 mH; XL @ fo = 2.000 ohm.

3. a. Draw the ac circuit at the fundamental frequency.
   b. Convert the fundamental excitation to a phasor.
   c. Solve for the desired phasor branch currents and/or voltages using ac circuit analysis.
   d. Convert the phasors back to instantaneous values.
4. Repeat step 3 for the i^{th} harmonic; $2 \le i \le N$.
5. Superimpose (add together) the dc and harmonic responses.

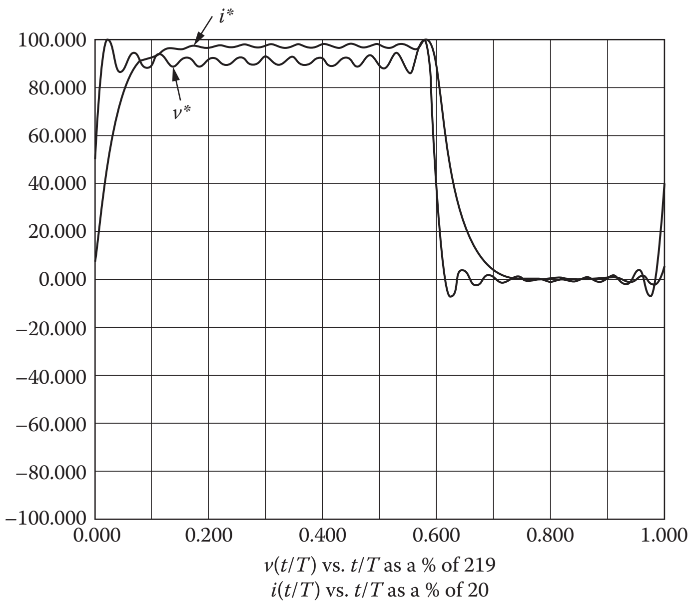


FIGURE 2.16 v* and i* for the circuit of Example 2.22.

There is the issue of how many harmonics we should use to get a "good" approximation to $v(t)$. The answer depends on the purpose to which the result will be put.

One "general purpose" rule is to compute the rms values of $v(t)$ and $v^*(t)$ and see how close they are. For example, applying this criterion to Example 2.20, we get

$$V^2 = 24000 \qquad V = 154.92 \quad \text{(exact value)}$$

$$(V^*)^2 = 23678 \qquad V^* = 153.89 \quad \text{(12 harmonics)}$$

$$(V^*)^2 = 23816 \qquad V^* = 154.32 \quad \text{(21 harmonics)}$$

Hence we are tempted to claim that we have captured 153.89/154.92 = 99.34% and 154.32/154.92 = 99.61% using twelve and twenty-one harmonics, respectively. Ultimately the number of harmonics needed is based on experience with a particular application.

2.7 The Transient Mode

The "transient" mode means that a circuit is operating in steady-state condition (the initial state), when at a given point in time (e.g., $t = 0$) a switch is thrown, after which all voltages and currents vary in time until they settle into a second steady-state condition (the final state). The trajectories in time taken by the voltages and currents form what is called the "transient solution." For this general condition, the L and C elements are not trivial, nor can we use the expediencies of dc and/or phasor analysis. We must use the general R,L,C models presented in Figure 2.3, which introduces integrals and derivatives.

Hence, the application of KCL and KVL results in a set of linear differential equations, which we subsequently must solve. There are two basic approaches to the solution. Since the system is linear, we can convert the problem into the frequency domain using Laplace transforms. This is what is normally done, provided the excitations are described by functions which are Laplace-transformable, which is usually the case. The second approach is to use the method of state variables, which is even more general.

A complete investigation of the general approach is beyond the scope of this book. However, there are two important and simple situations that are sufficient to demonstrate the general approach. These are sometimes called "transient" problems, and we will work through each case.

To fully understand the transients problem, we must introduce a new element: the (ideal) switch, as shown in Figure 2.17. Note that, simple as it is, the switch is a two-state device, and thus *nonlinear*, a situation we've never encountered before. Switches in effect partition a circuit into two circuits, each of which remains linear. One can think of a switch as a resistor, such that

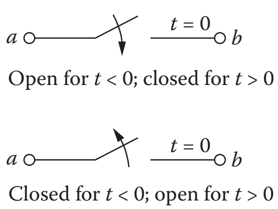


FIGURE 2.17 The ideal switch.

$R = 0$ (short) when closed, and $R = \infty$ (open) when opened. Of course, physical switches are never perfect shorts or perfect opens. Also, note that the ideal switch changes state in zero time, whereas physical switches may be fast but always require finite time to operate. Finally, note that the state of the ideal switch is undefined at $t = 0$. Switches can have multiple operations at any user-defined times.

We choose to constrain the transient circuits problem as follows:

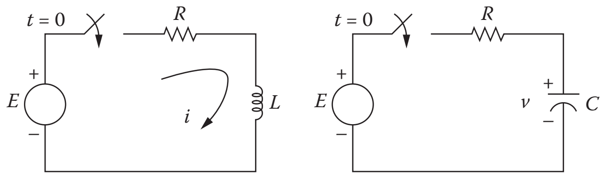
- All sources are dc.
- There is only one switch, which always operates at $t = 0$.
- There can be only one energy storage element (L or C).
- There must be at least one resistive (R) element.
- The current through inductors must be continuous in time; likewise, the voltage across capacitors must be continuous.[13]

If the energy storage element is inductance (L), we call this the "*RL* transients problem," and we normally solve for the inductor current.[14] If the energy storage element is capacitance (C), we call this the "*RC* transients problem" and normally solve for the capacitor voltage.[15]

The RL Transients Problem

Consider the circuit of Figure 2.18a. Applying KVL for $t > 0$:

$$E = R \cdot i + L \cdot \frac{di}{dt}$$



(a) The basic RL circuit (b) The basic RC circuit

FIGURE 2.18 **RL and RC transients.**

[13] See problem 2.25 for a defense of this point.

[14] Once the inductor current is known, we can solve for all other branch voltages and currents.

[15] Once the capacitor voltage is known, we can solve for all other branch voltages and currents.

Taking the Laplace transform:

$$\frac{E}{s} = R \cdot I(s) + L \cdot \{sI(s) - i(0)\}$$

$$(Ls + R)I(s) = \frac{E}{s} + Li(0) = \frac{E + sLi(0)}{s}$$

$$I(s) = \frac{E/L + si(0)}{s(s + R/L)} = \frac{E/R}{s} + \frac{-E/R + i(0)}{s + R/L}$$

Taking the inverse Laplace transform:

$$i(t) = \frac{E}{R} - \left(\frac{E}{R} - i(0)\right) \cdot e^{-t/\tau} \qquad \tau = \frac{L}{R}$$

Consider the final value of the current:

$$\lim_{t \to \infty}\left\{i(t)\right\} = \frac{E}{R} = i(\infty)$$

Substituting:

$$i(t) = i(\infty) - (i(\infty) - i(0)) \cdot e^{-t/\tau}$$

This is the general solution to the *RL* transients problem, and suggests the following methodology:

1. Draw the dc circuit for $t < 0$: the inductor is a short, and the switch is either a short or an open. Solve for the inductor current, which becomes $i(0^-) = i(0) = i(0^+)$.
2. The switch operates (changing state, going from a short to an open or vice versa). Looking into the circuit from the inductor's terminals, reduce the circuit to a single resistor (R) (turn off sources; *V*'s are shorts, and *I*'s are opens). Calculate $\tau = L/R$.
3. Draw the dc circuit for $t = \infty$ (the inductor is a short). Solve for the inductor current, which is $i(\infty)$.
4. Substitute values into

$$i(t) = i(\infty) - (i(\infty) - i(0)) \cdot e^{-t/\tau}$$

and simplify. An example will demonstrate.

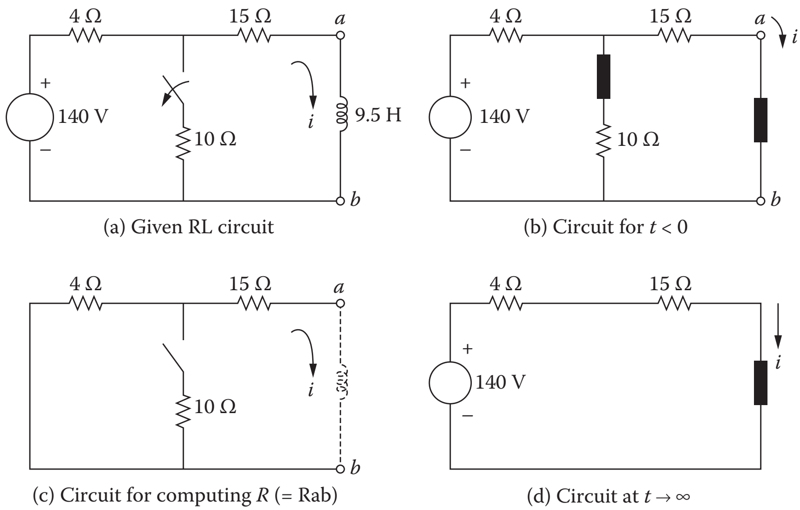


FIGURE 2.19 RL circuit for Example 2.20.

Example 2.23

Given the circuit of Figure 2.19a. Find and sketch the inductor current (i) over the range $-\tau < t < 5\tau$.

1. Draw the dc circuit for $t < 0$ (the inductor and switch are shorts). See the circuit of Figure 2.19b. We note that the 4 Ω resistor is in series with the parallel combination of the 15 Ω resistor and 10 Ω resistor. Solving for the **source** current:

$$i_S = \frac{140\ V}{4 + \dfrac{10 \cdot 15}{10 + 15}} = \frac{140\ V}{10} = 14\ A$$

Solving for the voltage across the 15 Ω resistor:

$$KVL: \quad v_{10} = 140 - 4 \cdot 14 = 84\ V = v_{15}$$

Solving for the inductor current:

$$\Omega L: \quad i = \frac{v_{15}}{15} = \frac{84}{15} = 5.6\ A = i(0^-) = i(0) = i(0^+)$$

2. The switch operates (opens). Looking out from the inductor's terminals, reduce the circuit to a single resistor (R). See Figure 2.19c. Calculate $\tau = L/R$.

$$R = 15 + 4 = 19\ \Omega$$

$$\tau = \frac{L}{R} = \frac{9.5\ H}{19\ \Omega} = 0.5\ s$$

3. Draw the dc circuit for $t = \infty$ (the inductor is a short). See Figure 2.19d. Solve for the inductor current, which is $i(\infty)$.

$$\Omega L: \quad i = \frac{140}{19} = 7.368\ A = i(\infty)$$

4. Substituting values:

$$i(t) = 7.368 - 1.768 \cdot e^{-2t}\ A$$

The RC Transients Problem

Consider the circuit of Figure 2.18b. Applying KCL for $t > 0$:

$$I = G \cdot v + C \cdot \frac{dv}{dt}$$

Taking the Laplace transform:

$$\frac{I}{s} = G \cdot V(s) + C \cdot \{sV(s) - v(0)\}$$

$$(Cs + G)V(s) = \frac{I}{s} + Cv(0) = \frac{I + sCv(0)}{s}$$

$$V(s) = \frac{I/C + sv(0)}{s(s + G/C)}$$

$$V(s) = \frac{I/G}{s} + \frac{-I/G + v(0)}{s + G/C}$$

Taking the inverse Laplace transform:

$$v(t) = \frac{I}{G} - \left(\frac{I}{G} - v(0) \right) \cdot e^{-G\,t/C}$$

$$v(t) = \frac{I}{G} - \left(\frac{I}{G} - v(0) \right) \cdot e^{-t/\tau} \qquad \tau = \frac{C}{G}$$

Converting from conductance (G) to resistance (R = 1/G):

$$v(t) = RI - (RI - v(0)) \cdot e^{-t/\tau} \qquad \tau = RC$$

Consider the final value of the voltage:

$$\lim_{t \to \infty} \{v(t)\} = RI = v(\infty)$$

Substituting:

$$v(t) = v(\infty) - (v(\infty) - v(0)) \cdot e^{-t/\tau}$$

This is the general solution to the RC transients problem, and suggests the following methodology:

1. Draw the dc circuit for $t < 0$: the capacitor is an open, and the switch is either a short or an open. Solve for the capacitor voltage, which is $v(0^-) = v(0) = v(0^+)$.
2. The switch operates (changing state, going from a short to an open or vice versa). Looking into the circuit from the capacitor's terminals, reduce the circuit to a single resistor (R) (turn off sources; V's are shorts, and I's are opens). Calculate $\tau = RC$.
3. Draw the dc circuit for $t = \infty$ (the capacitor is an open). Solve for the capacitor voltage, which is $v(\infty)$.
4. Substitute values into

$$v(t) = v(\infty) - (v(\infty) - v(0)) \cdot e^{-t/\tau}$$

and simplify. An example will demonstrate.

Example 2.24

Given the circuit of Figure 2.20a, find and sketch the capacitor voltage (v) over the range $-\tau < t < 5\tau$.

1. Draw the dc circuit for $t < 0$ (the capacitor is open, and the switch a short). See the circuit of Figure 2.20b. Solving for the source current:

$$KVL: \quad 20 = 5i_S + 2i_S - 50$$

$$i_S = \frac{70}{7} = 10\ A$$

Solving for the capacitor voltage:

$$KVL: \quad 50 - 2i_S + v = 0$$

$$v = -50 + 2(10) = -30\ V = v(0^-) = v(0) = v(0^+)$$

2. Looking into the circuit from the capacitor's terminals with the sources off:

$$R_{ab} = R = 5\Omega$$

$$\tau = RC = 5 \cdot (0.1) = 0.5\ s$$

See Figure 2.20c.

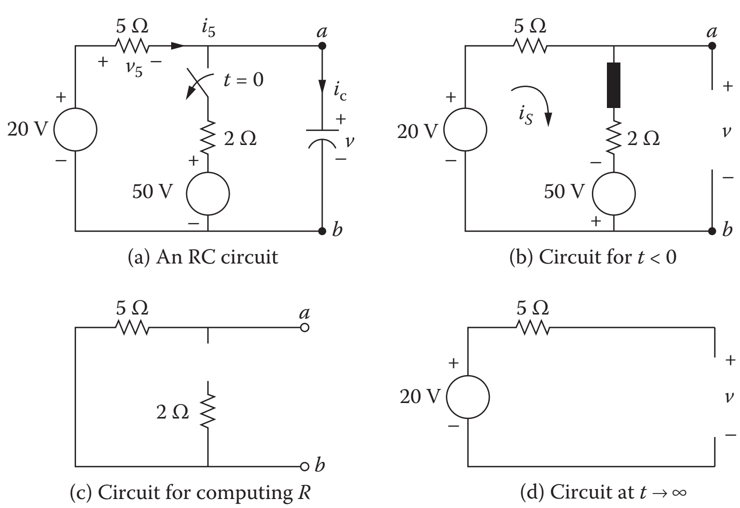


FIGURE 2.20 RC circuit for Example 2.16.

3. The dc circuit for $t = \infty$ is provided in Figure 2.20d. Solving for the capacitor voltage:

$$v(\infty) = 20\ V$$

4. Finally, substituting:

$$v(t) = v(\infty) - (v(\infty) - v(0)) \cdot e^{-t/\tau}$$

$$v(t) = 20 - 50 \cdot e^{-2t}\ V$$

Perhaps it seems overly restrictive to always solve for the inductor current in the RL case, and the capacitive voltage in the RC case. However, once these have been determined, it is straightforward to compute all branch voltages and currents. We demonstrate this point by continuing Example 2.24 in Example 2.25.

Example 2.25

Find all branch voltages and currents for $t > 0$ in the circuit of Example 2.24.

The capacitor:

$$v_C(t) = 20 - 50 \cdot e^{-2t}\ V$$

$$i_C = C \cdot \frac{dv_C}{dt} = 10 \cdot e^{-2t}\ A$$

The resistor:

$$v_5(t) = 20 - v_C = 20 - (20 - 50 \cdot e^{-2t}) = 50 \cdot e^{-2t}\ V$$

$$i_5 = \frac{v_5(t)}{5} = 10 \cdot e^{-2t}\ A$$

The source:

$$v_S(t) = 50\ V$$

$$i_S = 10 \cdot e^{-2t}\ A$$

Since the switch is open, the current through the 2 Ω resistor and the 50V source is zero. The voltage drop across the open switch is:

$$\begin{aligned} v_{\text{switch}} &= v_c + 50 \\ &= 70 - 50 \cdot e^{-2t} \end{aligned}$$

2.8 A dc Application: An Automotive Electrical System

As a young engineer makes the transition from academia into industry, he or she is typically confronted with a whole new world of technical jargon and symbolism. The engineer is tempted to conclude that "they taught me all the wrong stuff at Tech U. What I really need to know is the color code, wire gauge data, ampacity, and a host of other things that I ran into on my project last week." With more experience, the engineer realizes that this new material is supplementary, and quite compatible with the fundamentals of formal engineering education.

For example, look at Figure 2.21, which describes a typical automotive electrical system. Contrast this with the circuit diagram presented in Figure 2.22. Figure 2.21 and Figure 2.22 are two ways of looking at the same system in two different "languages."

A typical automotive electrical system has four major components:

1. *The alternator* is a three-phase ac generator (which we will discuss in more detail later) that supplies a three-phase rectifier (again, later) converting the output to dc. The combination can be modeled as a dc source, as shown in Figure 2.22. The diode is like a switch that is a short to current flow in the direction of the arrow, and an open to current flow in the opposite direction. This prevents power from flowing back into the alternator and driving it into the motor mode, once the engine is started. With the engine running, the alternator becomes the main electrical source, supplying the load and charging the battery.
2. *The battery* is an energy storage device, providing an electrical source to start the engine. When the engine is running, the battery interacts with the alternator to stabilize the system voltage. It also serves as an emergency power supply in case the alternator fails. The typical lead–acid 12 V car battery produces an open circuit voltage of 12.6 V when fully charged. The internal voltage (E_A) drops as the battery

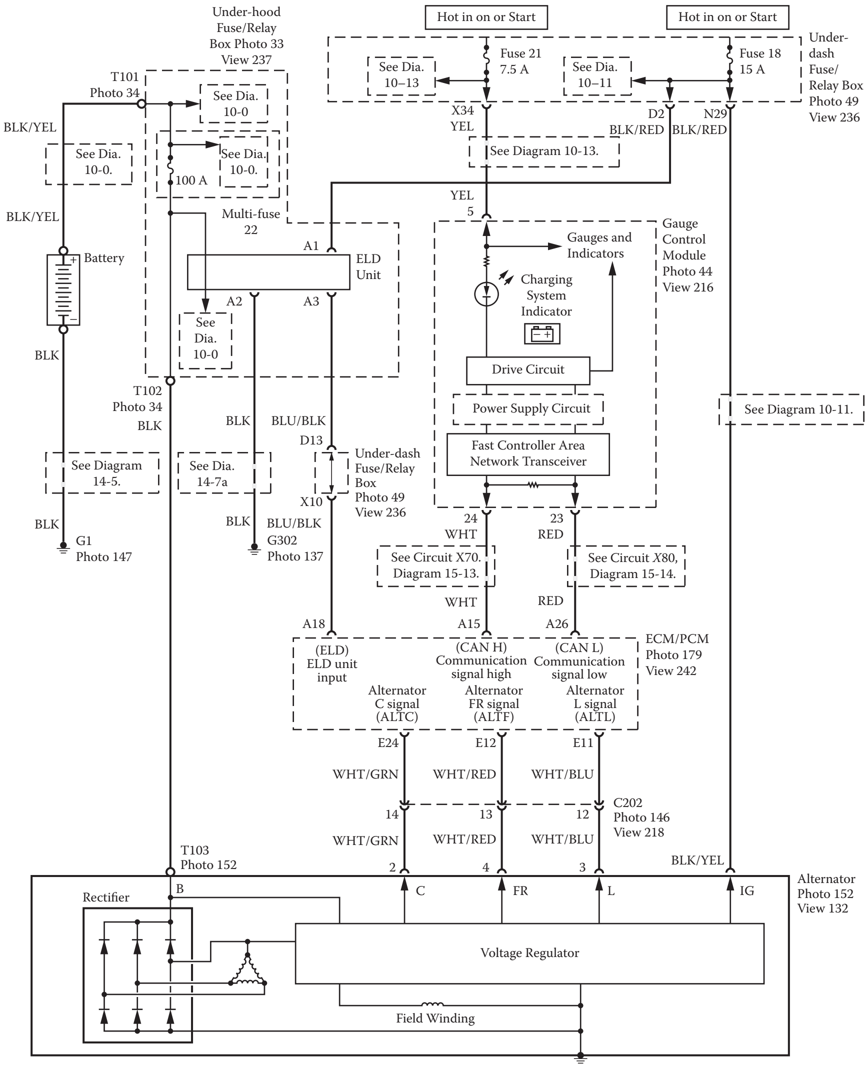


FIGURE 2.21 Typical wiring diagram.

loses charge. There is a rating called "cold cranking amps" (CCA), which is basically the short-circuit current under fully charged conditions. Another important rating is its charge capacity, usually stated in "amp-hours" (A-hr). One A-hr $= 60 \times 60 = 3600$ C.

3. *The voltage regulator* (VR) serves to keep the system voltage (V_L) near 12 V, and in some ways is the most fascinating part of the

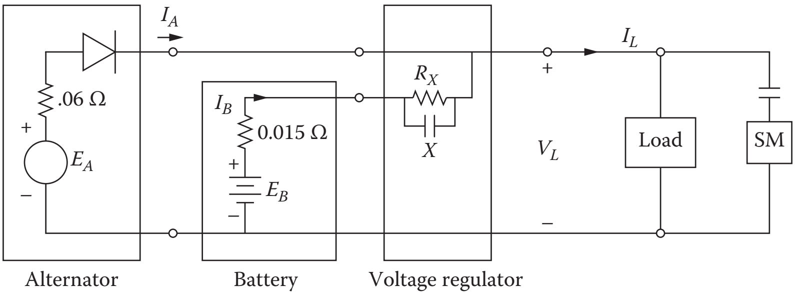


FIGURE 2.22 **A simplified automotive electrical system.**

system. It does this by controlling the alternator field current, which in turn controls the internal alternator voltage (E_A). Modern VR's are typically electronic devices and controlled by microprocessors (µP's), which store certain settings and receive system data provided by sensors. For example, specific inputs might include the battery current (I_B), the alternator current (I_A), and the system voltage (V_L). Settings might include the *maximum battery charging current*, *maximum I_A*, and *maximum and minimum V_L*.

4. *The load* is the total of the power requirements of all of the devices that require electricity to operate (e.g., the lights, the sound system, navigation aides, and instrumentation).

Before we begin our study of this system, we need to understand a component called the "relay." A relay is essentially a remote-controlled switch with two basic parts:

- *The coil:* a coil of wire wrapped around a ferromagnetic core
- *The contacts*: one or more pairs of conducting surfaces, such that when they touch, a short circuit, is formed, and when nontouching, they constitute an open circuit

When current flows through the coil, a force is produced to actuate a mechanism that moves the contacts. Refer to Figure 2.23. Contact pairs are one of two types:

- Normally open (NO), such that when the coil is deenergized, the contacts are open
- Normally closed (NC), such that when the coil is deenergized, the contacts are closed

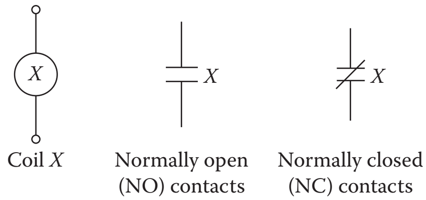


FIGURE 2.23 Electromechanical relay.

When the coil is energized, NO contacts close, and NC contacts open. There are three advantages to relays:

- A small current and voltage are used to switch a much larger current and voltage.
- The control circuit can be physically remote from the power circuit.
- Multiple switching actions can be performed simultaneously.

Relays are still used extensively in many applications, including automotive systems, although they can be replaced with solid-state electronic counterparts.

Let us examine the system in more detail, in the context of several engineering analyses.

System Data

Load

Lights

Headlights (low beam): 2×30 W

Headlights (high beam): 2×30 W

Day running lights: 2×30 W

Parking, taillights: 4×15 W

Flashers, turn signals: 4×25 W

Interior: 4×8 W

Map: 2×12 W

Glove box: 8 W

Trunk: 12 W

And so on

Maximum load (adjusting for intermittent operation): 900 W

Battery

$R_B = 0.015\ \Omega$

120 A-hr

$$E_\beta = 11.2 + 1.4 \cdot k_Q \quad 0.1 \le k_Q \le 1.0$$

$$k_Q = 0.1 \quad \text{battery at 10\% charge}$$

$$k_Q = 1.0 \quad \text{battery at 100\% charge}$$

Optimum charging rate: 20 A.[16]

Starter motor

$R_{SM} = 0.02\ \Omega$

Minimum starting current = 200 A

Alternator

12 – 20 V; 100 A

$R_A = 0.06\ \Omega$

The system must meet the following specifications:

1. Starting mode operation, engine off. Battery must start the car on 10% charge.
2. Normal mode, engine running, alternator operational
   a. The load voltage should be maintained from $12.3 < V_L < 12.6$ V.
   b. No components are overloaded.
3. Emergency mode, engine running, alternator inoperable, battery 80% charged. 100 mile range at half load; 55 mph.
4. Parking mode, standstill, battery 80% charged
   a. Parking lights on for 12 hours, battery drain to 10% charged.
   b. Flashers on for 12 hours, battery drain to 10% charged.

Our assignment:

Confirm that the given automotive system meets all operating specifications. Propose modifications to the system when specifications cannot be met.

Analysis follows. Refer to Figure 2.22. In particular, notice that $+\boldsymbol{I_B}$ flows *out of* the battery (discharging) and $-\boldsymbol{I_B}$ flows *into* the battery (charging);

[16] Actually the optimum charge rate varies with the battery charge status.

Check Spec 1

@ 10% charge:

$$E_B = 11.2 + 1.4 \cdot k_Q = 11.34 \text{ V}$$

$$I_{SM} == \frac{E_B}{R_B + R_{SM}} = \frac{11.34}{0.015 + 0.020} = 324 \text{ A}$$

324 A > 200 A at starting, so Spec 1 is met.

Check Spec 2a

The heaviest load on the alternator occurs at $V_L = 12.6$ V; full load, battery charging:

$$900W @ 12V:$$

$$I_L = \frac{900}{12} = 75\ A$$

$$KCL: \quad I_A = -I_B + I_L = 20 + 75 = 95\ A$$

The alternator current rating is 100 A (> 95 A). Also, if the load is suddenly removed, the alternator voltage rises to 12.6 + 95(0.06) = 18.2 V. Since the maximum alternator voltage rating is 20 V, the alternator is capable of a field current which produces an E_A value of 18.2 V.

Check Spec 2b

There is no operating condition that overloads the battery, alternator, or regulator. If the wiring is properly sized, we meet Spec 2b.

Check Spec 3

"Battery 80% charged" means $Q = 0.8(120) = 96$ A-hr. At 55 mph, we travel 100 miles in 100/55 = 1.818 hr. 80% load = 720 W. The average battery voltage is about 12 V; hence the current is about 720/12 = 60 A. In 1.818 hours, this amounts to a charge of 109.1 A-hr, which is greater than 96 A-hr. *Hence, we fail to meet Spec 3.*

Check Spec 4a

$Q = 96$ A-hr. Parking lights and taillights are 4 × 15 = 60 W. The average battery voltage is about 12 V; hence the current is about 60/12 = 5 A. In

12 hours, this amounts to a charge of 5 × 12 = 60 A-hr, which is less than 96 A-hr. Hence, we meet Spec 4a.

Check Spec 4b

Q = 96 A-hr. Flashers are 4 × 25 = 100 W. But they're on only half the time for an average load of 50 W. The average battery voltage is about 12 V; hence the current is about 50/12 = 4.2 A. In 12 hours, this amounts to a charge of 50.4 A-hr, which is less than 96 A-hr. Hence, we meet Spec 4b.

Recall that we failed to meet Spec 3. So what do we do? We schedule a meeting with the chief engineer to discuss our options. For example:

1. How important is the spec? Can we rewrite it to require an 80-mile range (which we can easily meet)?
2. Can we specify a larger battery? Cost? Weight?
3. How important is the 80% load requirement? How much load could be turned off without compromising safety or performance?
4. How important is the 80% charge requirement? Is it realistic to assume a near 100% charge? How close is "near"?
5. Any other options?

This is a situation that commonly occurs in engineering design. After much debate, our engineering team decides to recommend that management rewrite the specification to reduce the range requirement for an all-electric operation from 100 to 80 miles.

Let us consider one final issue. Suppose we are running in normal mode and the battery is very low (e.g., 10%Q).

$$\text{@ 10\% charge}$$

$$E_B = 11.2 + 1.4 \cdot k_Q = 11.34$$

$$-I_B == \frac{V_L - E_B}{R_B} = \frac{12.60 - 11.34}{0.015} = 84\ A$$

This is far in excess of the 20 A maximum charging current. The VR is permitted to drop the voltage to 12.3 V, but the charging current is still 64 A.

How can we simultaneously:

- Limit the charging current to 20 A?
- And satisfy the requirement that $12.3 < V_L < 12.6$ V?

We decide to insert a protective resistor R_X in series with the battery. But this requires that we address three more questions:

- What should be the size of R_X?
- How do we insert and remove R_X?
- What are the criteria for insertion and removal?

The size of R_X is determined by the lowest possible battery voltage (11.34 V). We reason that

@ 10% charge

$$R_X + 0.015 = \frac{12.30 - 11.34}{20} = 0.048\ \Omega$$

$$R_X = 0.033\ \Omega$$

With the contacts X closed, R_X is shorted (and therefore removed from the circuit). Opening the contacts X inserts R_X into the circuit. The system should work like this.

1. At the low(est) battery voltage (11.34V), we set $V_L = 12.3$ V. Contacts are open so that R_X is in the circuit. $(-I_B) = 20$ A flows into the battery.
2. As the battery charges, E rises. The VR responds by raising V_L, holding $(-I_B)$ at 20 A.
3. Eventually $V_L = 12.6$ V:

$$E = 12.6 - (0.048)20 = 11.64\ V$$

$$E = 11.2 + 1.4 \cdot k_Q = 11.64$$

$$k_Q = 0.3143$$

At this point, the battery is 31% charged. The VR cannot maintain a battery charging current of 20 A charging without V_L exceeding 12.6 V. If we switch out R_X:

$$(-I_B) = \frac{V_L - E_B}{R_B} = \frac{12.6 - 11.64}{0.015} = 64\ \text{A}$$

which is much too large. Switching out R_X *and* dropping V_L

$$V_L = (-I_B)R_B + E_B = (20)(0.015) + 11.64 = 11.94 \text{ V}$$

which is too low (recall $12.3 < V_L < 12.6$ V). So we must leave R_X in the circuit, and freeze V_L at its upper limit 12.6 V, which means that the charging current must drop.

4. The battery continues to charge, although $(-I_B)$ is dropping. Eventually $E_B = 12.3$ V.

$$(-I_B) = \frac{V_L - E_B}{R_B + R_X} = \frac{12.6 - 12.3}{0.015 + 0.033} = 6.25 \text{ A}$$

If we now switch out R_X:

5. The battery continues to charge until E reaches 12.6 V, at which point the battery is fully charged.

So, to summarize the operation:

- We want R_X in the circuit for $12.3 < V_L < 12.6$ V. and $-I_B = 20$ A.
- We want R_X out of the circuit for $V_L = 12.6$ V. and $-I_B < 6.25$ A.
- If R_X is in the circuit and $-I_B > 20$ A, the VR must lower V_L until $-I_B =$ 20 A.

The system should work like this. After the car is started, the VR raises V_L to 12.3 V with contact X closed, R_X out. Then:

1. If $-I_B < 20$ A, the VR raises V_L until $-I_B = 20$ A or until $V_L = 12.6$ V.
2. If $-I_B > 20$ A, then contact X opens, inserting R_X into the circuit, which will cause V_L to rise and $-I_B$ to drop (to a maximum value of 20A). For example, suppose the battery happens to be 17.71% charged, the load is 100%. Then:

$$E_B = 11.2 + 1.4 \cdot k_Q = 11.2 + 0.248 = 11.448 \text{ V}$$

$$@V_L = 12.6\,V \quad (-I_B) = \frac{12.6 - 11.448}{R_B + R_X} = \frac{1.152}{0.048} = 24 \text{ A}$$

$$@V_L = 12.3\,V \quad (-I_B) = \frac{12.3 - 11.448}{R_B + R_X} = \frac{0.852}{0.048} = 17.75 \text{ A}$$

We wish to sustain a maximum charging current of 20 A so that VR acts to control V_L.

$$V_L = 20(0.048) + 11.448 = 12.41 \text{ V}$$

As the battery charges, E_B rises, which causes V_L to also rise. When V_L reaches 12.6 V, the VR holds that value, and releases control on $-I_B$, allowing it to drop from 20 A. When $-I_B$ reaches 6.25 A, contact X closes, removing R_X from the circuit.

The VR, with the correct settings stored in its μP brain, can perform the necessary switching operations.

The foregoing example demonstrates how dc circuit analysis can be used to answer practical engineering questions. We will study the application of ac circuits in the next section.

2.9 An ac Example Application: The U.S. Residential Electrical System[17]

The U.S. electrical service provided to residences from the local electric utility typically arrives in the form of a cable, from either an overhead or underground source. The cable consists of three aluminum stranded conductors, two insulated (A,B) and one bare (N), which pass through a watt-hour meter into an electrical panel (see Figure 2.24a). The standard voltage is 120 V/240 V (V_{AN} = 120 V; V_{BN} = 120 V; V_{AB} = 240 *V*), and the standard frequency is 60 Hz. The bare conductor N is grounded (i.e., connected to a metal [ground] rod driven into the earth, and/or another structure which provides a low-resistance path to the earth) and connected to the ground bus[18] in the panel.

Conductors A and B are connected to a double-pole[19] switch called a "disconnect," which in turn is connected to a second double-pole switch called

[17] The systems and practices described in this section are for educational purposes only, and may differ in some details from those used by particular utilities and local contractors. Actual system design must be in conformity with all provisions of the National Electric Code, the National Electric Safety Code, and relevant local electrical codes.

[18] A "bus" is a substantial metal-conducting structure designed so that many conductors can easily be connected to it.

[19] The term "pole" refers to the number of contact pairs that are "made" (i.e., in contact with each other) or "broken" (not in contact with each other) in a switch. A single-pole switch has one pair of contacts. A two-pole switch has two pairs of contacts, each pair in a separate insulated circuit. In a multipole switch, the switching action is mechanically linked and synchronized. See Figure 2.24.

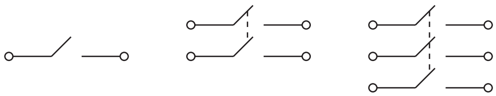


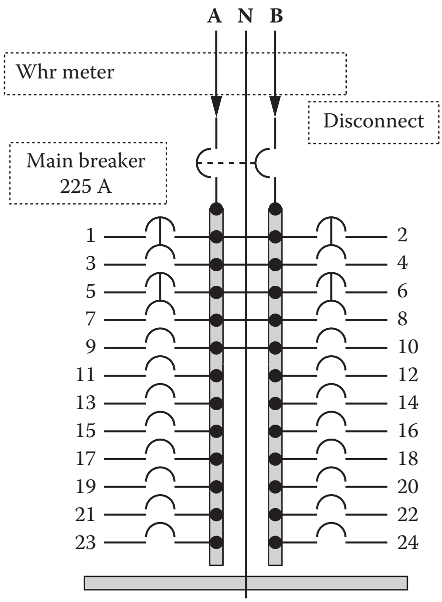
FIGURE 2.24 **Switches.**

"the main breaker." It may seem strange to have two switches in series. The main breaker is a switch that automatically operates when (a) an extremely large current flows ("instantaneous trip"), and/or (b) an excessive and sustained current flows, such that over time, the panel circuitry will overheat ("delayed trip"). As such, the breaker is designed to interrupt large currents, and as such, has no problem interrupting the panel load currents. On the other hand, the disconnect is a safety device, such that when open, it disconnects the panel circuitry (including the main breaker) from all electrical sources. If opened under load, serious arcing will occur, damaging the disconnect contacts. Electricians think, "Breaker first, then disconnect," if they want to isolate the panel from the electric service.

At the breaker secondary,[20] A and B are each connected to an insulated bus (i.e., "A" is insulated from "B" and "N," and "B" from "A" and "N"). The panel is designed so that single-pole breakers can be mechanically plugged in, such that the primary pole connects to Bus A (or B) and the secondary connects to a terminal block at which a 120 V branch circuit can be also connected. Double-pole breakers can also be accommodated to serve 240 V branch circuits. The 120 V circuits are commonly served with #14 conductors protected by 15 A breakers, or #12 conductors protected by 20 A breakers. The smaller the conductor size, the larger its "ampacity" (i.e., its current-carrying capacity). 240 V circuits are typically 30 A or 50 A, and are served with #10, #8, or #6 conductors; other sizes are also used.

We continue our study using a specific application. The panel is shown in Figure 2.25. The panel has a main breaker rated at 225 A and accommodates some 24 single-pole breakers. The design is such that a double-pole beaker fits in two single-pole spaces. Hence, a 24-slot panel might have four 240 V circuits and sixteen 120 V circuits, as shown in Figure 2.25a. A typical load assignment appears in Figure 2.25b.

[20] The terms "primary" and "secondary" usually refer to source and load sides, respectively. When the breaker is open, the primary is "hot" (i.e., energized); the secondary is "not hot."



(a) Panel circuitry

Cir	Description	Cir	Description
1	A/C	2	WH
3	A/C	4	WH
5	Range	6	Spare
7	Range	8	Spare
9	Lights Bd1; Ba1, Hall	10	Lights Bd2; Ba2, Hall, Fron
11	Lights Bd3; Foyer, Ba3, Hall	12	Lights GtRm, Din, Front
13	Lights, Kit	14	Garage
15	Bd1; Ba1, Hall, Back	16	Lights
17	Kitchen	18	Bd2; Ba2, Hall
19	Lights, Office, Den	20	Shop
21	Office, Den	22	Spare
23	Spare	24	Spare

(b) Circuit schedule

FIGURE 2.25 Residential 120/240 V electrical service.

Observe that the 240 V loads are essentially in parallel and can be combined by Tellegen's theorem as follows:

Circuit 1–3: A/C: 240 V, 5 kVA, pf = 0.8 lagging
Circuit 2–4: Water heater, 240 V, 4 kVA, pf = unity
Circuit 5–7: Range , 240 V, 4 kVA, pf = 0.9 lagging

Combining the load

$$\bar{S}_{13} = 5\angle 36.9^\circ = 4 + j3$$

$$\bar{S}_{24} = 4\angle 0^\circ = 4 + j0$$

$$\bar{S}_{57} = 4\angle 25.8^\circ = 3.6 + j1.744$$

$$\bar{S}_{AB} = \bar{S}_{13} + \bar{S}_{24} + \bar{S}_{57} = 11.6 + j4.744 = 12.53\angle 22.2^\circ$$

Similarly, the loads served on the AN and BN circuits can be combined, forming:

$$\bar{S}_{AN} = 5\angle 36.9^\circ = 4 + j3$$

$$\bar{S}_{NB} = 4\angle 0^\circ = 4 + j0$$

To determine the load currents, consider the circuit diagram in Figure 2.26. We compute

$$\bar{I}_{AB} = \left(\frac{\bar{S}_{AB}}{\bar{V}_{AB}}\right)^* = \left(\frac{12.53\angle 22.2^\circ}{0.24\angle 0^\circ}\right)^* = 52.21\angle -22.2^\circ \text{ A}$$

$$\bar{I}_{AN} = \left(\frac{\bar{S}_{AN}}{\bar{V}_{AN}}\right)^* = \left(\frac{5\angle 36.9^\circ}{0.12\angle 0^\circ}\right)^* = 41.67\angle -36.9^\circ \text{ A}$$

$$\bar{I}_{NB} = \left(\frac{\bar{S}_{NB}}{\bar{V}_{NB}}\right)^* = \left(\frac{4\angle 0^\circ}{0.12\angle 0^\circ}\right)^* = 33.33\angle 0^\circ \text{ A}$$

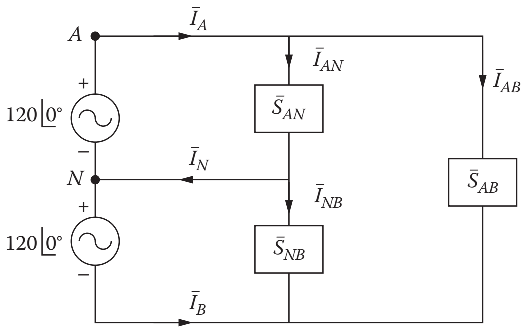


FIGURE 2.26 Circuit model.

To determine the panel input currents, again consider the circuit diagram in Figure 2.26. We compute

$$\bar{I}_A = \bar{I}_{AB} + \bar{I}_{AN} = 52.21\angle{-22.2°} + 41.67\angle{-36.9°} = 93.12\angle{-28.7°}\ \text{A}$$

$$\bar{I}_B = -(\bar{I}_{BA} + \bar{I}_{NB}) = -(52.21\angle{-22.2°} + 33.33\angle{0°}) = 84.02\angle{166.4°}\ \text{A}$$

$$\bar{I}_N = \bar{I}_{AN} - \bar{I}_{NB} = 41.67\angle{-36.9°} - 33.33\angle{0°} = 25.00\angle{-90°}\ \text{A}$$

It is clear, then, how ac circuit analysis can be used to study the performance of practical electric power systems. As we shall see, the approach can be extended to study the design and performance of the entire power grid.

2.10 Summary

We have studied the fundamentals of linear electric circuit theory, using five ideal elements:

- Active elements
  - The voltage source
  - The current source
- Passive elements
  - The resistor
  - The inductor
  - The capacitor

where the v,i relations for each element was defined (which we shall refer to as a generalized version of Ohm's law).

Interconnections of any number of these elements create what is called an "electrical circuit," the performance of which must conform to two basic laws:

1. Kirchhoff's current law (KCL) requires that the net current flowing into any node in a circuit must be equal to the net current flowing out.
2. Kirchhoff's voltage law (KVL) requires that the net voltage drop must be equal to the net voltage rise around any closed path traced out in a circuit.

Circuits can be said to operate in one of four possible modes (dc, ac, periodic, and transient), the first three of which are stable steady-state conditions:

"dc" means time invariant (i.e., the voltages and currents are not changing in time).
"ac" means all voltages and currents vary sinusoidally in time at the same frequency.
"Periodic" means that all voltages and currents vary arbitrarily in time over a fixed time interval called the "period" (T), but the pattern repeats over the next, and all subsequent, intervals of duration T.
"Transient" means that a circuit is operating in steady-state condition (the initial state), when at a given point in time (e.g., $t = 0$) a switch is thrown, after which all voltages and currents vary in time until they settle into a second steady-state condition (the final state).

Circuit analysis is the application of Kirchhoff's laws, and Ohm's law, to arbitrary interconnections between any combination of these elements, to the end of calculating all branch voltages, currents, and powers. This results in the solution of the following:

A set of independent algebraic equations (dc)
A set of independent algebraic vector (complex) equations (ac)

Using Fourier methods and superposition, the periodic case results in one dc circuit problem, and "N" ac circuit problems, where "N' is the number of harmonics used. The transient mode[21] requires the solution of two dc circuit problems, and the solution of a first-order linear differential equation.

[21] As constrained in Section 2.7.

The application of circuit theory extends far beyond the analysis of abstract circuits composed of ideal elements. Since virtually all electrical devices can be modeled as a circuit, from the very small (e.g., a transistor) to the very large (e.g., a 1000 MW generator), circuit analysis becomes a critically important system assessment and design tool in electrical engineering.

References

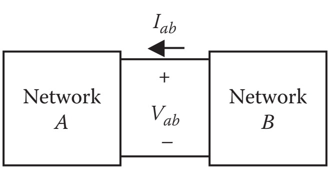
Gross, C.A. *Power System Analysis*, 2nd ed. New York: Wiley, 1986.
Hayt, William H., Jr., J.E. Kimmerly, and S.M. Durbin. *Engineering Circuit Analysis*, 7th ed. New York: McGraw-Hill, 2007.
Smith, Ralph J., and R.C. Dorf. *Circuits, Devices, and Systems*, 5th ed. New York: Wiley, 1992.

Problems

2.1. We observe that at a given point on a wire, 6.242×10^{20} electrons flow past left to right in 20 seconds.
   a. Find the current passing through the point.
   b. The current flows from left to right or right to left?

2.2. Two large metal plates are charged so that the voltage between them is "V." We remove one electron from the top plate and force it to the bottom plate, a process that requires 0.001602 pJ.
   a. Find V.
   b. Which plate is positive?

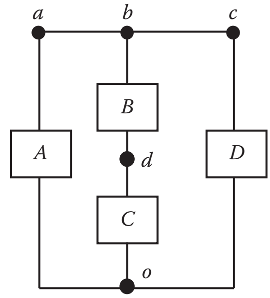
2.3 In the figure below, $V_{ab} = -80$ V and Network B *absorbs* 800 W.
   a. Find I_{ab}.
   b. Does Network A absorb or deliver power?



2.4. In the figure below, $V_{bd} = 7$ V; $V_{co} = 13$ V; $I_{do} = 3$ A; and $I_{ao} = -8$ A.
   a. How many nodes does the circuit have?
   b. How many branches does the circuit have?
   c. Find I_{db}.
   d. Find V_{ao}.

e. Find the power absorbed by each element.

f. Show that the powers in (f) add to zero.

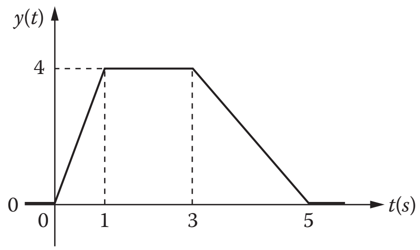


2.5. In the figure below, suppose $y(t)$ is the voltage across element X. Accurately sketch the current through if X is a:

a. 2 Ω resistor.

b. 2 H inductor.

c. 2 F capacitor.



2.6. In this figure, suppose $y(t)$ is the current through element X. Accurately sketch the voltage through if X is a:

a. 2 Ω resistor.

b. 2 H inductor.

c. 2 F capacitor.

2.7.

a. Define "dc" in a circuits context.

b. Prove that capacitors present open circuits to dc.

c. Prove that inductors present short circuits to dc.

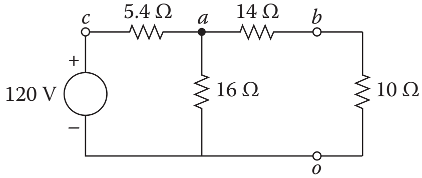
2.8.

a. Define "ac" in a circuits context.

b. Find the current through a 26.525 mH inductor if $v(t) = 169.7\cos(377t)$ V.

c. Find the current through a 0.26525 mF capacitor if $v(t) = 169.7\cos(377t)$ V.

2.9. In the figure below, find all branch voltages and currents.



2.10. Continuing Problem 2.9, find all branch powers. Show that Tellegen's theorem is satisfied.

2.11. Refer to the figure above.

a. Reduce the left side of the circuit (looking left from port b–0) to its Thevenin equivalent.

b. Use the Thevenin equivalent to compute V_{bo} and I_{bo}.

2.12. $v(t) = 650.54 \cos(377t - 20°)$. V.

Find V_{MAX}, f, ω, α, and T.

2.13. $v(t) = 650.54 \cos(377t - 20°)$ V.

a. Evalutate $V_{RMS} = \frac{V_{MAX}}{\sqrt{2}}$

b. Derive V_{RMS} from $V_{RMS} = \sqrt{\frac{1}{T} \int_{t_0}^{t_0+T} v(t)^2 \cdot dt}$

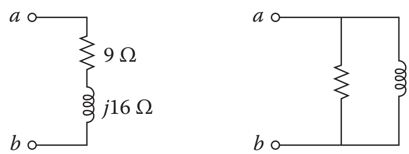
c. Evaluate the phasor $\bar{V}$

2.14. Complete the table for $f = 1$ kHz.

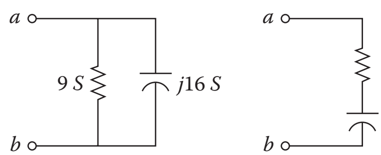
	$\bar{Z}$ (Ω)	$\bar{Y}$ (ms)
 100 Ω		
 31.83 mH		
 0.7958 μF		

2.15. Convert the circuits as indicated.

(a) Use $\bar{Z}$ for all elements.



(b) Use $\bar{Y}$ for all elements.

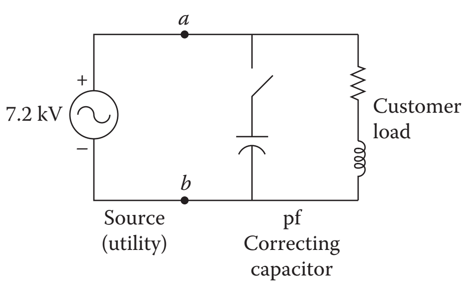


2.16. Consider a series circuit identical to Figure 2.11a, except that R = 8Ω, $L = 10.61$ μH, and $C = 0.2653$ mF.

   a. Draw the ac circuit.
   b. Solve for $\bar{I}, \bar{V}_R, \bar{V}_L$, and $\bar{V}_C$.

2.17. Continuing Problem 2.16, determine the complex power absorbed by each element, and show that Tellegen's theorem is satisfied.

2.18. There is a particularly important practical design situation that requires an understanding of ac circuits, complex power, and power factors. Utilities prefer that their customers operate at unity pf because this corresponds to maximum (real) delivered power at minimum current. To encourage maximum pf operation, the utility rate structure imposes a penalty for low pf operation. Hence, it is to a customer's advantage to "correct" their pf to near unity. This is called "the pf correction problem," and is presented here. A customer draws a single-phase load of 1000 kVA @ 7.2 kV; $pf = 0.6$ lagging. He wishes to install pf correcting capacitors as shown in the figure below to correct the metered pf to 0.92 lagging. Size the capacitors.



2.19. Repeat Problem 2.18 if the load is three-phase 12.47 kV 1000 kVA $pf = 0.6$ lagging.

2.20. Consider the waveform of Example 2.18.

a. What is the frequency of the sixth harmonic?

b. From Table 2.2, write the expression for the sixth harmonic.

c. Compute the RMS value of the sixth harmonic.

2.21. Consider a series circuit consisting of $R = 10\ \Omega$, $L = 20$ mH, $C = 50\ \mu$F, and a source such that

$$v(t) = 100\sqrt{2} \cdot \cos(\omega_0 t) + 30\sqrt{2} \cdot \cos(3\omega_0 t) \quad V$$

$$\text{where } \omega_0 = 1000\ rad / s$$

a. Draw the ac circuit at the fundamental frequency. Solve for the fundamental phasor inductor current and voltage.

b. Draw the ac circuit at the third harmonic frequency. Solve for the third harmonic phasor inductor current and voltage.

c. Find the RMS inductor current and voltage.

d. Find the current $i(t)$.

e. Find the average power delivered by the source.

2.22. Copy the *RL* circuit of Figure 2.18a, using 20 V, 4 Ω, and 8 H for *E*, *R*, and *L*, respectively.

a. Find the current and all voltages for $t > 0$, given that $i(0) = 0$.

b. Repeat (a) if $i(0) = -5$ A.

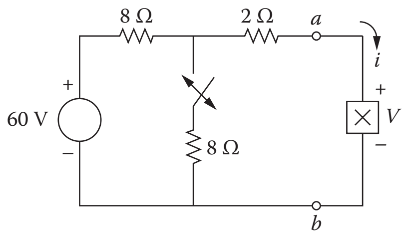
2.23. Copy the *RC* circuit of Figure 2.18b, using 20 V, 4 Ω, and 2 *F* for *E*, *R*, and *C*, respectively.

a. Find the current and all voltages for $t > 0$, given that $v_C(0) = 0$.

b. Repeat (a) if $v_C(0) = -10$ V.

2.24. It is argued that the current must be continuous through an inductor in the *R*–*L* case, and similarly the voltage must be continuous across a capacitor in the *R*–*C* case. Discuss the issue.

2.25. Given the circuit in the figure below, element *X* is a 2 H inductor. If the switch is *closed* at $t = 0$, find and sketch $i(t)$.



2.26. Given the circuit in the figure above, element X is a 2 H inductor. If the switch is *opened* at $t = 0$, find and sketch $i(t)$.

2.27. Given the circuit in the figure above, element X is a 0.5 F capacitor. If the switch is *closed* at $t = 0$, find and sketch $v(t)$.

2.28. Given the circuit in the figure above, element X is a 0.5 F capacitor. If the switch is *opened* at $t = 0$, find and sketch $v(t)$.

2.29. Consider the automotive electrical system described in Section 2.8.

a. Redraw the circuit of Figure 2.22 (omitting irrelevant parts).

b. The car is running at 55 mph, the battery is at 90% charge, and the electrical system delivers a load current of 60 A. Find the system voltage, the alternator current and internal voltage, and the battery current and internal voltage.

2.30. Consider a residential electrical system application, as discussed in Section 2.9. The loads are as follows:

AB: 12 kVA @ $pf = 0.8$ lagging

AN: 7 kVA @ $pf =$ unity

BN: 8 kVA @ $pf = 0.9$ leading

a. Draw an appropriate ac circuit to model the situation (two sources and three loads).

b. There are six phasor currents in the circuit of (a). Solve for all six.

This page intentionally left blank

Chapter 3

Electrical Energy

Most people have some notion of the concept of energy, but few can formally define it. Before we try, we might consider an even more fundamental property of matter (e.g., "mass"). Because it is one of the four fundamental concepts of physics,[1] mass is not rigorously defined. Informally, sometimes mass is defined as a measure of the extent of matter, or a measure of an object's inertia. Given the existence of a physical object with mass, we can assign to it a property called "energy," which is formally defined as the capacity to do work. Unfortunately, this definition does not immediately and fully clarify the nature of energy. There is of course no shortage of energy in the universe. Indeed, the universe is composed of only two entities: matter and energy![2] Broadly speaking, one can classify energy as being either potential or kinetic: that is, an object possesses energy by virtue of its position ("potential energy") or of its motion ("kinetic energy"). Energy can appear in various forms, including:

Mechanical
Thermal, or heat
Chemical
Electrical
Electromagnetic (radiant)
Sound energy
Nuclear energy

Humanity needs energy for lighting, heating, transportation, manufacturing, construction, and, sad to say, destruction. To direct energy to human needs, it is necessary to easily control it and convert it from one form to another, which prompts us to address a key question:

Why study energy in the electrical form?

Electrical energy is important because (1) it is relatively simple to convert into (and from) any other form, and (2) it is also simple to move it between two locations! Also, once energy is in the electrical form, we can subject it to incredibly complicated control. Think of the sophisticated control necessary to create a TV picture, refreshed at a rate of at least thirty times per second! Consider the conversion of the kinetic energy of falling water to electricity,

[1] The other three being length, time, and charge or current.
[2] Excluding consideration of a spiritual presence, which is beyond the scope of our study.

transporting it hundreds of miles away, and converting it back into light and heat in your home, all processes continuously happening at the speed of light!

When designing an electrical energy system, we must deal with four basic issues:

Conversion of energy into the electrical form ("generation")
Processing (conversion from ac to dc, or vice versa; changing voltage and/or current levels ("transformation"))
Transporting the energy over distance ("transmission")
Conversion of energy into some useful form ("utilization")

Our first problem is to consider electrical energy sources. We first consider natural sources of electricity.

3.1 Natural Sources of Electricity

Most sources of electrical energy in nature are either too diffuse or too expensive to capture to be practical. Bioelectricity occurs in life forms, but usually is quite small. A dramatic exception is the electric eel, which can develop over 300 volts and can deliver a lethal shock. A less violent example is the soft yellow light of the firefly, using bioelectricity in the form of electroluminescence.

Lightning is certainly spectacular and powerful; however, the energy content is modest. See Figure 3.1. Roughly estimating 100 MV, 20 kA, and 10 μs

FIGURE 3.1 Lightning. Source: http://en.wikipedia.org/wiki/File:Blesk.jpg.

for a medium-sized lightning bolt, the energy content would be

$$W = V \cdot I \cdot T = 100 \times 10^6 (20 \times 10^3)(10 \times 10^{-6}) = 20 \text{ MJ}$$

One kW-hr = 3.6 MJ, and is worth about \$0.15. Hence the energy in our bolt is valued at about 84 cents. Considering its unpredictable and intermittent nature, and the large area over which lightning is distributed, it would appear that the prospects of harnessing it for commercial purposes are slim.

So, if natural sources of electricity are not available, we must consider the conversion of other nonelectrical sources to the electrical form.

3.2 Electromagnetic-Mechanical (EMM) Energy Conversion

Recall two fundamental principles from physics:

Faraday's law: An electric circuit linked by a time-varying magnetic field will experience an induced voltage proportional to the time rate of change of that magnetic flux linkage, or:

$$v(t) = \frac{d\lambda}{dt}$$

λ = time-varying magnetic flux linking a circuit "C"

v = time-varying induced voltage around circuit "C"

The Lorentz force: A current flowing in a magnetic field will experience a force according to

$$\hat{F} = (\hat{i} \times \hat{B}) \cdot \ell$$

$\hat{F}$ = vector Lorentz force

$\hat{i}$ = vector current in conductor "C"

$\hat{B}$ = vector magnetic field (flux density)

ℓ = length of conductor "C" in the field

Exploiting these two principles, it is possible to design an apparatus (an "EMM machine") to convert mechanical to electrical energy.

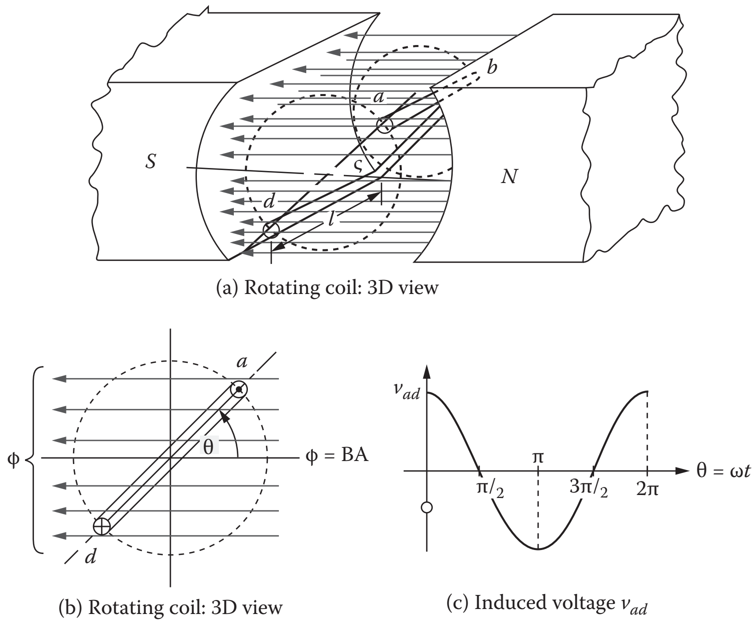


FIGURE 3.2 **Coil rotating in a magnetic field.**

Consider the situation in Figure 3.2a, which shows a rectangular three-sided coil abcd of radius r rotating at a constant angular velocity ω in a uniform magnetic field B. The coil dimensions into the page (ab and cd) are "l"; the coil dimensions in the plane of rotation (bc and ad) are $2r$. The magnetic flux is supplied by an external magnetic source, and flows right to left horizontally, from a north to a south pole. The same situation is depicted in the two-dimensional drawing of Figure 3.2b. The flux linking the coil depends on its angular position θ. Observe that

$$\lambda = 2Br\ell \cdot \sin(\theta)$$

$$\lambda = 2Br\ell \cdot \sin(\omega t)$$

Define:

$$A = 2r \cdot \ell = \text{cross-sectional area of the coil}$$

$$\phi_{MAX} = B \cdot A$$

From Faraday's law:

$$v_{ad} = \frac{d\lambda}{dt} = \phi_{MAX} \cdot \omega \cdot \cos(\omega t)$$

This provides great insight on how to design a machine that produces an ac voltage. Note that double-subscript notation means that the assigned positive sense of voltage is the drop from a to d (i.e., "a" is positive relative to "d"). The strength of the voltage will depend on the following:

- *The magnetic field B*. To supply substantive fields, we shall need to use ferromagnetic materials.
- *The size of the coil*. To produce substantive voltage, we need to use large coils (i.e., r, l should be large).

Now think of "abcd" in Figure 3.3a to be representing point locations (as opposed to conductors). Points abcd define a rectangular plane, in which we intend to wind a rectangular coil. Start with a single long flexible insulated

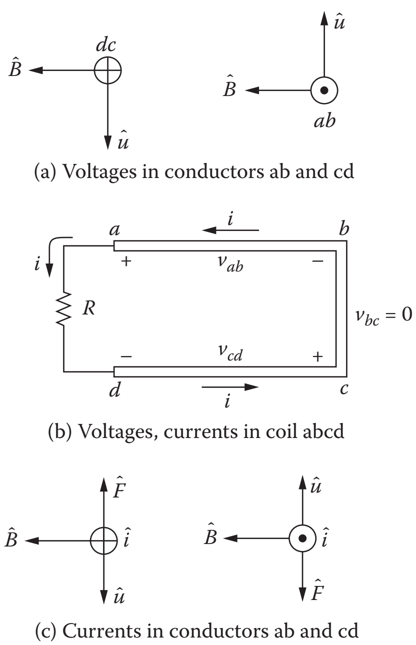


FIGURE 3.3 $\hat{u}, \hat{i}, \hat{F}$ in coil abcd at $\theta = 0$.

conductor with one end located at point a. Lay out a segment from a to b, bending at right angles at b toward c. Next, lay out a second segment from b to c, bending at right angles at c toward d. Next, lay out a third segment from c to d, and a fourth segment from d back to a. This constitutes one "turn." Repeat the procedure *N* times, using up the entire conductor and ending at d (the last turn has three sides; this last turn is closed by the external termination of the coil). Redefining a as the starting end of the conductor a, and d as the terminating end, we have formed an "*N*-turn" coil with terminals a–d. The effect is to multiply the flux linkage (and thus the voltage) by *N*:

$$\lambda = N \cdot \phi$$

$$v_{ad} = \frac{d\lambda}{dt} = B \cdot N \cdot A \cdot \omega \cdot \cos(\omega t) = N \cdot \phi_{MAX} \cdot \omega \cdot \cos(\omega t)$$

The radian frequency of the voltage is the same as the angular velocity of rotation (ω).

There is a second way to view the EMM voltage creation process. Instead of considering induced voltage as a consequence of time-varying flux linkage in a coil, suppose we think of moving a conductor through a magnetic field as causing a voltage to be created along its length. That is, if a conductor is moving through a magnetic field such that it intersects ("cuts through") the lines of flux, a voltage appears between the opposite ends of the conductor. In particular, consider conductor ab at $\theta = 0$. Observe that the opposite side of coil abcd is cd and the end connection (bc) connects ab in series with cd. Now imagine equal voltages are created in conductors ab and cd, such that

$$v_{ab} = v_{cd} = \frac{v_{ad}}{2} = \frac{2Br\ell\omega \cdot \cos(0)}{2} = B\ell r\omega = B\ell u$$

where $u = r\omega$ = translational velocity perpendicular to the field

This expression can be generalized as

$$\hat{v}_{ab} = (\hat{B} \times \hat{u}) \cdot \ell$$

$\hat{v}_{ab}$ = vector voltage along conductor ab

$\hat{B}$ = vector magnetic field (flux density)

$\hat{u}$ = vector lineal velocity of conductor ab

ℓ = length of conductor ab

Both ways of viewing the voltage creation process are correct, and both are useful.

We have managed to design a device that generates voltage, but what about current? Refer to Figure 3.3a. No current can flow since the "coil" has only three sides (side "ad" is open). Suppose we now terminate the coil in a resistor *R* (i.e., connect *R* from a to d).[3] In particular, consider the situation at $\theta = 0$. At this instant, the voltages are as indicated in Figure 3.3a and 3.3b (the "dot" in the circle depicting conductor ab means that the vector voltage points *out of* the page; the cross in cd means that the vector voltage points *into* the page). A current will now flow according to Ohm's law, around the closed path adcb, as shown in Figure 3.3b and 3.3c (the "dot" in the circle depicting conductor ab means that the vector current flows *out of* the page; the "cross" in cd means that the vector current flows *into* the page).

Example 3.1

The machine of Figure 3.2 has coil dimensions of r = 25 cm and l = 50 cm and rotates at 377 rad/s in a uniform magnetic field of 1.061 tesla. Determine v_{ad}:

$$v_{ad} = 2Br\ell\omega \cdot \cos(\omega t)$$

$$v_{ad} = 2(1.061)(0.25)(0.5)(377)\cos(377t) = 100\cos(377t)$$

We have a current flowing in a magnetic field; hence a Lorentz force must be created, which is

$$\hat{F} = (\hat{i} \times \hat{B}) \cdot \ell$$

Since at $\theta = 0$, and for the geometry of Figure 3.2, all vectors are mutually perpendicular:

$$F = i \cdot B \cdot \ell$$

[3] The following discussion assumes that the coil ac internal impedance is negligible compared with that of the external termination. The assumption is reasonable for the case discussed (i.e., the coil is rotating in air, and its conductors have high conductivity). In Chapter 4, a more general circuit model for a practical generator will be provided.

In fact, two forces (one on ab and one on cd, their directions predicted by the cross-products) are present, as shown in Figure 3.3c. Hence the torque on the one-turn coil is

$$T = 2 \cdot r \cdot F\cos(\theta) = 2iBr\ell \cdot \cos(\theta)$$

The torque on the N-turn coil is

$$T = 2iNBr\ell \cdot \cos(\theta)$$

Note that the direction of this EMM torque *opposes* rotation! It is hard to overstate the importance of this result! Had it come out in the reverse direction, we would live in a very strange universe indeed! Unopposed, the torque would have accelerated the coil, increasing the speed, the voltage, and hence the current. The increased current would have increased the torque, causing even faster acceleration, and very quickly the system would have accelerated to infinity!

But this didn't happen. The EMM torque opposes rotation, will decelerate the system, and, in fact, will bring the system to a complete stop. To keep the system running at a constant speed, we must externally apply an equal and opposite torque (i.e., a torque in the direction of rotation) using some external device. Apply too much external torque and the system speeds up; too little, and it slows down.

Continuing:

$$p = T \bullet \omega = 2iB\ell \cdot r \bullet \omega \cdot \cos(\theta) = 2iB\ell \cdot u \cdot \cos(\theta)$$

But $v_{ad} = 2Br\ell\omega \cdot \cos(\omega t) = 2B\ell \cdot u \cdot \cos(\theta)$

Then $p = v_{ad} \cdot i$

But this is precisely the same as the electrical power delivered to the external connection! That is, our machine has converted the mechanical power supplied in the form of externally applied torque into electrical power output with 100% efficiency! A practical device will have some coil resistance (which we have neglected), and it will take some energy to supply the requisite magnetic field. In addition, there will be some rotational losses (such as friction). Still, machines can be designed so that these losses are minimized and efficiencies in the high nineties can easily be achieved. Such EMM devices are called "generators," and they are an essential component in most of the technologies discussed in this chapter. Generators are covered in greater detail in Chapter 4.

Example 3.2

The machine of Example 3.1 has its coil terminated in a 10 Ω resistor. Determine the following:

A. The current, torque, and power
B. The *average* current, torque, and power

a.

$$i = \frac{v_{ad}}{R} = \frac{100\cos(377t)}{10} = 10\cos(377t)$$

$$T = 2iBr\ell\cos(\theta) = 2(1.061)(0.25)(0.50)(10\cos(377t))(\cos(377t))$$

$$T = 2.6525(\cos(377t))^2 = 1.3263 + 1.3263\cos(754t)$$

$$P = T\omega = 1000(\cos(377t))^2 = 500 + 500\cos(754t)$$

b.

$$I_{AVER} = \frac{1}{2\pi}\int_0^{2\pi} i \cdot d\theta = \frac{1}{2\pi}\int_0^{2\pi}(10\cos(\theta)) \cdot d\theta = 0$$

$$T_{AVER} = \frac{1}{2\pi}\int_0^{2\pi} T \cdot d\theta = \frac{1}{2\pi}\int_0^{2\pi}(2.652\cos^2(\theta)) \cdot d\theta = 1.326 \text{ Nm}$$

$$P_{AVER} = \frac{1}{2\pi}\int_0^{2\pi} P \cdot d\theta = \frac{1}{2\pi}\int_0^{2\pi}(1000\cos^2(\theta)) \cdot d\theta = 500 \text{ W}$$

The results of Example 3.2b illustrate an interesting point. Whereas the current and voltage average to zero, the power does not. The machine on average converts energy from the mechanical to electrical form at a rate of 500 W.

Before we leave the subject, note that if the external coil termination is an active network, current can be made to flow in the reverse direction (i.e., into the positive coil terminal). This means that the EMM torque reverses, as does the external torque required for dynamic equilibrium. Hence the power conversion process is reversed: power flows from the external electrical network, is converted into mechanical form, and is delivered to the external mechanical device. An EMM machine operating in this mode is called a "motor." We note that the device itself is an EMM machine; the mode of operation causes it to be viewed as a

1. "generator" if the energy conversion process is from mechanical to electrical; or
2. "motor" if the energy conversion process is from electrical to mechanical.

Most of the important sources of electrical energy to be discussed all employ the EMM generator. What distinguishes one from the others is what sort of device supplies the external mechanical torque. This device is traditionally called the "prime mover." Prime movers are one of two types: thermal and nonthermal.

3.3 Thermal Prime Movers

Thermal prime movers are typically turbines, converting hot pressurized gas flow into torque, or utilizing some other heat engine as the agent of torque production. The hot gas is either steam or natural gas. However, any type of thermal engine can be used for a prime mover.

3.3.1 Steam Turbine Prime Movers

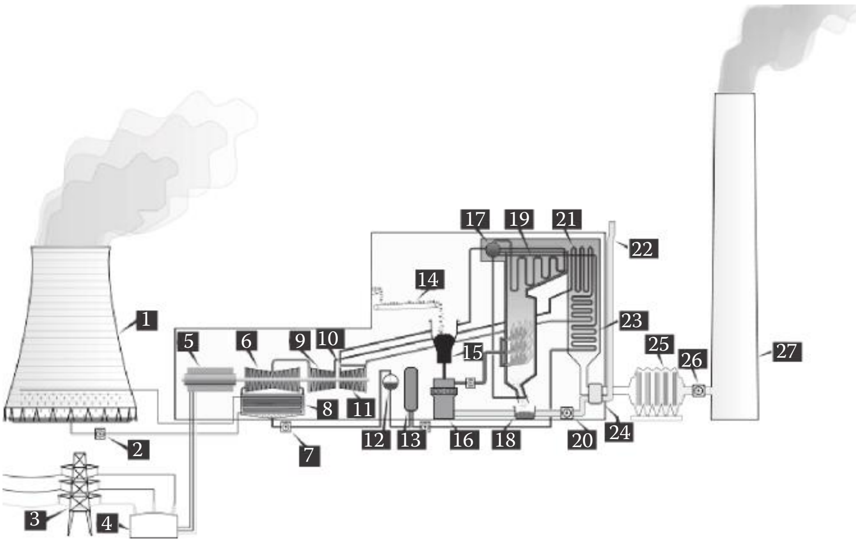
The majority of EMM electric generators use steam turbines as prime movers. A thermal energy source converts water in a boiler into steam, which is routed through one or more coaxial steam turbines. A typical arrangement is to use three stages (high, intermediate, and low pressure), as shown in Figure 3.4.

There are several possibilities for the heat source for the boiler.

3.3.1.1 Coal The largest single fuel source in the United States for heating boilers is coal, amounting to about 44.9% in 2009. Coal's advantages are that it is cheap and abundant. Its disadvantages are that it is a primary source of CO_2 and other greenhouse gas emissions, and that coal mining has a significant environmental impact. Most projections are that it will continue to play a major role for the foreseeable future.

The General James M. Gavin Plant on the Ohio River at Cheshire, Ohio, is a typical coal-fired power plant[4] (see Figure 3.5). Units 1 and 2 each rated at 26 kV and 1300 MW, for a total generating capacity of 2600 MW. The Gavin Plant is the largest generating station in the state of Ohio and has an average

[4] Material taken from an online educational study by Israel Urieli, *Engineering Thermodynamics,* Ohio University, with free use permission online. See http://www.ohio.edu/mechanical/thermo/Applied/Chapt.7_11/SteamPlant/GavinCaseStudy.html.



1. Cooling tower	10. Steam control valve	19. Superheater
2. Cooling water pump	11. High pressure steam turbine	20. Forced draught (draft) fan
3. Three-phase transmission line	12. De-aerator	21. Reheater
4. Step-up transformer	13. Feedwater heater	22. Combustion air intake
5. Electrical generator	14. Coal conveyor	23. Economizer
6. Low-pressure steam turbine	15. Coal hopper or silo	24. Air preheater
7. Boiler feedwater pump	16. Coal pulverizer	25. Precipitator
8. Surface condenser	17. Boiler steam drum	26. Induced draught (draft) fan
9. Internal pressure steam turbine	18. Bottom ash hopper	27. Flue gas stack

FIGURE 3.4 **Simplified coal-fired power plant. Source: http://en.wikipedia.org/wiki/Fossil_fuel_power_station.**

daily coal consumption of 25,000 tons at full capacity. The coal arrives by barge and is stored in the plant's coal yard. Conveyer belts carry the coal from the yard into the plant, where pulverizers grind the coal into a powder. The powdered coal is injected into the fire box of the steam generator, where it is burned, providing steam at high temperature and pressure.

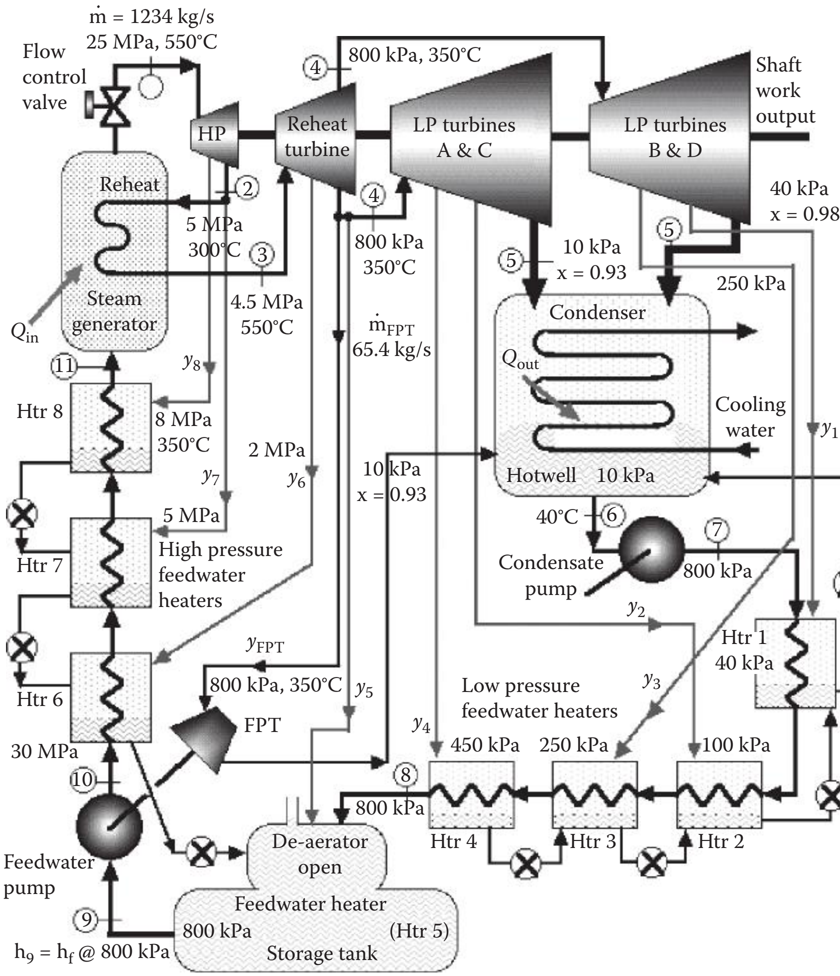


FIGURE 3.5 A simplified schematic diagram for one of the two Gavin generating units. Source: http://www.ohio.edu/mech anical/thermo/Applied/Cha pt.7_11/SteamPlant/GavinC aseStudy.html.

The two generators (#1 and #2) are on separate shafts (#1 and #2). The high-pressure turbine on shaft #1 exhausts steam to the reheater which provides steam to the reheat turbine on shaft #2. The exhaust steam from the reheat turbine separates into two paths, the first driving low-pressure turbines A and C on shaft #1, and the second driving low-pressure turbines B and D on shaft #2. Finally the steam flows into the condenser, where it condenses back into liquid water, releasing the heat of vaporization into a heat exchanger, heating the makeup water. The turbines convert the steam energies into torques, which are applied to the generator shafts. A schematic of the system is shown in Figure 3.5.

The feedwater to the boiler is supplied by a pump driven by a 65,000 hp motor, and its control is coordinated with the position of the main steam value. This provides control of the power output of the plant. Note that a power plant is an extremely complicated system that requires considerable engineering expertise for its design.

3.3.1.2 Natural Gas (Methane, CH_4) Natural gas is a large, and increasing, part of the fuel mix in the United States, amounting to about 23.4% in 2009. Like coal, the advantages of methane are that it is cheap and abundant. It is also cleaner than coal and most likely will play an even greater role for the foreseeable future. Natural gas can also be used directly as the working fluid in a gas turbine (as opposed to burning it in a boiler to produce steam). First using a gas turbine, and then using the waste heat to create steam which drives a steam turbine, is called a "combined cycle plant."

3.3.1.3 Oil Oil accounted for about 1% of electric-generating plant fuel in 2009. As it becomes scarcer, it will likely diminish in importance for this application.

3.3.1.4 Nuclear Nuclear reactions are the most abundant and powerful known sources of thermal energy in the universe. There are two basic processes:

1. *Fission*, the process where two or more atomic elements are broken apart to produce different (a) element(s), resulting in a mass defect. A mass defect is the difference between the total mass of the reactants before and after the reaction, the latter being less than the former.
2. *Fusion*, the process where two or more atomic elements combine to produce different (a) element(s), again with a mass defect.

In both cases, the mass defect is converted into energy according to Einstein's relation $E = \mathrm{mc}^2$.

Nuclear fission is the process of breaking large atoms (high atomic number) into smaller ones, creating a mass defect. See Figure 3.6. The element uranium (^{235}U) has a nuclear mass of 235.0439299 u and is a naturally occurring fissionable material. When a neutron of sufficient energy collides with

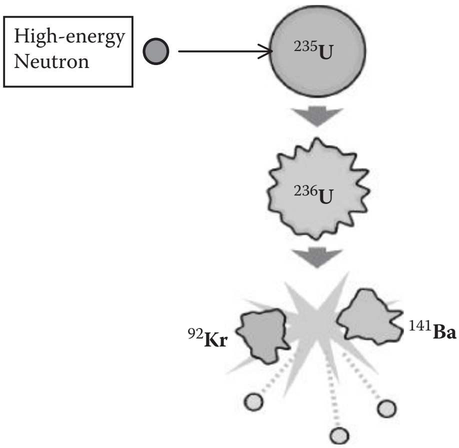


FIGURE 3.6 A standard fission nuclear reaction. Source: http://en.wikipedia.org/wiki/File:Nuclear_fission.svg.

the nucleus of ^{235}U, it splits into two atoms (krypton [^{92}K] and barium [^{141}Ba]), plus three neutrons. Computing the "before, after" masses:

"Before" … 235.0439299 (^{235}U) + 1.008665 (N) = 236.052595 u

"After" … 91.926156 (^{92}K) + 140.914411 (^{141}Ba) + 3 × 1.008665 (N) = 235.866562 u

u = 1 atomic mass unit = 1.661×10^{-27} kg

Note that we have lost mass. The mass defect is then:

$$\Delta m = 236.052595 - 235.866562 = 0.1860\, u = 0.3090 \times 10^{-27}\ \text{kg}$$

$$W = mc^2 = 0.3090 \times 10^{-27} (3 \times 10^8)^2 = 2.781 \times 10^{-11}\ \text{J}$$

The rate of reaction is determined by the relative density of neutrons. To slow down the rate, hollow tubes are placed in the reactor core by design, into which are placed "control" rods which have a large nuclear cross section to absorb neutrons. A typical control rod is made of an alloy of silver (80% silver, 15% indium, and 5% cadmium). Other materials, such as boron, are also used. As the rods are inserted further into the core, the neutron flux is reduced, and the nuclear reaction slows. In most reactor designs, for safety, control rods are mounted vertically above the core, and supported by electromagnets, rather than direct mechanical linkage. In the event of power failure, the control rods will fall due to gravity into the core and stop the reaction. Quickly shutting down a reactor in this way is called "scramming."[5]

It happens that U-238 is much more plentiful in nature than is U-235 (99.28% to 0.71%). Unfortunately U-238 is not fissionable. The so-called breeder reactor was developed to convert U-238 into a fissionable material. The breeder reactor core contains fissionable plutonium (Pu-239), surrounded by a layer of uranium-238. As the Pu-239 undergoes spontaneous fission, it releases neutrons which collide with U-238 to create more Pu-239. The process continues until all the U-238 is converted to Pu-239, which can be used as fuel in a conventional nuclear reactor. In other words, the breeder reactor creates fuel (Pu-239) as it operates. Two major disadvantages are that Pu-239 is extremely toxic and that it can be used in nuclear weapons.

The International Atomic Energy Agency (IAEA) reports that in 2006, nuclear power accounted for about 17% of the world's electric generation at 370 GW.[6] There were 435 operating nuclear reactors around the world. France had 78% of its generation from nuclear, with 54% for Belgium, 39% for the Republic of Korea, 37% for Switzerland, 30% for Japan, 19% for the United States, 16% for Russia, 4% for South Africa, and about 2% each for India and China. Growth in nuclear power is projected at about 9.4% for the next twenty-three years, matching the overall demand for electricity, with the highest growth areas centered in Asia. Present growth is greatest in Asia.

Nuclear fusion is the process which powers the sun. Several types of nuclear reactions occur in the stars. The most important in many stellar types, including our sun, is hydrogen fusion, or the so-called proton–proton chain reaction. Recall that a hydrogen atom consists of one electron and one proton, and that the sun is largely made of hydrogen. At high temperatures, the hydrogen ionizes, creating a "soup" of high-energy protons and electrons. Normally if two protons approach each other, they will repel each other and veer away, since both possess a positive charge. However, the solar environment is one of extremely high density and temperature. Thus many pairs of hydrogen nuclei

[5] Urban legend suggests that the control rods were hung by a rope above the reactor. In an emergency, a person with the title "Safety Control Rod Axe Man" (SCRAM) would take a fire axe and cut the rope.

[6] See the International Atomic Energy Agency (IAEA) report Energy, Electricity and Nuclear Power: Developments and Projections—25 Years Past and Future (Vienna: IAEA, 2007).

(protons) are approaching each other at extremely high speeds, resulting in many high-energy collisions. Some colliding pairs will fuse into deuterium (H_2) (heavy hydrogen, composed of one proton and one neutron), emitting a neutrino and a positron in the process.

These deuterium atoms present a large target to the high-speed protons, and will experience additional collisions, creating an isotope of helium (He3 … two protons and one neutron) and emitting a gamma ray.[7] In the third step of the reaction, two He3 atoms collide, producing the standard helium atom He4 (two protons and two neutrons) and two high-energy protons, which are now available for more collisions, and a sustained nuclear reaction.

Overall, starting with six protons, we end up with one helium nucleus and two protons, and also emit two positrons, two neutrinos, and two gamma rays. In essence, the nuclear reaction has fused four protons into one helium nucleus. Recall the following:

1 atomic mass unit (u) = 1.661×10^{-27} kg.
Electron/positron mass 0.000549 u
Proton mass = 1.007276 u
Neutron mass = 1.008665 u
Helium nucleus = 4.0015050 u

Note that the total mass of the helium nucleus is less than the sum of its constituent nucleons. Computing the "before and after" nuclear masses:

$4 \times 1.007276 = 4.029104$ u
$1 \times 4.0015105 = 4.0015050$ u

The difference is called the mass defect (Δm) which is converted into pure energy:

$$\Delta m = 4.029104 - 4.015050 = 0.013954\ u = 0.023178 \times 10^{-27}\ \text{kg}$$

$$W = mc^2 = 0.023178 \times 10^{-27} (3 \times 10^8)^2 = 0.2575 \times 10^{-11}\ \text{J}$$

which is the energy created by the conversion of hydrogen into a single helium nucleus. It is primarily this energy which powers the sun.

Presently there are no commercial fusion power plants. Since at least the 1960s, it has been routinely predicted that commercial fusion was about thirty

[7] Tritium (two neutrons and one proton, or H^3) is also produced, providing a similar reaction path to creating helium.

years in the future, prompting some pundits to say, "Fusion is the energy source of the future, and always will be!" Still, since fusion is the "ultimate" solution to clean energy, research continues. The fundamental problem to achieving a controlled fusion reaction is to continuously confine an extremely high-density plasma of elementary particles at extremely high temperatures (well in excess of 10,000,000 K). If this can be done at all, the necessary equipment required is currently horrendously expensive.

3.3.1.5 Heat Rate Clearly efficiency is important when considering energy conversion from thermal to electrical form. Efficiency is defined as

$$\eta = \text{Efficiency} = \frac{\text{power output}}{\text{power input}}$$

where input and output must be in the same units. A value of unity (100%) would mean that all of the thermal input was converted to the electrical output, with no losses. But for thermal generating units, the input and output are not normally available in the same units. A more convenient (but more mysterious) way of describing the efficiency of a thermal power plant is through the use of the term "heat rate," defined as

$$\text{Heat Rate} = HR = \frac{\text{thermal power input (Btu/h)}}{\text{electrical power output (kW)}} = \frac{3414}{\eta}$$

The "3414" figure accounts for the different units used for input and output. Observe that a *lower* heat rate means *higher* efficiency. A heat rate of 3414 Btu/kW-hr would correspond to an efficiency of 100%.

Example 3.3

A coal-fired power plant has a heat rate of 9500 Btu/kW-h. What is the plant efficiency?

$$HR = \frac{3414}{\eta} = 9500$$

$$\eta = \frac{3414}{9500} = 0.3594 = 35.94\%$$

The thermal energy content of some selected fuels are as follows:

Gasoline: 48 MJ/kg
Natural gas: 37 MJ/m^3 @ STP (Standard Temperature and Pressure)
Coal: 32 MJ/kg
Oil: 44 MJ/kg
U^{235}: 71 × 10^6 MJ/kg

Example 3.4

The coal-fired plant of Example 3.3 has an output of 700 MW. How many tons of coal are burned per hour?

$$\text{HR} = 9500 \text{ Btu/kW} \cdot \text{hr}$$

$$\text{Input} = 9500 \text{ Btu/hr} \times 700{,}000 \text{ kW} = 6.65 \times 10^9 \text{ Btu/hr}$$

$$1 \text{ Btu} = 0.001055 \text{ MJ}$$

$$\text{Input} = 7.016 \times 10^6 \text{ MJ/hr}$$

$$\text{Input} = \frac{7016 \times 10^3 \text{ MJ/hr}}{32 \text{ MJ/kg}} = 219.2 \times 10^3 \text{ kg/hr} = 219.2 \text{ metric tons/hr}$$

3.3.1.6 Biomass "Biomass" is a term for products derived from living organisms, including wood from trees, harvested grasses, plant residues, and animal waste material. It is a traditional and important source of fuel, its key advantage being that it is continuously replenished. The U.S. Department of Energy (DOE) has noted that biomass recently exceeded hydropower as the largest domestic source of renewable energy. Currently, the worldwide installed capacity of electric generation from biomass is about 35 GW, which is expected to double every ten years through 2030, and is also expected to increase as a percentage of total generation, according to the DOE.

From one point of view, bioenergy production is CO_2 neutral since its combustion releases the same CO_2 that was removed from the atmosphere when it was formed.[8] Biomass can be burned directly or can be converted into a combustible gas. Municipal waste can be considered as a part of biomass, but is sometimes put in a separate category.

[8] Of course, the same is true of coal and oil; it's just that the latter did their extractions eons ago.

3.3.1.7 Solar Thermal Energy (STE) Stepping out of an air-conditioned building into the sunshine on a warm summer day makes us immediately aware that the sun is a source of energy. Solar power density onto a surface perpendicular to the incident solar radiation on the surface of the earth is about 1 kW/m, so estimating an average human face to be about 15 × 25 cm, or 0.0375 m^2, we feel about 37 W warming our smiling upturned faces.

The sun's total output power is estimated at a fantastic 386 yottawatts (386×10^{24} W). To put this number in perspective, the world's total annual energy consumption in 2008 was about 474 exajoules (474×10^{18} J). This means that the sun puts out enough energy to supply over 814,000 times the world's annual energy needs *every second*!

Solar thermal collectors may be defined as low-, medium-, or high-temperature types. Low- and medium-temperature collectors are generally flat plates. High-temperature collectors use mirrors or lenses to concentrate sunlight on a smaller area, and are those used for large-scale electric power production. An estimated 600 MW of solar thermal power was online in 2009, another 400 MW is under construction, and over 14,000 MW is under development. The major disadvantage is the requirement of large land areas, due to the inherent low power density of solar radiation. See Figure 3.7 for a typical STE installation.

FIGURE 3.7 A solar thermal installation. Source: http://www.photovoltaicsolarenergy.com.

3.3.1.8 Geothermal Energy Geothermal energy is that which is extracted from subterranean heat stored in the earth. This energy originates from the original formation of the planet, from radioactive decay of minerals, and from solar energy absorbed at the surface. It has been used for space heating and bathing since ancient Roman times. Worldwide, about 11 GW megawatts of electric generation from geothermal sources are online in twenty-four countries, with some 28 GW used for other purposes, such as heating and desalination.

Geothermal power is environmentally friendly and sustainable. Some greenhouse gases are released from geothermal wells but much less so than that related to the combustion of fossil fuels. At appropriate locations, geothermal power is also cost-effective.

Nowhere has geothermal energy been exploited more than in Iceland, where it supplies heating for over 87% of the population at low cost. The Svartsengi Geothermal Station, one of five major plants, is located in Keflavik, Iceland, and produces 76.5 MW of electrical generation and about 80 MW of hot water for space heating. Surplus mineral-rich water from the plant supplies the Blue Lagoon, a tourist bathing resort (see Figure 3.8). Total electric generation from geothermal sources is over 200 MW and growing. Iceland is also rich in hydroelectric power, with an installed capacity of 1155 MW. Table 3.1 shows the major geothermal sites worldwide.

FIGURE 3.8 The Svartsengi Geothermal Power Station Source: http://en.wikipedia.org/wiki/Svartsengi_Power_Station.

Table 3.1 Geothermal Energy Production: The Top Ten Countries in Geothermal Energy Utilization, 2005

Country	GWh Electric	Country	GWh Thermal
United States	17,917	China	12,605
Philippines	9,253	Sweden	10,001
Mexico	6,282	United States	8,678
Indonesia	6,085	Turkey	6,901
Italy	5,340	Iceland	6,806
Japan	3,467	Japan	2,862
New Zealand	2,774	Hungary	2,206
Iceland	1,483	Italy	2,099
Costa Rica	1,145	New Zealand	1,969
Kenya	1,088	Brazil	1,840
Sum of top ten	54,834	Sum of top ten	54,125
All other	1,952	All other	19,979
World total	56,786	World total	75,943

Source: Glitnir Bank (2007), United States: Geothermal Energy, Market Report, September 2007.

3.3.2 Nonturbine Thermal Prime Movers

Conventional internal combustion, or diesel, engines can be used as prime movers for electric power generation. These are usually small stand-alone units for emergency residential applications, medium-sized units to supply power at commercial construction sites, or medium-to-large units that serve as back-up power supplies for critical facilities (such as hospitals, communications, and police). They are typically noisy, low in efficiency, expensive, and in need of routine maintenance.

Another important application is that of emergency electric generators for residential applications, which range in size from about 2.5 to 7.5 kW and use traditional reciprocal internal combustion engines. Larger units are used for power at construction sites, and are more likely to use diesel engines. While such generators provide an important need, they constitute a negligible fraction of the total grid energy.

3.4 Nonthermal Prime Movers

Nonthermal prime movers are typically turbines, converting fluid flow into torque, the fluids being either water or air (wind).

3.4.1 Hydroelectric Generation

A conventional hydroelectric generating facility uses a dam to create a reservoir of water, which supplies a controlled flow through turbines that drive EMM generators. See Figure 3.9.

Recall from basic physics the following expression:

$$W = mgH = \text{gravitational potential energy}$$

$$m = \text{mass positioned at height “H” above a reference plane}$$

$$g = \text{acceleration of gravity} = 9.807 \text{ m/s}^2$$

Now apply this concept to a reservoir of water whose surface is located at height “H” above the surface of a second lower reservoir. When applied to a hydroelectric facility, this height “H” is called a “head.” Suppose we run a channel (a “penstock”) from the upper to the lower reservoir. Water will flow down the penstock, such that

$$P = power = \frac{dW}{dt} = \frac{d}{dt}(mgH) = gH\left(\frac{dm}{dt}\right) = gH\rho\left(\frac{dvol}{dt}\right)$$

$$\frac{dm}{dt} = \text{mass flow rate, kg/s}$$

$$\frac{dvol}{dt} = \text{volumetric flow rate, m}^3/\text{s}^2 = \rho \cdot \frac{dm}{dt}$$

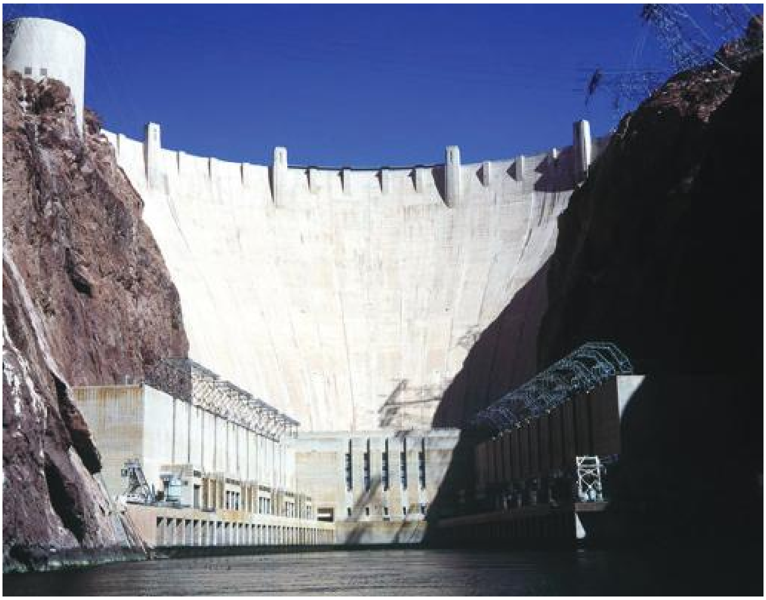
$$\rho = \text{ density of water} = 1000 \text{ kg/m}^3$$

Example 3.5

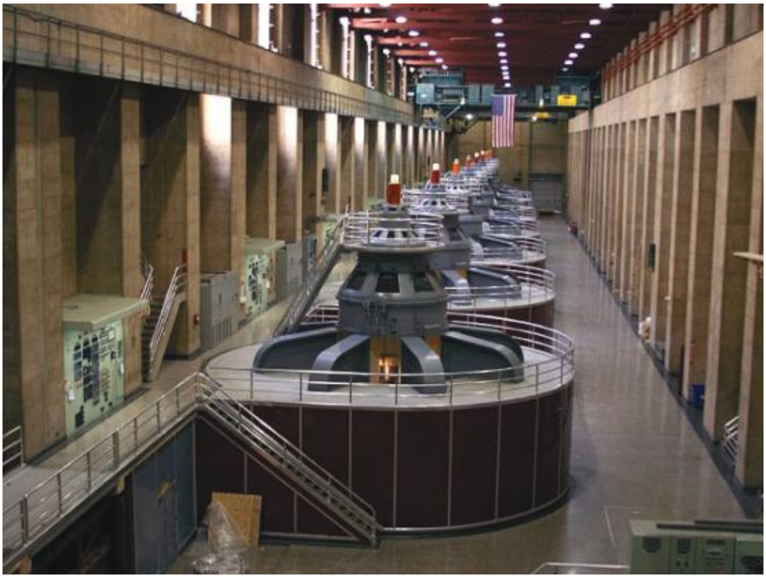
The head and flow at Niagara Falls, New York, are about 53 m and 6000 m^3/s, respectively. Determine the hydropower.

$$P = gH\rho\left(\frac{dvol}{dt}\right)$$

$$P = 9.807(53)(1000)(6000) = 3119 \text{ MW}$$



(a) Hoover Dam



(b) Generating Hall: Hoover Dam

FIGURE 3.9 A hydroelectric generating facility. a. Hoover Dam. b. Generating Hall, Hoover Dam. Source: http://www.usbr.gov/lc/hooverdam/images.

The device that converts the kinetic energy of the moving water to torque is called a "hydraulic turbine" and is connected to the generator shaft to become the prime mover for hydroelectric generation. See Figure 3.10 for three types of hydraulic turbines used for hydroelectric power generation.

The Three Gorges Dam on the Yangtze River in China is the centerpiece of the world's largest hydroelectric facility, which is the largest electricity-generating plant of any kind. When it reaches full capacity in 2011, its thirty-two 700 MW generators, plus two smaller 50 MW units, will supply a total capacity of 22.5 GW.

In 2006, hydroelectric generation (777 GW) accounted for about 19% of the world's total generating capacity (4012 GW). Hydroelectric generation features zero hydrocarbon emissions and zero fuel costs. However, its environmental impact is far from negligible. The Three Gorges Dam

(a)

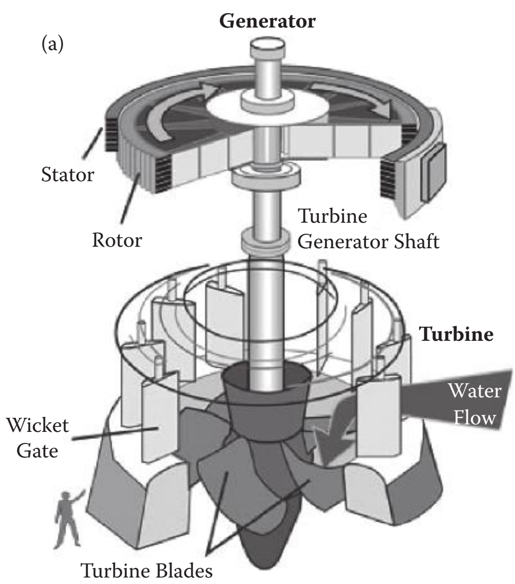


FIGURE 3.10 Three types of hydraulic turbines used in hydroelectric generation. a. Kaplan Turbine (low head: $2 < H < 40$ m); b. Francis Turbine (medium head: $10 < H < 350$ m); c. Pelton Turbine (high head: $50 < H < 1300$ m) (b and c on following page). Source: a. http://upload.wikimedia.org/wikipedia/commons/a/a4/W ater_turbine.jpg. b. http://upload.wikimedia.org/wikipedia/commons/4/44/Francis_Runner_grandcoulee.jpg. c. http://upload.wikimedia.org/wikipedia/commons/d/d9/Peltonturbine-1.jpg.

(b)

(c)

FIGURE 3.10 **(Continued).**

project is said to have displaced 1.3 million people, and has been dubbed an "environmental disaster," flooding archaeological and cultural sites and causing significant ecological changes, including an increased risk of landslides.

3.4.2 Tidal Power Generation

Another form of hydroelectric power captures the daily flow of tidal currents, using turbines appropriate for low-head hydro. The world's first and largest tidal generating facility is the Rance Power Station, located on an estuary of the Rance River in Brittany, France. The installation is owned and operated by Électricité de France; has a peak rating of 240 MW; and features an average difference of 8 m between high and low tide, with a maximum difference of 13.5 m. See Figure 3.11.

The greatest tidal elevation difference in the world is observed in the Bay of Fundy, Nova Scotia, Canada, measured at average and maximum values of 13.9 and 16.6 m, respectively, causing over 100 billion tons of water to flow

FIGURE 3.11 Rance Tidal Barrage, 240 MW power station, Brittany, France. Source: http://www.reuk.co.uk/La-Rance-Tidal-Power.

in and out of the bay in a little over 6 hours. The push to green energy has sparked renewed interest in this location as a source of tidal power generation. A large tidal difference in elevation is necessary for commercial viability, restricting the technology to just a few sites. Also, altering tidal flows at one location has an impact on neighboring sites, with ecological consequences. Research on this resource is ongoing.

3.4.3 Wave Generation

Deep-water wave power resources are enormous, and are estimated to be in excess of 2 TW. Locations with the most potential for wave power are in the north and south temperate zones, since the prevailing winds in these regions are very predictable. Water has a density over 800 times that of air, permitting much smaller energy conversion devices. To date there are very few commercial installations, but this resource shows great promise, and research is ongoing. See Figure 3.12 for several wave machine designs.

3.4.4 Wind Generation

A mass of moving air (wind) has kinetic energy, which, if passed through a properly designed turbine, can be converted into torque. See Figure 3.13 for two turbine designs, horizontal wind turbine (HAWT) and vertical wind turbine (VAWT).

Consider the standard three-dimensional right-handed *xyz* Cartesian coordinates (*y* up in the vertical direction, *x* to the right, and *z* out of the page). Now consider the kinetic energy of a mass of air *dm* moving along the x-axis at velocity *u*.

$$dW = \frac{1}{2} dm \cdot u^2 = \text{kinetic energy of mass } dm$$

$$u = \text{horizontal velocity of mass } dm \quad \text{along the x-axis}$$

Now consider the mass to be passing through a circular area of radius *r* in the *y*–*z* plane in time *dt*.

$$P = \frac{dW}{dt} = \frac{1}{2}\left(\frac{dm}{dt}\right) \cdot u^2 = \text{power of mass } dm \text{ moving through the area}$$

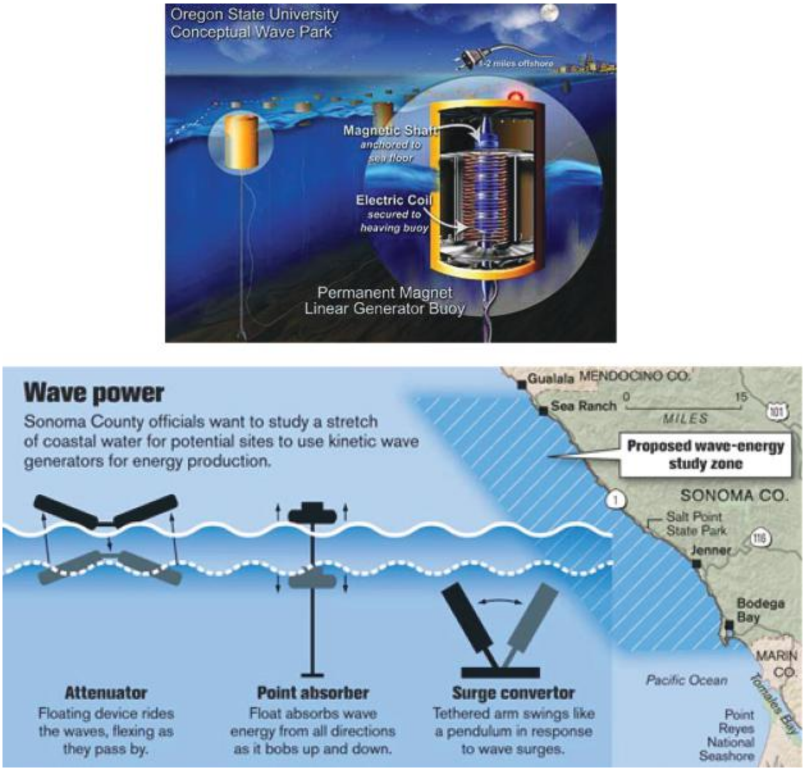


FIGURE 3.12 **Several wave machine designs. Top: Oregon State University/NSF Conceptual Wave Park. Bottom: From *ESRI Nature* 8, no. 11, Design Science News, Buckminster Fuller Institute. Next page: Top: http://upload.wikimedia.org/wikipedia/commons/c/cc/Pelamis_bursts_out_of_a_wave.JPG; Bottom: http://en.wikipedia.org/wiki/File:WaveDragon.JPG.**

$$dW = \frac{1}{2} dm \cdot u^2 = \text{kinetic energy of mass } dm$$

$$u = \text{horizontal velocity of mass } dm \quad \text{along the x-axis}$$

The density of air (ρ) is somewhat complicated since it depends on the definition of "ideal" air, the temperature, the altitude, and the water vapor content. It is approximately 1.2 kg/m^3 at sea level and room temperature, a value which is sufficiently accurate for our purposes. If one could design a "perfect" wind

FIGURE 3.12 Continued.

FIGURE 3.13 **Two wind turbine designs. Left: A horizontal wind turbine (HAWT); and right: a vertical wind turbine (VAWT). Source: Left: http://en.wikipedia.org/wiki/File:Windenergy.jpg. Right: http://upload.wikimedia.org/wikipedia/commons/3/3c/Darrieus-windmill.jpg.**

turbine, one could convert all of this power into torque, and the air velocity on the downstream side would be zero. A study of the aerodynamics of the conventional turbine reveals that the maximum theoretical limit on turbine efficiency is 59.26%.[9] In general:

$$P_m = \eta \cdot \frac{1}{2}(\pi \rho r^2) \cdot u^3 = \text{mechanical power on the turbine shaft in Nm}$$

$$\eta = \text{turbine efficiency} < 0.5926$$

Example 3.6

The Vestas V90 3 MW wind turbine features a three-blade design with a radius of 45 m and a rated wind speed of 15 m/s. Compute the efficiency.

[9] Called the "Betz limit," after the German physicist Albert Betz, who first derived it in 1926.

$$P_m = \eta \cdot \frac{1}{2}(\pi\rho r^2)\cdot u^3 = \eta \cdot \frac{\pi(1.2)(45)^2(15)^3}{2} = (25.77\times 10^6)\cdot\eta$$

$$\eta = \frac{6}{25.77} = 0.2329 = 23.29\%$$

3.5 Chemical to Electric Energy Conversion

It is no accident that electrical engineering at many institutions began in chemistry departments. As early as 1794, Alessandro Volta had discovered that a chemical interaction causes electricity, and he had developed the first battery by 1800.

3.5.1 The Primary Cell

A typical primary ("nonrechargeable") cell is a manganese dioxide (MnO_2) and zinc (Zn) alkaline cell, which utilizes a manganese dioxide cathode and a zinc anode that becomes oxidized. At the anode:

$$\mathrm{Zn} + 2\mathrm{MnO_2} + \mathrm{H_2O} \rightarrow \mathrm{ZnO} + 2\mathrm{MnOOH}$$

At the same time, the anode is consuming hydroxyl ions and producing water:

$$\mathrm{Zn} + 2\mathrm{OH}^- \rightarrow \mathrm{H_2O} + \mathrm{ZnO} + 2e^-$$

When the circuit is closed, the free electrons from the zinc anode electrode will flow through the external circuit to the cathode electrode. The cell is exhausted (completely discharged) when all of the zinc has been converted to zinc oxide. The process is not reversible, hence the cell is not rechargeable.

3.5.2 The Secondary Cell

Suppose two plates of dissimilar metals ("electrodes") are immersed in an appropriate conducting fluid (the "electrolyte"). In general, we observe that a voltage will appear between the electrodes. If a closed circuit is made by externally connecting a resistor between the electrodes, a sustained electrical current will flow, with a corresponding sustained chemical reaction at each electrode. Such a structure is called a "cell." One of the oldest cells that is still commonly used is the "lead–acid" cell, using a lead (Pb) electrode, a lead dioxide (PbO_2) electrode, and a strong (35%) sulfuric acid (H_2SO_4) solution as the electrolyte.

At the PbO_2 electrode, lead sulfate and water are produced, along with two free electrons and three hydrogen ions, making this the positive electrode (the anode), with an electrochemical potential of 1.685 V from anode to electrolyte:

$$PbO_2 + 2H_2SO_4 \rightarrow PbO_2 + 3H^+ + 2e^- + HSO_4^- \rightarrow PbSO_4 + 2H_2O$$

At the Pb electrode, lead sulfate and water are also produced, along with two free electrons and one hydrogen ion, an excess of one free electron, making this the negative electrode (the cathode), with an electrochemical potential of −0.356 V from cathode to electrolyte:

$$Pb + 2H_2SO_4 \rightarrow PbSO_4 + 2H_2O + H^+ + 2e^-$$

For the entire cell:

$$PbO_2 + Pb + 2H_2SO_4 \rightarrow 2PbSO_4 + 2H_2O$$

The total electrochemical potential of −0.356 V from anode to cathode is therefore $1.685 - (-0.356) = +2.041$ V. If an external connection is made such that external current flows from anode to cathode, eventually both electrodes are completely converted to lead sulfide, and the electrolyte to water, at which point we say that the cell is "completely discharged." If current is forced into the anode from an externally applied source, the reaction is reversed, and we say the cell is "charging." Because of this property, wet cells are also called "rechargeable." The charge–discharge cycle can be repeated indefinitely. Various materials can be used in secondary cells.

3.5.3 The Battery

A battery is a number of identical cells, permanently connected in series and/or parallel. A primary battery uses primary cells; a secondary battery uses secondary cells. The more descriptive terms "nonrechargeable" and "rechargeable" are commonly used instead of "primary" and "secondary," respectively.

An example is the familiar 12 V car battery, which uses six lead–acid cells connected in series. A cell typically consists of two sets of eight plates, each a little larger than a large postcard and interleaved with, but not touching, each other. One set is made of lead–oxide and forms the positive anode:

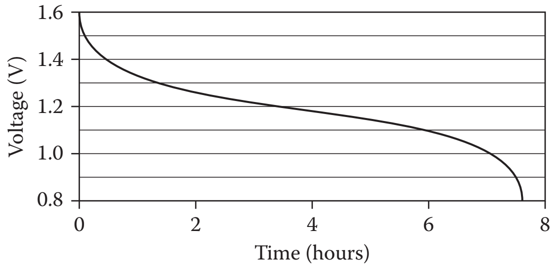


FIGURE 3.14 **Voltage versus time for a AA 1.5 volt alkaline battery at 20°C constant current of 250 mA. Source: http://data.energizer.com/PDFs/alkaline_appman.pdf.**

the other is made of lead and forms the negative cathode. The electrodes are immersed in the electrolyte (a 35% sulfuric acid [H_2SO_4] solution). The conduction path is the short distance of a few mm through the electrolyte between opposite plates, and is dispersed over a large area, permitting high current flow.

Connecting six such cells in series produces a total electrochemical voltage of $6 \times 2.041 = 12.246$ V (nominally 12 V).

Car batteries have three ratings[10] that are important to consumers:

Voltage: The voltage rating is the nominal open-circuit voltage at full charge. For the standard 12 V car battery, the actual voltage would be about 12.3 V.

"Cold-cranking amps" (CCA): This is the minimum current the battery can deliver at 0°F with the terminal voltage maintained at a minimum of 7.2 V.

Amp-hours (A-hr): This is a measure of stored charge in a fully charged battery (1 A-hr = 3600 coulombs).

A second example is a typical 1.5 V alkaline AA battery, which consists of a single MnO_2–zinc cell. A plot of battery voltage versus time at 20°C and a constant current of 250 mA is shown in Figure 3.14. The plot shows that the voltage slowly drops from 1.5 V to about 1.0 V in a little over 7 hours, after which the battery voltage drops rapidly.

The familiar 9 V alkaline battery has six such cells in series.

[10] Of course, the warranty period and its particulars are also important.

Batteries come in dozens of ratings and sizes, which are defined in IEC standards. The relevant U.S. standard is ANSI C18.1, *Dry Cells and Batteries: Specifications*, which was developed by a committee of the National Electronic Distributors Association (NEDA). Both standards cover physical specifications, safety, and differing types of primary and rechargeable cells. As usual, manufacturers also have their own systematic identification of cell types. A tabulation of various battery types is provided in Table 3.2.

Table 3.2 Some Battery Types

Nonrechargeable Batteries	Rechargeable Batteries
• Alkaline battery	• Dr30 battery
• Aluminum battery	• Flow battery
• Atomic battery	• Vanadium redox battery
• Optoelectric nuclear battery	• Zinc–bromine flow battery
• Nuclear microbattery	• Fuel cell
• Bunsen cell	• Lead–acid battery
• Chromic acid cell	• Deep cycle battery
• Clark cell	• VRLA battery
• Daniell cell	• AGM battery
• Dry cell	• Gel battery
• Earth battery	• Lithium–ion battery
• Galvanic cell	• Air-fueled lithium–ion battery
• Grove cell	• Lithium–ion polymer battery
• Leclanché cell	• Lithium–iron–phosphate battery
• Lemon battery	• Lithium–sulfur battery
• Lithium battery	• Lithium–titanate battery
• Lithium air battery	• Molten salt battery
• Mercury battery	• Nickel–cadmium battery
• Molten salt battery	• Nickel–cadmium battery, vented cell type
• Nickel oxyhydroxide battery	
• Oxyride battery	• Nickel–hydrogen battery
• Organic radical battery	• Nickel–iron battery
• Paper battery	• Nickel–metal hydride battery
• Reserve battery	• Low self-discharge NiMH battery

(*continued*)

Table 3.2 Some Battery Types (Continued)

Nonrechargeable Batteries	Rechargeable Batteries
• Silver–oxide battery	• Nickel–zinc battery
• Solid-state battery	• Organic radical battery
• Voltaic pile	• Polymer-based battery
• Penny battery	• Polysulfide bromide battery
• Trough battery	• Rechargeable alkaline battery
• Water-activated battery	• Sodium–sulfur battery
• Weston cell	• Super iron battery
• Zinc–air battery	• Zinc–bromine flow battery
• Zinc–carbon battery	• Zinc matrix battery
• Zinc–chloride battery	

Source: http://en.wikipedia.org/wiki/List_of_battery_types.

A cursory study of Table 3.3 reveals that lithium–ion batteries have superior energy density capabilities. Hence it is not surprising that Li–ion batteries are a prime consideration for electric and hybrid electric vehicles (EVs and HEVs). Advantages include the following:

- High energy density
- High cell voltage (3.7 V)
- Low self-discharge rates

Table 3.3 Relative Energy Storage Ratings for Some Battery Types

Battery Type	Cost $ per Wh	Wh/kg	Joules/kg	Wh/liter
Lead–acid	$0.17	41	146,000	100
Alkaline long-life	$0.19	110	400,000	320
Carbon–zinc	$0.31	36	130,000	92
NiMH	$0.99	95	340,000	300
NiCad	$1.50	39	140,000	140
Lithium–ion	$4.27	128	460,000	230

Source: http://www.allaboutbatteries.com/Battery-Energy.html.

- Satisfactory operation over a large range of temperature
- Low maintenance

The major disadvantage is high cost. Fortunately, there is an economy of scale as the market for EVs and HEVs expands. Li–ion battery design criteria vary depending on the application (all electric or hybrid electric). For the EV, we design for maximum energy density; the HEV requires a maximum power density design, resulting in a significantly cheaper battery. For a detailed authoritative study of costs, consult *Costs of Lithium-Ion Batteries for Vehicles*, published by the Center for Transportation Research, Argonne National Laboratory, DOE.

3.5.4 The Fuel Cell

It takes more energy to manufacture primary batteries than that which is contained in the battery, and once the battery is discharged, it must be replaced. The secondary battery can be recharged, but this requires a recharging device, which in turn requires more electrical energy than that which is stored in the battery. Thus one can argue that batteries are not "sources" at all, but rather energy storage devices, like capacitors and inductors. If it were possible to continuously chemically refresh the electrolyte and the electrodes, we could exploit the electrochemical reaction to make a true electrical source.

It happens that there is such a device, called the "fuel cell," whose history is almost as old as that of the battery. The principle of the fuel cell was published by German scientist Christian Friedrich Schönbein in 1838 and demonstrated by Welsh scientist Sir William Robert Grove in 1839. In 1955, W. Thomas Grubb, a General Electric (GE) chemist, further refined the original fuel cell design, which itself was further improved by GE chemist Leonard Niedrach. This became known as the "Grubb–Niedrach fuel cell," which, in cooperation with NASA and McDonnell Aircraft, led to its use on the *Gemini* spacecraft project, which was the first commercial use of a fuel cell. Later, a cell using potassium hydroxide as the electrolyte and compressed hydrogen and oxygen as the reactants was developed, producing electricity and sterile drinking water as a byproduct, making it a near-ideal power supply for spacecraft.

Like batteries, there are many possible choices for electrolyte and electrode materials, and growth in the technology continues. Another key part of the fuel cell design is a necessary internal membrane through which ions must pass between electrodes. In the hydrogen fuel cell, air may be substituted for the second reactant because of its high oxygen content. The hydrogen fuel cell is particularly attractive because its waste product is water. One kg of hydrogen has a yield of 140.4 MJ when combined with oxygen, with a productivity at a maximum efficiency of about 70% electric (98.3 MJ); the remaining 30% converted to thermal energy (42.1 MJ), producing 9 kg of water.

Example 3.7

A hydrogen–oxygen fuel cell is designed for an electric vehicle (EV) application to produce a maximum output of 75 kW.

A. Determine the maximum mass flow of hydrogen
  A hydrogen mass flow rate of 1 g/s creates a total power of 140.4 kJ/s or 98.3 kJ/s (98.3 kW) electric.
  75 kW corresponds to (75/98.3)(1) = 0.763 g/s.

B. If it takes a fuel cell output of 20 kW at a cruising speed of 55 mph to maximize the range, what mass of hydrogen fuel is required for a range of 400 miles?
  20 kW corresponds to (20/98.3)(1) = 0.2035 g/s
  55 mph = 24.5872 m/s
  400/55 = 7.273 hours = 26.182 ks
  m = 26.182 ks (0.2035 g/s) = 5.328 kg

C. How much water is produced by the cell in the situation described in (b)?

$M_{H20} = 9 \times 5.328 = 47.95$ kg

Larger (200 kW and higher) stationary fuel cell systems have been developed for cogeneration applications in hospitals, universities, and commercial applications. Perhaps the most active area of fuel cell development currently is for the hybrid electric (HEV) and/or the all-electric (EV) vehicle.

For EV applications, the overall efficiency is around 64% (chemical to electrical to mechanical), which still exceeds traditional internal combustion engine propulsion. The disadvantages include the following:

- A commercially viable source of hydrogen fuel
- A commercially viable network of refueling stations

- Fuel cell costs
- Durability of the membrane over a large temperature range
- Hydration of the membrane over a large temperature range
- Safety issues associated with on-board hydrogen storage and refueling
- Bulk and weight

To date, fuel cells make a small, but significant and growing, fraction of the overall mix of the total electrical energy supply.

3.6 Photovoltaic Energy Conversion

If electromagnetic radiation impinges on a semiconductor (P–N) junction, a voltage is produced, a phenomenon known as the "photovoltaic effect," which was discovered by Alexander Bequerel in 1839.[11] These P–N junctions are arranged into structures called "solar cells," which in turn are interconnected and mounted in assemblies called "solar panels."[12] A solar panel array is shown in Figure 3.15.

The most important technical information relating to a solar panel is its so-called I–V characteristic, which is a plot of its terminal current (I) versus its

FIGURE 3.15 Solar panel array. Source: http://en.wikipedia.org/wiki/File:Solar_panels_in_Ogiinuur.jpg.

[11] The discussion of the physics of photovoltaics will be discussed later.
[12] It is assumed that sunlight will be the source of the incident electromagnetic radiation.

terminal voltage (V). The I–V characteristic depends on the insolation level, defined as the solar power density of incident solar radiation. The average solar insolation at the outer limits of the earth's atmosphere is approximately 1.366 kW/m^2.[13] At the earth's surface, standard test conditions (STC) are specified as receiving 1.0 kW/m^2 insolation of spectral type AM1.5 (defined by convention to simulate sunlight) and 25°C cell temperature. I–V characteristics are typically provided at STC.

Example 3.8

The I–V characteristic for a typical solar panel appears in Figure 3.16a.[14] There are several important ratings and other data, including:

V_{OC} = the open circuit panel voltage = 22.5 V
I_{SC} = the short circuit panel current = 3.63 A
V_{MP} = the panel voltage at the maximum power point = 18.0 V
I_{MP} = the panel current at the maximum power point = 3.33 A
$P_{\text{MAX}} = V_{MP} I_{MP} = 60$ W
Length × width × depth = 80 × 55 × 3.5 cm
Weight = 5.5 kg

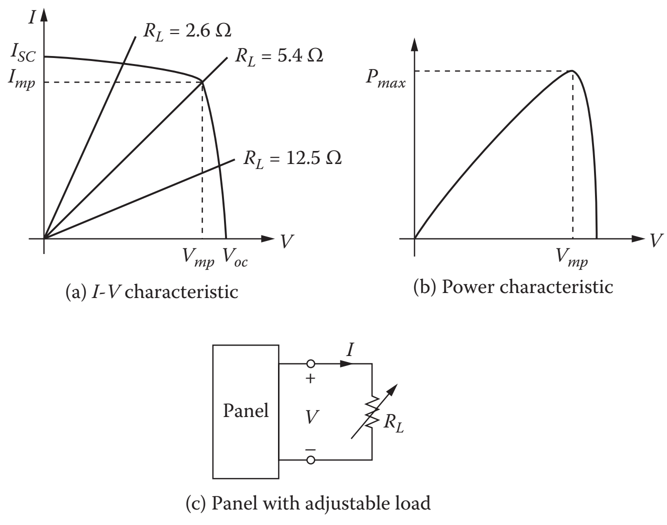


FIGURE 3.16 Solar panel characteristics.

[13] Data from weather satellite observations.
[14] The I–V characteristic is at STC.

a. Suppose the panel is terminated in an adjustable resistor R_L, as shown in Figure 3.16c. Determine the panel voltage and current for a given value of R_L.

METHODOLOGY

The panel is perceived as a nonlinear source, with the given I–V characteristic shown in Figure 3.16a. Ohm's law requires that V = $R_L I$, which, if plotted in I–V coordinates, is a straight line emanating from the origin with slope $1/R_L$ (the "load" line). Thinking graphically, for a given R_L, observe that two constraints must be simultaneously met: (1) we must be on the panel I–V characteristic, and (2) we must be on the load line. Thus, we must be at the intersection of the two curves. We will consider three values of R_L. Plotting the load line in Figure 3.16b, and reading the intersection points:

R_L	V	I	P
Ohm	Volt	Ampere	Watt
2.6	9	3.46	31
5.4	18	3.33	60
12.5	20	1.60	32

We can show by picking values around 5.4 Ω that this is the maximum power point.

b. Determine the power density in watts/m².

$$Area = l \cdot w = 0.80(0.50) = 0.40 \text{ m}^2$$

$$S = \frac{P}{Area} = \frac{60}{0.40} = 150 \text{ W/m}^2 \text{ (electric)}$$

c. How many panels are required for a 5 kW application?

$$N_P = \frac{5000}{150} = 33.33; \quad \text{round up to 34 panels}$$

d. Determine the roof area and weight required for a 5 kW application.

$$Area = 34(0.40) = 13.6 \text{ m}^2$$

$$Weight = 34(5 \text{ kg}) = 170 \text{ kg}$$

As of 2010, solar photovoltaics generate over 21 GW of grid-connected electric power globally, which is currently a small fraction of the 4800 GW total electric power-generating capacity. However, it is currently the fastest growing generation technology, and, as the costs drop, it is destined to become a major resource.

3.7 Thermal to Electric Energy Conversion

3.7.1 The Thermoelectric Effect

It is possible to convert thermal energy directly to the electric form exploiting a principle called the thermoelectric, or Seebeck, effect. When two sides of a junction of dissimilar metals are at different temperatures, a voltage across the junction is created. Conversely, when an external voltage is applied across the junction, it creates a temperature difference known as the cooling, or Peltier, effect. This effect can be used to generate electricity, to measure temperature, to cool objects, or to heat or cook them.

Thermoelectric devices have the advantages of simplicity and no moving parts. Disadvantages are low efficiency, high cost, and bulk for high outputs. Research and development continue, but to date thermoelectric devices have not added substantially to the total electrical energy supply.

3.7.2 Magnetohydrodynamics (MHD)

It is recommended that the reader make a simple sketch. Draw (or visualize) a three-dimensional (xyz) right-handed Cartesian coordinate system (x to the right, y up, and z out of the page). Now visualize two parallel conducting plates in the horizontal (xz) plane. A conducting ionized hot plasma is flowing in the positive x direction (left to right) between the plates, and also through a magnetic field which is oriented along the z axis in the negative z direction (into the page). There will be a Lorentz force[15] on each plasma ion in the y direction, that is, mutually perpendicular to the flow direction (x) and the magnetic field ($-z$), according to the following:

$$\hat{F} = q(\hat{v} \times \hat{B})$$
$$\hat{F} = \text{vector Lorentz force}$$
$$\hat{v} = \text{vector velocity of } q$$
$$\hat{B} = \text{vector magnetic field (flux density)}$$
$$q = \text{ion charge}$$

[15] The Lorentz force will be discussed in greater detail in Section 3.5.

The force on positive ions will be up, and it will be down on negative ions, forcing a positive charge to collect on the upper plate and a negative charge to collect on the lower plate. Thus we have a dc voltage between the plates, which can deliver a current to an externally connected load.

The residual plasma heat can be used to make steam and drive a turbine to make the overall system more efficient. The emergence of "super" magnets (those producing extremely high magnetic fields using superconducting coils) has spurred continuing research into MHD. The application of MHD as a component of the generation mix for electric utilities is negligible at this time.

3.8 Summary

Controlled energy is vital to the development of human civilization. Because of its controllability, energy in the electric form is particularly important. Since "natural" electrical energy is not abundant, we must consider conversion from other forms, which include the following:

- Electromagnetic-mechanical (EMM) energy conversion
  - Thermal prime movers
    - Steam turbine prime movers
      - Coal
      - Natural gas (methane and CH_4)
      - Oil
      - Nuclear
        - Fission
        - Fusion
      - Biomass
      - Solar thermal energy
      - Geothermal energy
    - Nonturbine prime movers
  - Nonthermal prime movers
    - Hydroelectric generation
    - Tidal power generation
    - Wave generation
    - Wind generation
- Chemical to electric energy conversion
  - The primary cell
  - The secondary cell

The battery
The fuel cell
Photovoltaic energy conversion
Thermal to electric energy conversion
The thermoelectric effect
Magnetohydrodynamics (MHD)

Which of these conversion methods is to be selected depends on economics, environmental impact, the availability of fuels, political considerations, and the current state of related technologies. All are important and contribute to the future of the planet. A summary of the world's electric generation capacity as of 2006 is shown in Table 3.4.

Electrical engineering is primarily concerned with two major applications:

Control of energy
Control of information

Table 3.4 Installed Electric Generation Capacity as of January 1, 2006 (GW)

Country	Thermal	Nuclear	Conventional Hydroelectric Geothermal/ Solar	Wind, and Biomass	Total
United States	761.603	77.821	100.334	24.996	964.754
North America	835.600	160.193	115.044	27.957	1,138.794
Antarctica	0.000	0	0	0	0.000
Central and South America	81.941	128.479	3.025	6.666	220.112
Europe	457.574	164.642	135.745	52.227	810.188
Eurasia	237.740	67.928	38.638	0.239	344.544
Middle East	135.904	8.731	0	0.029	144.664
Africa	86.034	22.043	1.800	0.381	110.258
Asia and Oceania	917.397	224.744	82.572	19.162	1,243.875
World Total	2,752.191	776.760	376.824	106.661	4,012.435

Source: U.S. Energy Information Administration, *International Energy Annual Report 2006*, Table 6.4.

An electrical energy (or power) system has four major parts:

Generation
Transformation
Transmission
Utilization

We will consider generation and transformation in Chapter 4.

References

Hinrichs, Roger A. *Energy*. London: Sanders College Publishing, 1992.
Kiehne, H. A., ed. *Battery Technology Handbook*, 2nd ed. New York: Marcel-Dekker, 2001.
Kirby, Richard S., Sydney Withington, Arthur B. Darling, and Frederick G. Kilgour. *Engineering in History*. New York: McGraw-Hill, 1956.
Patel, Mukund R. *Wind and Solar Power*. New York: CRC Press, 1999.
Reece, Holtzapple. *Foundations of Engineering*. New York: McGraw-Hill, 2003.
Wright, Paul H. *Introduction to Engineering*, 2nd ed. New York: Wiley, 1994.

Problems

3.1. Define energy and power, and provide SI units.

3.2. What are the advantages of electrical energy?

3.3. Write a short one-page paper on one of the following:
- Lightning
- Bioelectricity
- Electroluminescence

3.4. Electromagnetic-mechanical (EMM) energy conversion is based on two principles: Faraday's law and the Lorentz force equation.
  a. State Faraday's law.
  b. Provide the Lorentz force equation.

3.5. The machine of Figure 3.2 has one-turn coil dimensions of $r = 20$ cm and $l = 40$ cm and rotates at 314.2 rad/s in a uniform magnetic field of 0.3979 tesla.
  a. Determine v_{ad}.
  b. Suppose the one-turn coil of (a) is replaced by a ten-turn coil. Determine v_{ad}.

3.6. Continuing Problem 3.5b, we terminate the coil in a 12.5 Ω resistor.
  a. Find the current.
  b. Find the instantaneous power.

c. Find the average power.
d. Find the EMM torque.
e. What is the direction of the EMM torque?

3.7. Derive the expression for heat rate.

3.8. A coal-fired 500 MW power plant uses coal with an energy content of 32 MJ/kg. How much coal is required daily if the plant HR = 8000 Btu/kW-hr?

3.9. Explain the following terms:
a. Nuclear fusion
b. Nuclear fission
c. Mass defect

3.10. A conventional nuclear fission reaction with U235 (mass 235.044 u) results in a mass defect of

$$m = 0.1860\ u$$

a. Determine the energy released in the fission of 1 kg of U235.
b. Determine the mass of coal with equivalent energy content (32 MJ/kg).

3.11 Write a short one-page paper on geothermal energy in one of the following countries:
a. Iceland
b. United States

3.12. The head at Three Gorges Dam, China, is about 80.6 m, and the maximum flow for one of the thirty main 700 MW generators is 950 m^3/s. Determine the hydropower.

3.13. The head at the French 240 MW Rance Tidal Power Station has a maximum value of 13.5 m. Assuming an efficiency of 90%, determine the corresponding volumetric water flow.

3.14. Write a short (one-page) paper on wave generation.

3.15. A potential site for a wind turbine has an average sustained wind speed of 14 m/s. A conventional three-blade 1 MW HAWT is selected, with an efficiency of 40%. Determine the blade length.

3.16. Explain the difference between primary and secondary electrochemical cells.

3.17. a. Explain the difference between a battery and an electrochemical cell.
   b. Tests show that a fully charged electrochemical cell has an open-circuit voltage of 2 V, a short-circuit current of 10 A, and a charge of 2 A-hr. Explain (in terms of diagrams of series or parallel combinations of these cells) how you would make the following batteries:
      1. A 12 V battery with a short-circuit current of 10 A
      2. A 6 V battery with a short-circuit current of 20 A
   c. Determine the charge rating of each battery in part (b).

3.18. A standard lead–acid 12 V car battery has a CCA rating of 100 A. Arguing that the battery has a Thevenin equivalent circuit, determine the Thevenin elements.

3.19. A hydrogen–oxygen fuel cell is designed for an electric vehicle (EV) application to produce a maximum electrical output of 80 kW.
   a. Determine the maximum mass flow of hydrogen, assuming a maximum efficiency of 65%.
   b. If it takes a fuel cell output of 10 kWe at a cruising speed of 55 mph, what mass of hydrogen fuel is required for a range of 350 miles?
   c. How much water is produced by the cell in the situation described in (b)?

3.20. An ideal solar panel has a rectangular I–V characteristic such that

V_{OC} = the open-circuit panel voltage = 25 V
I_{SC} = the short-circuit panel current = 10 A
Length × width × depth = 1.5 × 80 × 4 cm
Weight = 13 kg

   a. Determine the coordinates of the maximum power point and P_{MAX}.
   b. Design a 250 V 5 kW array using the panels of (a). The array is to fit in a rectangular space of 3 × 8 m. Make a diagram showing electrical interconnections.
   c. Determine the array weight.

3.21. We wish to build a solar thermal electric-generating facility in a desert location with a maximum solar insolation level of 1.2 kW/m^2. If the maximum plant electrical output is to be 100 MW, and the overall thermal-electrical efficiency is 40%, how much collector area will be required?
3.22. Write a short one-page paper on the thermoelectric effect.
3.23. Make a drawing of the MHD system discussed in Section 3.7.2.

Chapter 4

Electrical Generation and Transmission

In Chapter 2 we studied the electric circuit operating in the ac mode. We built on that study in Chapter 3, considering basic sources of controllable energy and how it is converted into the electrical form, with special emphasis on electromagnetic-mechanical (EMM) energy conversion. We expand on that here, considering electrical generation in greater detail, and how electrical energy gets to the point of usage.

There are two standard power frequencies used worldwide: 50 Hz and 60 Hz, 50 Hz being the most common. 60 Hz is used throughout North America and a few other countries. Japan uses both frequencies.

It happens that, as we deal with larger amounts of power, a standard circuit configuration known as "three-phase" is used, operating in a mode referred to as "balanced." Hence, our first order of business is to develop and understand the balanced N-phase circuit, with particular emphasis on the three-phase configuration.

4.1 Polyphase ac Circuits

Consider the circuit shown in Figure 4.1, operating in the ac mode. The circuit is designed to supply electrical power from an electrical source (on the left) to a load (on the right), and as such, it is viewed to have three components: the source, the transmission line, and the load.

The limit to the amount of power we can transfer is determined by the maximum values of voltage (V) and current (I) that are practical. The voltage that is selected will define the insulation required, not just for the line, but for the source and load as well. Hence, it is important to have some standard values for compatibility between components. Some standard U.S. voltages are shown in Table 4.1.

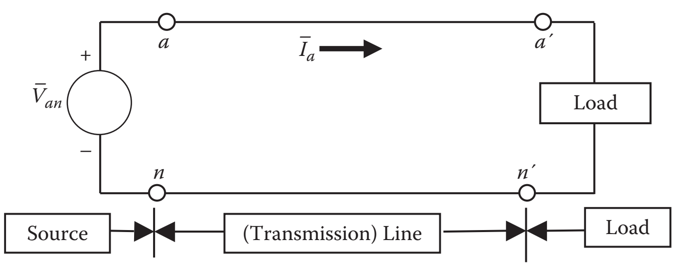


FIGURE 4.1 A basic electrical power transmission line.

Table 4.1 Some Standard U.S. Power Voltages

Secondary Distribution	120/208 V	240 V	277/480 V		
Primary Distribution	2.4/4.16 kV	7.2/12.47 kV	20/34.5 kV		
Transmission	115 kV	230 kV	345 kV	500 kV	765 kV

Now that the system voltage level has been selected, we are constrained to operate the line at or near that value. For a given voltage, the maximum power then becomes proportional to the maximum current to be allowed. The current-carrying capacity ("ampacity") of a conductor is determined by thermal issues. If the conductor overheats, its insulation system may be damaged, as well as the conductor itself. Also, a fire hazard is created.

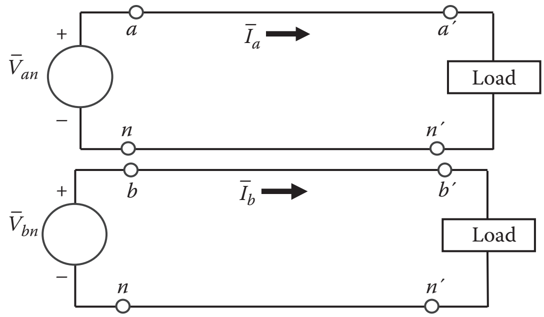
The ampacity of a conductor is determined by its size (more specifically, its cross-sectional area (A)) and its electrical conductivity (σ). The two common materials used for power conductors are copper (Cu) and aluminum (Al), whose important properties are shown in Table 4.2.

Copper is the better conductor on a per volume basis because of its higher conductivity. However, aluminum is better on a per weight basis because of its lower density. Also, aluminum conductors are cheaper.

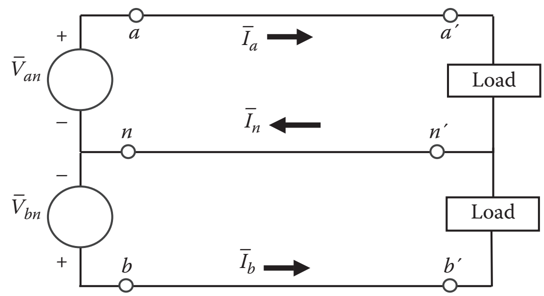
Now suppose that we have selected a standard voltage and, for a given maximum power transfer, the current and corresponding conductor, with

Table 4.2 Some Properties of Copper and Aluminum

Copper	
Atomic number	29
Density	8.9 g/cm^3 at 20°C
Melting point	1083°C
Boiling point	2595°C
Conductivity	59.6 S/µm at 20°C
Aluminum	
Atomic number	13
Density	2.7 g/cm^3 at 20°C
Melting point	660.4°C
Boiling point	2467°C
Conductivity	35.0 S µm at 20°C



(a) Two circuits; double the capacity of Fig. 4.1



(b) A circuit with common return

FIGURE 4.2 **Circuits for transferring power. a. Two circuits; double the capacity of Figure 4.1. b. A circuit with common return.**

cross-sectional area "A." We say we are using "2A worth" of conductors: one to carry the current from source to load, and a second to carry the return current.

To transfer more power without changing the voltage, we obviously need a larger conductor (i.e., a larger "A"). Is there a limit to how large "A" can be? Yes. In constructing the line, we must transport the conductors to the construction site, which requires that we must roll the conductor onto a spool, which imposes an upper limit on "A." Our only option (at the selected voltage) is to add a circuit, as shown in Figure 4.2a, which requires "4A" conductors.

It occurs to us that we could modify the circuit to have a common return path, using three conductors instead of four, and shown in Figure 4.2b. But further thought reveals that the return conductor would have to be of size 2A since it would carry twice as much current, and the design needs a total of A + 2A + A = 4A conductors, which is no better than the two-circuit design.

But what if the voltages a–n and b–n are 180° out of phase? For equal loads, the currents a–a' and b–b' would now be 180° *out of phase*, and the current n'–n would be zero, requiring a return (or "neutral") conductor of zero area! The design requires only 2A. This is the simplest version of a form which we shall refer to as "N-phase" power.

Theoretically speaking, we need no neutral at all as long as the loads are balanced (i.e., equal)! However, in practice, the loads are never exactly balanced, so a neutral is provided to insure that the proper voltage is applied to each load. The size of the neutral depends on the anticipated degree of unbalance. The worst case would occur if one phase carried maximum current, and the other zero. So a conservative design would size the neutral the same as a phase conductor, or "A." This would reduce the advantage of the N-phase for N = 2 from "2A" to "3A." Still it represents a 25% savings in conductors, and is commonly used. An example seems in order.

Example 4.1

Given the system of Figure 4.2b, suppose the load a–n is 12 kVA @ pf = 0.866 lagging, as is the b–n load.

a. Given the voltage:

$$\overline{V}_{an} = \overline{V}_{bn} = 120\angle 0^\circ$$

Find all currents. We compute the following:

$$\theta = \cos^{-1}(pf) = 30^\circ;$$

$$\overline{S} = 100\angle 30^\circ = 86.6 + j50$$

$$\overline{I}_a = \overline{I}_b = \left(\frac{\overline{S}}{\overline{V}_{an}}\right)^* = 100\angle -30^\circ \quad A$$

$$\overline{I}_n = \overline{I}_a + \overline{I}_b = 200\angle -30^\circ \quad A$$

b. Given the voltage:

$$\overline{V}_{an} = 120\angle 0^\circ \qquad \overline{V}_{bn} = 120\angle 180^\circ$$

Find all currents. We compute the following:

$$\bar{I}_a = \left(\frac{\bar{S}}{\bar{V}_{an}}\right)^* = 100\angle{-30°} \quad A$$

$$\bar{I}_b = \left(\frac{\bar{S}}{\bar{V}_{bn}}\right)^* = 100\angle{150°} \quad A$$

$$\bar{I}_n = \bar{I}_a + \bar{I}_b = 0 \quad A$$

We are now prepared to define a balanced N-phase transmission line for delivering power from source to load. Such a line is composed of N + 1 conductors, the first "N" called "phase conductors" and the last called the "neutral conductor." For the balanced case, the phase-to-neutral voltages and the phase currents are equal in magnitude, 360°/N shifted in phase.[1]

In theory, "N" can be any integer. There are applications where N = 6, 12, and 24 are used. However, in an overwhelming majority of cases, N = 3 is used, and this has become a de facto world standard. We deal with the balanced[2] three-phase case in the next section.

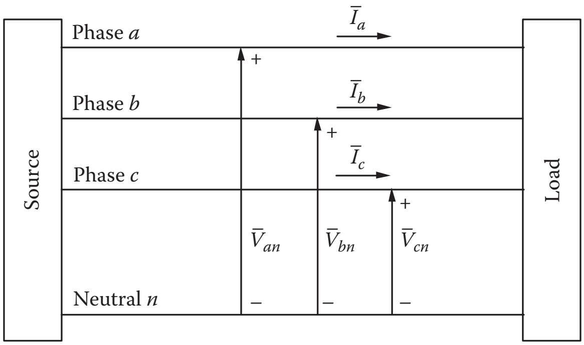
4.2 The Balanced Three-Phase Circuit

Consider the circuit shown in Figure 4.3a, operating in the ac mode. The conductors a, b, and c are referred to as "phase" conductors; conductor "n" is called the "neutral," and is typically grounded.[3]

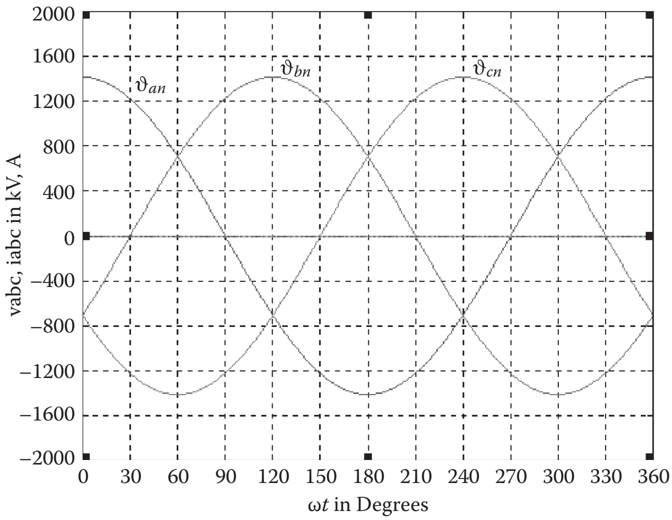
[1] The word "phase" as used in electrical engineering can be confusing. Most commonly, it means "phase angle," as in reference to the phase of a voltage or current. Sometimes it refers to a *winding*, such as a three-phase motor, transformer, or generator. Here, "phase" means "conductor," but as differentiated from a "neutral," which is also a conductor.

[2] In practice, no N-phase system is ever exactly "balanced." However, in normal operation, N-phase systems are *nearly* balanced, and analysis based on this assumption is close enough for most engineering purposes. Analysis of unbalanced operation is justified in some cases, but is considerably more complicated and beyond the scope of our study.

[3] Electrical grounding means connection to the earth, or any large conducting structure that is designated to be "ground." For example, for an automotive system, ground is the chassis of the vehicle; for an aircraft system, the airframe; for a ship, the hull; and so on.



(a) The basic three-phase circuit



(b) Phase Voltage Plots: sequence abc.

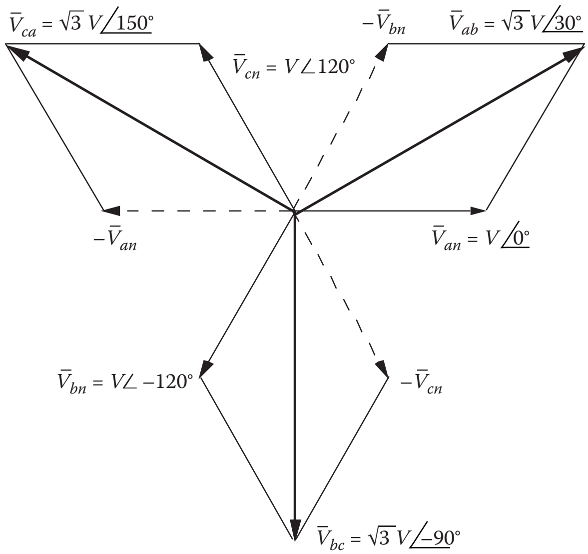
FIGURE 4.3 **The basic three-phase situation.** **(Continued on p. 160).**

The balanced phase voltages are

$$v_{an}(t) = V_{\max}\cos(\omega t) = \sqrt{2}\cdot V\cdot\cos(\omega t)$$

$$v_{bn}(t) = V_{\max}\cos(\omega t - 120°) = \sqrt{2}\cdot V\cdot\cos(\omega t - 120°)$$

$$v_{cn}(t) = V_{\max}\cos(\omega t + 120°) = \sqrt{2}\cdot V\cdot\cos(\omega t + 120°)$$



(c) Phase Diagram: Sequence abc.

FIGURE 4.3 **The basic three-phase situation. (Continued)**

and are plotted in Figure 4.3b. The corresponding phasors are

$$\bar{V}_{an} = V\angle 0°$$

$$\bar{V}_{bn} = V\angle -120°$$

$$\bar{V}_{cn} = V\angle +120°$$

The phasor diagram of Figure 4.3c shows the relationships between phasor voltages in the balanced case.[4] Note that the RMS phase voltage is "V." Recall that in an ac circuit, the phase location of one phasor is arbitrary. We arbitrarily define the phase "a" voltage to be reference (i.e., zero phase).

[4] Again, the adjective "balanced," when applied to three-phase voltages or currents means "equal in magnitude, 120° separated in phase." Balanced means equal when applied to phase impedances or powers.

It is important to determine the phase-to-phase, or so-called "line," voltages. Using KVL, we get

$$\bar{V}_{ab} = \bar{V}_{an} + \bar{V}_{nb} = \bar{V}_{an} - \bar{V}_{bn} = V\sqrt{3}\,\angle 30^\circ$$

$$\bar{V}_{bc} = \bar{V}_{bn} - \bar{V}_{cn} = V\sqrt{3}\,\angle -90^\circ$$

$$\bar{V}_{ca} = \bar{V}_{cn} - \bar{V}_{an} = V\sqrt{3}\,\angle 150^\circ$$

The term "line voltage" (V_L) is commonly used for the magnitude of V_{ab}, V_{bc}, and/or $V_{ca.}$. For the balanced case, observe that

$$V_L = V\sqrt{3} = V_{ab} = V_{bc} = V_{ca}$$

The corresponding phasors are shown in Figure 4.3c. It is common to use line voltage when designating the voltage of a three-phase system. That is, if a single voltage is given in a three-phase situation, it is understood to be the line (phase-to-phase) RMS value, which is the custom in American industry. Again, balanced operation, sequence abc, and V_{an} as a phase reference will be assumed unless specifically stated otherwise.

There is a concept called "phase sequence," that is associated with polyphase systems.[5] For the balanced three-phase system, there are two possibilities: abc and acb. Taking the phase "a" voltage as reference:

Either "b" lags "a" by 120° (and "c" leads "a" by 120°), or
"c" lags "a" by 120° (and "b" leads "a" by 120°).

If "b" lags "a" by 120°, we say the phase sequence is *abc*; if "c" lags "a" by 120°, we say the phase sequence is *acb*. In situations where the phase sequence is unknown, we shall always default to sequence abc. Both sequences are used in industry, and it is important to know the phase sequence in a given application.

[5] Phase sequence is sometimes called "phase order" or "phase rotation."

Example 4.2

Given a 480 V, three-phase system, find all six voltages:[6]

$$V_L = 480\,V; \quad V = \frac{480}{\sqrt{3}} = 277.1\,V$$

$$\bar{V}_{ab} = 480\angle +30°\,V \qquad \bar{V}_{an} = 277.1\angle 0°\,V$$

$$\bar{V}_{bc} = 480\angle -90°\,V \qquad \bar{V}_{bn} = 277.1\angle -120°\,V$$

$$\bar{V}_{ca} = 480\angle 150°\,V \qquad \bar{V}_{cn} = 277.1\angle +120°\,V$$

Balanced three-phase loads are typically modeled with three equal impedances, connected in one of two symmetrical connections: the wye or the delta, which are shown in Figure 4.4.

Observe that the neutral leads to an open circuit in the delta connections. That is, if the source is four wire (abcn), there is no place to connect the neutral to the delta load, forcing the neutral current to always be zero.

Example 4.3

Given a 480 V, three-phase system, find the currents if the following apply:

a. The load is wye connected, with $Z_Y = 8.66 + 5\ \Omega$.

$$\bar{I}_a = \frac{\bar{V}_{an}}{\bar{Z}_Y} = \frac{277.1\angle 0°}{8.66 + j5} = \frac{277.1\angle 0°}{10\angle 30°} = 27.71\angle -30°\ A$$

$$\bar{I}_b = \frac{\bar{V}_{bn}}{\bar{Z}_Y} = \frac{277.1\angle -120°}{10\angle 30°} = 27.71\angle -150°\ A$$

$$\bar{I}_c = \frac{\bar{V}_{cn}}{\bar{Z}_Y} = \frac{277.1\angle 120°}{10\angle 30°} = 27.71\angle 90°\ A$$

$$\bar{I}_n = \bar{I}_a + \bar{I}_b + \bar{I}_c = 0$$

[6] An unfortunate and embarrassing coincidence. The symbol "V" is commonly used both for voltage and simultaneously for the abbreviation for "volt."

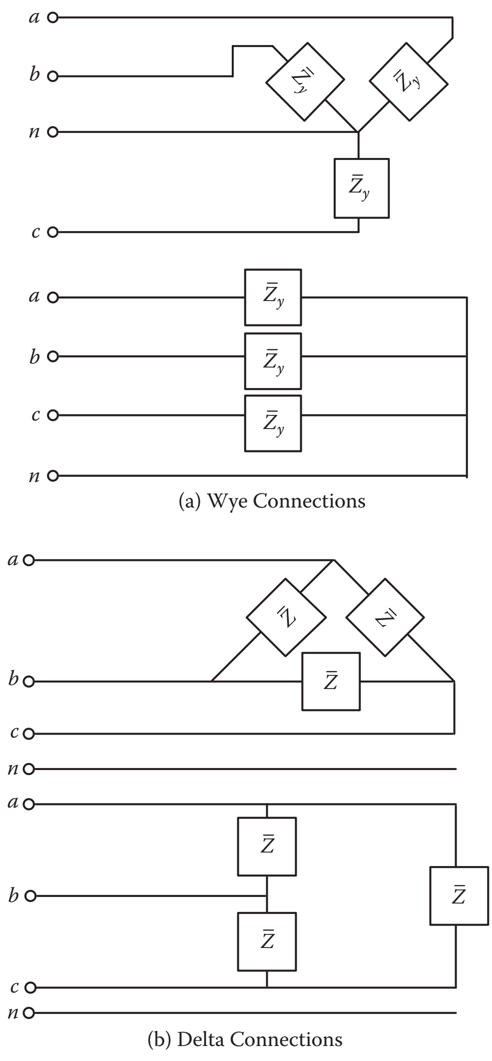


FIGURE 4.4 The two balanced three-phase connections.

Observe that the neutral current is *always* zero for the balanced case! Thus, the neutral conductor is optional; results are the same with or without the neutral conductor.[7]

b. The load is delta connected, with $Z_Y = 25.98 + 15\ \Omega$.

We observe that there are six currents to be found:

$$\bar{I}_{ab} = \frac{\bar{V}_{ab}}{\bar{\bar{Z}}} = \frac{480\angle 30^\circ}{25.98 + j15} = \frac{480\angle 30^\circ}{30\angle 30^\circ} = 16\angle 0^\circ\ A$$

$$\bar{I}_{bc} = \frac{\bar{V}_{bc}}{\bar{\bar{Z}}} = \frac{480\angle -90^\circ}{30\angle 30^\circ} = 16\angle -120^\circ\ A$$

$$\bar{I}_{ca} = \frac{\bar{V}_{ca}}{\bar{\bar{Z}}} = \frac{480\angle 150^\circ}{30\angle 30^\circ} = 16\angle 120^\circ\ A$$

$$\bar{I}_a = \bar{I}_{ab} - \bar{I}_{ca} = 16\angle 0^\circ - 16\angle 120^\circ = 27.71\angle -30^\circ\ A$$

$$\bar{I}_b = \bar{I}_{bc} - \bar{I}_{ab} = 16\angle -120^\circ - 16\angle 0^\circ = 27.71\angle -150^\circ\ A$$

$$\bar{I}_c = \bar{I}_{ca} - \bar{I}_{bc} = 16\angle 120^\circ - 16\angle -120^\circ = 27.71\angle 90^\circ\ A$$

The term “line current” (I_L) is commonly used for the magnitude of I_a, I_b, and/or I_c. These are also called the “phase currents.”[8]

If the line currents are the same (magnitude and phase) for the wye and delta cases, the loads are said to be “equivalent,” which was the case in Example 4.3. That is, in general, the condition for wye–delta equivalence in the balanced case is

$$\bar{Z} \ = 3\bar{Z}_Y$$

If indeed for every wye there is an equivalent delta, and vice versa, we needn’t know the actual connection in a given application. Example 4.4 demonstrates.

7 In practical applications, the presence of the neutral conductor does make a difference. The voltages are never exactly balanced, so a small neutral current will flow in the four-wire case. Also, in the “real world,” the voltages are never perfectly sinusoidal and harmonic currents will flow in the neutral.

8 A key issue in three-phase circuits is to understand the terminology, which can be confusing. We would avoid the adjective “line” if we could. However, its use in industry is so entrenched that this is not realistic.

Example 4.4

In a given 480 V three-phase system, the phase "a" current is measured to be

$$\bar{I}_a = 27.71\angle -30°\ A$$

Given that the load is balanced, find its three-phase impedance.

Solution:
Actually, the problem is incorrectly stated. In a three-phase system, there is no such thing as "three-phase impedance." There is only impedance per phase equivalent wye, or impedance per phase equivalent delta, and each option requires three impedances.

$$\bar{Z}_Y = \frac{\bar{V}_{an}}{\bar{I}_a} = \frac{277.1\angle 0°}{27.71\angle -30°} = 10\angle 30° = 8.66 + j5\ \Omega$$

$$\bar{Z} = 3\bar{Z}_Y = 3(8.66 + j5) = 25.98 + j15\ \Omega$$

A final observation on the consequences of wye–delta equivalence. Any proof based on the assumption of a wye connection is also valid for the delta case, and vice versa. Hence, we needn't check both cases. Our default choice will always be wye because we need only three currents, as opposed to six for the delta case.

Power in balanced three-phase systems is a straightforward extension of the single-phase case. Consider a wye-connected load. We write

$$\bar{S}_a = \bar{V}_{an} \cdot \bar{I}_a{}^* = (V\angle 0°) \cdot (I\angle -\theta)^* = V \cdot I\angle \theta$$

$$\bar{S}_b = \bar{V}_{bn} \cdot \bar{I}_b{}^* = (V\angle -120°) \cdot (I\angle -120° - \theta)^* = V \cdot I\angle \theta$$

$$\bar{S}_c = \bar{V}_{cn} \cdot \bar{I}_c{}^* = (V\angle +120°) \cdot (I\angle +120° - \theta)^* = V \cdot I\angle \theta$$

$$\bar{S}_{1\phi} = \bar{S}_a = \bar{S}_b = \bar{S}_c = V \cdot I\angle \theta$$

$$\bar{S}_{1\phi} = P_{1\phi} + jQ_{1\phi} = V \cdot I \cdot \cos(\theta) + jV \cdot I \cdot \sin(\theta)$$

$$P_{1\phi} = V \cdot I \cdot \cos(\theta)$$

$$Q_{1\phi} = V \cdot I \cdot \sin(\theta)$$

By Tellegen's theorem, the total load complex power is

$$\bar{S}_{3\phi} = \bar{S}_a + \bar{S}_b + \bar{S}_c = 3\bar{S}_{1\phi} = 3P_{1\phi} + j3Q_{1\phi} = P_{3\phi} + jQ_{3\phi}$$

$$P_{3\phi} = 3V \cdot I \cdot \cos(\theta)$$

$$Q_{3\phi} = 3V \cdot I \cdot \sin(\theta)$$

Recalling that

$$V = \frac{V_L}{\sqrt{3}} \qquad I = I_L$$

Then:

$$S_{3\phi} = \sqrt{3} \cdot V_L \cdot I_L$$

$$P_{3\phi} = \sqrt{3} \cdot V_L \cdot I_L \cdot \cos(\theta)$$

$$Q_{3\phi} = \sqrt{3} \cdot V_L \cdot I_L \cdot \sin(\theta)$$

Although a wye-connected load was assumed, the results are valid for the delta-connected load as well.

Example 4.5

Using the data from Examples 4.2, 4.3, and 4.4, determine the complex load power.

Solution:

Assuming a wye connection, in phase a:

$$\bar{S}_a = \bar{V}_{an} \cdot \bar{I}_a{}^* = (277.1\angle 0°) \cdot (27.71\angle -30°)^*$$

$$= 7.68\angle 30° \text{ kVA} = 6.651 \text{ kW} + j3.84 \text{ kvar}$$

The result is also valid for a delta connection.

Continuing Example 4.5:

$$\bar{S}_a = \bar{S}_b = \bar{S}_c = \bar{S}_{1\phi} = 6.651 \text{ kW} + j3.84 \text{ kvar}$$

which is true in general for balanced three-phase systems. The total load complex power is then

$$\bar{S}_{3\phi} = 3\bar{S}_{1\phi} = 19.75 \text{ kW} + j11.52 \text{ kvar} = 23.04\angle 30^\circ \text{ kVA}$$

Power in balanced three-phase systems is a straightforward extension of the single-phase case.

The three-phase power triangle is similar to the a, b, and c triangles, except it is three times as large. Power factor and the notion of leading and lagging remain the same. Example 4.6 will summarize.

Example 4.6

A 480 V system serves a 23.04 kVA load @ pf = 0.866 lagging.

a. Find the load powers.
b. Find the line current.

Solution:

a. $S_{3\phi} = 23.04 \text{ kVA}$

$$\theta = \cos^{-1}(0.866) = +30^\circ$$

$$\bar{S}_{3\phi} = 23.04\angle 30^\circ = 19.75 \text{ kW} + j11.52 \text{ kvar}$$

$$P_{3\phi} = 19.75 \text{ kW}$$

$$Q_{3\phi} = 11.52 \text{ kvar}$$

b. $$I_L = \frac{S_{3\phi}}{\sqrt{3}V_L} = \frac{23.04}{\sqrt{3}(0.48)} = 27.71\ A$$

We consider one final power issue. Consider the instantaneous "a" phase power.

$$p_a(t) = v_{an}(t)\cdot i_a(t) = (\sqrt{2}V)\cdot(\sqrt{2}I_L)\cdot\cos(\omega t+\alpha)\cos(\omega t+\beta)$$

$$p_a(t) = 2V\cdot I_L\cdot\left(\frac{\cos(2\omega t+\alpha+\beta)+\cos(\alpha-\beta)}{2}\right)$$

$$p_a(t) = V\cdot I_L\cdot\cos\theta + V\cdot I_L\cdot\cos(2\omega t+\alpha+\beta)$$

Deriving the "b" and "c" phase instantaneous powers:

$$p_b(t) = V\cdot I_L\cdot\cos\theta + V\cdot I_L\cdot\cos(2\omega t+\alpha+\beta+240^\circ\ )$$

$$p_c(t) = V\cdot I_L\cdot\cos\theta + V\cdot I_L\cdot\cos(2\omega t+\alpha+\beta-240^\circ\ )$$

Adding the phase powers:

$$p_{3\phi}(t) = p_a(t)+p_b(t)+p_c(t) = 3\cdot V\cdot I_L\cdot\cos\theta + 0$$

$$p_{3\phi}(t) = \sqrt{3}\cdot V_L\cdot I_L\cdot\cos\theta$$

which is constant in time!

Example 4.7

A 20.78 kV, three-phase system serves a 174.9 kVA load @ pf = 0.866 lagging.

a. Formulate and plot the phase "a" instantaneous power.

$$I_L = \frac{S_{3\phi}}{\sqrt{3}\cdot V_L} = \frac{174.9}{\sqrt{3}\cdot(20.19)} = 5\ A$$

$$V = \frac{V_L}{\sqrt{3}} = \frac{20.78}{\sqrt{3}} = 12\ kV$$

$$p_a(t) = V\cdot I_L\cdot\cos\theta + V\cdot I_L\cdot\cos(2\omega t+\alpha+\beta)$$

$$p_a(t) = 12(5)(0.866) + 12(5)\cos(2\omega t-30^\circ)$$

$$p_a(t) = 51.96 + 60\cos(2\omega t-30^\circ)\ kW$$

The plot is shown in Figure 4.5a.

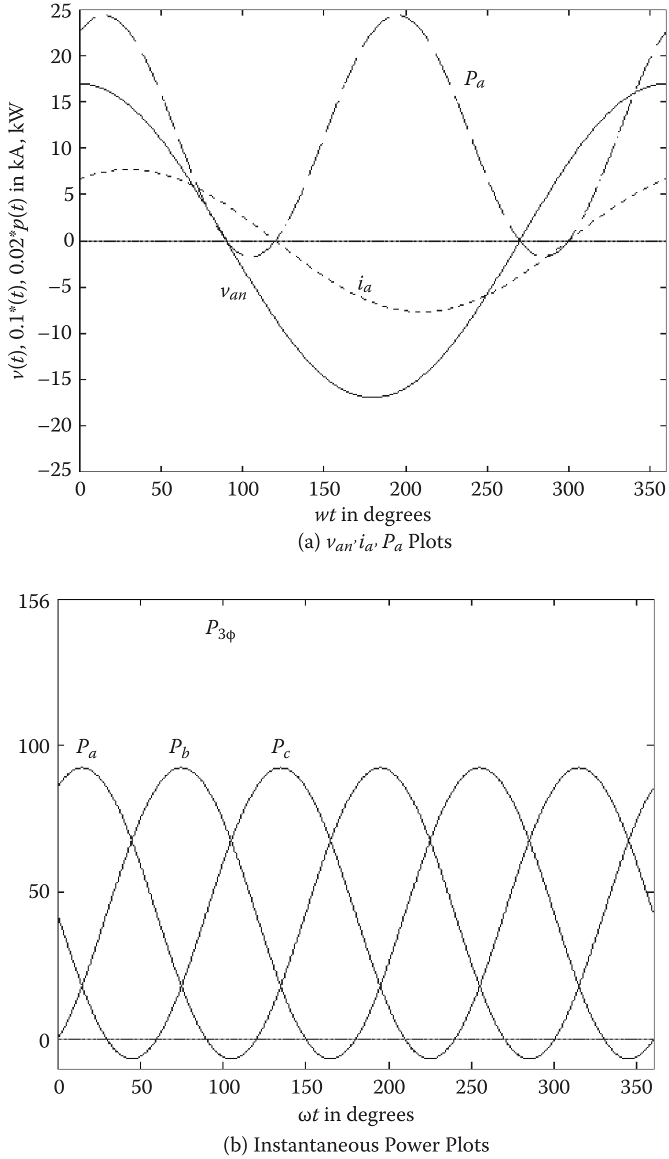


FIGURE 4.5 Plots for Example 4.7.

b. Formulate and plot all three instantaneous-phase, and three-phase, powers

$$p_a(t) = 51.96 + 60\cos(2\omega t - 30°)\ kW$$

$$p_b(t) = 51.96 + 60\cos(2\omega t + 90°)\ kW$$

$$p_c(t) = 51.96 + 60\cos(2\omega t - 150°)\ kW$$

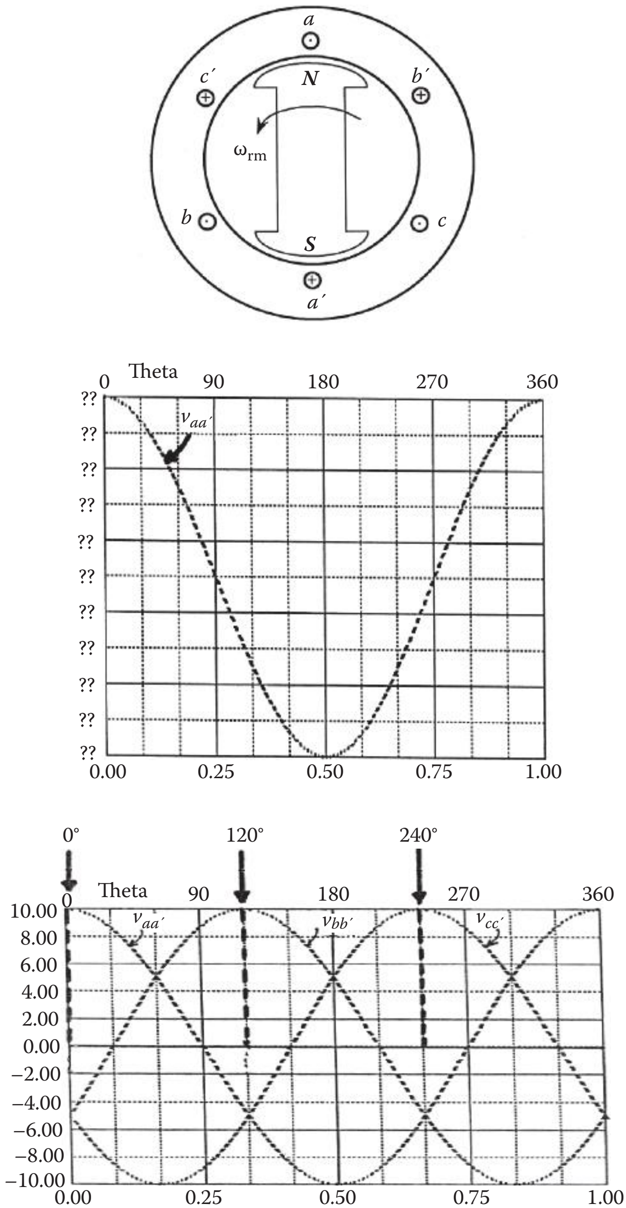
$$p_{3\phi}(t) = p_a(t) + p_b(t) + p_c(t) = 3V \cdot I_L \cdot \cos\theta = 156\ kW$$

The plots are shown in Figure 4.5b.

4.3 The Traditional ac Generator: The Three-Phase Synchronous Machine

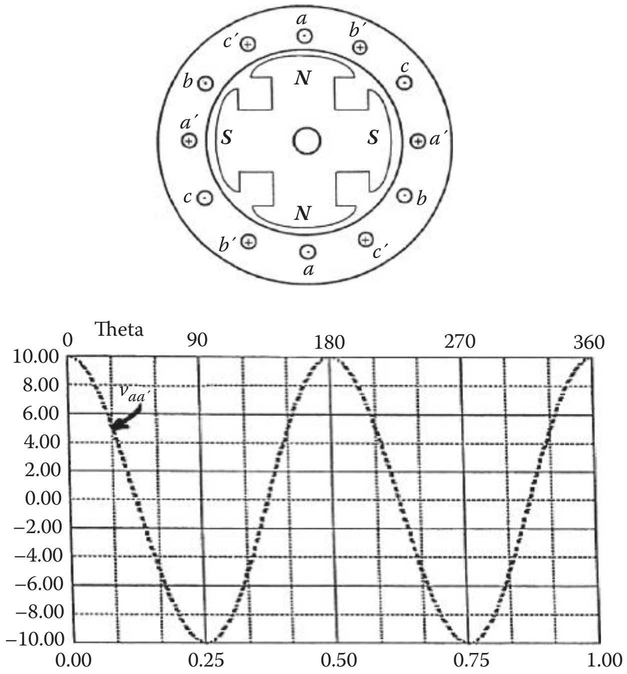
Energy sources for the production of electrical energy were presented in Chapter 3. The most important of these provided mechanical power in the form of a turbine which applied torque to a rotating shaft. We now investigate the device that converts this mechanical power into electrical power, namely, the ac generator. More accurately, the device is called a three-phase synchronous machine. The machine becomes a "generator," if it operates in the generator mode (i.e., it converts power from the mechanical to the electrical form). It is a "motor," if it operates in the motor mode (i.e., it converts power from the electrical to the mechanical form).

In Chapter 3, we learned that if one spins a coil in a magnetic field, an ac voltage is induced in the coil. The reverse is also true; if one spins a magnet in a coil, one induces an ac voltage in the coil, as shown in Figure 4.6a. The magnet is mounted on a rotating structure referred to as the *rotor*, and the coil mounted on a stationary structure referred to as the *stator*. The waveform will be sinusoidal if the change in the coil magnetic flux linkage is sinusoidal, a condition that is achieved by design of the coil and the magnet. Note that the voltage goes through one cycle as the rotor executes one revolution. Hence, if the frequency of the voltage is to be 60 *cycles* per second (or 60 Hz), the rotor must be turning at 60 *revolutions* per second



(a)

FIGURE 4.6 The basic ac generator. a. The basic two-pole ac generator. b. The basic four-pole ac generator.



(b)

FIGURE 4.6 (Continued)

(or 60 × 60 revolutions per minute, or rpm). The SI unit for angular velocity is rad/s. Therefore:

$$\omega_r = (rpm)\frac{2\pi}{60} = 2\pi f$$

ω_r = speed of the rotor in "mechanical" rad/s

f = frequency of the induced stator voltage, Hz

We also note that

$$\omega_e = (rpm)\frac{2\pi}{60} = 2\pi f$$

ω_e = radian frequency of the induced stator voltage, "electrical" rad/s

$\omega_e = \omega_r$

Apparently, there are two kinds of “radians” involved here. The mechanical, or spatial, radian refers to the rotor’s angular position in space. The electrical, or temporal, radian refers to points in time, or phase, in the voltage waveform. There is a one-to-one correlation between the rotor’s spatial position and the phase location in the induced voltage’s waveform.

If we define the rotor angle θ to be measured from the phase “a” coil magnetic axis to the centerline of the rotor field north pole, the induced stator voltage will be at a positive maximum when θ is zero. Since rotation is counterclockwise, positive angle θ is measured CCW as well. If we mount a second coil (b) on the stator, located at 120° CCW from coil “a,” the rotor north pole will align “b” when $\theta = 120°$. But this will happen in 120° in a later phase! That is to say that the voltage induced in “b” will lag that in “a” by 120°! Likewise, in a third coil “c,” located at 240° CCW from coil “a,” the voltage induced in “c” will lag that in “a” by 240°!

Label the coil terminals symmetrically as a, a’, b, b’, c, and c’. Then connect a’, b’, and c’ to each other and label the common terminal “n.” Note that we have formed in essence a balanced three-phase voltage source!

We still have a fundamental problem. For a given frequency, only one speed works. If the frequency is to be 60 Hz:

$$\omega_e = \omega_r = 2\pi f = 377\ rad/s = 3600\ rpm$$

While there are some prime movers (e.g., steam turbines) that can be designed to operate efficiently at this speed, many others require lower operating speeds. We consider redesigning the rotor, as shown in Figure 4.6b. Note that with four poles, the induced voltage executes *two* complete cycles for one rotor revolution. Hence, we have to turn the rotor only half as fast to induce 60 Hz voltages, or 1800 rpm. For a “N_P” pole design:

$$\omega_r = \frac{\omega_e}{N_P/2} = \frac{4\pi f}{N_P}$$

where “N_P” is any even integer.

Example 4.7

A large 500 MVA generator is to be designed for a modern nuclear power plant in the United Kingdom, the steam turbine for which operates most efficiently at a speed of 1492 rpm. How many poles would you recommend for the generator design?

The UK grid frequency is 50 Hz. For two-, four-, and six-pole designs:

$$\omega_r = \frac{4\pi f}{N_P} = \frac{4\pi(50)}{2} = 314.2\ rad/s = 3000\ \text{rpm}$$

$$\omega_r = \frac{4\pi f}{N_P} = \frac{4\pi(50)}{4} = 157.1\ rad/s = 1500\ \text{rpm}$$

$$\omega_r = \frac{4\pi f}{N_P} = \frac{4\pi(50)}{6} = 104.7\ rad/s = 1000\ \text{rpm}$$

The best choice is clearly four poles because it corresponds to a speed closest to 1492 rpm.

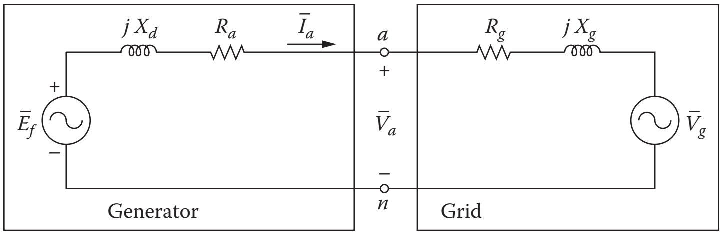
We see how the machine generates voltage, but what about current? We terminate the stator in a balanced three-phase circuit, referred to as "the grid"; hence, the currents will be balanced three-phase. Focus only on phase "a," and rely on three-phase symmetry to deal with phases "b" and "c." The situation is as shown in Figure 4.7a.[9]

Typically, the generator impedance is significantly greater than the external system impedance. Also, $X_d >> R_a$, so the simplified circuit of Figure 4.7b is usually adequate for our purposes, and the one we will use by default. However, don't forget that the circuit of Figure 4.7a is our best model, and the only one that explains some issues.

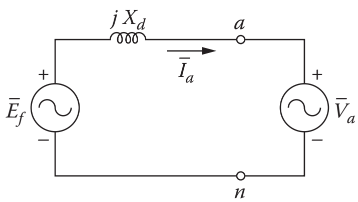
Example 4.8

A 100 MVA, 15 kV, 60 Hz generator operates at rated conditions and a pf = 0.866 lagging. $X_d = 2.5\ \tilde{\Omega}$. Find all values in the circuit of Figure 4.7b.

[9] You might wonder why a source is included and not simply an impedance. If the machine is operating in isolation where it serves a dedicated passive load, the source is not needed. But if we are in a utility environment, where a given generator is connected into the grid, remember that there are thousands of other generators that must also be included in the external model.



(a) Phase *a* equivalent circuit



(b) Simplified circuit

FIGURE 4.7 **Stator equivalent circuits for a generator in a utility environment.**

$$\bar{V}_a = \bar{V}_{an} = \frac{15}{\sqrt{3}}\angle 0^\circ = 8.660\angle 0^\circ \; kV$$

$$I_a = I_L = \frac{S_{3\phi}}{\sqrt{3}V_L} = \frac{100}{\sqrt{3}(15)} = 3.849 \; kA$$

$$\theta = \cos^{-1}(0.866) = 30^\circ \quad \bar{I}_a = 3.849\angle -30^\circ \; kA$$

$$\bar{E}_f = jX_d \cdot \bar{I}_a + \bar{V}_{an} = j2.5(3.849\angle -30^\circ) + 8.660\angle 0^\circ = 15.84\angle 31.7^\circ \; kV$$

Example 4.8 demonstrates several important general concepts:

- One always assumes a *balanced* three-phase situation unless specifically directed otherwise.
- Three-phase equipment is always rated in terms of three-phase power. Thus, "100 MVA" is read as a three-phase rating.
- Three-phase equipment is always rated in terms of *line* voltage (V_{ab}).
- The same pf conventions (e.g., leading and lagging) apply to each phase in the three-phase case.

The stator circuit model of Figure 4.7b captures the relations between machine current and voltage, but we still must deal with power conversion issues. The power transfer from the machine to the system (measured at the machine terminals) is[10]

$$\overline{S}_a = \overline{V}_a \cdot \overline{I}_a{}^* = \overline{V}_a \cdot \overline{E}_f \left(\frac{\overline{E}_f - \overline{V}_a}{jX_d} \right)^*$$

$$\overline{S}_a = \frac{E_f \cdot V_a}{X_d} \sin(\delta) + j\left(\frac{E_f \cdot V_a}{X_d} \cos(\delta) - \frac{V_a^2}{X_d} \right)$$

where $\overline{E}_f = E_f \angle\delta$

$$P_{3\phi} = 3\frac{E_f \cdot V_a}{X_d} \sin(\delta)$$

$$Q_{3\phi} = 3\frac{E_f \cdot V_a}{X_d} \cos(\delta) - 3\frac{V_a^2}{X_d}$$

Example 4.9

Continuing Example 4.8, calculate the complex power transferred from the generator to the grid in two ways.

Using the equivalent circuit:

$$\overline{S}_a = \overline{V}_{an} \cdot \overline{I}_a{}^* = 8.660\angle 0°(3.849\angle{-30°})^*$$

$$\overline{S}_a = 28.87 + j16.67$$

$$P_{3\phi} = 3(28.87) = 86.60\ MW$$

$$Q_{3\phi} = 3(16.67) = 50.00\ \text{Mvar}$$

Using the power transfer equations:

$$P_{3\phi} = 3\frac{E_f \cdot V_a}{X_d} \sin(\delta) = 3\frac{(15.84)\cdot(8.66)}{2.5} \sin(31.7°) = 86.60\ MW$$

$$Q_{3\phi} = 3\frac{E_f \cdot V_a}{X_d} \cos(\delta) - 3\frac{V_a^2}{X_d} = 3\frac{(15.84)\cdot(8.66)}{2.5} \cos(31.7°) - 3\frac{(8.66)^2}{2.5}$$

$$= 50\ \text{Mvar}$$

[10] Technically, the stator leakage flux and the phase resistance must be negligible for the developed power and torque expressions to be rigorously correct.

We now see that the power P flows from the source E_f to the source V_a (the grid). What is still not clear is where the power (or energy) came from. That is, how did the source E_f physically get the power? To address this question, we need to consider the magnetic fields caused by the rotor and stator quantities. Consider the two-pole machine.

The rotor field appears to be simple. If the rotor contains a two-pole permanent magnet, the field leaves the north pole, passes into the stator structure (half going left and half going right), and recombines 180° later, crossing the air gap back into the south pole of the rotor. There are three-phase synchronous machines with permanent magnet fields. However, those designed for utility generators always have wound fields: the field is produced by a dc current I_f flowing in windings around the field poles. Hence, the rotor field strength can be controlled by controlling the field current I_f. Mathematically, it may be expressed as

$$\Im_R(\theta,t) = N_f \cdot I_f \cdot \cos(\omega t - \theta + \gamma)$$

which says that the rotor field is of fixed amplitude, sinusoidally distributed, and rotating in space. $\Im$ represents the magnetic field (i.e., the mmf).

The stator field is not nearly as clear. First, one must understand that each of the three stator coils, or phases, produces a magnetic field due to the current flowing in each. For example, the field due to phase “a” is of the form:

$$\Im_a(\theta,t) = N \cdot i_a \cdot \cos(\theta) = N \cdot I\sqrt{2}\cos(\omega t + \beta) \cdot \cos(\theta)$$

For phases “b” and “c”:

$$\Im_b(\theta,t) = N \cdot I\sqrt{2}\cos(\omega t + \beta - 120°) \cdot \cos(\theta - 120°)$$

$$\Im_c(\theta,t) = N \cdot I\sqrt{2}\cos(\omega t + \beta + 120°) \cdot \cos(\theta + 120°)$$

The total stator field is obtained by adding the abc components:

$$\Im_S(\theta,t) = \Im_a(\theta,t) + \Im_b(\theta,t) + \Im_c(\theta,t) = \frac{3}{2} N \cdot I\sqrt{2}\cos(\omega t - \theta + \beta)$$

which says that the stator field is of fixed amplitude, sinusoidally distributed, and rotating in space. Thus, it is structurally identical to the rotor field!

So here’s the bottom line: there are two rotating magnetic fields in the air gap (i.e., the space between the rotor and stator), each with N_P poles and sinusoidally distributed in space. The speed of rotation of the stator field is called "synchronous speed" and is

$$\omega_s = \frac{4\pi f}{N_P}$$

The speed of the rotor field is the same as the speed of the rotor and is ω_r. Under steady-state conditions, the two fields rotate at the same speed (i.e., are "synchronized," thus the name "synchronous machine"):

$$\omega_r = \omega_s$$

The energy conversion process in the machine occurs through the electromagnetic torque produced by the interaction of the rotor and stator fields!

The rotor field ($\hat{F}$) and stator field ($\hat{S}$) can be represented as spatial vectors and can be added as vectors to produce a resultant field ($\hat{R}$):

$$\hat{R} = \hat{F} + \hat{S}$$

From Figure 4.5b:

$$\overline{V}_a = \overline{E}_f + (-jX_d \cdot \overline{I}_a)$$

It can be shown that

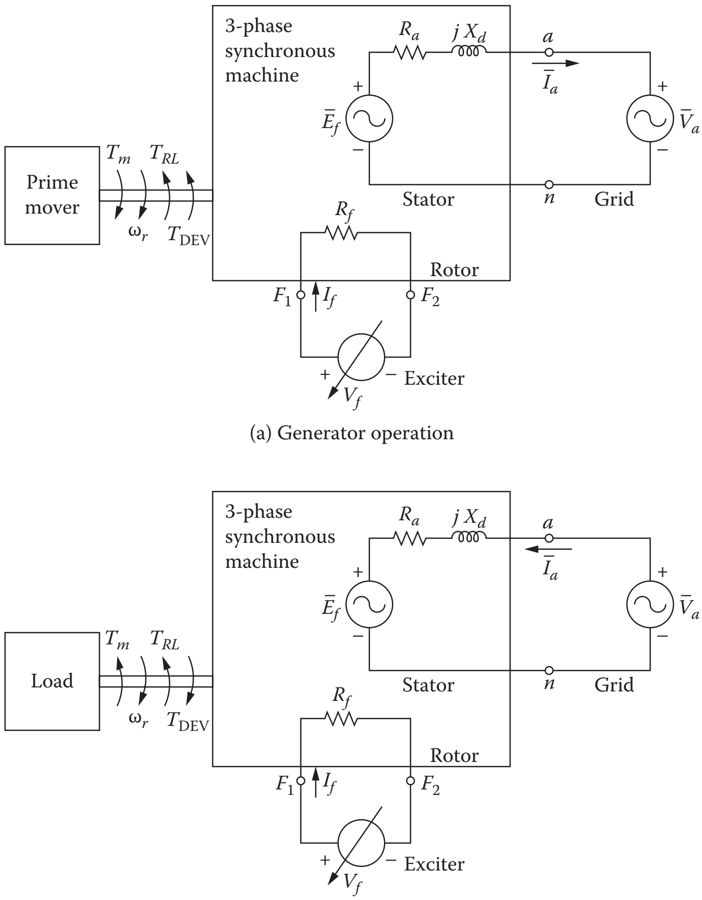
- the voltage E_f is induced by, and is proportional to, the field F.
- the voltage $X_d I_a$ is induced by, and is proportional to, the field S.
- the voltage V_a is induced by, and is proportional to, the field R.
- the *spatial* angle δ between $\hat{F}$ and $\hat{R}$ is equal to the *phase* angle δ between $\overline{E}_f$ and $\overline{V}_a$.

Although the energy conversion process occurs through the interaction of the machine's internal magnetic fields, it may be computed from the stator equivalent circuit of Figure 4.5. Therefore, the developed power and torque are[11]

$$P_{DEV} = P_{3\phi} = 3\frac{E_f \cdot V_a}{X_d}\sin(\delta)$$

$$T_{DEV} = \frac{P_{DEV}}{\omega_r}$$

[11] The power at the terminals is also the internal power since the inductor X_d absorbs zero P.



(a) Generator operation

(b) Motor operation

FIGURE 4.8 **Synchronous machine model.**

The complete model for the three-phase synchronous machine is provided in Figure 4.8. For generator mode operation (Figure 4.8a), we write

$$\bar{E}_f = jX_d \cdot \bar{I}_a + \bar{V}_{an}$$

The rotor field winding:

$V_f = R_f \cdot I_f$

$V_f =$ exciter dc voltage, applied to the rotor field winding, V

$I_f =$ the dc rotor field winding current, A

$R_f =$ the dc rotor field winding resistance, Ω

The dc rotor field current I_f produces the ac stator voltage E_f:

$$E_f = K_f \cdot I_f$$

$K_f =$ the linearized magnetization characteristc coefficient, V/A

Newton's Second Law applied at the shaft:

$$T_m - (T_{DEV} + T_{RL}) = J \cdot \frac{d\omega_r}{dt}$$

$T_m =$ Prime Mover Torque, Nm

$T_{DEV} =$ Electromagnetic developed Torque, Nm

$T_{RL} =$ Rotational Loss Torque, Nm

$J =$ Mass polar moment of inertia of all rotating parts, kg-m^2

$\omega_r =$ shaft speed, rad/s

At synchronous operation:

$$\omega_r = \omega_s = \frac{4\pi f}{N_p}$$

$$T_m = T_{DEV} + T_{RL}$$

For motor mode operation (Figure 4.8b), we reverse the positive sense of stator current and torques.

$$\bar{V}_a = jX_d \cdot \bar{I}_a + \bar{E}_f$$

$$T_{DEV} = T_m + T_{RL}$$

We will perform comprehensive generator and motor analyses in the next section.

4.4 The Pumped Storage Application

Consider the nature of an electric utility. The utility converts energy in non-electrical form (such as thermal, hydro, wind, etc.) into a processed electrical form (balanced, three-phase ac constant voltage). Its customers expect to buy as much of, or as little of, the product in real time as they choose, and expect it to be available 24 hours per day, 365 days per year.

We can interpret this customer expectation as a sort of mission statement for an electric utility, adding that safety is also a prime consideration, as is economical operation. Since at any given instant, energy may be thought of as being represented by power, we are led to the operational consideration:

How much generation (i.e., generated power) should be available at a given instant in time?

The power delivered to the customers is described as "load," and it is clear that we must generate at least this much power, at a minimum. But power requires current, which must flow through the various conducting paths to the loads. Resistance is encountered, producing RI^2 losses, which must also be supplied by the generators.

Furthermore, utilities are electrically interconnected with their neighbors. It may be that utility A needs a certain amount of additional power P_X at a particular time T_X for a duration D_X, and that its neighbor (utility B) happens to have surplus power at that time. A and B enter into a contract wherein B agrees to supply A with P_X at time T_X, at cost C_X, and for duration D_X. This is a win–win situation: A gets power at a price lower than it would cost to generate it internally, and B can sell its excess generation and avoid taking it offline.

Finally, utilities must maintain a certain safety margin ("spinning reserve") to maintain service in the event of unexpected loss of generation ("forced outage") or inadvertent damage to the system (e.g., lightning). This is another strong argument for interconnected systems (forming "the grid"), making the grid much more reliable than any single utility.

So as to the question of how much generation is needed, the answer is as follows:

Enough to meet the load, the loss, the intertie commitments, and the spinning reserve requirements.

If there ever was a perfect, beautiful, and elegant alignment of the theoretical and "real" worlds, this is it! We see immediately that the outputs of all of the generators must match the P, Q requirements of the total load, plus losses, before we draw a single circuit diagram, much less write any circuit equations. Why?

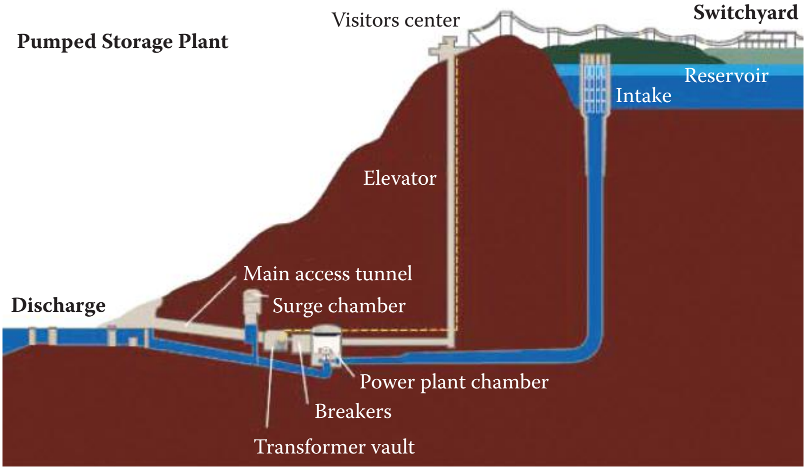


FIGURE 4.9 Typical pumped storage arrangement.

Tellegen's theorem!

Utilities buy and sell power to each other constantly, hour by hour, over the course of the day and are prepared to spend millions of dollars based on this insight!

Therefore, it is necessary to accurately estimate the load to be served at any given instant in time. It is the nature of electrical load that

- it varies in time over the course of a day.
- it varies seasonally. In the south, the annual peak is typically in the summer, due to air conditioning; in the north the peak may be in the winter, due to heating.
- it also varies over the week, dropping considerably on the weekends.
- at any given instant, load is correlated to local weather conditions.

It takes some time (e.g., from tens of minutes to hours) to bring thermal power plants on- and offline, whereas hydro plants can be brought on- and offline much faster (in seconds to minutes). Therefore, hydro units are very desirable in adjusting for hourly and weekly load fluctuations. Unfortunately, many systems have few hydro resources. Such systems find that they must have the bulk of their generating capacity in thermal form, and must keep it online even when it is not economical to do so. For these systems, a concept called "pumped storage" may be practical.

What is required for pumped storage is a reservoir at high elevation (e.g., a lake or reservoir on top of a mountain) and a water supply (e.g., a river) at a suitable location in the system, as shown in Figure 4.9. The idea is to use the stored energy in the upper reservoir to generate during the peak load period and refill the reservoir by pumping during an off-peak period. This would reduce the amount of thermal generation required to be online, resulting in substantial operating savings.

Refer to Figure 4.10, which is the weekday load cycle for a fictitious System X, which has only thermal units (i.e., no hydro resources are available). Note that the peak load requirement occurs at about 5:00 pm (17:00) and is 8 GW. If losses, spinning reserve, and intertie commitment requirements force us to maintain a 12.5% margin, we need 9 GW of generation online, some of which may be required to run all day due to thermal considerations. Note that for most of the day we have far too much generation online, which is wasteful. If we could add temporary hydro generation to get us over the hump from 16:00 to 20:00, the base load generation requirements would drop from 9 to 8 GW, with substantial operating savings. As luck would have it, we have an appropriate site (a mountain, the top of which is available for the construction of a reservoir, and a river near its foot).

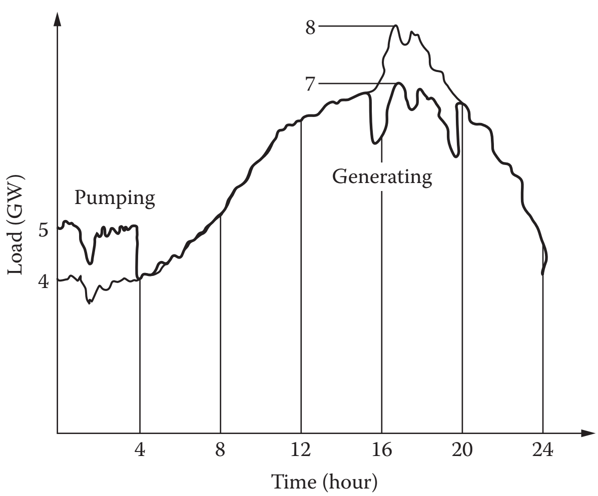


FIGURE 4.10 **Daily weekday load cycle for System X.**

The pumped storage facility is to operate as follows:

- At minimum load conditions (weekdays 24:00–04:00, and weekends), the facility operates in the pumping mode, moving water from the river to the upper reservoir. The power sources are the online thermal units.
- At maximum load conditions (weekdays 16:00–20:00), the facility operates in the generating mode, using the stored energy in the upper reservoir to supply electric power to the grid.

We will use this pumped storage application to serve as an incentive for synchronous machine analysis. To flesh out our study with numbers, assume System X has a pumped storage facility (PSF) designed to supply 1000 MW for 4 hours per day for 5 days (Monday through Friday) weekly to the grid. The location is 90 miles north of a major metropolitan area. An average head of 300 m is available. "Head" is the vertical distance from the lower to the upper reservoir, which has an average depth of 10 m. There are four identical pumping and generating units, each with a synchronous machine with the following data:

Stator ratings: 17 kV 60 Hz 24-pole 300 MVA, Xd = 1 Ω

Rotor field data: Rf = 0.5 Ω, Kf = 20 V/A, 24 poles

$$\omega_s = \frac{4\pi f}{N_P} = \frac{377}{12} = 31.42\ rad/s = 300\ rpm$$

We choose to operate the units at rated stator conditions and the system has an efficiency of 85% when either pumping or generating.

Example 4.10

For system PSF operating in the generating mode, injecting 1000 MW into the grid, find the pf, the stator current, and the field current for one unit.

If the unit operates at rated output, $S_{3\phi} = 300$ MVA. Also $P_{3\phi} = 250$ MW. Assuming that the system is at near peak load, we would want the unit to inject $Q_{3\phi}$ *into* the grid.[12]

[12] When the grid is heavily loaded, its voltage tends to sag. It can be shown that insertion of Q into the grid tends to raise the voltage. Likewise, at light load conditions, the grid voltage tends to rise, and removal of Q tends to lower the voltage.

$$\bar{S}_{3\phi} = 250 + jQ_{3\phi} = 300\angle\theta$$

$$Q_{3\phi} = \sqrt{S_{3\phi}^2 - P_{3\phi}^2} = \sqrt{(300)^2 - (250)^2} = 165.8\ M\,\text{var}$$

$$\bar{S}_{3\phi} = 250 + j165.8 = 300\angle 33.6^\circ; \quad pf = \cos(33.6^\circ) = 0.8333\ \text{lagging}$$

The leading, lagging terminology is slightly different for generators. When we say a generator is operating at lagging pf, we mean that its load appears inductive: the generator delivers Q to the load, which in this case is the grid.

$$I_a = \frac{S_{3\phi}}{\sqrt{3}\cdot V_L} = \frac{300}{\sqrt{3}\cdot(17)} = 10.19\ kA; \quad \bar{I}_a = 10.19\angle -33.6^\circ\ kA$$

$$\bar{E}_f = jX_d\cdot\bar{I}_a + \bar{V}_a = j1\cdot(10.19\angle -33.6^\circ) + \frac{17}{\sqrt{3}}\angle 0^\circ = 17.63\angle 28.8^\circ\ kV$$

Continuing with our analysis:

$$I_f = \frac{E_f}{K_f} = \frac{17.63}{20} = 0.8815\ kA = 881.5\ A$$

$$V_f = R_f I_f = (0.5)(881.5) = 440.7\ V$$

Example 4.10 raises some interesting questions. How do we control the unit so that it injects 250 MW into the grid? The generator shaft is driven by a hydraulic turbine. Water flows from the upper reservoir down a tunnel called a "penstock" through adjustable gates called "wicket gates," then through the turbine, and finally exiting into the lower reservoir. Imagine the unit initially at standstill, and the generator disconnected from the grid. We open the wicket gates a little, so the water begins to accelerate the turbine and generator rotor. Since the stator current is zero, the electromagnetic developed torque is zero, and the turbine torque is used for pure acceleration. When we reach synchronous speed (300 rpm), we adjust the generator field until the generator voltage matches the grid voltage. At this point we close the generator circuit breaker, connecting the generator to the grid, a process called "synchronizing the generator."

After synchronizing, we continue to open the wicket gates, which applies more torque to the turbine blades, and hence the generator rotor, which accelerates slightly above 300 rpm. Recall that the electromagnetic (developed) torque is

$$T_{DEV} = \frac{P_{DEV}}{\omega_r} = 3\frac{E_f \cdot V_a}{\omega_r X_d}\sin(\delta)$$

As the rotor moves ahead of the revolving stator field, this increases the angle δ, increasing the developed torque, which "pushes back" until it balances the turbine torque, and the rotor once more stabilizes at a speed of 300 rpm.

It is of interest to note why the stator field may be assumed to rotate at precisely 300 rpm independent of what the turbine generator is doing. The speed of the rotating stator field is determined by the grid frequency, which is determined by the combination of all of the synchronous machines in the external system. Hence, the terminal grid behaves like an "infinite" second synchronous machine.[13]

Therefore, the generator rotor will stabilize at a speed of 300 rpm. Using values from Example 4.10, suppose the rotor field voltage has been adjusted to 440.7 V.

$$I_f = \frac{V_f}{R_f} = \frac{440.7}{0.5} = 881.5\ A$$

$$E_f = K_f I_f = 20(881.5\ A) = 17.63\ kV$$

Using these values:

$$T_{DEV} = 3\frac{E_f \cdot V_a}{\omega_r X_d}\sin(\delta) = 3\frac{(17.63)\cdot(9.815)}{(31.42)(1)}\sin(\delta) = 16.52\sin(\delta)\ MNm$$

$$P_{DEV} = T_{DEV}\cdot\omega_r = 16.52(31.42)\sin(\delta) = 519.06\sin(\delta)\ MW$$

Solving for δ:

$$P_{DEV} = 519.06\sin(\delta) = 250\ MW$$

$$\delta = 28.8°$$

[13] In other words, a machine with infinite inertia. Actually the inertia is very large, and is essentially the sum of the inertias of all of the generators connected to the grid.

To summarize:

Wicket gates are set to 83.33% open, causing a real power (P) flow of 250 MW. (*Assumes that fully open corresponds to 300 MW and power is proportional to WG setting.*)

Exciter is set to $V_f = 440.7$ V, causing a reactive power (Q) flow of 165.8 *Mvar*. (*Combined with a WG setting of 83.33%.*)

Note that the unit is capable of operating at much higher power levels; however, the excessive current would cause unacceptable stator overheating. The point is:

Real power output (P) is controlled by position of the wicket gates!
Reactive power output (Q) is controlled by the exciter.

To follow up on the second point, suppose we decreased the exciter voltage to 236.5 V. How would the unit react? Well, the field current would change to the following:

$$I_f = \frac{V_f}{R_f} = \frac{236.5}{0.5} = 473\ A = 0.473\ kA$$

This means that

$$E_f = K_f I_f = 20(0.473) = 9.459\ kV$$

Since the WG setting was not changed, P_{DEV} remains at 250 MW. Therefore:

$$P_{DEV} = 3\frac{E_f \cdot V_a}{\omega_r X_d}\sin(\delta) = 3\frac{(9.459)\cdot(9.815)}{(1)}\sin(\delta) = 250\ MW$$

$$\delta = 63.8°$$

The current is

$$\overline{I}_a = \frac{\overline{E}_f - \overline{V}_a}{jX_d} = \frac{9.459\angle 63.8° - 9.815\angle 0°}{j1} = 10.19\angle + 33.6°\ kA$$

So look what's happened! Adjusting the field voltage has changed the phase of the stator current![14] In fact, it changed from a lagging to a leading position!

[14] In general, the current magnitude will change as well. It didn't in this case because of the particular field voltage value we selected.

The output complex power is

$$V_f = 440.7\ V \quad \overline{S}_{3\phi} = 250 + j165.8 = 300\angle 33.6°; \quad pf = 0.8333 \text{ lagging}$$

$$V_f = 236.5\ V \quad \overline{S}_{3\phi} = 250 - j165.8 = 300\angle -33.6°; \quad pf = 0.8333 \text{ leading}$$

The reactive power flow has changed from an *output* of 165.8 *Mvar* to an *input* of 165.8 *Mvar*. In general:

Reactive power output (Q) is controlled mainly by the exciter setting!
To increase Q, increase the excitation; to decrease Q, decrease the excitation!

We continue our study via Example 4.11.

Example 4.11

For system PSF operating in the pumping mode, absorbing 1000 MW from the grid, find the pf, the stator current, and the field current for one unit.

If the unit operates at rated output, $S_{3\phi} = 300$ MVA. Also $P_{3\phi} = 250$ MW. Assuming that the system is at near minimum load, we might want the unit to absorb $Q_{3\phi}$ from the grid:

$$\overline{S}_{3\phi} = 250 + jQ_{3\phi} = 300\angle\theta$$

$$Q_{3\phi} = \sqrt{S_{3\phi}^2 - P_{3\phi}^2} = \sqrt{(300)^2 - (250)^2} = 165.8\ M\text{var}$$

$$\overline{S}_{3\phi} = 250 + j165.8 = 300\angle 33.6°; \quad pf = \cos(33.6°) = 0.8333 \text{ lagging}$$

The leading, lagging terminology for motors is the same as for passive loads. When we say a motor is operating at lagging pf, we mean that it appears inductive to the system: the system delivers Q to the motor.

$$I_a = \frac{S_{3\phi}}{\sqrt{3}\cdot V_L} = \frac{300}{\sqrt{3}\cdot(17)} = 10.19\ kA; \quad \overline{I}_a = 10.19\underline{/-33.6°}\ kA$$

$$\overline{E}_f = \overline{V}_a - jX_d\cdot\overline{I}_a = \frac{17}{\sqrt{3}}\underline{/0°} - j1\cdot(10.19\underline{/-33.6°}) = 9.459\underline{/-63.8°}\ kV$$

$$I_f = \frac{E_f}{K_f} = \frac{9.459}{20} = 0.473\ kA = 473\ A$$

$$V_f = R_f \cdot I_f = 0.5(473) = 236.5\ V$$

Example 4.11 raises some interesting issues. How do we control the unit so that it pumps and absorbs 250 MW + j165.8 Mvar from the grid? If we pump, the water flow in the penstock must reverse, which means that the turbine must reverse.[15] Hence the synchronous machine rotor must reverse its direction of rotation, which means we (1) must disconnect it from the grid, (2) stop it, (3) start it in the opposite direction, and (4) synchronize it to the grid.

If we close the gates and open the stator breaker, the turbine generator will stop, due to the machine and turbine losses and the residual water in the turbine chamber. We now must start the machine as a motor, and in the reverse direction. Reversing the synchronous machine is surprisingly simple. The direction of rotation of the rotating stator field is determined by the phase sequence. This is easily reversed by interchanging any two phases (e.g., exchange phase "b" with "c," and "c" with "b"). As a motor, the rotor will follow the stator field (i.e., turn in the same direction). We now must start the machine as a motor. There are several methods. For example:

- With the stator breaker open, use an auxiliary motor (e.g., a "pony" motor) to accelerate the machine at no load up to speed. Since the load is basically inertial, the pony motor need be only a small fraction of the size of the synchronous machine.
- With the stator breaker open, use the dc exciter as a dc motor to accelerate the machine at no load up to speed. This works only if the exciter is a dc machine.
- The synchronous machine damper windings are specially designed to allow the machine to start as an induction motor; the stator breaker is closed and the unit starts.

The machine is synchronized to the grid and inherently runs as a motor at no load. As the gates are open, the rotor attempts to slow down. As the rotor

[15] It is at least conceptually possible to design a nonreversing turbine and reverse the flow with a complicated system of gates. However, the simplest and cheapest solution is to reverse the turbine.

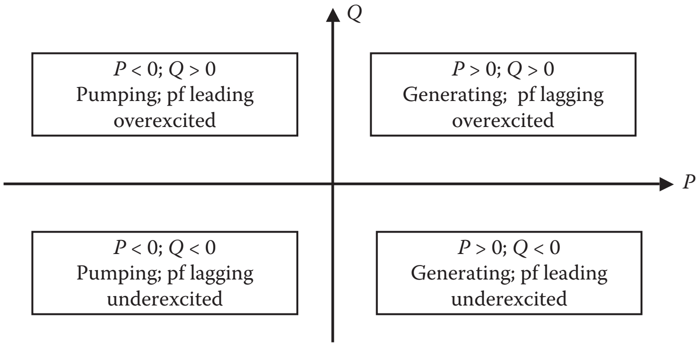


FIGURE 4.11 **Synchronous machine operating modes in a pumped storage facility.**

field lags the stator field (i.e., again the angle δ opens, and the subsequent developed torque pulls the rotor forward). The more the gates are opened, the greater the angle δ and the developed torque. The volumetric water flow is controlled by position of the wicket gates! The greater the flow, the more "P" the motor must absorb from the grid.

As for the generator case, reactive power input (Q) is controlled by the exciter voltage!

Zero reactive power flow is the unity pf case, which corresponds to an exciter voltage of V_{f0}. We say that the machine is "overexcited" if $V_f > V_{f0}$, and "underexcited" if $V_f < V_{f0}$. Figure 4.11 summarizes these ideas.

Example 4.12

System PSFX generates for $5 \times 4 = 20$ hours per week supplying 1000 MW to the grid. How long must we pump to replace the water in the reservoir?

$$W_{IN} = 100(20) = 2000\ MWhr$$

$$W_{STORED} = \frac{2000}{0.85} = 2353\ MWhr$$

Hence the energy extracted from the grid when pumping:

$$W_{OUT} = \frac{2353}{0.85} = 2768\ MWhr$$

Pumping with an input of 1000 MW:

$$T_{PUMP} = \frac{2768}{1000} = 27.68\ hr$$

Recall that we pump from midnight until 4:00 a.m. on weekdays (5 × 4 = 20 hr), leaving 7.68 hr (round up to 8 hr) for pumping on weekends.

4.5 Some Basic Magnetics

Generators are electromagnetic-mechanical devices, and to fully understand them, we would need to understand their magnetic properties as well as their electrical and mechanical characteristics. Still, it was possible to understand them on an operational basis without going into magnetics in any significant depth. Such is not the case for our next major power system component, the transformer. To understand the basics of transformers, we need to understand some basic magnetic concepts.

Consider the simple *electrical* circuit of Figure 4.12a, which shows a 1.5 V AA battery terminated in a high-resistance wire of cross-sectional area A in m^2, length l in m, and conductivity Ω in siemen/m. The wire resistance in Ω is

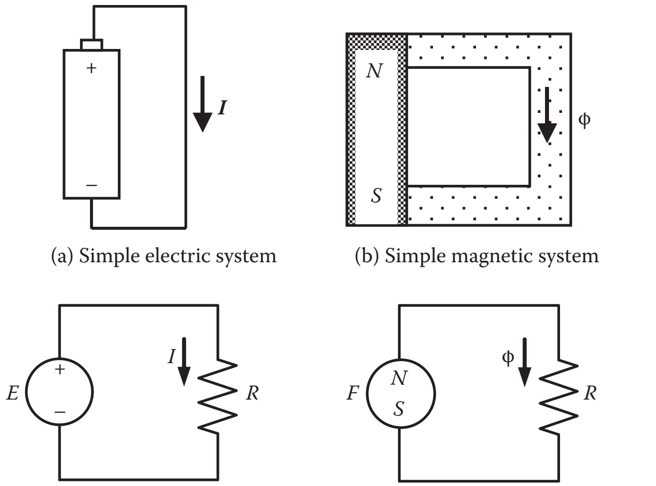
$$R = \frac{\ell}{\sigma A}$$

The battery has a property called "electromotive force" (emf), symbolized by E in volts.[16]

To solve for the current in amperes, we use Ohm's law:

$$I = \frac{E}{R}$$

[16] There was a time in EE when all source voltages were designated as "E's," representing "electromotive force" or "emf," whereas opposing passive voltages were "V's." This practice has long since been deprecated, because of its awkwardness. Still, in terms of the fundamentals of electromagnetic physics, there is a distinction between "E" and "V," and it suits our purposes to resurrect the concept of emf.



(a) Simple electric system (b) Simple magnetic system

(c) Electric circuit diagram (d) Magnetic electric circuit diagram

FIGURE 4.12 **Analogous electric and magnetic systems.**

In particular, note that the positive direction of current flow is out of the source positive terminal, around the external circuit, and into the source negative terminal.

Now consider the simple *magnetic* circuit of Figure 4.12b, which shows a permanent magnet terminated in a ferromagnetic core of cross-sectional area A in m^2, length l in m, and permeability μ in henry/m. The core reluctance in Ω is

$$\Re = \frac{\ell}{\mu A}$$

The magnet has a property called "magnetomotive force" (mmf), symbolized by F in ampere-turns. To solve for the magnetic flux in webers, we use "magnetic" Ohm's law:

$$\phi = \frac{\Im}{\Re}$$

In particular, note that the positive direction of flux flow is out of the magnetic north pole, around the external core, and into the magnetic south pole. (Magnetic) flux is a mathematical construct which represents the structure of the magnetic field, usually depicted pictorially as a pattern of lines. In fact, "lines of flux" is a common expression, and an early unit of flux was the

Table 4.3 The Electric-Magnetic Circuit Analogy

Electrics	Magnetics
Emf (E), volt	Mmf (F), ampere-turns
Current (I), ampere	Flux (ϕ), weber
Current density (J), A/m^2	Flux density (B), tesla
Resistance (R), Ω	Reluctance (R), ampere per Wb
Conductivity (σ), Σ/m	Permeability (μ), H/m
+,−	N, S
At any node ΣI's = 0 (KCL)	At any node Σf's = 0
Around any loop ΣV's = 0 (KVL)	Around any loop SF's = 0

"line." The SI unit of flux is the weber.[17] Another magnetic concept is flux density (B). Assuming uniform flux density:

$$B = \frac{\phi}{A} \quad Wb/m^2 \quad \text{or tesla, T}$$

It is clear that there is a direct analogy between dc electric circuits and dc magnetic circuits. The essentials are summarized in Table 4.3.

Analogies are convenient as a mnemonic to help us remember relationships between the two domains. However, they are no substitute for the underlying science. The relation between mmf, flux, and reluctance is not true *because of* the analogy to the dc electric circuit. Rather, the analogy helps us remember the proper relations.

We have noted before that there are interactions between the electrical and magnetic domains: electrical "stuff" can produce magnetic "stuff," and vice versa. In particular, we don't need a permanent magnet to produce a magnetic field: a current-carrying coil will also produce mmf, as shown in Figure 4.13b.

In this case, the mmf produced by the current is

$$\Im = N \cdot I \qquad \text{ampere-turns}$$

Another important magnetic concept is *flux linkage* (λ):

$$\lambda = N \cdot \phi \qquad \text{weber-turns}$$

[17] 10^8 lines = 1 weber (Wb).

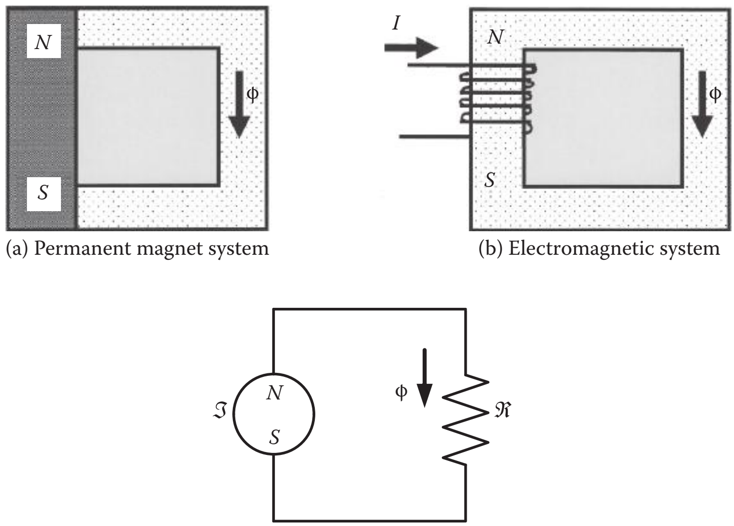


FIGURE 4.13 Two equivalent magnetic systems.

which assumes that all of the core flux links each turn (i.e., no flux leaks out of the core). Now if the coil current is time-varying, so is the mmf, and hence the flux and flux linkage. The coil represents a part of an electrical circuit which is linked by a time-varying magnetic field. Faraday's Law predicts an induced voltage in the coil, according to

$$v = \frac{d\lambda}{dt} \quad \text{volt}$$

The coil inductance is defined as

$$L \triangleq \frac{\lambda}{i} \quad \text{henry, H}$$

The following example will be particularly important.

Example 4.13

The coil in the magnetic circuit of Figure 4.13b has zero resistance and is supplied with a current of

$$i = I\sqrt{2}\sin(\omega t)$$

Determine the voltage, convert v and I to phasors, and compute the coil inductance and complex impedance.

$$\Im = N \cdot i = N \cdot I\sqrt{2}\,\sin(\omega t)$$

$$\phi = \frac{\Im}{\Re} = \frac{N \cdot I\sqrt{2}}{\Re} \cdot \sin(\omega t)$$

$$\lambda = N\phi = \frac{N^2 \cdot I\sqrt{2}}{\Re} \cdot \sin(\omega t)$$

$$v = \frac{d\lambda}{dt} = \frac{N^2 \cdot I\sqrt{2} \cdot \omega}{\Re} \cdot \cos(\omega t) = L \cdot I\sqrt{2} \cdot \omega \cdot \cos(\omega t)$$

where $$L = \frac{\lambda}{i} = \frac{N^2}{\Re}$$

Converting v and I to phasors:

$$\bar{I} = I\,\underline{/-90^\circ}$$

$$\bar{V} = \omega L \cdot I\,\underline{/0^\circ}$$

Computing the impedance:

$$\bar{Z} = \frac{\bar{V}}{\bar{I}} = \omega L\,\underline{/+90^\circ} = j\omega L$$

We realize that the system of Example 4.13 is basically an inductor (i.e., a passive component whose major electrical property is its inductance). Of course, in the real world, the resistance of the coil is not zero, but it is small. We now appreciate why the symbol chosen for inductance is a coil. Coils inherently have large inductance. Also, the results of Example 4.13 will prove to be particularly important in the next section as we consider transformers.

Pay close attention to the direction in which the current flows around the core in Figure 4.13b. The current flows CCW as viewed from the top: across the front left to right, then into the page, then across the back right to left, and then out of the page. Encircle the core with your right hand with your fingers pointing in the direction in which the current is flowing. Your thumb will point up, which reveals the direction in which the magnetic field (ϕ) is forced

FIGURE 4.14 **IEEE insignia.**

to flow (CW in the plane of the paper). This procedure is what we shall call the “right-hand rule.”[18]

The IEEE insignia (see Figure 4.14) presents the right-hand rule: if the circular CCW arrow is interpreted as current, the vertical “up” arrow represents the corresponding magnetic field direction. Another reason to join the IEEE: your IEEE pin can be used as a “cheat sheet” for the right-hand rule!

For the situation illustrated, this means flux (ϕ) is forced out of the top of the coil which by definition becomes a north (N) pole. Flux (ϕ) flows CW around the core and enters the bottom of the coil which by definition becomes a south (S) pole.

Now in practical situations in electrical circuits, it is important to keep track of which is the positive terminal for all sources. For example, AA 1.5 V batteries have the positive/negative terminals marked as +/–. 12 V car battery terminals are marked +/–, and the relevance is obvious to anyone who has tried to jump-start a car with a dead battery. This raises the question:

Is magnetic N,S “polarity” important for magnetic sources, and if so, what system is used to indicate polarity?

The answer is “Yes (usually),” and several notations are used. For permanent magnets, the poles are marked “N,S.” If the N,S markings are missing, one can determine the magnetic polarity experimentally by moving the north pole of a second magnet of known polarity near pole “X” of the specimen

[18] There are several versions of the right-hand rule in magnetics.

and observing the interacting force: if the force is attractive pole X is S; if repulsive, X is N.

When there are two coils positioned on a core such that a flux path exists that passes through both, magnetic polarity is of particular importance. This is recorded by way of the so-called dot convention. Refer to Figure 4.15, which shows two coils on a common core. The dots are determined by the following procedure:

1. Define either coil to be #1. We pick the top coil.
2. Place a dot at either terminal of Coil #1. We dot terminal 1.
3. Define a current I_1 into the dot.
4. Determine the direction of the corresponding $\Im_1$ by the right-hand rule. We get "up."
5. Define a current I_2 into either terminal of Coil #2. We pick terminal 3.
6. Determine the direction of the corresponding $\Im_2$ by the right-hand rule. We get "up."
7. Compare the results of Steps 4 and 6. The mmf's will either be in the same or opposite directions.
   a. If the same, terminal 3 will be dotted.
   b. If opposite, terminal 4 will be dotted.

For the situation of Figure 4.15, the mmfs are in the same direction; hence the second dot is at terminal 3. So there are four possibilities (two

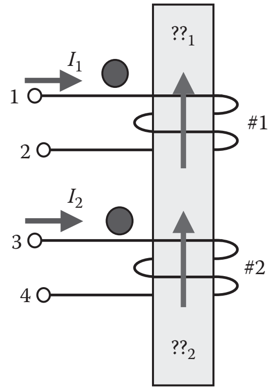


FIGURE 4.15 The dot convention.

right and two wrong):

Correct dot locations: 1 and 3, 2 and 4
Incorrect dot locations: 1 and 4, 2 and 3

The significance of the dots will be made clear in the next section.

4.6 Power Transformers

It happens that it is most economical to design large generators to operate at medium-level voltages (e.g. 23 kV), whereas we need much higher voltages (e.g. 500 kV) to move bulk power over long distances. Also, most loads operate at much lower voltages (e.g. 480 V). Thus, we need a simple device that can convert voltages from one level to another at high efficiency. Such a device is called a "transformer." Transformers come in all sizes from a few watts (such as those used to power your laptop or scientific calculator) to hundreds of megawatts (such as those used to interface large utility generators to the grid). The latter types are called "power" transformers. A large power transformer is shown in Figure 4.16.

Figure 4.17a shows a basic three-winding transformer in cross-section. We abbreviate "three-winding" as 3W. The core is a closed ferromagnetic structure of high permeability (μ)[19]. The windings are electrically insulated conductors of high conductivity (σ)[20]. An *ideal* transformer is one for which the core has infinite m and the windings have infinite σ.

$$Lim\{R\}_{\sigma\to\infty} = Lim\left\{\frac{\ell}{\sigma A}\right\}_{\sigma\to\infty} = 0$$

$$Lim\{\Re\}_{\sigma\to\infty} = Lim\left\{\frac{\ell}{\mu A}\right\}_{\mu\to\infty} = 0$$

This means the core has zero reluctance and all windings have zero resistance. Note that all dots are defined as positive and all currents defined into the dots. Refer to this convention as the canonical assignment of winding voltages and currents.

[19] Permeability is a measure of how receptive a material is to the passage of a magnetic field through it.

[20] Conductivity is a measure of how receptive a material is to the passage of an electric current through it.

FIGURE 4.16 A large power transformer. Source: http://images.google.com/images?q=photos+of+power+transformers

For the ideal 3W transformer of Figure 4.17b operating in the ac mode:

$$\Im_1 + \Im_2 + \Im_3 = 0$$

$$N_1 \cdot i_1 + N_2 \cdot i_2 + N_3 \cdot i_3 = 0$$

Transforming to phasors:

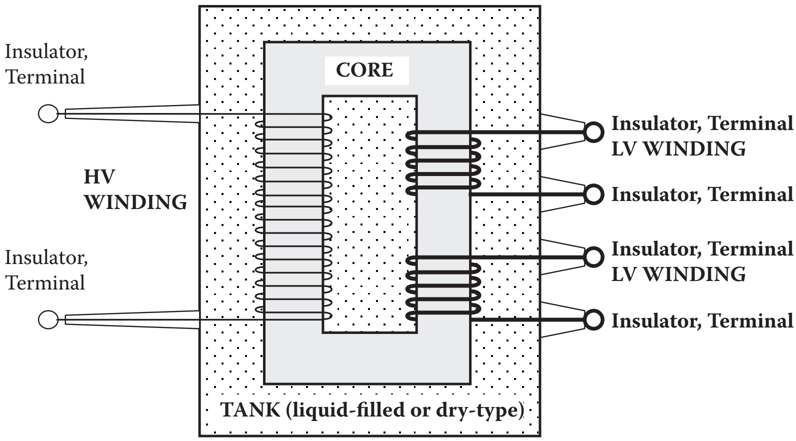
$$N_1 \cdot \overline{I}_1 + N_2 \cdot \overline{I}_2 + N_3 \cdot \overline{I}_3 = 0$$

Now consider the following:

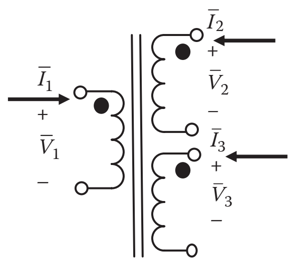
$$v_1 = \frac{d\lambda_1}{dt} = \frac{d}{dt}(N_1 \cdot \phi) = N_1 \cdot \frac{d\phi}{dt}$$

$$v_2 = N_2 \cdot \frac{d\phi}{dt}$$

$$v_3 = N_3 \cdot \frac{d\phi}{dt}$$



(a) Cross-sectional view



(b) Symbol. Cannonical V, I assignments

FIGURE 4.17 The basic three-winding transformer.

Combining:

$$\frac{v_1}{N_1} = \frac{v_2}{N_2} = \frac{v_3}{N_3}$$

Transforming to phasors:

$$\frac{\bar{V}_1}{N_1} = \frac{\bar{V}_2}{N_2} = \frac{\bar{V}_3}{N_3}$$

For the foregoing equations to be valid, canonical assignments for voltage polarity, current directions must be used: dots positive, currents defined positive into the dots. Solving for the primary current:

$$\bar{I}_1 = -\frac{N_2 \cdot \bar{I}_2}{N_1} - \frac{N_3 \cdot \bar{I}_3}{N_1}$$

Conjugating:

$$\bar{I}_1^* = -\frac{N_2 \cdot \bar{I}_2^*}{N_1} - \frac{N_3 \cdot \bar{I}_3^*}{N_1}$$

Multiplying by the primary voltage:

$$\bar{V}_1 \cdot \bar{I}_1^* = -\frac{N_2 \cdot \bar{V}_1 \cdot \bar{I}_2^*}{N_1} - \frac{N_3 \cdot \bar{V}_1 \cdot \bar{I}_3^*}{N_1}$$

But,

$$\bar{V}_2 = \frac{N_2 \cdot \bar{V}_1}{N_1} \quad \text{and} \quad \bar{V}_3 = \frac{N_3 \cdot \bar{V}_1}{N_1}$$

Therefore:

$$\bar{V}_1 \cdot \bar{I}_1^* = -\bar{V}_2 \cdot \bar{I}_2^* - \bar{V}_3 \cdot \bar{I}_3^*$$

It follows that

$$\bar{S}_1 + \bar{S}_2 + \bar{S}_3 = 0$$

To summarize, the basic equations of the 3W ideal transformers are

$$N_1 \cdot \bar{I}_1 + N_2 \cdot \bar{I}_2 + N_3 \cdot \bar{I}_3 = 0$$

$$\frac{\bar{V}_1}{N_1} = \frac{\bar{V}_2}{N_2} = \frac{\bar{V}_3}{N_3}$$

$$\bar{S}_1 + \bar{S}_2 + \bar{S}_3 = 0$$

where

$$\bar{S}_1 = \bar{V}_1 \cdot \bar{I}_1{}^* = P_1 + jQ_1$$

$$\bar{S}_2 = \bar{V}_2 \cdot \bar{I}_2{}^* = P_2 + jQ_2$$

$$\bar{S}_3 = \bar{V}_3 \cdot \bar{I}_3{}^* = P_3 + jQ_3$$

The extension to "*n*" windings should be clear, as should the contraction to two windings. This says that the total complex power into an n-winding ideal transformer is zero. Stated another way, the complex power flowing in equals the complex power flowing out, or the device has zero P and/or Q losses.

Observe that since

$$\frac{V_i}{V_j} = \frac{N_i}{N_j} \quad i, j = 1, 2, \text{ or } 3$$

supplying the voltage ratings is in effect supplying the turns ratios, which is the normal commercial practice.

An example would be useful.

Example 4.14

Given a 3W transformer rated at 7.2 kV/120 V/120 V and 50 kVA/25 kVA/25 kVA, number the windings 1, 2, and 3, respectively.

a. Find the turns ratios between all windings.

$$\frac{N_1}{N_2} = \frac{V_{1Rated}}{V_{2Rated}} = \frac{7200}{120} = \frac{60}{1} \qquad \frac{N_1}{N_3} = \frac{V_{1Rated}}{V_{3Rated}} = \frac{7200}{120} = \frac{60}{1}$$

$$\frac{N_2}{N_3} = \frac{V_{2Rated}}{V_{3Rated}} = \frac{120}{120} = \frac{1}{1}$$

b. Find the rated current for all windings.

$$I_{1Rated} = \frac{S_{1Rated}}{V_{1Rated}} = \frac{50}{7.2} = 6.944\ A$$

$$I_{2Rated} = \frac{S_{2Rated}}{V_{2Rated}} = \frac{25}{0.12} = 208.3\ A$$

$$I_{3Rated} = \frac{S_{3Rated}}{V_{3Rated}} = \frac{25}{0.12} = 208.3\ A$$

It is necessary that the canonical assignments of the winding voltages and currents, as shown in Figure 4.17b for the foregoing equations, be valid. The consequences of the canonical assignments include the following:

- All winding voltages are in phase.
- The sum of the winding mmfs is zero (which is sort of like Kirchhoff's current law [KCL], except that the concept of mmf replaces that of current).
- The sum of the input complex powers is zero.

A practical transformer has high, but not infinite, permeability, which gives rise to leakage flux and magnetizing current, as well as core losses. Likewise, the windings have high, but not infinite, conductivity, which means that the windings have nonzero resistance. These effects are included in more sophisticated circuit models of power transformers. However, the *ideal* transformer will be sufficiently accurate for our purposes. That said, we must never forget that for some studies winding impedance and losses cannot be ignored, and better models must be used. Since we will approximate all transformers as ideal, we will interpret the term "transformer" as "ideal transformer."

For the 2W transformer, it is common to refer to the winding to which a source is connected as the primary (winding); the load side is called the secondary (winding). If the source side is the high-voltage side, the application is said to be "step down"; if the source is on the low-voltage side, the application is said to be "step up." For the 3W case, the third winding is called the "tertiary," which typically supplies a load.

We intend to use the transformer of Example 4.14 to serve the residential application of Section 2.9. The source is the local electric utility distribution system, which in the United States is typically three-phase 12.47 kV. Note that V_{an} = 7.2 kV is the voltage applied to our transformer, which makes this application step down.

The two 120 V secondary windings are interconnected as shown in Figure 2.21. It is not uncommon for two or more residences to be served from one transformer. When this is the case, the multiple loads are in parallel.

Doing a rough estimate of the load, we get 5 + 4 + (5 + 4 + 4) = 22 kVA. Hence, one 50 kVA transformer can easily serve two homes. From Section 2.9, the loads for one residence are available. For two homes, we will double these values:

$$\bar{S}_{AN} = 2 \times 5\angle 36.9^\circ = 8 + j6$$

$$\bar{S}_{NB} = 2 \times 4\angle 0^\circ = 8 + j0$$

$$\bar{S}_{AB} = 2 \times 12.53\angle 22.2^\circ = 23.2 + j9.488$$

Recomputing the load currents:

$$\bar{I}_{AB} = \left(\frac{\bar{S}_{AB}}{\bar{V}_{AB}}\right)^* = 104.42\angle -22.2^\circ \ A$$

$$\bar{I}_{AN} = \left(\frac{\bar{S}_{AN}}{\bar{V}_{AN}}\right)^* = 83.33\angle -36.9^\circ \ A$$

$$\bar{I}_{NB} = \left(\frac{\bar{S}_{NB}}{\bar{V}_{NB}}\right)^* = 66.67\angle 0^\circ \ A$$

Recomputing the input currents for two panels:

$$\bar{I}_A = \bar{I}_{AB} + \bar{I}_{AN} = 186.2\angle -28.7^\circ \ A$$

$$\bar{I}_B = -(\bar{I}_{BA} + \bar{I}_{NB}) = 168.04\angle 166.4^\circ \ A$$

$$\bar{I}_N = \bar{I}_{AN} - \bar{I}_{NB} = 50.04\angle -90^\circ \ A$$

Finding all currents in all transformer windings:

$$\bar{I}_2 = -\bar{I}_A = -186.2\angle -28.7^\circ \ A$$

$$\bar{I}_3 = \bar{I}_B = 168.04\angle 166.4^\circ \ A$$

$$N_1 \cdot \bar{I}_1 + N_2 \cdot \bar{I}_2 + N_3 \cdot \bar{I}_3 = 0$$

$$\bar{I}_1 = -\left(\frac{N_2 \cdot \bar{I}_2}{N_1} + \frac{N_3 \cdot \bar{I}_3}{N_1}\right) = \frac{-\bar{I}_2 - \bar{I}_3}{60}$$

$$\bar{I}_1 = \frac{186.2\angle -28.7^\circ - 168.04\angle 166.4^\circ}{60}$$

$$\bar{I}_1 = 5.853\angle -21.5^\circ \ A$$

We observe that all transformer currents are well within ratings. As a check, let us compute complex powers. First, the primary input power:

$$\begin{aligned}\bar{S}_1 &= \bar{V}_1 \cdot \bar{I}_1^* \\ &= 7.2(5.853\angle{-21.5^\circ})^* = 42.14\angle{+21.5^\circ}\ kVA \\ &= 39.21\ kW + j15.44\ kvar\end{aligned}$$

Now the secondary output power:

$$-\bar{S}_2 = \bar{V}_2 \cdot (-\bar{I}_2)^* = 0.12(186.2\angle{-28.7^\circ})^* = 19.6 + j10.73$$

$$-\bar{S}_3 = \bar{V}_3 \cdot (-\bar{I}_3)^* = 0.12(-168.04\angle{166.4^\circ})^* = 19.6 + j4.74$$

$$\begin{aligned}-\bar{S}_2 - \bar{S}_3 &= 19.6 + j10.73 + 19.6 + j4.74 \\ &= 39.21\ kW + j15.47\ kvar\end{aligned}$$

Observe that

$$\bar{S}_1 = -\bar{S}_2 - \bar{S}_3$$

Our analysis is extremely conservative for several reasons:

- It is highly unlikely that all connected loads in a home will be on at the same time.[21]
- It is unlikely that all homes served from one transformer will have their loads on at the same time.
- Loads that cycle on and off automatically (e.g., water heaters, heating systems, refrigerators, etc.) are unlikely to all be on at the same time.

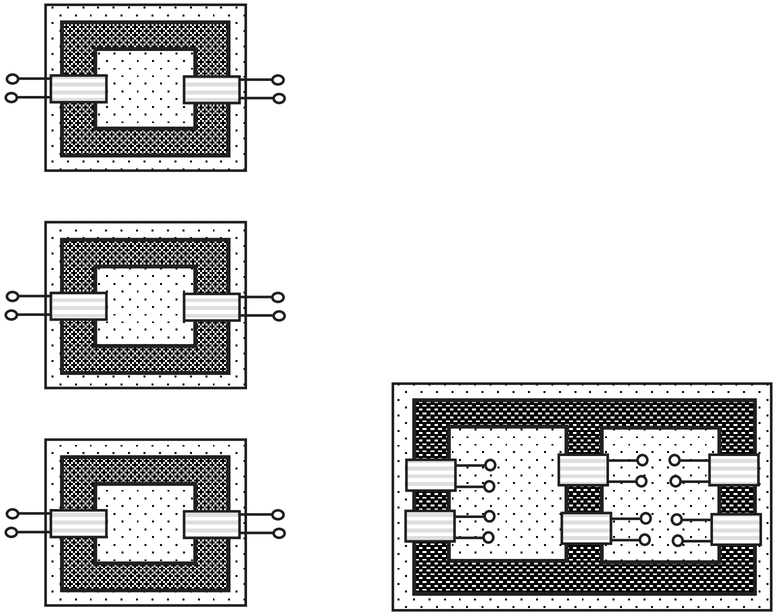
Utility engineers have studied the situation in some detail and account for it in their design. Still, we are beginning to appreciate how ac circuit theory can be used to assess some very practical problems.

As we move to higher power levels, again we realize that power systems are three-phase. So how do we transform three-phase power?

There are two possibilities:

- We use three single-phase transformers (3–1 f Xfmrs).
- We use one three-phase transformer (1–3 f Xfmr).

[21] Unless one has one or more teenagers living at home.



(a) three 1-phase 2W Transformers (b) one 3-phase 2W Transformer

FIGURE 4.18 Transforming 3-phase power.

See Figure 4.18. The primary is now a set of three windings, which are either wye or delta connected, as is the case with the secondary, tertiary, and so on. We will confine our study to the simpler of the two, namely, the use of a single three-phase transformer, and ignore wye–delta connection details.

Recall the pumped storage situation of Section 4.3. There were four identical pumping and generating units, each with a synchronous machine with the following data:

Stator ratings: 17 kV 60 Hz 24-pole 300 MVA, $X_d = 1\ \Omega$
Rotor field data: $R_f = 0.5\ \Omega$, $K_f = 20$ V/A, 24 poles

Suppose the facility is served by several 230 kV power transmission lines. Each unit is to be connected into the transmission system with a three-phase transformer. Each transformer should be rated at 17 kV: 230 kV 300 MVA, where it is understood that the voltage ratings are *line* values and the power ratings are *three-phase*.

Actually, every power plant requires something called "station service," so the unit transformers may be 3W[22]. The tertiary serves the local plant needs, which may include lights and HVAC, and some heavier loads such as pumps and cranes. A 3W transformer will require *three* power ratings, one for each set of windings.

4.7 Power Transmission Lines

We now have some sense of how electric power is produced, and how the voltage can be transformed to lower or higher levels. Usually sites suitable for power production are remote from where the power is most needed. We next must consider how we move the power from generating to load centers.

There are two possibilities:

1. *Radiation*: Electromagnetic energy can be radiated from point A to point B through space.
2. *Conduction*: Electrical energy can be channeled from point A to point B by providing a conducting path, called a "power transmission line."

The radiation option is appealing because it requires no structures between points A and B, which after all can be a considerable distance. However, there are problems:

The output of the generators is inherently conductive. Conversion to radiation form would require additional hardware of a cost comparable to that of the generator.

At the receiving end, conversion from radiation to conduction form again would require very expensive hardware.

The distance would be limited to line of sight, after which repeater stations would have to be built.

The beam must be focused, with an energy density in the MW/m^2 range. There would have to be some safety systems in place to interrupt the beam if any object were to enter the beam corridor.

[22] Since the transformers are three-phase devices, they would actually have nine windings (three per phase). Still, in transformer jargon, these are called "three-winding transformers," once more reminding us of the importance of "knowing the lingo."

If the beam was suddenly interrupted, this would constitute a severe disturbance on the grid. The transmitted power would be forced to seek out alternate flow paths.

Given the state of today's technology, the power transmission line option is clearly the most practical. There are four basic ways to build a line.

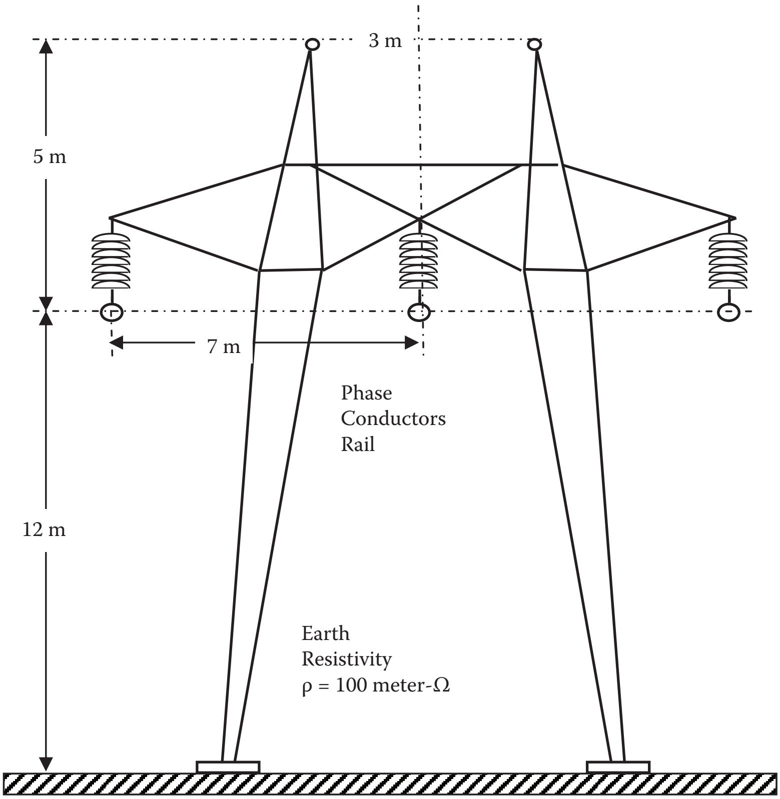
- Overground, which requires dedicated land use and creates crossover problems with highways, rivers, and railroads.
- Undersea, which may be practical in certain locations (e.g., for service to islands).
- Underground, which is practical in certain high-population and high–load density locations, such as the New York City area.
- Overhead, which has an economic advantage and permits multiple line uses.

All four schemes are used, with the overhead option being by far the most prevalent.

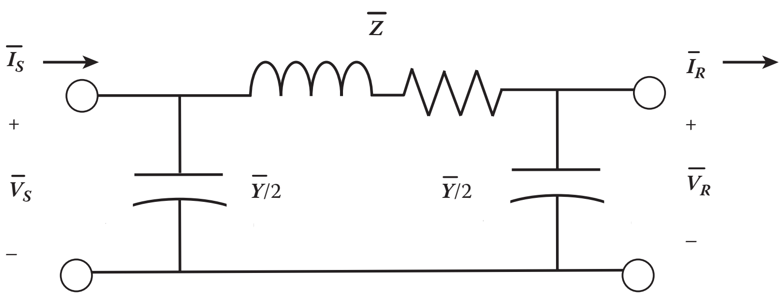
Figure 4.19a shows a typical overhead three-phase power transmission line in cross-section at a tower. The phase conductors (a, b, and c) are typically bare-stranded aluminum and hang from suspension insulators, which insulate them from the tower structure.

The overhead neutrals (n) are typically bare-stranded aluminum also and serve two purposes:

- The neutrals provide a path for current flow if the load is unbalanced. They are grounded (i.e., are connected to the tower structure and to copper-clad rods driven into the earth at each tower.) Sometimes one or more bare copper conductor(s) (called "counterpoises") are buried along the line's corridor and become part of the neutral system.
- The neutrals provide overhead shielding for lightning protection. Since the line is outdoors, is tall, and has a huge footprint, it is virtually certain to be struck by lightning. If the strike impinges on the neutral, it typically bleeds off to ground with no significant disturbance to the system or damage to the line.



(a) Cross-sectional view at a tower



(b) Per-phase equivalent circuit

FIGURE 4.19 Three-phase power transmission line.

The design of the towers and distance between towers (span) is a structures problem, and is typically solved by civil engineers.

Figure 4.18b shows a per-phase equivalent circuit that accounts for the most important electrical properties of the line. Lines in normal operation are very close to balanced. For our work, we shall assume that they are always exactly balanced (i.e., phase voltages and currents are equal in magnitude, 120° separated in phase, with a phase sequence abc). Hence we need focus only on phase "a," and can use symmetry if we're interested in phases b and c. Some key issues:

- "S,R" stand for "sending, receiving end," respectively.
- Voltages are phase to neutral values, and referenced to phase "a."
- Currents are phase (line) values, and referenced to phase "a."
- The powers are one-third of the total transmission powers.
- The series impedance $Z = R + jX$: R accounts for the phase conductor resistance, and the line's real power losses; and X accounts for the magnetic field surrounding the phase conductors, and corresponding voltage loss.
- The shunt admittance $Y = jB$: B accounts for the electric field between the phase conductors, the neutrals, and ground, and the corresponding current loss.

Z and Y values can be computed from conductor data and the line geometry (conductor spacing, and length).

Like the power transformer and the generator, the power transmission line has a thermal rating, which is

$$S_{3\phi rated} = \sqrt{3} \cdot V_{Lrated} \cdot I_{Lrated}$$

Data for a 230 kV power transmission line (TLX) is given in Table 4.4.

We will use TLX and the circuit of Figure 4.18b to investigate several issues. For example, the thermal rating of TLX is

$$S_{3\phi rated} = \sqrt{3} \cdot V_{Lrated} \cdot I_{Lrated} = \sqrt{3}(230)(1) = 398.4\ MVA$$

Table 4.4 Data for Power Transmission Line TLX

Ratings: VLine: 230 kV; ILine:1000 A				
Earth Resistivity: 100.0 m-ohm; Line Length: 160.9 km = 100 mi				
Conductor Radius Gmr Resistance kcmil Ampac.				
==				
Phase RAIL 1.480 cm 1.1735 cm 0.07332 ohm/km 954.0 1000				
Neut. AWELD 0.526 cm 3.5966 mm 0.59155 ohm/km				
Phase Conductors/bundle = 1				
Bases: 3ph Power = 100.0 MVA; Line Voltage = 230.0 kV				
Conductor Location	X position	Y position		
---------------------(m)------(ft)------------(m)------(ft)--				
Phase a	-7.00	-22.97	12.00	39.37
Phase b	0.00	0.00	12.00	39.37
Phase c	7.00	22.97	12.00	39.37
Neutral #1	-3.00	-9.84	17.00	55.77
Neutral #2	3.00	9.84	17.00	55.77
Series Z = 11.8 + j80.3 ohms				
Shunt Y/2 = j0.2637 mS				

The line serves rated load at rated voltage, pf = 0.866 lagging. Determine the receiving end voltage, current, and power.

$$\bar{V}_R = \frac{230}{\sqrt{3}} = 132.8\angle 0°\ kV \quad \theta = \cos^{-1}(0.866) = 30° \quad \bar{I}_R = 1\angle{-30°}\ kA$$

$$\bar{S}_R = \bar{V}_R \cdot \bar{I}_R{}^* = 115\ MW + j66.4\ Mvar$$

$$\bar{S}_{R(3\phi)} = 3 \cdot \bar{S}_R = 345\ MW + j199.2\ Mvar$$

Determine the sending end voltage and current from the circuit of Figure 4.18b.

$$\bar{I}_Z = \bar{I}_R + \frac{\bar{Y}}{2} \cdot \bar{V}_R$$

$$= 1\angle{-30^\circ} + j(0.2637)(0.1328\angle{0^\circ}) = 0.9830\angle{-28.23^\circ}\ kA$$

$$\bar{V}_S = \bar{V}_R + \bar{Z} \cdot \bar{I}_Z$$

$$= 132.8\angle{0^\circ} + (11.8 + j80.3)(0.9830\angle{-28.23^\circ}) = 191.4\angle{19.55^\circ}\ kV$$

$$\bar{I}_S = \bar{I}_Z + \frac{\bar{Y}}{2} \cdot \bar{V}_S$$

$$= 0.9830\angle{-28.23^\circ} + j(0.2637)(0.1914\angle{19.55^\circ}) = 0.9462\angle{-26.18^\circ}\ kA$$

Determine the line losses and the efficiency.

$$P_{LOSS} = 3 \cdot I_Z^2 \cdot R = 3(0.983)^2(11.8) = 34.2 \text{ MW}$$

$$\eta = \frac{P_{OUT}}{P_{IN}} = \frac{P_{OUT}}{P_{OUT} + P_{LOSS}} = \frac{345}{345 + 34.2} = 91\%$$

Determine the sending end complex power.

$$\bar{S}_S = \bar{V}_S \cdot \bar{I}_S *$$

$$= (191.4\angle{19.55^\circ})(0.9462\angle{-26.18^\circ})* = 126.4\ MW + j129.7\ Mvar$$

$$\bar{S}_{S(3\phi)} = 3 \cdot \bar{S}_S = 379.2\ MW + j389.1\ Mvar$$

$$S_{S(3\phi)} = 543.3 \text{ MVA}$$

Observe that

$$P_{S(3\phi)} - P_{R(3\phi)} = 379.2 - 345 = 34.4\ MW$$

which checks with P_{LOSS}!

As we reflect on our results, we note some surprises.

- The sending end current (946 A) is actually less than the receiving end current (1000 A), whereas the apparent power is greater (543.3 versus 398.4 MVA), which is counterintuitive! This happens because part of the load reactive current was supplied by the line capacitors, and hence need not flow from the sending end.

- The voltage along the line varies much more than we might have assumed. Starting from 230 kV at the load end, it was 331 kV at the source end, some 47% higher 100 miles away! This suggests that we may have a voltage control problem on the grid, which in fact is the case. For a unity pf load, the sending end voltage would have been a significantly lower value of 283 kV, which suggests that much of the problem was caused by Q flow on the line.
- The line efficiency was only 91%, which suggests that we may not want to operate a line at its full thermal capacity. For example, if we drop the load to 135 MVA, the sending end voltage is a more reasonable 258 kV and the efficiency is nearly 97%! This is because while the current is proportional to load, the losses vary as the square of the current.

4.8 Summary

We have learned that an electrical power system has four major components:

- Generators
- Transformers
- Transmission lines
- Loads

We also discovered that the system is typically three-phase, and operates in the ac mode, at a frequency of 50 or 60 Hz, the latter being standard in the United States.

Generators provide the electrical source, and in most cases are large three-phase synchronous machines. These machines are driven by either steam or hydraulic turbines, at speeds which determine the system frequency. We realize that all the generators must be synchronized (i.e., run at the same electrical speed). Generators are modeled with equivalent circuits, which are invaluable in assessing the overall system steady state and dynamic performance. The pumped storage application was a particularly useful example of how generators are used in power systems.

It is necessary to efficiently change, or "transform," voltage levels throughout power systems. The device which is used for that purpose is the electric power transformer, or simply the transformer. Thousands of transformers are deployed throughout the system: when used to raise the voltage, they are described as "step up," and when used to lower the voltage, as "step down." The current changes in the opposite direction, the transformer being essentially a constant power device, after the losses are accounted for.

Because load and generation centers tend to be remote from each other, there is the problem of moving bulk electrical energy over large distances. The most common solution is the overhead three-phase power transmission line. As for all power system components, the equivalent circuit provides an excellent mathematical model for investigating operational and design problems.

It remains for us to consider the power system loads, which in many ways are the most complex part because of their number and diversity. We will investigate loads in Chapter 5.

References

Beaty, H. W., and J. L. Kirtley. *Electric Motor Handbook*. New York: McGraw-Hill, 1998.

Boldea, I. *The Electric Generators Handbook*, vols. 1 and 2. Boca Raton, FL: CRC Press, 2006.

Fink, D. G., and H. W. Beaty, eds. *Standard Handbook for Electrical Engineers*, 14th ed. New York: McGraw-Hill, 1999.

Fitzgerald, A. E., C. Kingsley Jr., and S. D. Umans. *Electric Machinery*. 6th ed. New York: McGraw-Hill, 2003.

Flanagan, W. M. *Handbook of Transformer Applications*. New York: McGraw-Hill, 1986.

Gross, C. A. *Power System Analysis*, 2nd ed. New York: Wiley, 1986.

Gross, C. A. *Electric Machines*. Boca Raton, FL: CRC Press, 2007.

Laithwaite, E. R. *A History of Linear Electric Motors*. London: Macmillan, 1987.

National Electrical Manufacturers Association. *Motors and Generators* (Publication No. MG1-1972). New York: NEMA, 1972.

Problems

4.1. a. Given a 208 V three-phase system, find all six voltages for a phase sequence of abc.

b. Draw the corresponding phasor diagram.

4.2. Repeat Problem 4.1 for a phase sequence of acb.

4.3. Given a 208 V three-phase system, find the currents if

a. the load is wye connected, with $\bar{Z}_Y = 8.66 + 5\ \Omega$.

b. the load is delta connected, with $\bar{Z}_\delta = 25.98 + 15\ \Omega$.

4.4. a. Three 12 Ω inductors are connected in wye. Find the equivalent delta elements.

b. Three 12 Ω capacitors are connected in delta. Find the equivalent wye elements.

4.5. In a balanced 480 V three-phase system, the line current is measured to be 15 A. Find $\bar{Z}_Y$ and $\bar{Z}$.

4.6. For the system of Problem 4.3, determine the complex load power.

4.7. Consider a three-phase, 12.47 kV system supplying a 600 kVA load, pf = 0.8 lagging. Determine the instantaneous power (a) per phase, and (b) in total.

4.8. Consider a six-pole ac generator, Determine synchronous speed for operation on a

a. 60 Hz system.

b. 50 Hz system.

4.9. A 150 MVA 13.8 kV 60 Hz generator operates at rated conditions. X_d = 1.5 Ω. Draw the equivalent circuit and find all values if the pf is

a. unity.

b. 0.866 lagging.

c. 0.866 leading.

4.10. For the situation described in Problem 4.9, determine the complex power flowing from the generator into the grid for a pf of

a. unity.

b. 0.866 lagging.

c. 0.866 leading.

4.11. For the situation described in Problem 4.9, show that the complex power flowing from the generator into the grid for a pf of 0.866 lagging can be calculated from

$$\bar{S}_{3\phi} = 3\frac{E_f \cdot V_a}{X_d}\sin(\delta) + j3\left(\frac{E_f \cdot V_a}{X_d}\cos(\delta) - \frac{V_a^2}{X_d}\right)$$

4.12. A pumped storage application is similar, but not identical, to that discussed in Section 4.4. The system features four identical units each rated as follows:

Stator ratings: 17 kV 60 Hz 24-pole 290 MVA, $X_d = 1$ Ω

Rotor field data: $R_f = 0.5\ \Omega$, $K_f = 20$ V/A, 24 poles

$$\omega_s = \frac{4\pi f}{N_P} = \frac{377}{12} = 31.42\ rad/s = 300\ rpm$$

The system operates at an overall efficiency of 86% efficiency when either pumping or generating. The facility is scheduled to inject 1000 MW into the grid for 4 hours per day, 5 days per week.

a. If the unit operates at rated conditions, determine the two possible operating pf's and their corresponding field current values.
b. How long must we pump to refill the reservoir, if we pump at 1000 MW?
c. If we pump at rated conditions, determine the two possible operating pf's and corresponding field current values.

4.13. Example 4.13 demonstrated that a coil wrapped around a magnetic core is a practical design for an inductor. Given a toroidal core with cross-section 1 cm^2, mean length 14 cm, and $\mu = 5000\mu_0$, find the number of turns required to make a 50 mH inductor.

4.14. Given a 1ϕ3W transformer rated at 7.2kV/120V/120V and 75kVA / 37.5 kVA/ 37.5 kVA, number the windings 1, 2, and 3.

a. Find the turn ratios between all windings.
b. Find the rated current for all windings.

4.15. The transformer described in Problem 4.13 supplies four identical efficiency apartments, each with the following connected loads:

AN	120V	6 kVA	pf = unity
BN	120V	5 kVA	pf = 0.80 lagging
AB	240V	7 kVA	pf = 0.85 lagging

Draw the circuit and solve for all currents, assuming a worst-case condition of all four units at maximum load.

4.16. Three identical transformers as described in Problem 4.13 are interconnected to form a three-phase bank. The bank is supplied from a 12.47 kV three-phase system, with the HV primaries connected in wye, and serves a small apartment building with twelve units. Each unit has a connected load as described in Problem 4.14. The bank secondary supplies three one-phase panels (Panel A … phase A, Panel B … phase B, and Panel C … phase C), each panel supplying four units. Solve for all currents, assuming a worst-case condition of all twelve units at maximum load.

4.17. Each phase of a 345 kV three-phase power transmission line is made up of two 900 A conductors.

a. Determine the rated line current.
b. Compute the thermal MVA rating of the line.

4.18. The line of Problem 4.16 is 100 miles long, and has the following equivalent values:

$$\bar{Z} = 5.6 + j60\ \Omega \qquad \frac{\bar{Y}}{2} = j0.35\ mS$$

The line serves a load of 1076 MVA @ 345 kV. Compute the sending end voltage, current, power, and pf, if the load pf is

a. unity.
b. 0.866 lagging.
c. 0.866 leading.

This page intentionally left blank

Chapter
5

Electrical Loads

An electrical power system may be defined as having four major components:

- Generators
- Transformers
- Transmission lines
- Loads

In Chapter 4 we studied the generation, transformation, and transmission of electrical energy. We now turn our attention to the utilization of electrical energy. Electric utility engineers refer to these energy sinks as "loads" on the system. Loads are the most complex part of a power system because of their number and diversity. They range from a watt or two (e.g., a plug-in night light) to megawatts (e.g., an electrified rail system). On even a medium-sized system, they number into the millions and are spread over the entire utility service area. Our objective in this chapter is to investigate a few of the more common load types. Bear in mind that entire books have been written on each load type, so our study must be concise and limited in scope. Still, this material should serve as a useful starting point for more in-depth work, should the reader have the need and interest.

Loads may be divided into two categories:

- Static (those which have no moving parts)
- Rotating (e.g., motors)

5.1 Modeling Loads

Loads are frequently modeled in one of two ways:

- Constant (complex) power
- Constant impedance

It is common to describe loads as "kVA (MVA) @ a given pf and voltage." To explain how these data can be converted into complex power and/or complex impedance is best demonstrated with an example. Let us first review the meaning of "leading or lagging pf" terminology. "Lagging" means that the current *lags* the voltage, which is the case when the load is *inductive*. *Inductive* loads *absorb* reactive power ($Q > 0$). "Leading" means that the current *leads* the voltage, which is the case when the load is *capacitive*. *Capacitive* loads *deliver* reactive power ($Q < 0$).

Example 5.1

Given a 100 kVA 7.2 kV 0.8 pf lagging load:

a. Compute the corresponding complex power.

$$\theta = \cos^{-1}(0.8) = \pm 36.9°$$

We choose "+36.9°" because of the meaning of "lagging."

$$\bar{S} = 100\underline{/36.9°}\text{ kVA} = 80\text{ kW} + j60\text{ k var}$$

b. Compute the corresponding complex impedance.

$$\bar{I} = \left(\frac{\bar{S}}{\bar{V}}\right)^* = \left(\frac{100\underline{/36.9°}}{7.2}\right)^* = 13.89\underline{/-36.9°}\text{ A}$$

$$\bar{Z} = \frac{\bar{V}}{\bar{I}} = \left(\frac{7.2}{13.89\underline{/-36.9°}}\right) = 0.5184\underline{/36.9°} = 414.7 + j311\ \Omega$$

At first there would seem to be no difference in the two models. But now suppose the operating voltage increases 10%. For the constant power model, the current *decreases* 10%; for constant impedance, the current *increases* 10%. An example illustrates.

Example 5.2

For the load of Example 5.1, the load voltage increases to 7.92 kV. Compute the current if the load is modeled as

a. constant power.

$$\bar{I} = \left(\frac{\bar{S}}{\bar{V}}\right)^* = \left(\frac{100\angle 36.9°}{7.92}\right)^* = 12.63\angle -36.9°\text{ A}$$

b. constant impedance.

$$\bar{I} = \frac{\bar{V}}{\bar{Z}} = \left(\frac{7.92}{0.5184\angle 36.9°}\right) = 15.28\angle -36.9°\text{ A}$$

It clearly makes a difference as to which approximation is used. It is straightforward to extend the procedure to the balanced three-phase case.

Example 5.3

Given a 300 kVA 12.47 kV 0.8 pf lagging three-phase load:

a. Compute the corresponding complex power.

$$\theta = \cos^{-1}(0.8) = 36.9°$$

$$\bar{S}_{3\phi} = 300\angle 36.9° \text{ kVA} = 240 \text{ kW} + j180 \text{ kvar}$$

b. Compute the corresponding complex impedance.

$$I_L = \frac{S_{3\phi}}{\sqrt{3}V_L} = \frac{300}{\sqrt{3}(12.47)} = \left(\frac{100\angle 36.9°}{7.2}\right)^* = 13.89\angle -36.9° \text{ A}$$

$$\bar{I}_a = 13.89\angle -36.9° \text{ A}$$

$$\bar{Z}_Y = \frac{\bar{V}_{an}}{\bar{I}_a} = \left(\frac{7.2}{13.89\angle -36.9°}\right) = 0.5184\angle -36.9° \ k\Omega = 414.7 + j311\,\Omega$$

$$\bar{Z} = 3 \cdot \bar{Z}_Y = \frac{\bar{V}_{an}}{\bar{I}_a} = 1244 + j933\,\Omega$$

Example 5.3 raises some issues which can cause confusion. When an engineer says that we have a 12.47 kV three-phase system, just what voltage is "12.47 kV"? After all, there are at least six possibilities:

$$V_{an}; V_{bn}; V_{cn}; V_{ab}; V_{bc}; V_{ca}$$

Well, unless otherwise specified, the speaker assumes that the system is balanced, which reduces the possibilities to two since $V_{an} = V_{bn} = V_{cn}$ and $V_{ab} = V_{bc} = V_{cb}$. Again, unless otherwise specified, the speaker means the following:

$$V_{ab} = V_{bc} = V_{ca} = V_L = 12.47 \text{ kV}$$

It follows that

$$V_{an} = V_{bn} = V_{cn} = \frac{V_L}{\sqrt{3}} = \frac{12.47}{\sqrt{3}} = 7.2\ \text{kV}$$

The next issue is the apparent power (300 kVA). Is it to be interpreted as per phase or three-phase? Well, think about an automobile with a 100 hp four-cylinder engine. We understand that 100 hp is the total rating of the engine (25 hp per cylinder). And so it is with three-phase apparatus. The power rating is *always* total, or three-phase. Thus:

$$S_{3\phi} = 300\ \text{kVA} \qquad S_{1\phi} = \frac{S_{3\phi}}{3} = 100\ \text{kVA}$$

Finally, there is the term "three-phase impedance." What would this be? Impedance by definition is a port (two-terminal) concept. But "three-phase" implies three (or four) terminals (abcn). We are forced to conclude that there is no such thing as "three-phase impedance." In a three-phase situation, *three* impedances are necessary, which must be equal for the balanced case. Some experimentation concludes that there are only two possible ways to connect three equal impedances to a balanced three-phase source such that the currents will be balanced: the wye and the delta (see Figure 4.4 for the proper connections). The details of computing these were presented in Example 5.3.

Static loads are usually modeled as constant impedance. We will investigate two important types of electrical static loads: heating and lighting.

5.2 Electric Heating

Conversion of electric energy into heat is simple and straightforward. Recall for a resistor:

$$P = R \cdot I^2$$

where the energy associated with P is heat. Thus, if we want to design an electric "heater," we design a resistor (i.e., the "heating element") that can operate at high temperature with no damage to itself. The heating element must be electrically and thermally insulated from human contact for safety reasons.

Radiation heaters emit infrared radiation (IR) through air or space onto an absorbing surface, where it is converted to heat and/or reflected, warming the surface, as opposed to the air. This style of heater is particularly useful in areas which unheated air flows through. They are ideal for task-specific heating. They lack overheating protection.

Convection heaters heat the surrounding air, which either circulates through the heater by natural convection, or is forced to circulate by means of a fan, in which case they are called "forced convection" heaters. They are usually thermostatically controlled to prevent overheating. Most home-heating systems are forced convection, with the fan referred to as the "blower."

Storage heating systems heat a storage element (thermal bricks, water, etc), usually converting electric energy at a time of day when electricity is cheaper, and using the stored thermal energy when needed. The familiar home water heater is a stored energy system.

Underfloor heating systems use insulated high-resistance electric cables as the heating element, and are laid out in a grid pattern under a floor. They heat the floor, which heats the adjacent air, which rises and diffuses the room with warm air. Sometimes a transformer is used to step the voltage down to a low level (< 50 V) for an extra safety margin. The system provides the best thermal gradient from floor to ceiling. Frank Lloyd Wright is said to have commented on "the indescribable comfort of being warmed from below." Unfortunately it is somewhat expensive, particularly if the space must be cooled as well. The system is thermostatically controlled with sensors in the floor. The floor covering needs to be compatible, avoiding a highly insulating material such as thick carpets.

Heat pumps are devices designed to move heat energy from Point A to Point B. If heat is to flow from a hot location to a cold location, nature will take care of things for us. However, if we want heat to move "uphill," we must be a bit more ingenious.

Consider a fluid confined to a closed path, called the "working" fluid. The fluid flows through a structure permitting easy heat transfer between the fluid and local environment (a "heat exchanger"). Recall that the heat of vaporization of a fluid is the amount of energy per unit mass required to convert from the liquid to the gas state. Heat flows from the environment to the working fluid as we convert from the liquid to the gas state ("boiling"), creating a so-called cold point. Likewise, heat flows from the working fluid to the environment as we convert from the gas to the liquid state ("condensing"), creating a so-called hot point.

So, for residential heating, we want the cold point outside, and the hot point inside: for cooling, the hot point outside, and the cold point inside. To move the fluid around the loop between locations, we must do work on the fluid, a task which is done by a motor-driven pump (the "compressor").

The heating mode degrades considerably at very low outside temperatures, so most units have auxiliary resistive strip heaters to add additional heat. It is possible to route the outside air through underground ducts, prewarming (or precooling) the air and raising the efficiency considerably.

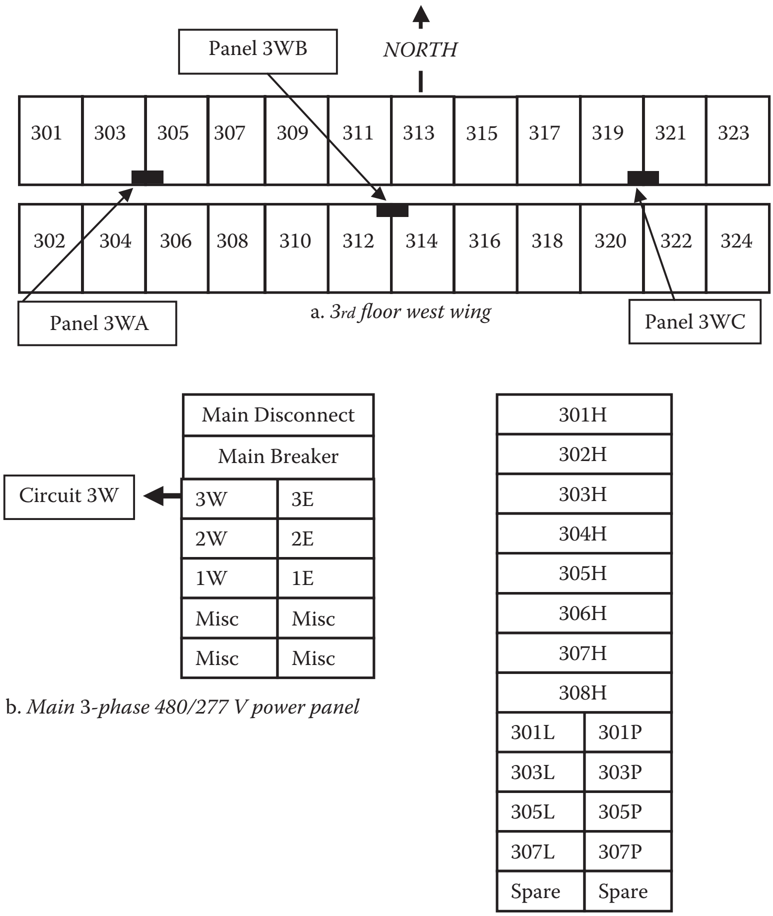
Example 5.4[1]

A three-floor, 144-room motel is rectangular, with two wings running east and west. The front faces south, and the back north. Each floor has an identical room layout of forty-eight rooms, with twenty-four rooms on each wing. See Figure 5.1a. Each room has a single-phase 240 V 4 kW heating unit. Also, each room is to have two 120 V circuits: one for lighting and one for outlets (TV, coffee maker, etc.) with a total connected load of 2 kW. Considering only room requirements, design an electrical distribution system. Assume that all loads are unity pf.

Power enters the building through underground conduits from a pad-mounted 12.47 kV Δ :480/277 VY, three-phase transformer, providing 480/277 V three-phase to the main power panel (see Figure 5.1b), located in a secure electrical closet on the first floor and in the center of the building. A raceway containing six three-phase 480 V circuits (1W, 1E, 2W, 2E, 3W, and 3E) runs up the center of the building, branching west, or east at the appropriate floor (e.g., circuit 3W runs through a raceway above the hall ceiling down the west wing on the third floor).

Let's start on the third floor. Circuit 3W supplies three single-phase transformers (3WA, 3WB, and 3WC), each rated at 480 V: 120 V/120 V, located in the ceiling cavity above panels 3WA, 3WB, and 3WC, respectively.

[1] The design presented in this example is for academic purposes only. It is not intended, or even recommended, as a template for a practical design. Practical designs must consider restrictions imposed by the National Electric Code, the National Electric Safety Code, and all other relevant building codes. No code restrictions have been considered here, nor have any economic considerations.



a. *3rd floor west wing*

b. *Main 3-phase 480/277 V power panel*

c. *240/120 V panel 3WA*

FIGURE 5.1 **Third-floor west wing layout.**

The phases are rotated according to the following:

Transformer 3WA: Primary voltage $V_{ab} = 480$ V
Transformer 3WB: Primary voltage $V_{bc} = 480$ V
Transformer 3WC: Primary voltage $V_{ca} = 480$ V

This is to balance the load, as they are throughout the building on all floors and all wings. Panel 3WA will supply a cluster of eight rooms (rooms 301–308). Each room will have a dedicated 240 V heater circuit (e.g., 301H in room 301). A lighting circuit (301L) serves rooms 301–302, as does the power circuit 301P. The latter includes a ground fault interrupter (GFI) at the outlet for the coffee maker and hair dryer. Note that each power and lighting circuit serves two rooms.

The connected load for each room is

Heating: 4 kVA
Lighting: 1 kVA (includes TV, etc.)
Power: 1 kVA
Total: 6 kVA (all loads assumed to have unity pf)

The connected load for each room is as follows:

Connected load for Panel 3WA (or 3WB or 3WC) is 8 × 6 = 48 kVA. However, this is **connected** load. The probability that this entire load will be simultaneously on at any given time is near zero. We account for this by multiplying by a "diversity factor" (DF), where DF ≤ 1. There are tables with recommended DF values for various applications. We select a conservative value of 0.75, producing an estimated maximum load value of 0.75(48) = 36 kVA. Hence we size the single-phase transformer 3WA and panel 3WA at 37.5 kVA (a standard size).

The three-phase 480 V circuit 3W capacity then should be 3 × 36 = 108 kVA. The corresponding current should be

$$I_L = \frac{S_{3\phi}}{\sqrt{3}V_L} = \frac{108}{\sqrt{3}(0.48)} = 130\text{ A}$$

The six main 480 V three-phase circuits (3W, 3E, 2W, 2E, 1W, and 1E) should each have an ampacity of 130 A, with equal circuit breaker ratings. The second- and first-floor designs replicate this design, such that the total power requirement is 6 × 108 = 648 kVA.

Checking our total:

$$S_{TOT} = 144\text{ rooms} \times 6\text{ kVA per room} \times 0.75\text{ (DF)} = 648\text{ kVA}$$

However, this is connected load. The probability that all this load will be on simultaneously is very small. We'll assume that 70% of this figure is conservative, or 605 kVA.

However, we haven't considered all the load. How about the following?

Office
Lobby

Pool
Hot tub
Restaurant
Laundry
Halls
Parking lot
Security system
Exercise room
Outside advertising lights?

Suppose all this adds up to another 352 kVA of diversified load.

$$S = 648 + 352 = 1000 \text{ kVA}$$

The main transformer rating, then, is 12.47 kV/480 V three-phase 1000 kVA.

The main three-phase power panel rating is 480 V 1000 kVA, with main circuit breaker and disconnect ratings of 1200 A.

Observe that the basic design of commercial electrical systems is a straightforward application of fundamental circuit concepts. In particular, notice how useful Tellegen's theorem was. Example 5.4 presented many useful design concepts. However, be aware that an acceptable, economic, and practical design must comply with all relevant codes, which were not considered here.

5.3 Electric Lighting

One of the earliest applications of electric power was lighting, beginning with Edison's invention of the electric incandescent lamp in 1879. For a dramatic view of electric lighting today, see Figure 5.2.

Since visible light is a form of electromagnetic radiation, like any field phenomena, we quantify it in terms of flux, and the illumination in the terms of flux density:

$$E = \frac{\phi}{A} = \text{illumination, in lumen/m}^2 (lux)$$

$$\phi = \text{luminous flux, lumen}$$

$$A = \text{area, m}^2$$

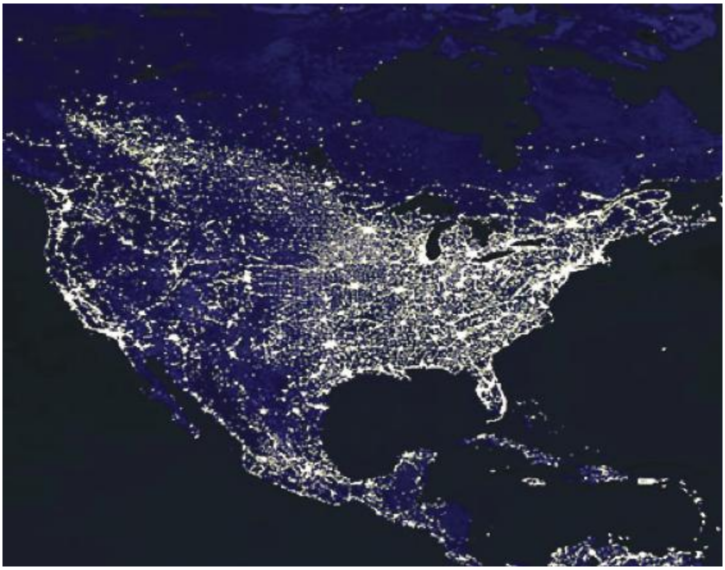


FIGURE 5.2 Satellite photograph at night. Source: http://www.ecoprint s.net/blog/envi ronment/usa_night_satellite.jpg.

If A is in ft^2, then E is in foot-candles (fc). A light source can be thought of as an energy conversion device with input power (P, in watts). If all the input was converted with 100% efficiency to light, one watt is equivalent to 683 lumens. In general:

$$Efc = \frac{\phi}{P} = \text{efficacy, lumen/watt}$$

$$P = \text{power input to the light source, W}$$

Efficacies for a few common light sources are listed in Table 5.1. Accounting for lumen source output degradation and room geometry, we produce the design equation:

$$E = \frac{CU \cdot LDF \cdot \phi}{A}$$

where:

E: flux density (flux per unit area), foot-candles (fc)
ϕ: light output, lumens

Table 5.1 Efficacies for a Few Common Light Sources

Source	Efficacy (Lumen/W)
Candle	0.3
40 W incandescent	13
100 W incandescent	18
Projection lamps	35
White LED	150
Fluorescent	50–100
High-pressure sodium	85–150
Low-pressure sodium	100–200

A: Area to be illuminated, ft^2
LDF: Lumen Depreciation Factor
CU: Coefficient of utilization

An example application follows.

Example 5.5

Design the lighting for an office space of 15’ × 25’ with a 9-foot ceiling using Lithonia LB440A fixtures. The work surface is assumed to be 30” (desktop) above the floor. The $CU = 0.61$ and the $LDF = 0.73$. The recommended light level for general office space is 100 fc.

$$E = \frac{CU \cdot LDF \cdot \phi}{A} = \frac{(0.61)\cdot(0.73)\cdot\phi}{(15)\cdot(25)} = 100$$

$$\phi = 84213 \text{ lumens}$$

Data for the Lithonia LB440A fixture (120V):

		Description:
Manufacturer:	Lithonia	This luminaire uses four 4’ standard fluorescent T12 tubes, is ceiling mounted, and has a wrap-around prismatic lens.
Model:	LB440A 4’F40T12	
Lamps: 4 each	34 W	
Lumens per Lamp:	2,650	
Total Lumens:	10,600	

Source: http://www.lightcalc.com/cutable.html

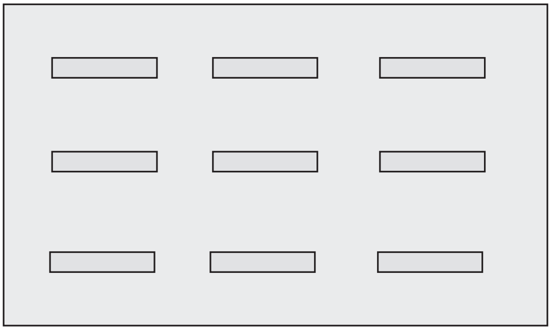


FIGURE 5.3 **Finished design for Example 5.5.**

Thus we need a minimum of 84213/10600 = 7.9 fixtures. We round up to nine to make a more even pattern of lighting, as shown in Figure 5.3. Each fixture has four lamps, which absorb 34 W each, for a total of 4 × 34 = 136 W. Since the fixture has an electronic ballast, we assume it is pf corrected to unity. Adding in 4 W for ballast losses, we get 140 W. We serve the room from a one-phase 120 V 20 A circuit. The required current is (9 × 140)/120 = 10.5 A. We serve the room from a one-phase 120 V 20 A circuit, a conservative design, because we want to avoid excessive voltage drop and provide for future expansion.

In addition to light intensity, lighting design must take into account many other issues, including color rendition, esthetics, economics, and maintenance. It is indeed part art, part science.

5.4 Electric Motors

Another broad application of the use of electrical energy is to convert it to mechanical form to do mechanical work. Think of a construction site where we need to lift 1000 lbs of material up to the fifth floor. We could give ten persons 100 lbs each to carry up eight flights of stairs. Or construct a block and tackle and use a mule. But, wouldn't it be great if one person could just flip a switch, raise the load some 60 feet, and swing it into position by pushing a lever with our little finger?

Or consider the problem of moving 100 people 100 miles. They could walk (it took the Roman legions about four days). Or they could go by stagecoach (about three days travel, and we'd need 20–25 stagecoaches given the size of the modern American). But, wouldn't it be great to have a vehicle automatically accelerate, cruise, and stop, arriving at our destination in 40 minutes or so? Motors can do these things, and more, and are clearly worthy of our serious consideration.

There are dozens of types of electric machines. However, all are electromagnetic energy conversion machines, and as such, they can operate in two modes:

- A generator, converting energy from mechanical to electrical form
- A motor, converting energy from electrical to mechanical form

When we focus on the latter mode, we refer to the machine as a "motor." We will consider three types of motors:

- Three-phase synchronous motors
- Three-phase induction motors
- Single-phase motors

Consider Newton's Second Law of Motion applied to the motor shaft (see Figure 5.4a).

$$T_{DEV} - T_L - T_{RL} = J \cdot \frac{d\omega_r}{dt}$$

T_{DEV} = Electromagnetic developed torque, Nm

T_L = Load torque, Nm

$T_{RL} = K_{RL} \cdot \omega_r$ = Rotational loss torque, Nm

J = Mass polar moment of inertia of all rotating parts, kg-m^2

ω_r = shaft speed, rad/s

The SI unit for angular velocity is rad/s. However, the use of revolutions per minute (rpm) is common in motor work.[2] To convert:

$$\text{rpm} = \frac{60(\text{rad/s})}{2\pi}$$

[2] Technically we should say "rev/min." However, "rpm" is so entrenched in industry that we defer to common usage. May the reader forgive our transgression.

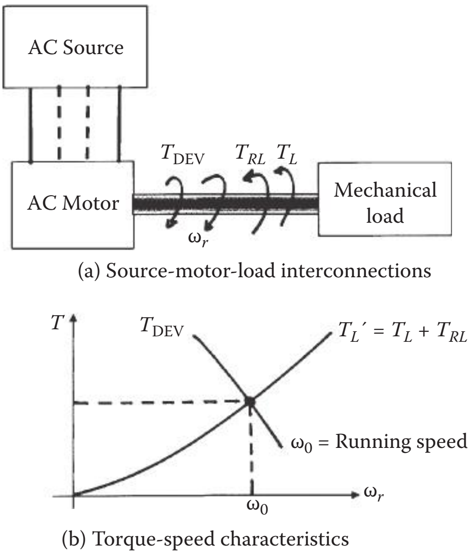


FIGURE 5.4 Motor-load system.

Example 5.6

Convert 188.5 rad/s to rpm.

$$\omega = \frac{60(\text{rad/s})}{2\pi} = \frac{60(188.5)}{2\pi} = 1800 \text{ rpm}$$

A basic fact:

> All motors run at a speed defined by the intersection of the motor and load torque speed characteristics

Thus, to determine the speed, we need

- The motor torque-speed characteristic ($T_{DEV} - \omega_r$)
- The load torque-speed characteristic ($T^1_L - \omega_r$)

See Figure 5.4b.

5.4.1 Three-Phase Synchronous Motors

Recall that we have already studied the three-phase synchronous machine in Chapter 4. We review its operation via Example 5.6.

Example 5.7

A given load is driven by a 480 V four-pole 100 hp three-phase synchronous motor, the data for which are as follows:

```
----------------------------Ratings----------------------------
VLine = 480 V; ILine = 120.3 A; S3ph = 100.0 kVA
Freq = 60 Hz; poles = 4; sync speed = 1800.0 rpm
Rotor type: permanent magnet; EF @ 1800 rpm = 300.0 V
==============================================================
*** Equivalent circuit values (R, x in ohms) ***
Ra = 0.0460; TRL = KRL*wrm; KRL = 0.0281 Nm-s/rad
Xd = 2.500;
```

Load: $T_L = 0.01 \cdot \omega_r^2$ Nm

Motor: $E_f = 300$ V; $X_d = 2.5\ \Omega$

Find the load torque, speed, current, efficiency, and pf.

The motor runs at only one speed:

$$\omega_r = \omega_s = \frac{4\pi f}{N_P} = \frac{4\pi(60)}{4} = 188.5 \text{ rad/s}$$

In Chapter 4, we neglected rotational and stator winding losses. Here, we will consider these.

Continuing:

$$T_L = 0.01 \cdot \omega_r^2 = 0.01(188.5)^2 = 355.3 \text{ Nm}$$

$$T_{RL} = 0.0281 \cdot \omega_r = 0.0281(188.5) = 5.3 \text{ Nm}$$

$$T_{DEV} = T_L + T_{RL} = 360.6 \text{ Nm}$$

$$P_{DEV} = T_{DEV} \cdot \omega_r = 67.98 \text{ kW}$$

$$P_{DEV} = 3\frac{E_f\ V_{an}}{X_d}\sin(\delta) = 3\frac{300(277.1)}{2.5}\sin(\delta) = 99.76\sin(\delta) \text{ kW}$$

$$99.76 \cdot \sin(\delta) = 67.98$$

$$\delta = 43° \ (\bar{E}_f \text{ lags } \bar{V}_{an} \text{ in the motor mode})$$

Computing the current:

$$\overline{I}_a = \frac{\overline{V}_{an} - \overline{E}_f}{jX_d} = \frac{277.1\angle 0° - 300\angle -43°}{j2.5}$$

$$\overline{I}_a = 85.03\angle -15.74° \ A$$

$$pf = \cos(15.74°) = 0.9625 \text{ lagging}$$

Computing the powers:

$$\overline{S}_{3\phi} = 3\overline{V}_{an} \cdot \overline{I}_a{}^* = 3(277.1)(85.03\angle -15.74°)^*$$

$$\overline{S}_{3\phi} = 72.42 \text{ kW} + j12.55 \text{ kvar}$$

$$P_{\text{IN}} = 72.42 \text{ kW}$$

$$P_{\text{OUT}} = T_L \cdot \omega_r = 355.3(188.5) = 67.97 \text{ kW}$$

$$\eta = \frac{67.97}{72.42} = 92.48\ \%$$

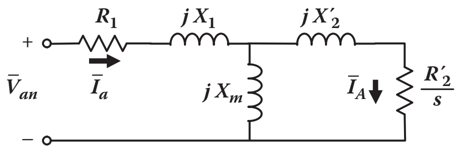
5.4.2 Three-Phase Induction Motors

We now consider the three-phase induction motor. The induction machine's stator is identical to that of the synchronous machine. However, the rotor is quite different. Think of the rotor as being a structural replica of the stator, containing three-phase windings wound for N_P poles, and shorted.[3]

Remember that we need two interacting magnetic fields to produce torque. We have the stator field, due to the applied voltage. But how about the rotor field? Surely no rotor currents flow if the windings are shorted! This is true as the rotor turns at synchronous speed, making the rotor stationary with respect to the rotating stator field. If the rotor currents are zero, there is no rotor field. No rotor field, no torque! Hence, the rotor must slow down due to the load torque.

Now as the rotor slows down, there is relative motion between the rotor and the rotating stator field. By Faraday's Law (i.e., "induction"), small voltages

[3] While there are such things as wound rotor induction machines, most have cage rotors (think squirrel cage). It's just easier to think in terms of windings. In terms of operation, the two designs are essentially the same.



R_1 = stator per phase resistance, Ω
X_1 = stator per phase leakage reactance, Ω
R_2 = rotor per phase resistance, Ω
X_2 = rotor per leakage reactance, Ω
$\alpha = N_1/N_2$ = stator to rotor turns ratio
$R'_2 = \alpha^2 \cdot R_2$ = rotor resistance reflected into the stator, Ω
$X'_2 = \alpha^2 \cdot X_2$ = rotor reactance reflected into the stator, Ω
X_m = per phase magnetizing reactance, Ω
s = slip
$\bar{I}'_a$ = stator phase a phasor current, A
$\bar{I}'_A$ = rotor phase A phasor current (reflected into stator), A
$\bar{V}_{an}$ = stator phase a-n phasor voltage (externally applied), V

FIGURE 5.5 **Circuit diagram for a three-phase induction motor.**

are induced in the rotor windings. But the windings are shorted! Hence, small voltages create big currents, which in turn create a substantial rotor field. We now have the requisite two fields to produce torque.

The circuit model for the three-phase induction motor is provided in Figure 5.5. Note the variable "s," which stands for "slip," which is a measure of the relative velocity of the rotor with respect to the stator field.

$$s = \text{slip} = \frac{\omega_s - \omega_r}{\omega_s}$$

The circuit of Figure 5.5 is a reasonably accurate model for the wound-rotor polyphase induction machine, operating under balanced three-phase ac excitation at constant speed. However, it is in serious error for the cage rotor device. Nonetheless, it is frequently used to predict performance for both machines. The purpose of the equivalent circuit is to provide a mathematical model which may be used to assess machine performance.

We will consider what we can determine from the model by means of a comprehensive example. The general problem is to determine "everything," given a machine of known circuit values and known terminations. "Everything" includes the slip, all currents, all powers, all torques, the power factor, and the efficiency.

Data for an example 460 V 100 hp three-phase wound-rotor induction motor are provided in Table 5.2. The motor is operating at balanced rated

Table 5.2 Data for a Three-Phase Induction Motor

*** Three-Phase Induction Motor Data ***
-------------------------Ratings-------------------------
Line Voltage = 460 Volts; Horsepower = 100 Hp
Stator Frequency = 60 Hz; No. of Poles = 4
Design Class = W; Synchronous Speed = 1800.0 Rpm
Rx' = 0.0000 ohm; Ns/Nr = 2.5
====================== Test Data ======================
-- Blocked Rotor(@ 60 Hz) -|--- No Load Values ---
Line Voltage = 84.82 | 460.0 Volts
Line Current = 125.5 | 35.09 Amperes
3 Ph Power = 4000.2 | 2394.2 Watts
*** Equivalent Circuit Values (in stator ohms @ Rated Freq) ***
R1 = 0.04232; X1 = 0.19044; Xm = 7.406
R2' = 0.04232; X2' = 0.19044; Jm = 0.9299 kg-m sq
Rotational Loss Torque = TRL = 0.06299 × (speed in rad/s) Nm

Note: a. Per-phase wye equivalent circuit for a three-phase wound-rotor induction motor.
b. Circuit constants and variables.

stator voltage @ 60 Hz, and running at 1770 rpm with the rotor shorted. We shall find "everything."

First, find synchronous speed.

$$\omega_s = \frac{4\pi f}{N_P} = \frac{4\pi(60)}{4} = 188.5 \text{ rad/s} = 1800 \text{ rpm}$$

R_1 = stator per phase resistance, Ω

X_1 = stator per phase leakage reactance, Ω

R_2 = rotor per phase resistance, Ω

X_2 = rotor per phase leakage reactance, Ω

$\alpha = N_1/N_2$ = stator to rotor turns ratio

$R_2' = \alpha^2 \cdot R_2$ = rotor resistance reflected into the stator, Ω

$X_2' = \alpha^2 \cdot X_2$ = rotor reactance reflected into the stator, Ω

X_m = per phase magnetizing reactance, Ω

$s = \text{slip}$

$\bar{I}_a = \text{stator phase a phasor current, A}$

$\bar{I}_A^{/} = \text{rotor phase A phasor current (reflected into stator), A}$

$\bar{V}_{an} = \text{stator phase a-n phasor voltage (externally applied), V}$

Now find the slip.

$$s = \text{slip} = \frac{\omega_s - \omega_r}{\omega_s} = \frac{1800 - 1770}{1800} = 0.01667$$

$$\omega_r = 1770\left(\frac{188.5}{1800}\right) = 185.36 \text{ rad/s}$$

Determination of the machine currents is a key problem. Applying Kirchhoff's voltage law (KVL):

$$\bar{V}_{an} = \frac{460}{\sqrt{3}}\angle 0° = (R_1 + jX_1)\bar{I}_a + jX_m\left(\bar{I}_a - \bar{I}'_A\right)$$

$$0 = jX_m\left(\bar{I}'_A - \bar{I}_a\right) + \left(\frac{R'_2}{s} + jX'_2\right)\bar{I}'_A$$

which can be solved for the currents. This is complicated but straightforward. Results are

$$\bar{I}_a = 107.4\angle -26.46° \quad \text{A}$$

$$\bar{I}'_A = 99.33\angle -7.98° \quad \text{A}$$

Next we shall determine the power losses.

$$\text{Stator winding loss} = \text{SWL} = 3 \cdot I_a^2 \cdot R_1 = 1.465 \text{ kW}$$

$$\text{Rotor winding loss} = \text{RWL} = 3 \cdot I_A^{/2} \cdot R_2^{/} = 1.253 \text{ kW}$$

$$\text{Rotational loss} = P_{RL} = K_{RL} \cdot \omega_r^2 = 2.164 \text{ kW}$$

$$\text{Sum of losses} = \Sigma L = \text{SWL} + \text{RWL} + P_{RL} = 4.882 \text{ kW}$$

Observe that

$$\frac{R_{2x}^{/}}{s} = R_{2x}^{/} + \left(\frac{1-s}{s}\right) R_2^{/}$$

Multiplying by $3 \cdot I_A^{/2}$:

$$3 \cdot I_A^{/2} \cdot \frac{R_2^{/}}{s} = 3 \cdot I_A^{/2} \cdot R_2^{/} + 3 \cdot I_A^{/2} \cdot \left(\frac{1-s}{s}\right) R_2^{/}$$

$$3 \cdot I_A^{/2} \cdot \frac{R_2^{/}}{s} = \text{RWL} + P_{DEV}$$

P_{DEV} is the so-called developed power or that power which is converted from electrical to mechanical form. This is the power that the machine was designed to convert. Again:

$$P_{DEV} = 3 \cdot I_A^{/2} \cdot \left(\frac{1-s}{s}\right) R_2^{/}$$

For our example situation:

$$P_{DEV} = 3 \cdot I_A^{/2} \cdot \left(\frac{1-s}{s}\right) R_2^{/} = 73.902 \text{ kW}$$

For motor operation:

$$P_{OUT} = P_{DEV} - P_{RL} = 73.902 - 2.164 = 71.738 \text{ kW}$$

Converting to horsepower:

$$HP_{OUT} = \frac{P_{OUT}}{0.746} = 96.163\ hp$$

The input power, power factor (pf), and efficiency (h):

$$P_{IN} = 3 \cdot V_{an} \cdot I_a \cdot \cos(\theta_a) = 76.919 \text{ kW}$$

$$pf = \cos(\theta_a) = 0.8952 \text{ lagging}$$

$$\eta = \frac{71.738}{76.919} = 93.63\ \%$$

Computing the torques:

$$\text{Developed Torque} = T_{DEV} = \frac{P_{DEV}}{\omega_r} = \frac{3\, I_A'^2\, (1-s)R_2'/s}{(1-s)\cdot\omega_s}$$

$$T_{DEV} = \frac{3\cdot I_A'^2\cdot(R_2'/s)}{\omega_s} = 398.7\ N\cdot m$$

$$\text{Rotational Loss Torque} = T_{RL} = K_{RL}\cdot\omega_r = 0.063(185.36) = 11.68\ N\cdot m$$

$$T_m = T_{OUT} = T_{DEV} - T_{RL} = 387.0\ N\cdot m$$

We have demonstrated how to find "everything," given a motor with known data, known terminations, and the speed, an exercise which we refer to as a "comprehensive analysis." Observe that knowing the speed was critical, since the speed determines the slip.

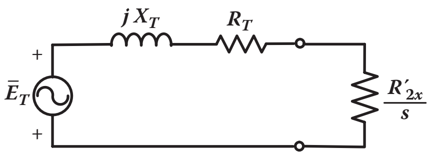
If the speed changes, so does the slip, and everything else (the currents, the powers, and the torques, as well as the efficiency and power factor).

So what if the speed is not known? To perform a comprehensive analysis, we must either know the speed or know enough about the mechanical load to determine the speed. Knowing the load means knowing the load torque–speed characteristic.

Observe that the most difficult part of the comprehensive analysis was solving for the currents, since we had to solve a two-mesh ac circuits problem. We note that T_{DEV} is a function of the *rotor* current (I_A') only. Since we are going to solve the problem repeatedly, there is motivation to simplify the current calculations by simplifying the circuit model to the one-mesh circuit shown in Figure 5.6a. We apply Thevenin's theorem to the stator section. The reduction equations are supplied in Figure 5.6b. Then:

$$\bar{I}_A' = \frac{\bar{E}_T}{\left(R_T + \frac{R_2'}{s}\right) + jX_T}$$

$$T_{DEV} = \frac{3\cdot I_A'^2\cdot\left(\frac{R_{2x}'}{s}\right)}{\omega_s}$$



(a) Circuit diagram

$$\bar{Z}_T = R_T + jX_T = jX_2' + \frac{jX_m(R_1 + jX_1)}{R_1 + j(X_1 + X_m)}$$

$$\bar{E}_T = \frac{jX_m}{R_1 + j(X_1 + X_m)}(\bar{V}_{an})$$

(b) Thevenin values

FIGURE 5.6 **Simplified per-phase wye equivalent circuit for a three-phase wound-rotor induction motor.**

Example 5.8

Determine the Thevenin equivalent circuit for the example machine of Table 5.2.

$$\bar{Z}_T = R_T + jX_T = jX_2' + \frac{jX_m(R_1 + jX_1)}{R_1 + j(X_1 + X_m)} = 0.04022 + j0.3763\ \Omega$$

$$\bar{E}_T = \frac{jX_m}{R_1 + j(X_1 + X_m)}(\bar{V}_{an}) = 258.9\angle\alpha \quad \text{V}$$

Example 5.9

Use the Thevenin equivalent circuit for the example machine of Table 5.2 to compute the developed torque at a speed of 1770 rpm.

$$s = slip = \frac{\omega_s - \omega_r}{\omega_s} = \frac{1800 - 1770}{1800} = 0.01667$$

$$\frac{R_2'}{s} = \frac{0.04232}{0.01667} = 2.539\ \Omega$$

$$\bar{I}_A' = \frac{\bar{E}_T}{\left(R_T + \frac{R_2'}{s}\right) + jX_T} = \frac{258.9\angle\alpha}{(0.04022 + 2.539) + j0.3763} = 99.33\angle\beta \quad \text{A}$$

$$T_{DEV} = \frac{3 \cdot I_A^{/2} \cdot (R_2^/ / s)}{\omega_s} = 398.7\ N \cdot m$$

The results agree with those computed from the two-mesh equivalent circuit, which of course they must. Note that the Thevenin method is much simpler.

The Thevenin circuit offers one more advantage. Suppose we were interested in the maximum developed torque and the speed at which it occurs. Consider the following:

$$T_{DEV} = \frac{3 \cdot I_A^{/2} \cdot (R_2^/ / s)}{\omega_s}$$

Since ω_s is constant, T_{DEV} maximizes when the numerator maximizes. But this is the power dissipated in the resistor R_2'/s. The maximum power transfer theorem says that maximum power is the power dissipated in the resistor R_2'/s when

$$\left.\frac{R_2^/}{s}\right|_{s=s_{MT}} = Z_T$$

$$s_{MT} = \frac{R_2^/}{Z_T}$$

Once the slip is known, the speed and torque can be computed.

Example 5.10

Use the Thevenin equivalent circuit for the example machine of Table 5.2 to compute the maximum developed torque and the slip and speed at which it occurs.

$$s_{MT} = \frac{R_{2x}^/}{Z_T} = \frac{0.04232}{0.3785} = 0.1118$$

$$\omega_{MT} = (1 - s_{MT}) \cdot \omega_{MT} = (1 - 0.1118) \cdot 1800$$

$$\omega_{MT} = 1599 \text{ rpm}$$

$$\bar{I}'_A = \frac{\bar{E}_T}{(R_T + Z_T) + jX_T} = \frac{258.9\angle\alpha}{(0.04232 + 0.3785) + j0.3763} = 458.6\angle\beta \quad \text{A}$$

$$T_{\text{MAX}} = \frac{3 \cdot I_A'^2 \cdot \left(\dfrac{R'_{2x}}{s}\right)}{\omega_s} = \frac{3 \cdot (458.6)^2 \cdot (0.3785)}{188.5} = 1267\ N \cdot m$$

We now turn our attention to solving for the speed with a known load and motor. Our strategy is to use a "cut and try" approach:

1. Pick a speed (98.5% of synchronous is a reasonable first guess).
2. Compute the load (T_m') and motor (T_{DEV}) torques.

3a. If $T_{DEV} > T_m'$, pick a slower speed (in small increments, say, 0.2%).

3b. If $T_{DEV} < T_m'$, pick a faster speed.

4. Repeat steps 1–3 until $T_{DEV} = T_m'$ (to within about 0.1%).

We will demonstrate using our example motor driving a known load, which is

$$T_L = 0.01 \cdot \omega_r^2 \ \text{ Nm} \qquad T_{RL} = 0.063 \cdot \omega_r \ \text{ Nm}$$

$$T'_L = T_L + T_{RL} = 0.063 \cdot \omega_r + 0.01 \cdot \omega_r^2$$

We'll use the Thevenin equivalent to compute T_{DEV} (implemented by computer).

```
Enter initial guess for speed < 1764.0>...
DW:=K*SRPM*((Tdev-TRL)-TM)/TRAT Enter "K" <0.0200>...
  wrm    Tdev     TL'     DW=change in speed
============================================================
 1764.0   471.24   352.88   Enter DW < 10.77>...
 1774.8   339.06   357.13   Enter DW < -1.64>...
 1773.1   359.81   356.48   Enter DW < 0.30>...
 1773.4   356.00   356.60   Enter DW < -0.05>...
 1773.4   356.68   356.58   Enter DW < 0.01>...
Solution converged for this load!
```

See how the method was implemented. Starting with a guess of 98% of synchronous speed, we get 1764. At this speed, the computer computed 471 and 353 for the developed and load torques, respectively. In the physical world, the system would accelerate to a faster speed, so our second guess would be greater than 1764. Let's speed up by 10.8 rpm to 1774.8 for our second guess.

At this speed, the computer computed 339 and 357 for the developed and load torques, respectively, meaning that we overshot the run speed. The system would decelerate to a slower speed, which our algorithm recommends to be 1774.8 − 1.6 = 1773.1. The method converges to a solution within 0.1 rpm in three more iterations.

After determining that the equilibrium running speed is 1773.4 rpm, once the speed is determined, we can execute a comprehensive analysis. Results are as follows:

```
Speed = wrm = 1773.4 rpm = 185.7 rad/s; slip = 0.01479
 LOAD shaft speed = wL = 1773.4 rpm = 185.7 rad/s
 Stator voltage, frequency: 460.0 V; 60.0 Hz
Operating mode: MOTOR
Currents in (stator) amperes ... Ia(stator), IA'(rotor)
 Ia = 96.979 @ -27.71 deg; IA' = 88.477 @ -7.07 deg
 Torques in newton-meters (RL = rotational loss):
 Developed = 356.56; TRL = 11.697
 TL = 344.88;
Powers in kW:
 Mech. output = 64.044 (85.850 hp)
 Stator input = 68.404; developed = 66.216
Losses in kW:
 SWL = 1.194; RWL = 0.994; PRL = 2.172; ΣL = 4.360
Eff = 93.63 %; PFactor = 88.53 %;
```

Note that T_{DEV}, T_L', ω_s, and ω_r were all positive in this analysis. Positive direction of rotation is originally defined by ω_s, that is, the direction of rotation of the stator field, which in turn is defined by the phase sequence of the applied stator voltages and the physical location of the three stator phase windings. For example, refer to Figure 4.6. If the phase sequence of the applied stator voltages is abc, the rotating field rotates from phase winding a to phase b to phase c of CCW.

Suppose that ω_s is CCW in this section. Once positive rotation is defined, the positive senses of T_{DEV}, T_L', and ω_r are defined in Figure 5.4a, T_{DEV}, ω_s, and ω_r being CCW, and T_L' being CW. Now, what if the phase sequence of the applied voltages were reversed? The direction of rotation of the stator field

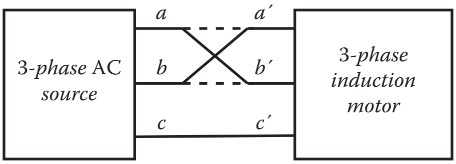


FIGURE 5.7 **Reversing a three-phase induction motor.**

would reverse to CW, as would the positive senses of that T_{DEV}, T_L', and ω_r, and the system would run backward![4] It is simple to reverse the phase sequence of the applied voltages: just reverse any two phase connections. Therefore:

> To reverse the direction of a three-phase induction motor, reverse any two phase connections, as shown in Figure 5.7.

We return to the standard situation of Figure 5.4, where T_{DEV}, T_L', ω_s, and ω_r are all positive in this analysis, and positive rotation is CCW. Also recall that

$$P_{DEV} = T_{DEV} \cdot \omega_r$$

And so if T_{DEV} and ω_r are positive, this means that P_{DEV} flows from machine to load, as shown in Table 5.3.

For speeds in the range of $0 < \omega_r < \omega_s$ equilibrium points can be found where T_{DEV}, T_L', ω_s, and ω_r are all positive in this analysis, and positive rotation is CCW. Since power is converted from the electrical to the mechanical form, we say the machine is operating "in the motor mode."

But now suppose that $\omega_r > \omega_s$. It follows that the slip is negative. If the slip is negative:

$$T_{DEV} = \frac{3 \cdot I_A'^2 \cdot \left(\frac{R_{2x}'}{s}\right)}{\omega_s} < 0$$

which means that torque and speed are in opposite directions, and P_{DEV} is negative! Negative P_{DEV} indicates that power now flows from the "load" into

[4] This assumes, of course, that the load torque–speed characteristic is symmetrical about the origin, as most loads are. For example, suppose a fan requires T_o at ω_0. Typically it also requires $-T_o$ at $-\omega_0$.

Table 5.3 Operating Modes of the Three-Phase Induction Machine

Speed, Slip	Torque	Power Flow	Mode
$0 \leq \omega_r \leq \omega_s$ $1 \geq s \geq 0$	$T_{DEV} \geq 0$	Machine to load	Motor
$\omega_r > \omega_s$ $s < 0$	$T_{DEV} < 0$	Load to machine	Generator
$\omega_r < 0$ $s > 1$	$T_{DEV} > 0$	Machine to load	Braking

the machine, and after losses, from the machine stator, into the external three-phase electrical source. Since power flows from the mechanical port (i.e., the shaft) and exits through the electrical port (i.e., the stator source), the machine is now a generator by definition.

But how can a "load" do this? Well, most "loads" can't. However, a few can. Take a windmill, for example. In still air, the machine "motors," drawing power for the electrical source and rotating the windmill blades. But if the wind starts to blow with sufficient force, it will exert torque in the direction of rotation and the direction of energy flow will reverse. So it seems we were prejudiced by calling the termination a "load." Thinking more generally, perhaps we should call it a "mechanical thing." The key issue is that *if* the "load" is capable of providing torque in the direction of rotation, then the machine automatically becomes a generator.

Finally suppose that $\omega_r < 0$. The corresponding slip is greater than one. The developed torque remains positive but the speed is negative, so that energy flows *into* the machine from the mechanical port. But since all resistors in the circuit model are positive, power must also flow *into* the machine at the electrical port! Clearly, the machine is not a generator or a motor, since it is not converting power either from electrical to mechanical or from mechanical to electrical form!

So where is all this energy going? ***Into heat!***

If we stay in this mode long, the machine will quickly overheat and suffer severe damage. And what shall we call this mode? Since the developed torque opposes rotation, we decide on the term "braking." One way to get into the mode is to be operating normally in the motor mode, and then suddenly reverse the phase sequence.[5] There will be a dramatic deceleration, since both the load and T_{DEV} oppose rotation. This is called "plugging" and results in an emergency stop. Again, be aware that this is extremely hard on the machine and should be used only when studies reveal that the machine will not be permanently damaged.

Example 5.11

Given a six-pole 60 Hz three-phase induction motor, determine the operating mode at speeds of −400 rpm, 800 rpm, and 1600 rpm.

$$\omega_s = \frac{60f}{P/2} = \frac{60(60)}{6/2} = \frac{3600}{3} = 1200 \text{ rpm}$$

$$\omega_r = -400 \text{ rpm} \quad \omega_r < 0 \qquad \text{Braking}$$

$$\omega_r = 800 \text{ rpm} \quad 0 < \omega_r < 1200 \qquad \text{Motor}$$

$$\omega_r = 1600 \text{ rpm} \quad \omega_r > 1200 \qquad \text{Generator}$$

5.4.3 Single-Phase Motors

Although three-phase motors are the most common motor in industry for loads over 5 kW, for small loads, single-phase motors are normally used. In terms of numbers, they outnumber three-phase devices by at least ten to one. Compared to three-phase machines, they are

- smaller,
- less efficient, and
- noisier.

There are many types of single-phase motors, the most common of which is the single-phase induction motor. The "pure" single-phase induction motor (i.e., one with only one stator winding) suffers from one huge disadvantage: it develops no starting torque and hence cannot start itself. To overcome this

[5] Not unlike throwing a car into reverse while cruising down the interstate.

problem, a second ("auxiliary") winding is added, creating a "split-phase" motor. The split phase family includes the following:

- The basic split-phase motor
- The capacitor-start split-phase motor
- The capacitor-start capacitor run split-phase motor

The capacitors are added to enhance starting and running torque.

Equivalent circuits for the split-phase motor do exist, and analysis is similar to that for the three-phase induction motor. However, these are significantly more complicated and beyond the scope of our study. We consider a simple example.

Example 5.12

A blower in a residential HVAC system is driven by a 1/3 hp 120 V four-pole single-phase capacitor-start motor. It runs at rated output with an efficiency of 88% and a pf of 0.85 lagging.

Find the input current.

$$P_{OUT} = 0.3333 \text{ hp} = 248.7 \text{ W}$$

$$P_{IN} = \frac{P_{OUT}}{\eta} = \frac{248.7}{0.88} = 282.6 \text{ W}$$

$$S_{IN} = \frac{P_{IN}}{pf} = \frac{282.6}{0.85} = 332.5 \text{ VA}$$

$$I = \frac{S_{IN}}{V} = \frac{332.5}{120} = 2.771 \text{ A}$$

In system load studies, it is common to model motors as constant power loads, and assume that they are operating at their nameplate ratings.

5.4.4 Gearing

Sometimes motors are connected to their loads through a gear box, or through a system of pulleys, as shown in Figure 5.8.

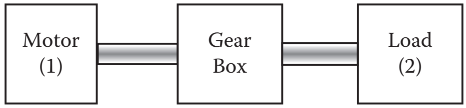


FIGURE 5.8 Driving a load through a gear box.

Neglecting gear box losses:

$$\text{Gear Ratio} = GR = \frac{\omega_1}{\omega_2} = \frac{T_2}{T_1}$$

$$P_1 = T_1 \cdot \omega_1 = T_2 \cdot \omega_2 = P_2$$

Example 5.13

A carousel ("merry-go-round") is to be driven at 4 rpm by an 1800 rpm 50 hp motor, which develops 50 hp at top speed.

a. Specify the requisite GR for the gear box.
b. Determine the torque at the motor and load shafts.

$$\text{Gear Ratio} = GR = \frac{\omega_1}{\omega_2} = \frac{1800}{4} = \frac{450}{1}$$

At the motor shaft:

$$P_1 = 50(0.746) = 37.3\ \text{kW}$$

$$T_1 = \frac{P_1}{\omega_1} = \frac{37.3}{0.1885} = 197.9\ N \cdot m$$

At the load shaft:

$$T_2 = 450(T_1) = 89\ kN \cdot m$$

A gear box may be thought of as a sort of "mechanical transformer," transforming a high-torque, low-speed load to a low-torque high-speed motor, or vice versa. For dynamic situations, the load inertia is important. Considering a pure inertial load:

$$T_2 = J_2 \cdot \frac{d\omega_2}{dt}$$

$$T_1 = \frac{T_2}{GR} = \frac{J_2}{GR} \cdot \frac{d\omega_2}{dt} = \frac{J_2}{GR} \cdot \frac{d(\omega_1/GR)}{dt}$$

$$T_1 = \left(\frac{J_2}{GR^2}\right) \cdot \frac{d\omega_1}{dt} = J_1 \cdot \frac{d\omega_1}{dt}$$

where $J_1 = \frac{J_2}{GR^2}$

Example 5.14

Continuing Example 5.13, suppose the inertia of the carousel is 101250 kg-m^2. Compute the inertia seen by the motor.

$$J_1 = \frac{J_2}{GR^2} = \frac{101250}{(450)^2} = 0.5 \text{ kg-m}^2$$

5.4.5 Speed Control

As we have learned, a motor will run in the steady state at the intersection of the motor and load torque–speed characteristics. However, there are many applications where speed control is necessary. For example, consider a hybrid electric vehicle (HEV). It must stop, back up, and run forward over a continuous range of speeds. Since motors must run at the intersection of the motor and load torque–speed characteristics, there are only two control options:

- Alter the load characteristic.
- Alter the motor characteristic.

Typically, the second option is the most practical. Note that both the synchronous and induction motors run at or near synchronous speed. Also recall that

$$\omega_s = \frac{4\pi f}{N_P}$$

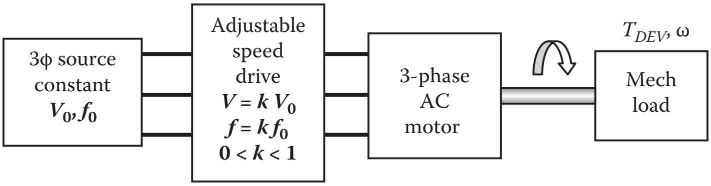


FIGURE 5.9 Ac motor–ASD systems for variable speed operation.

A fairly obvious strategy is to control the frequency. But there's a problem. Ac motors are designed to operate best if the internal magnetic fields are at or near an optimum design value. It happens that the field levels are directly proportional to voltage magnitude, and inversely proportional to voltage frequency. Hence, if we lower the frequency, we should lower the magnitude by the same percentage to maintain constant internal fields. This gives rise to the "constant volts per hertz" principle of ac motor speed control. Specifically, to run slower, decrease V and f; and, to run faster, increase V and f. This is achieved by inserting a "black box," (i.e., a motor-adjustable speed drive, or ASD), between the source and the motor, as shown in Figure 5.9.

If the input to the ASD is V_0 @ f_0, the output is kV_0 @ $k\,f_0$, where $0 < k < 1$.

We will define our drive to be ideal, and power invariant, like the ideal transformer.[6]

Example 5.15

The source in Figure 5.9 is 480 V three-phase and 60 Hz. What is the drive output for $k = 0.5$?

Solution:

Drive output ($k = 0.5$) 240 V, 30 Hz

Next we shall consider motor performance when supplied by an ASD.

5.4.5.1 Three-Phase Synchronous Motors

Recall that

$$T_{DEV} = \frac{P_{DEV}}{\omega_r} = \frac{P_{DEV}}{\omega_s} = 3\frac{E_f\ V_{an}}{X_d \cdot \omega_s}\sin(\delta)$$

[6] Practical drives have efficiencies in the mid- to upper 90s. It is possible to design a drive to operate at unity pf at all loads.

But if the motor input is

$$V_{an} = k \cdot V_0$$

$$f = k \cdot f_0$$

$$0 < k < 1$$

then

$$E_f = k \cdot E_0$$

$$\omega_s = k \cdot \omega_0$$

$$X_d = k \cdot X_0$$

and

$$T_{DEV} = 3\frac{E_f\ V_{an}}{X_d \cdot \omega_s}\sin(\delta) = 3\frac{(k \cdot E_0)(k \cdot V_0)}{(k \cdot X_0)(k \cdot \omega_0)}\sin(\delta)$$

$$T_{DEV} = 3\frac{E_0\ V_0}{X_0 \cdot \omega_0}\sin(\delta)$$

which says that the developed maximum torque is independent of k![7]

Example 5.16

A given load is driven by a 480 V four-pole 100 hp three-phase 60 Hz synchronous motor controlled by a motor drive, data for which is:

Load: $T_L = 0.01 \cdot \omega_r^2$ Nm

Motor: $E_f = 300\ V$; $X_d = 3\ \Omega$ (@ 60 Hz); 100% efficient

At half speed (900 rpm; k = 0.5) find the ASD input and output and the load torque.

[7] We have to cheat a little at very low speeds. $k = 0$ would imply $f = 0$ (which is dc) and $V = 0$, at which point the model breaks down (resistance cannot be ignored). Practically speaking, ASDs do not permit "k" to drop much below 0.1.

Solution:

ASD input: 480 V, 60 Hz
ASD output (k = 0.5): 240V, 30 Hz

Now for k = 0.5:

$$\omega_r = \omega_s = \frac{4\pi f}{N_P} = \frac{4\pi(30)}{4} = 94.25 \text{ rad/s (900 rpm)}$$

$$E_f = 0.5(300) = 150 \text{ V}$$

$$X_d = 0.5(3) = 1.50\ \Omega$$

$$V_{an} = 0.5\left(\frac{480}{\sqrt{3}}\right) = 138.6 \text{ V}$$

$$T_L = 0.01 \cdot \omega_r^2 = 88.83\ Nm$$

$$P_L = T_L \cdot \omega_r = 8.372 \text{ kW}$$

$$P_{DEV} = 3\frac{E_f\ V_{an}}{X_d}\sin(\delta) = 3\frac{150(138.6)}{1.25} \cdot \sin(\delta) = 49.88\sin(\delta) \text{ kW}$$

Computing the motor current:

$$\bar{I}_a = \frac{\bar{V}_{an} - \bar{E}_f}{jX_d} = \frac{138.6\angle 0° - 150\angle -9.66°}{j1.25}$$

$$\bar{I}_a = 21.46\angle 20.3°\ A$$

$$pf = \cos(20.3°) = 0.9379 \text{ leading}$$

Computing the power:

$$\bar{S}_{3\phi} = 3\bar{V}_{an} \cdot \bar{I}_a{}^* = 3(138.6)(21.46\angle 20.3° \text{ A})^*$$

$$\bar{S}_{3\phi} = 8.368 \text{ kW} - j3.096\ k\text{ var}$$

On the source (60 Hz) side of the ASD:

$$\bar{I}_a{}^* = \left(\frac{\bar{S}_{3\phi}}{3\bar{V}_{an}}\right)^* = \left(\frac{8.368 - j3.096}{3(277.1)}\right)^* = 10.73\angle 20.3° \text{ A}$$

5.4.5.2 Three-Phase Induction Motors Performance of the three-phase induction motor using an ASD is similar to that of the synchronous motor. Recall the Thevenin equivalent circuit model (refer to Figure 5.6).

Neglecting resistance R_T:[8]

$$\text{Approximation \#1:} \quad \bar{Z}_T = R_T + jX_T \simeq jX_T$$

$$\text{Approximation \#2:} \quad X_T \simeq Z_T; \quad R_T = 0$$

For maximum torque:

$$X_T = \frac{R'_{2x}}{s_{MT}}$$

$$\bar{I}'_A = \frac{\bar{E}_T}{X_T + jX_T}; \qquad I'_A = \frac{E_T}{X_T\sqrt{2}}$$

$$T_{\text{MAX}} = \frac{3 \cdot I'^2_A \cdot \left(\dfrac{R'_{2x}}{s}\right)}{\omega_s} = \frac{3 \cdot E_T^2 \cdot (X_T)}{2X_T^2\omega_s} = \frac{3 \cdot E_T^2}{2X_T\omega_s}$$

But if the motor input is: $V_{an} = k \cdot V_0 \quad f = k \cdot f_0$,

$$\text{then} \quad V_{an} = k \cdot V_0$$

$$E_T = k \cdot E_0$$

$$\omega_s = k \cdot \omega_0$$

$$X_T = k \cdot X_0$$

$$\text{and} \quad T_{\text{MAX}} = \frac{3 \cdot E_T^2}{2X_T\omega_s} = \frac{3 \cdot (kE_0)^2}{2(kX_0)(k\omega_0)} = \frac{3 \cdot E_0^2}{2X_0\omega_0}$$

which says that the maximum developed torque is independent of k![9]

[8] Note that there are two ways we can make the Thevenin impedance purely inductive. The author favors approximation #2 because it results in a closer match to T_{MAX} and ω_{MT} for $k = 1$.

[9] Again, not strictly true. At low speeds (and hence low X_T), R_T will no longer be negligible. However, at low speeds the load torque is usually low as well, and the degradation in T_{DEV} is not that important.

Example 5.17

Compare the maximum torque computed in Example 5.9 with that computed neglecting R_T.

From Example 5.9:

$$T_{\text{MAX}} = 1267\ N \cdot m \quad @\,1599\ \text{rpm}$$

Neglecting R_T:

$$s_{MT} = \frac{R'_{2x}}{X_T} = \frac{0.04232}{0.3784} = 0.1118$$

$$\omega_{MT} = (1 - s_{MT}) \cdot \omega_{MT} = (1 - 0.1118) \cdot 1800$$

$$\omega_{MT} = 1599\ \text{rpm} \qquad T_{\text{MAX}} = \frac{3 \cdot E_0^2}{2X_0\omega_0} = \frac{3 \cdot (258.9)^2}{2(0.3784)(188.5)} = 1404\ Nm$$

The speed for maximum torque checks exactly with Example 5.9 (1599 rpm). The maximum torque is about 10% higher (1404 versus 1269 Nm).

Example 5.18

The load of Example 5.16 is driven by a 460 V four-pole 100 hp three-phase 60 Hz induction motor controlled by a motor drive, data for which are as follows:

Load: $T_L = 0.01 \cdot \omega_r^2$ Nm

Motor @ 60 Hz: $X_1 = X'_2 = 0.20\ \Omega; \quad X_m = 8\ \Omega;$

$R_1 = 0; \quad R'_2 = 0.05\ \Omega; \quad K_{RL} = 0$

$R_T = 0; \quad X_T = 0.3951\ \Omega;$

At half speed (900 rpm; k = 0.5), find the ASD input and output and load torque.

Solution:

ASD input: 460 V, 60 Hz
ASD output (k = 0.5): 230 V, 30 Hz

By "cut and try":

$$f = 30 \text{ Hz}; \ \omega_s = 900 \text{ rpm}; \ \omega_r = 892.6 \text{ rpm}; \ s = 0.00819;$$

$$T_{DEV} = T_L = 87.4 \text{ Nm}.$$

From the circuit (motor input, ASD output @ 30 Hz):

$$\bar{V}_a = 132.8\angle 0° \text{ V} \qquad \bar{I}_a = 39\angle -58° \text{ A} \qquad \bar{I}_A' = 21.2\angle -2° \text{ A}$$

$$\bar{S}_{3\phi} = 8.24 \text{ kW} + j13.18 \ k\text{var}$$

$$T_{DEV} = 87.4 \ Nm;$$

The ASD input @ 60 Hz:

$$\bar{V}_a = 265.6\angle 0° \text{ V} \qquad \bar{I}_a = 19.5\angle -58° \text{ A}$$

$$\bar{S}_{3\phi} = 8.24 \text{ kW} + j13.18 \ k\text{var}$$

The induction motor analysis was more complicated because we had to find the speed by "cut and try" using the more complex circuit model. Results were similar to those obtained with the synchronous motor, with one exception. The pf was *leading* for the synchronous case and *lagging* for the induction case.

We have considered a "normal" situation, that is, a motor driving a load in a specific direction, which we shall define as "forward." But this is only one of four possibilities. The machine could have been running in either direction (forward or backward), and could have been converting power from electrical to mechanical form, or vice versa (motor or generator). These possibilities are referred to as "operating modes," and are shown graphically in Figure 5.10.

Ac machines can inherently operate in all four modes. However, ac motor drives must be specially designed to do so, and are designated as one-, two-, or four-quadrant drives, referring to what modes of operation they permit. Reversing direction is not too difficult for an ASD, since it only requires

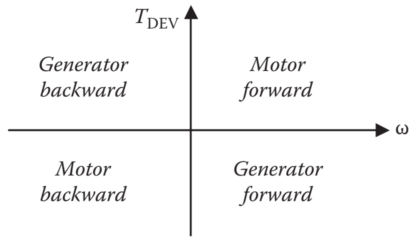


FIGURE 5.10 Ac motor-ASD operating modes.

reversing the phase sequence of the output voltage. Changing from motor to generator mode generally requires reversal of current direction, which in turn requires additional semiconductor components. The elevator application is an example of where motor-drive systems are required to operate in all four modes. We will investigate the elevator application in greater detail in Section 5.5.

5.4.6 Dynamic Performance

The foregoing analyses have been for motor-load systems running at constant speed which implies an equilibrium situation between motor and load torques:

$$T_{DEV} = T_L + T_{RL} = T_L^{/}$$

For simplicity, we've lumped the externally connected load with the rotational loss torque to form an equivalent load torque T_L'.

But rotating systems do not always run at constant speed: for example, they must be started and stopped. When this is the case, the equation of motion becomes the following:[10]

$$T_{DEV} - T_L^{/} = J \cdot \frac{d\omega}{dt}$$

If the speed is time-varying, so are the currents, voltages, powers, torques, and so on in the equivalent circuit. At first it would appear that we must abandon ac circuit analysis and use transient analysis, replacing phasors and impedance with differential equations. Rigorously speaking, this is true. However, it turns out that the relevant mechanical time constants are much slower than the relevant electrical time constants, the former being in tenths of seconds and

[10] In this section, we will use "ω" for rotor speed (rather than "ω_r") because we need the subscript position to key speed with time.

the latter in milliseconds, so we can still use ac analysis, albeit "time-varying ac."[11] Hence, the equivalent circuits and equations derived therefrom are still applicable.[12] We will restrict our dynamic study to the starting problem of a three-phase induction motor.

You might think that solving the equation of motion is no big deal. After all, it's only a first-order differential equation. Why not use Laplace transforms? However, on closer inspection we realize that the torques are both *nonlinear* functions of speed. Therefore, the equation of motion is a *nonlinear* differential equation and, as such, will require numerical methods for solution. One possible approach is to use the trapezoidal method.

Consider:

$$T_A = \text{accelerating torque} = J \cdot \frac{d\omega}{dt} = T_{DEV} - T_L'$$

$$\frac{d\omega}{dt} = \frac{T_A}{J}$$

$$J \cdot \Delta\omega \simeq T_A \cdot \Delta t$$

Discretize all variables:

$$T_i = T_A @ t = t_i$$

$$\omega_i = \omega @ t = t_i$$

$$\Delta t_i = t_{i+1} + t_i$$

$$\Delta\omega_i = \omega_{i+1} + \omega_i$$

$$\omega_0 = 0 \quad t_0 = 0$$

We can either fix $\Delta\omega$ and compute Δt or vice versa. We shall initially fix $\Delta\omega$. For example:

$$\Delta\omega = \frac{\omega_s}{100}$$

[11] Somewhat of an oxymoron, like a "three-wheel bicycle." This means that within a computation time step, voltages and currents may be represented as phasors with fixed magnitudes and phases; in the next step, the phasor magnitudes and phases will be updated.

[12] There are high-performance, low-inertia situations (like robotics) where this approach is not valid. For these applications, we require true transient motor models.

The average accelerating torque in the interval ω_i to ω_{i+1} is

$$\frac{T_{i+1} + T_i}{2}$$

The time to accelerate from ω_i to ω_{i+1} is

$$\Delta t_i = \frac{2J}{(T_{i+1} + T_i)} \Delta\omega$$

This works well at first. But as we approach steady-state speed, T_A approaches zero. For example, suppose at i = m, $T_A < 0.10T_{RATED}$. The time elapsed is

$$\sum_{i=1}^{m} \Delta t_i = t_m$$

So now change the strategy. Set Δt to some fixed value and compute $\Delta\omega$:

$$\Delta\omega_i = \frac{(T_{i+1} + T_i)}{2J} \Delta t$$

One possibility: keep Δt at its computed value in the mth step.

So when do we stop computing? When T_A equals zero. But this could take forever! The speed could asymptotically approach its final value! So realistically we need to stop when the system acceleration is near zero (e.g., stop when $T_A < 0.001T_{\text{RATED}}$). Suppose this happens after k steps. The total time to start is

$$\sum_{i=1}^{m} \Delta t_i + k \cdot \Delta t$$

Applying this approach to an example, consider the machine of Table 5.2 connected to the load used in Section 5.3.2, namely:

$$T_L = 0.01 \cdot \omega_r^2 \text{ Nm}$$

$$T_{RL} = 0.063 \cdot \omega_r \text{ Nm}$$

$$T_L' = T_L + T_{RL} = 0.01 \cdot \omega_r^2 + 0.063 \cdot \omega_r$$

The machine and load inertias are

$$J = J_M + J_L = 0.9299 + 4.6495 = 5.5794 \text{ kg-m}^2$$

Table 5.4 A Dynamic Analysis: Motor Starting

```
Load Torque (Nm) = TL = A0 + A1*WL + ... + AN*WL**N
  WL = LOAD shaft speed in rad/s; Order = 2
  A0 = 0.000; A1 = 0.000000;      A2 = 0.01
  Gear Ratio = GR = Load/Motor Speed = 1.0000; J's in
   kg-m sq
  MOTOR Inertia (JM) = 0.9299;   LOAD Inertia (JL) =
   4.6495
     JT = JM + sqr(GR) × JL = 0.9299 + 4.6495 = 5.5794
```

Time (s)	Speed (rpm)	Current (A)	Tdev	Tm
0.00	0	689.3	304	0
0.33	180	687.5	336	4
0.64	361	685.0	376	14
0.92	540	681.7	425	32
1.18	721	677.0	490	57
1.41	900	669.8	575	89
1.62	1082	657.8	694	128
1.79	1262	635.5	865	175
1.93	1442	585.7	1105	228
2.04	1623	441.1	1264	289
--				
2.33	1773	89.8	362	344.7

Implementing the method by computer, we get the values in Table 5.4. A practical application that pulls together the topics previously discussed is an elevator, which is the subject of the next section.

5.5 An Example Application: The Elevator

Consider a fifteen-passenger elevator that is required to serve a ten-floor building. Data are provided in Figure 5.11. Our objective is to select an appropriate motor for this application.

Our first priority is to establish a coordinate system. We define x to be the vertical position of the elevator relative to the building, with $x = 0$ corresponding to the elevator located on the first floor. "Up" shall be positive, and will correspond to forward rotation of the motor.

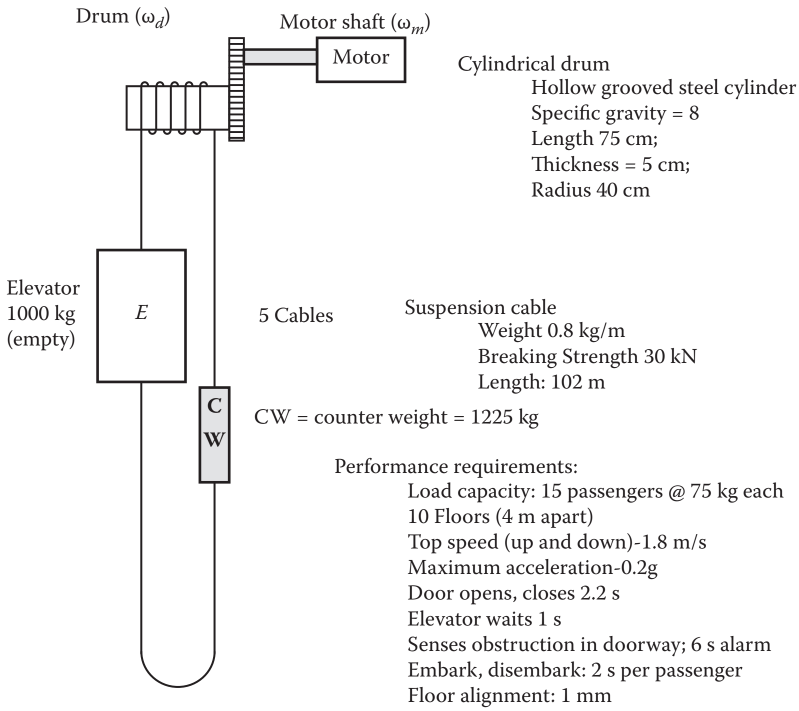


FIGURE 5.11 **An Elevator Application.**

To simplify matters, assume that all passengers weigh 75 kg. We decide to size the counterweight so that the system is balanced with three passengers:

On the elevator:

$$M_{CW} = M_E + 3(75) = 1000 + 225 = 1225\ kg$$

which means that the system is in neither the motor nor the generator mode when three passengers are on the elevator. Observe that if more than three are onboard, the machine must do work on the load to raise the elevator, and hence we are in the motor mode (energy flows from electrical to mechanical). When less than three are onboard and going up, the load must do work on the machine to raise the elevator, and hence we are in the generator mode (energy flows from mechanical to electrical).

The following terms are defined:

MF = motor forward (going up, more than three passengers)
GF = generator forward (going up, less than three passengers)

MB = motor backward (going down, less than three passengers)
GB = generator backward (going down, more than three passengers)
BF = balanced forward (going up, three passengers)
BB = balanced backward (going down, three passengers)

"Motor" mode means that energy is flowing from the electric power supply into the system; "generator" mode means that energy is flowing from the system into the electric power supply.

We will consider the important and practical problem of selecting the machine and drive ratings (either the synchronous or induction machine will work). We first consider steady-state speed requirements. What is the worst case (maximum load)?

Maximum load is fifteen passengers. The MF mode is slightly worse than the GB mode because in the MF mode, the electrical system has to supply the load plus the motion losses, whereas in the GB, the motion losses subtract from the mechanical input.

For the MF case:

$$M_{\mathrm{LOAD}} = M_E + 15(75) - 1225 = 1000 + 225 = 900 \text{ kg}$$

$$F_{\mathrm{LOAD}} = M_{\mathrm{LOAD}} \cdot g = 900(9.807) = 8826\ N$$

$$P_{\mathrm{LOAD}} = F_{\mathrm{LOAD}} \cdot v = 8826(2) = 17.65 \text{ kW} = 23.67 \text{ hp}$$

Rounding up to the nearest standard size, a 25 hp machine would be adequate to handle the maximum steady-state load.

Now consider the speed. The maximum elevator speed is 1.8 m/s. The corresponding drum angular velocity is

$$\omega_d = \frac{v_e}{r_d} = \frac{v_e}{0.4} = 2.5 v_e = 4.5 \text{ rad/s}$$

Since the drum gear has four hundred teeth, and the motor shaft gear ten teeth, the gear ratio is

$$GR = \frac{\text{DrumTeeth}}{\text{GearTeeth}} = \frac{400}{10} = \frac{40}{1}$$

The motor shaft turns through forty revolutions for each drum rotation.

$$\omega_r = 40\omega_d = 180 \text{ rad/s} = 1719 \text{ rpm}$$

Hence a four-pole 60 Hz ac machine would be quite satisfactory.

Now consider the dynamics. We have to accelerate mass from zero to 1.8 m/s at 0.2 g, which will require substantially more torque, and therefore more motor current. There are two kinds of motion to consider:

- Translational, which includes the elevator, passengers, counterweight, and cable[13]
- Rotational, which includes the drum and the motor rotor

First, the translational elements:

$$M_{\text{TRAN}} = M_E + M_P + M_{CW} + M_{\text{CABLE}}$$

$$M_{\text{TRAN}} = 1000 + 15(75) + 1225 + 0.8(102)(5)$$

$$M_{\text{TRAN}} = 1000 + 1125 + 1225 + 408 = 3758 \text{ kg}$$

As we convert from translational to rotational motion, the rotational equivalent to mass is the mass polar moment of inertia (J). The mass M_{TRAN} may be considered a point mass located at the drum radius rotating about the drum axis. Therefore:

$$J_{\text{TRAN}} = M_{\text{TRAN}} \cdot r^2 \cdot = 601.3 \text{ kg} \cdot \text{m}^2$$

For the drum:

$$M_{\text{DRUM}} = \pi\rho d\left(r_{\text{outside}}^2 - r_{\text{inside}}^2\right)$$

$$M_{\text{DRUM}} = \pi(8000)(0.75)[(0.4)^2 - (0.35)^2] = 707 \text{ kg}$$

The drum is an annular cylinder:

$$J_{\text{DRUM}} = \frac{1}{2} M_{\text{DRUM}}\left(r_{\text{outside}}^2 + r_{\text{inside}}^2\right)$$

$$= 0.5(707)[(0.40)^2 + (0.35)^2]$$

$$= 99.85 \text{ kg-m}^2$$

The total load inertia is therefore:

$$J_L = J_{\text{DRUM}} + J_{\text{TRAN}}$$

$$= 601.3 + 99.85 = 701.2 \text{ kg} \cdot \text{m}^2$$

[13] Technically, the cable should be divided into two parts: that which is wrapped around the drum (rotational), and that which is hanging (translational). However, as long as the total cable mass is included, it doesn't matter since the cable moves as a rigid body.

Reflecting the inertia to the motor shaft:

$$J_m = \left(\frac{1}{GR^2}\right) \cdot J_L$$
$$= \left(\frac{1}{40^2}\right) \cdot 701.2 = 0.4382 \text{ kg} \cdot \text{m}^2$$

Estimating the inertia[14] of a 25 hp motor at 0.18 kg-m^2, the total system "J" on the motor shaft is about

$$J = 0.44 + 0.18 = 0.62 \text{ kg} \cdot \text{m}^2$$

Now, the maximum acceleration on the elevator is specified to be 0.2 g = 0.2 (9.807) = 1.9614 m/s^2, which translates to a drum, motor angular accelerations of

$$\alpha_{\text{DRUM}} = \left(\frac{a_{\text{DRUM}}}{r}\right) = \frac{1.9614}{0.4} = 4.9035 \text{ rad/s}^2$$

$$\alpha_{\text{MOTOR}} = 40 \cdot \alpha_{\text{DRUM}} = 196.14 \text{ rad/s}^2$$

Converting to torque:

$$T_{\text{ACCEL}} = J \cdot \alpha_{\text{MOTOR}}$$
$$= (0.62) \cdot (196.14) = 121.6 \text{ N} \cdot \text{m}$$

Adding the full-load steady-state torque:

$$T = T_{\text{ACCEL}} + T_{FL}$$
$$= 121.6 + \left(\frac{17.65}{0.180}\right) = 121.6 + 98.1 = 219.7 \text{ N} \cdot \text{m}$$

In many applications, this might not be a major problem, since we don't stop and start that often, and starting would be a fairly benign transient. But an elevator is a very demanding application, in that we are constantly starting and stopping! So our motor would be in a more or less constant state of overload. Hence, we need a bigger motor. But how big?

[14] Actually, we should contact the manufacturer to obtain accurate J values, as well as all data, for whatever motor we select.

In a motor for a given speed rating, rated power is proportional to rated torque. So, since 98 Nm was rated for a 25 hp motor, (220/98)25 = 56 hp would be a reasonable estimate for motor size. The next larger standard size is 75 hp, which has a estimated J of 0.54 kg-m^2. So let's recalculate the accelerating torque for a 75 hp motor.

$$J = 0.44 + 0.54 = 0.98 \text{ kg} \cdot \text{m}^2$$

$$\begin{aligned} T_{\text{ACCEL}} &= J \cdot \alpha_{\text{MOTOR}} \\ &= (0.98) \cdot (196.14) = 192.2 \text{ N} \cdot \text{m} \end{aligned}$$

$$\begin{aligned} T &= T_{\text{ACCEL}} + T_{FL} \\ &= 192.2 + 98 = 290.2 \text{ N} \cdot \text{m} \end{aligned}$$

Hence, we estimate the following:

$$HP_{\text{RATED}} = \left(\frac{290.2}{98} \right) 25 = 74.04 \text{ hp}$$

We choose to select a 75 hp motor, which should be more than adequate to handle the worst-case steady-state and dynamic performance requirements.

The elevator application clearly requires a four-quadrant drive. Drives are capable of ramping up the applied voltage, where the rate of rise is user defined. For this case, the maximum acceleration is specified at 0.2 g = 1.96 m/s^2. Thus, the time to reach top speed (1.8 m/s) is

$$T = \frac{v}{a} = \frac{1.8}{1.96} = 0.92 \text{ s}$$

Assuming the facility has 480 V three-phase 60 Hz power available, the drive input ratings would be correspondingly 480 V three-phase and 60 Hz (motor mode). If the motor rating is 460 V 60 Hz four-pole 75 hp, the drive output would be 460k V 60k Hz three-phase $0 < k < 1$. The rate setting would be 460/.92 = 500 V/s.

5.6 An Example Application: High-Speed Rail (HSR)

Most industrialized countries have a fully developed rail industry as part of their transportation system. As the pressures of increasing population mount, rail transportation is likely to become more important worldwide, with

FIGURE 5.12 **The Japanese MLX01 MagLev Experimental Train.**

particular emphasis on high-speed rail (HSR). HSR may be defined as public transport by rail in speeds at excess of 200 km/h (124 mph). A lower figure is sometimes used in the United States, and can be as low as 90 mph. HSR offers significant advantages in speed and efficiency. Research continues in Japan, Germany, France, China, and the United States. Figure 5.12 shows the Japanese MLX01 MagLev Experimental Train which recently set the HSR world's speed record in excess of 500 km/hr.

HSR is most economical for high traffic volume and medium distances (i.e., for trips from one to three hours or about 150–600 km). HSR beats both air and car transportation in this range, with faster check-in, security screening, luggage handling, embarking, and disembarking.

The drive motors are electromagnetic linear induction motors (LIMs) or linear synchronous motors (LSMs), which are similar to their rotational counterparts, except the stator is "cut and rolled out" along the track. The "rotor" becomes the "translator," since its motion is translational. The "rotating" stator magnetic field becomes the "translating" magnetic field. In the rotating machine, opposite stator coil sides are 180 electrical degrees apart, which is the angular distance the field travels in one half-cycle.

Correspondingly, in the linear machine, the lineal distance between stator coil sides (i.e., the coil pitch) is the lineal distance the field travels in one half-cycle. The translator can be on either structure: the stationary track or the moving train car.

Example 5.19

A linear motor for an HSR application is to be designed. If top speed is to be 500 km/hr @ 60 Hz, what should be the stator coil pitch (cp)?

$$500 \text{ km/hr} = \frac{500 \times 10^3}{3.6 \times 10^3} = 138.9 \text{ m/s}$$

$$1/2 \text{ cycle} \rightarrow 1/120 \text{ s}$$

$$cp = (138.9)\left(\frac{1}{120}\right) = 1.157 \text{ m}$$

Instead of producing torque (T_{DEV}), thrust (F_{DEV}) is produced. The HSR dynamic equation of motion is as follows:[15]

$$F_{dev} - (F_{AD} + F_G) = M\frac{dv}{dt}$$

where

F_{dev} = EM developed force (thrust), N

$F_{AD} = k_{ad}v^2$ = Aerodynamic Drag, N

v = velocity, m/s

$k_{ad} = \dfrac{c_d \rho_d A}{2}$ = Drag constant, kg-m/s^2 $\quad$ A = Cross-sectional area, m^2

c_d = Drag coefficient, $0.26 < c_d < 0.30$ for HSR vehicles

(can be considerably larger for multicar trains)

ρ_d = Air density, 1.3 kg/m^3 @ 0 deg C; 1 atm.

$F_G = MgG_R$ = Gravitational gradient force, N

G_R = Grade, %

$M = M_V + M_L$ = vehicle + payload mass, kg

[15] Neglecting low-speed suspension drag.

Example 5.20

Given an HSR vehicle of the following design:

Cross-section height × width = 4.0 × 3.8 m
Length = 15 m
Empty vehicle weight = 10 t = 10,000 kg
Payload = 40 passengers + luggage = 40 × (75 + 75) = 6000
Drag coefficient = 0.40; air density = 1.2 @ 23°C

For a six-car train:
Performance requirements:

Top speed on level surface = 500 km/hr
Top speed climbing 6% grade = 450 km/hr

Neglect suspension drag at high speed. Determine the following:

a. The thrust and power required at top speed on a level surface
b. The thrust and power required at top speed on grade
c. If the LIM is three-phase 12.47 kV pf = 0.85 lagging 94% efficient, compute the requisite rated current.

Solution:

$$k_{ad} = \frac{c_d \rho_d A}{2} = \frac{0.40(1.2)\,(4.0)(3.8)}{2} = 3.648 \text{ N-s}^2/\text{m}^2$$

Converting km/hr to m/s:

$$\frac{1000(500 \text{ km/hr})}{3600} = 138.9 \text{ m/s} \qquad \frac{1000(450 \text{ km/hr})}{3600} = 125 \text{ m/s}$$

a. Aerodynamic Drag @ 500 km/hr $= k_{ad} v^2 = 3.648(138.9)^2 = 70.38$ kN

Power $= F_{AD} \cdot v = 70.38\ (138.9) = 9.776$ MW

b. Aerodynamic Drag @ 450 km/hr $= k_{ad} v^2 = 3.648(125)^2 = 57.00$ kN

Gravitational gradient force $= F_G = MgG_R = 16000(9.802)(0.06)$

$= 9.410$ kN

Power $= (F_{AD} + F_G) \cdot v = (66.41)\ (125) = 8.301$ MW

c. Rated current:

$$S_{3\phi} = \frac{P_{3\phi}}{\eta \cdot pf} = \frac{9.776}{0.94(0.85)} = 12.24 \text{ MVA}$$

$$I_L = \frac{S_{3\phi}}{\sqrt{3} \cdot V_L} = \frac{12.24}{\sqrt{3} \cdot (12.47)} = 566.5 \text{ A}$$

Should HSR become established, it would constitute a major component of the national electric power system grid and introduce a new set of technical operating and design problems.

5.7 An Example Application: The Hybrid Electric Vehicle (HEV)

The hybrid electric vehicle (HEV) is generally understood to be a vehicle propelled by a conventional gasoline or diesel engine and one or more electric motor(s).[16] All of the major car manufacturers either offer, or plan to offer, HEVs. The main advantages of HEVs include higher gas mileage and longer warranties. Government incentives such as tax credits and access to high-occupancy vehicle (HOV) lanes further encourage the use of HEVs.

The major disadvantage of the HEV is its higher initial cost. The handling experience is somewhat different, which may require some adjustment by drivers. Also HEVs are much quieter, and thus may introduce a safety hazard to pedestrians and bicyclists.

Assume the following HEV data to expedite our discussion:

- Net weight 800 kg; cargo capacity: four passengers plus luggage, or 400 kg
- Drag coefficient 0.25
- Full auxiliary power (lights, ac, etc.): 4 kW
- Regenerative braking employed
- Dc power bus voltage = 255 v
- Dc auxiliary bus voltage = 50 v

[16] There are various design possibilities. We shall assume one electric motor, applying torque to the drive shaft along with the internal combustion engine.

- Wheel + tire radius = 32 cm
- Aerodynamic drag coefficient = 0.22
- Aerodynamic cross-section = 2 m^2
- CVT transmission: 3.00 > gr > 0.70; gr = motor shaft speed/drive shaft speed (3 @ low speed; 0.7 @ high speed)

Performance criteria are as follows:

- Top speed 120 km/hr on 6% grade; 100% auxiliary power, and battery charging @ 5 A
- Cruising: 100 km/hr; 100% auxiliary power; battery charging @ 5 A
- Cruising range: 600 km
- Cruising range on battery power alone; 2% grade; 50% aux power; for 75 km
- Acceleration: 0–50 km/hr in 3 s
- Must climb 25% grade
- Range 600 km @ 100 km/hr on 0% grade; 100% auxiliary power, and battery charging @ 5 A

We shall investigate sizing the electric motor and battery capacity. The basic propulsion equation is

$$M\frac{dv}{dt} = F_T - F_D - F_{RR} - F_G$$

where

v = velocity of vehicle relative to road surface, m/s

F_T = propelling force (Thrust), N

F_D = aerodynamic drag force, N

F_{RR} = tire rolling resistance force, N

F_G = component of gravitational force (on grade) = $M \cdot g \cdot \sin(\theta)$, N

θ = angle of grade (0 = level)

M is the total mass of the vehicle, plus cargo in kg. Therefore:

$$M = 800 + 400 = 1200 \text{ kg}$$

Rolling resistance:

$$F_{RR} = K_{RR} \cdot M \cdot g$$
$$= (0.01)(1200)(9.807) = 117.7 \quad N$$

The drag force is

$$F_D = \frac{1}{2}\rho \cdot A \cdot C_D \cdot v^2$$

where

v = velocity of vehicle relative to road surface, m/s

ρ = density of air = 1.3 kg/m^3

C_D = aerodynamic drag coefficient = 0.25

A = cross-sectional area of vehicle, m^2 = 2.0

$$F_D = \frac{1}{2}(1.3) \cdot (2) \cdot (0.25) \cdot v^2 = 0.325 \cdot v^2$$

We must size the electric motor to cruise with 50% auxiliary power, up a 2% grade. Drag @ 100 km/hr (27.78 m/s):

$$F_D = 0.325 \cdot (27.78)^2 = 250.8\ N$$

Gravitational force @ 2% grade:

$$F_G = Mg \cdot \sin(\theta)$$
$$F_G = (1200)(9.807)(0.02) = 235.4\ N$$

Requisite thrust:

$$F_T = F_D + F_{RR} + F_G$$
$$F_T = 250.8 + 117.7 + 235.4 = 603.9 \text{ Nm}$$
$$P_T = F_T \cdot v = \frac{603.9(27.78)}{1000} = 16.78 \text{ kW}$$

Estimating the transmission efficiency to be 95%:

$$P_M = \frac{P_M}{0.95} = \frac{16.78}{0.95} = 17.66 \text{ kW}$$

Adding 50% auxiliary power and battery charging:

$$P = P_M + P_{AUX} + P_{BAT}$$

$$P = 17.66 + 4 + 1.28 = 22.94 \text{ kW or } 30.7 \text{ hp}$$

Rounding up, 35 hp should be adequate to satisfy the all-electric cruising requirement.

Consider the top-speed requirement: 120 km/hr on 6% grade. Drag @ 120 km/hr (33.33 m/s):

$$F_D = 0.325 \cdot (33.33)^2 = 361.1\ N$$

Gravitational force @ 6% grade:

$$F_G = Mg \cdot \sin(\theta)$$

$$F_G = (1200)(9.807)(0.06) = 706.1\ N$$

Requisite thrust:

$$F_T = F_D + F_{RR} + F_G$$

$$F_T = 361.1 + 117.7 + 706.1 = 1184.9 \text{ Nm}$$

$$P_T = F_T \cdot v = \frac{1184.9(33.33)}{1000} = 39.49 \text{ kW}$$

Estimating the transmission efficiency to be 95%:

$$P_M = \frac{P_M}{0.95} = \frac{39.49}{0.95} = 41.57 \text{ kW}$$

Adding 100% auxiliary power and battery charging:

$$P = P_M + P_{AUX} + P_{BAT} = 41.57 + 4 + 1.28 = 46.85 \text{ kW or } 62.79 \text{ hp}$$

Hence the minimum size of the gasoline engine would be 63 − 35 = 28 hp (or 63 hp operating on engine power alone) just to operate at top speed. However, to meet acceleration requirements, the engine must be much larger.

Consider the acceleration specification: 0–50 km/hr in 3 s on level surface.

$$a = \frac{dv}{dt} = \frac{(50000/3600) - 0}{3} = 4.63 \text{ m/s}^2$$

By Newton's Second Law:

$$F = ma = (1200)(4.63) = 5556\ N$$

Including drag @ 50 km/hr (13.89 m/s) and rolling resistance:

$$F = 5556 + 62.7 + 117.7 = 5736\ N$$

Convert to torque on the drive shaft:

$$T = F \cdot r = (5736)(0.32) = 1836\ Nm$$

In a motor, torque is proportional to current (and thrust is proportional to torque). We have established that a thrust of 604 N corresponded to rated motor power (and hence rated current). If we allowed 200% of rated current under maximum acceleration, the motor could contribute about 1200 N of thrust to the total required amount of 5736, or about 20%. This would suggest that the engine be roughly five times the rating of the motor, or about $5 \times 35 = 175$ hp. The problem is actually more complicated, and a more thorough analysis is required. Still, this gives a rough estimate of the ratings of the drive power components, and checks with commercial engineering designs by Ford, GM, Honda, Hyundai, and Toyota.

Finally, consider the battery size. The specification is that the HEV must have a range of 75 km on battery power alone cruising at 50% auxiliary power.

Requisite thrust:

$$F_T = F_D + F_{RR} + F_G$$

$$F_T = 250.8 + 117.7 + 0 = 368.5 \text{ Nm}$$

$$P_T = F_T \cdot v = \frac{368.5(27.78)}{1000} = 10.24 \text{ kW}$$

Estimating the transmission efficiency to be 95%:

$$P_M = \frac{P_M}{0.95} = \frac{10.24}{0.95} = 10.77 \text{ kW}$$

Adding 50% auxiliary power:

$$P = P_M + P_{AUX}$$

$$P = 10.77 + 2 = 12.77 \text{ kW}$$

At cruising speed (100 km), it should take 0.75 hr to go 75 km. Therefore, the battery capacity should be 0.75(12.77) = 9.58 kW-hr (round up to 10 kW-hr). At an energy density of 200 W-hr/kg, such a battery would weigh 50 kg.

The foregoing analysis provides illustrations of some of the interactions between the electrical and mechanical aspects of the system.

5.8 Summary

Electrical loads are incredibly diverse in type and size, ranging from milliwatts to megawatts. Still, loads can be divided into two broad categories:

- Static (those which have no moving parts)
- Rotating (i.e., motors)

To consider the impact on power systems, loads are normally modeled in one of two ways:

- Constant impedance
- Constant power

We discussed several major static load types, including electric lighting and heating. Three types of motors were discussed, including three-phase synchronous, three-phase induction, and single-phase devices. Also, three important applications (elevators, HSR, and HEVs) were considered.

As we noted in Chapter 1, there are two broad applications of electrical phenomena: energy and information processing. We have completed our study of energy processing. In Chapter 6, we begin our study of information processing via electrical devices, an area frequently called "electronics."

References

American Public Transportation Association. *High-Speed Rail Investment Background Data Report*. February 1. Washington, DC: American Public Transportation Association, 2011.

Barney, G. C. *Elevator Traffic Handbook: Theory and Practice*. London: Spon Press, 2003.

Beaty, H. W., and J. L. Kirtley. *Electric Motor Handbook*. New York: McGraw-Hill, 1998.

Ehsani, M., Y. Gao, S. E. Gay, and A. Emadi. *Modern Electric, Hybrid Electric, and Fuel Cell Vehicles; Fundamentals: Theory and Design*. Boca Raton, FL: CRC Press, 2005.

Erikson, C. J. *Handbook of Electric Heating for Industry*. Piscataway, NJ: IEEE, 1994.

Fink, D. G., and H. W. Beaty, eds. *Standard Handbook for Electrical Engineers*, 14th ed. New York: McGraw-Hill, 1999.
Fitzgerald, A. E., C. Kingsley Jr., and S. D. Umans. *Electric Machinery*, 6th ed. New York: McGraw-Hill, 2003.
Francfort, J. *Advanced Vehicle Testing Activity: Hybrid Electric Vehicle Testing*. Washington, DC: U.S. Department of Energy, FreedomCAR and Vehicle Technologies, Electric Drive Transportation Association Conference, Vancouver, Canada, December 2005.
General Specification for Lift, Escalator and Passenger Conveyor Installation in Government Buildings of the Hong Kong Special Administrative Region 2002 Edition. Hong Kong: Hong Kong Special Administrative Region, 2002.
Gieras, J. F. *Linear Induction Drives*. Oxford: Clarendon Press, 1994.
Gieras, J. F., and Z. J. Piech. *Linear Synchronous Motors: Transportation and Automation Systems*. Boca Raton, FL: CRC Press, 1999.
Gross, C. A. *Electric Machines*. Boca Raton, FL: CRC Press, 2007.
Gross, C. A. *Power System Analysis*, 2nd ed. New York: Wiley, 1986.
Laithwaite, E. R. *A History of Linear Electric Motors*. London: Macmillan, 1987.
Miles, L. *Heat Pumps*: *Theory and Service*. Clifton Park, NY: Delmar Cengage, 1993.
National Electrical Manufacturers Association. *Motors and Generators* (Publication No. MG1-1972). New York: NEMA, 1972.
Orfeuil, M. *Electric Process Heating*. Columbus, OH: Battelle Press, 1987.
Sauer, H. J. *Heat Pump Systems*. Malabar, FL: Krieger, 1990.
U.S. Public Interest Research Groups. *Study a Track Record of Success: High Speed Rail around the World and its Promise for America*, Fall. Washington, DC: US PIRG, 2010.
Unnewehr, L. E., and S. A. Nasar. *Electric Vehicle Technology*. New York: Wiley, 1982.
Zorbas, D. *Electric Machines: Principles, Applications, and Control Schematics,* St. Paul, MN: West, 1989.

Problems

5.1. Given a 200 kVA 7.2 kV 0.85 pf lagging load:
   a. Compute the corresponding complex power.
   b. Compute the corresponding complex impedance.

5.2. For the load of Problem 5.1, the load voltage increases to 7.92 kV. Compute the current if the load is modeled as
   a. constant power.
   b. constant impedance.

5.3. Given a 600 kVA 12.47 kV 0.85 pf lagging three-phase load:
   a. Compute the corresponding complex power.
   b. Compute the corresponding wye and delta complex impedances.

5.4. Consider two heating units, A and B. Ratings are as follows:

Unit A: single-phase 480 V 3 kW
Unit B: three-phase 480 V 5 kW

a. Compute the current for Unit A.
b. Compute the current for Unit B.
c. Twelve Unit A's and five Unit B's are served from a three-phase 480 V circuit.
    1. Should the Unit A's be connected in wye or delta?
    2. What is the total circuit current?

5.5.

a. Design the lighting for a classroom of 16' × 30' with a 9-foot ceiling using Lithonia LB440A fixtures (see Section 5.4). The work surface is assumed to be 30" (desktop) above the floor. The CU = 0.65, and the LDF = 0.72. The recommended light level for classrooms is 150 fc.
b. The lighting system is to be served from a three-phase 208/120 V circuit. Provide details.

5.6. Given a three-phase six-pole 60 Hz synchronous motor:

a. Determine synchronous speed in rpm and rad(s).
b. Repeat (a) for a three-phase six-pole 60 Hz induction motor.

5.7. A given load is driven by a 480 V four-pole 100 hp three-phase synchronous motor, data for which are as follows:

$$\text{Load:} \quad T_L = 0.008 \cdot \omega_r^2 \text{ Nm}$$

$$\text{Motor:} \quad E_f = 400 \text{ V}; \quad X_d = 2.5\ \Omega$$

Find the load torque and speed and motor current and pf.

5.8. A given load is driven by a 460 V four-pole 60 Hz 100 hp three-phase induction motor, data for which are as follows:

$$R_1 = R_2' = 0.04\ \Omega; \quad X_1 = X_2' = 0.20\ \Omega$$

$$X_m = \text{large}; \quad T_{RL} = 0.07 \cdot \omega_r \quad N \cdot m$$

For a speed of 1775 rpm, perform a comprehensive analysis (find the slip, currents, powers, torques, efficiency, and pf).

5.9. The motor of Problem 5.8 drives the following load:

$$T_L = 0.008 \cdot \omega_r^2 \quad N \cdot m$$

a. Find the Thevenin equivalent circuit.
b. Find the motor-load speed (by trial and error).

5.10. A residential clothes washer is driven by a 1/4 hp 120 V four-pole single-phase capacitor-start motor. It runs at rated output with an efficiency of 87% and a pf of 0.86 lagging. Find the input current.

5.11. A given load requires 950 Nm @ 360 rpm. It is to be driven by a 480 V four-pole three-phase 60 Hz synchronous motor through a gear box. Size the motor, and specify the gear ratio.

5.12. A given load is driven by a 480 V four-pole 100 hp three-phase 60 Hz synchronous motor controlled by a motor drive, data for which are as follows:

Load: $T_L = 0.008 \cdot \omega_r^2$ Nm

Motor: $E_f = 400\,V; \quad X_d = 2.5\,\Omega$

At half-speed (900 rpm), find the drive input, output, load torque, frequency, motor current, and pf.

5.13. Consider the elevator application discussed in Section 5.4. Determine the operating mode (MF, MB, GF, and GB) if there are
a. ten passengers, going up.
b. ten passengers, going down.
c. one passenger, going up.
d. one passenger, going down.

5.14. Consider the elevator application discussed in Section 5.4. Assume the machine selected is 90% efficient in all modes.
a. There are ten passengers, going up, at 2 m/s. Determine the power output of and input to the machine (i.e., the motor). Identify the machine output and input as either electrical or mechanical. Is the machine a motor or a generator?
b. There are ten passengers, going down, at 2 m/s. Determine the power output of and input to the machine (i.e., the generator).

Identify the machine output and input as either electrical or mechanical. Is the machine a motor or a generator?

5.15. Consider the elevator application discussed in Section 5.4. Assume we select a 460 V 75 hp 60 Hz four-pole synchronous machine with the following data:

$$E_f = 300 \text{ V} \quad X_d = 2.3\ \Omega \quad @\ f = 60 \text{ Hz } (k = 1)$$

Show that this machine can easily meet the worst-case torque requirement.

5.16. Given a HSR car of the following design:

Cross-section height × width = 4.0 × 3.8 m
Length = 15 m
Empty vehicle weight = 10 t = 10000 kg
Payload = 40 passengers + luggage = 40 × (75 + 75) = 6000 kg
Drag coefficient = 0.4 (for a six-car train)
Air density = 1.2 @ 23°C

Consider a train of six such vehicles. Compute the drag power loss at

a. 100 km/hr.
b. 250 km/hr.
c. 500 km/hr.

5.17. Given the HSR system of Problem 5.16, a 12.47 kV three-phase LIM supplies the thrust and operates at 93% efficiency, and pf = 0.87 on a level surface at 500 km/hr.

5.18. Given the HEV system of Section 5.7:

a. Find the thrust and power required to propel the vehicle at 55 mph on a level surface.
b. Find the thrust and power required to propel the vehicle at 55 mph up a 3% grade.
c. Find the thrust and power required to propel the vehicle at 55 mph down a 3% grade.

5.19. Continuing Problem 5.18:

a. Find the thrust required to accelerate the vehicle to 30 mph in 3 seconds on a level surface. Approximate the drag as constant at its 20 mph value.

b. Suppose the battery size is 100 kg, with an energy density of 200 Whr/kg. What is the HEV range on battery power alone @ 55 mph on a level surface? Assume a transmission efficiency of 95% and auxiliary power of 2.4 kW.

This page intentionally left blank

Chapter 6

Semiconductor Devices

This chapter first introduces the basics of semiconductor physics. Following this, the most common semiconductor devices are introduced, including diodes, transistors, and integrated circuits. Applications of these devices are discussed. The chapter concludes with a discussion of how to read a typical data sheet.

6.1 Semiconductor Fundamentals

Before discussing specific types of devices, it is essential to know the basic concepts of modern semiconductor materials. The study of semiconductors is a branch of solid-state physics. Much of the understanding we have of how useful devices, such as transistors, are fabricated and how they work has been gleaned from the work of twentieth-century theoretical and experimental physicists.

6.1.1 Bandgap, Temperature, and Free Electrons

As far as electrical properties are concerned, there are three types of materials in the world: insulators, conductors, and semiconductors. Insulators are materials such as glass and rubber which have very high resistance to the flow of current. Conductors, typified by most metals, have very low resistance. Semiconductors are materials whose natural resistance is neither very high nor very low. Another characteristic of semiconductors is that their resistance is very sensitive to added impurities and to temperature. The most common semiconductor is the element silicon in crystalline form. Other common semiconductor materials are alloys such as gallium arsenide, indium phosphide, and the tertiary alloy mercury cadmium telluride (popularly nicknamed mercad).

Crystalline semiconductors have an important property called the bandgap energy, sometimes just called the bandgap. The bandgap energy is the amount of energy required to break a covalent bond within the crystal. This determines a number of important material properties, including the resistivity at room temperature and how much the resistivity changes with changes in temperature. The higher the temperature, the more the atoms vibrate, and the more bonds will be broken. For each broken bond, one electron is set free and is available to carry current. Therefore, the more broken bonds, the lower the resistance.

At absolute zero temperature, there is no vibration, so there will be no broken bonds, and the semiconductor will act like a perfect insulator. However, at room temperature or any normal temperature of operation, there will be many broken bonds, and thus many free electrons.

6.1.2 Energy Bands

Electrons that are free (not involved in crystal bonds) are said to be in the conduction band. The conduction band is a range of energies that extends from the top of the energy bandgap to the vacuum level, where they would be free from the crystal entirely. Electrons that are involved in covalent bonds between atoms are said to be in the valence band. The valence band is a range of energies that extends downward from the bottom of the bandgap to the most tightly bound electrons. The reason why the electron energies in the conduction and valence bands are spread out is due to a principle of quantum mechanics known as the Pauli exclusion principle. This fundamental law states that no two particles can have the same quantum states. This in turn implies that no two electrons in a crystal can have the same energy level. Accordingly, each electron in the crystal must have its own energy level. These energy levels spread out to form the bands that are observed. A helpful analogy is to picture people attending a concert in an auditorium. Since only one person can occupy each seat, they must spread out. People will typically take the closest available open seats. In the crystal, electrons will occupy the lowest energy states first, in keeping with the principle of statistical mechanics that says a physical system will tend to minimize its overall energy. An energy band diagram is shown in Figure 6.1.

6.1.3 Holes

When a covalent bond between atoms is broken in a pure semiconductor crystal, an electron is set free. It enters the conduction band. At the same time, its

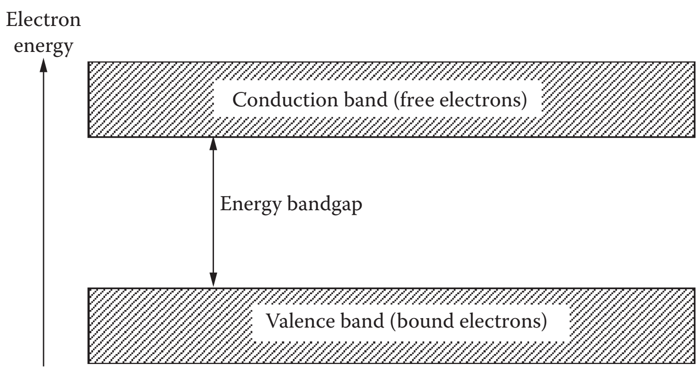


FIGURE 6.1 Energy band diagram for a semiconductor crystal.

absence leaves behind a vacancy in the valence band, which is referred to as a hole. The hole carries a positive charge of equal magnitude to the electron. If there are no holes in the valence band, such as at absolute zero of temperature, the valence band electrons cannot move, since there are no available energy levels for them to occupy other than the ones they are already in. However, when there are holes, the valence band electrons can move and thereby conduct current. It is convenient to represent this process as the movement of the holes, although actually it is the combined effect of all the valence band electrons in motion. It is somewhat like describing the activity in a parking lot in terms of the apparent motion of empty spaces rather than the motion of the cars. This analogy works best if we consider all the cars to be identical, as are electrons.

It is important to understand that holes carry current independently from the conduction band (free) electrons. When an external electric field is applied, holes flow in the opposite direction from electrons, but because holes and electrons are oppositely charged, the hole and electron currents actually add together. Physically, these two components of current arise from the movement of electrons in the two separate energy bands.

6.1.4 Doping

In almost all electronic devices, such as transistors, silicon is the semiconductor used. Henceforth, unless stated otherwise, we will restrict our discussion to silicon.

It is helpful to refer to a section of the periodic table shown in Figure 6.2. Silicon is in Group IV of the periodic table, since it has four valence electrons. This means that the most natural arrangement of silicon atoms into a crystal is for each atom to have four symmetrically arranged nearest neighbors connected by covalent bonds. In most electronic devices, a pure semiconductor is not used. Instead, the silicon is doped n-type or p-type by adding trace amounts of elements from Column III (e.g., boron, or B) or Column V (e.g., phosphorus, or P) of the periodic table.

When these dopant atoms are introduced into the silicon crystal, each atom substitutes for a silicon atom in a random position within the crystal. At typical doping levels, the dopant concentration is on the order of 1 part per million or less. In other words, at most one silicon atom in a million is replaced with a dopant atom. When a Group V dopant substitutes for a silicon atom, it releases

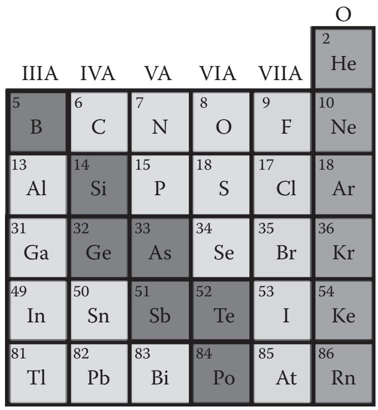


FIGURE 6.2 **Section of the periodic table of the elements around silicon—element 14.**

one of its outer shell (valence) electrons, thus becoming singly ionized. This is illustrated in Figure 6.3.

The remaining four valence electrons satisfy the covalent bonds previously occupied by the silicon atom it replaced. The dopant ion now has a positive charge, and is in a fixed location in the crystal. However, the electron it released is now free. It becomes part of the "free electron sea" within the crystal, and can move under the influence of an electric field.

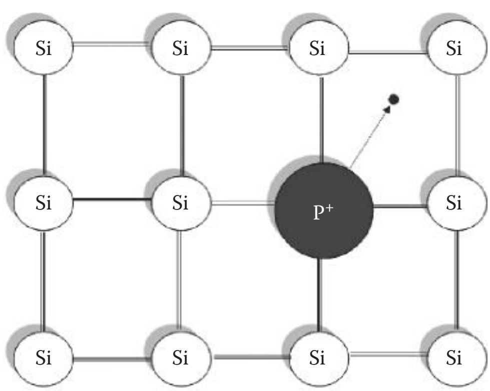


FIGURE 6.3 **Silicon crystal doped n-type with phosphorus from Group V. The small black dot represents a free electron.**

When a Group III dopant substitutes for a silicon atom, it attracts and keeps a free electron from the electron sea in order to have the four valence electrons it needs to satisfy the four covalent bonds. The Group III atom thereby becomes a negatively charged ion. By taking an electron from the sea of free electrons, it leaves a positively charged vacancy in the sea, which is known as a hole. The behavior of the hole is fully consistent with treating it as a free particle having positive charge and finite mass (though different from that of an electron).

For device manufacturing, the starting material is highly purified silicon with dopant atoms added. This material is called the substrate. If the dopants are from Group V of the periodic table (typically phosphorus, arsenic, or antimony), the silicon is called "n-type" since it contains free electrons (one per dopant atom) which are negatively charged. If the dopant atoms are from Group III of the periodic table (typically boron), the silicon is called "p-type" since it contains free holes (one per dopant atom) which are positively charged. It is important to understand that even though the free electrons and holes are charged, the material as a whole is charge neutral since each electron or hole originated from a lattice atom that became ionized with the opposite charge when the carrier left it. The charge on the lattice ions exactly cancels the charge on the free carriers.

The doping process establishes a known value of electrical conductivity and establishes either holes or electrons as the majority carriers. In n-type material, electrons are the majority carriers, and holes are the minority carriers. In p-type material, the situation is reversed. When an electric field is applied, both types of carriers will conduct current, but the majority carrier current will typically be many orders of magnitude larger than the minority carrier current.

6.1.5 The Semiconductor p–n Junction

A remarkable phenomenon occurs when a region of n-type material adjoins a region of p-type material. The resulting p–n junction is the basis for all modern transistors and many other types of electronic devices. A p–n junction is illustrated in Figure 6.4.

In order to explain the behavior of a p–n junction in qualitative, readily understandable terms, one may, somewhat apologetically, use an analogy involving human culture. Suppose we observe a border between two nations with very different cultures, and in fact allow there to be some enmity or

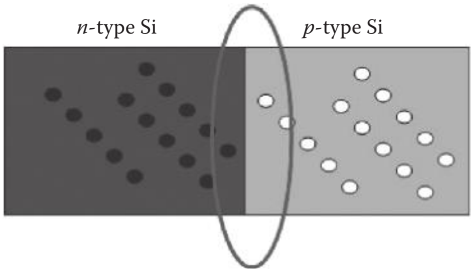


FIGURE 6.4 **A p–n junction. The n-type silicon on the left contains many electrons and few holes. The p-type silicon on the right contains many holes and few electrons. The physics of the junction region is crucial to the operation of all solid-state electronics, including diodes, transistors, and computer chips.**

significant tension between the peoples. Often we find a heavily guarded "no-man's land" or demilitarized zone (DMZ) surrounding the border, which serves as a buffer. We would observe that the population density in the DMZ is extremely low, since it would be highly dangerous for ordinary citizens to be found there. There is an imposed tendency for people on each side to stay "at home." But now suppose that for some reason the guarding of the DMZ is radically reduced. We would then likely observe a significant flow of people across the border in each direction for various reasons—perhaps curiosity, reuniting with family, economics, or the like. In fact, it could be that this border traffic would remain high indefinitely once people from other nations realized that this was an easy and convenient place to traverse through and perhaps circumvent other, more heavily guarded border crossings.

Astoundingly similar behavior is observed at a p–n junction, although of course the particles involved are not sentient beings.[1] In equilibrium, that is, with no external fields applied and with uniform temperature throughout, there is no net traffic across the junction. There is a detailed balance between diffusion (movement away from high concentration to low concentration driven by random thermal motion) and drift (movement along electric field lines due to the Lorenz force). The net current is zero. However, surrounding the p–n junction, there is a permanent double-layer boundary referred to

[1] So far as we know!

by the two synonymous names "space charge region" or "depletion region." Both names refer to the fact that there are very few free charges in the region. However, there is a fixed charge due to the dopant ions, and this fixed charge is of opposite sign on each side of the junction (+ on the n-side and – on the p-side). This fixed charge separation establishes an electric field, which in turn forces a built-in voltage to exist. The built-in field serves as the "guard" for the DMZ in our analogy. If a charged particle strays into the space charge region, it will be driven home forcefully by the electric field. This is illustrated in Figure 6.5.

If an external voltage is applied, the resulting electric field will either strengthen the guard (increase the electric field) or weaken the guard (reduce the electric field) depending upon the polarity. In the former case, there is no significant change; still no current will flow. However, in the latter case, diffusion will cause significant current to flow, just as people will eagerly cross the border when restrictions are removed in our analogy.

This discussion provides the basis for understanding why the p–n junction allows current to flow in one direction but not the other.

Subsequent sections will show how p-type and n-type semiconductors are used to create specific devices, including diodes and transistors.

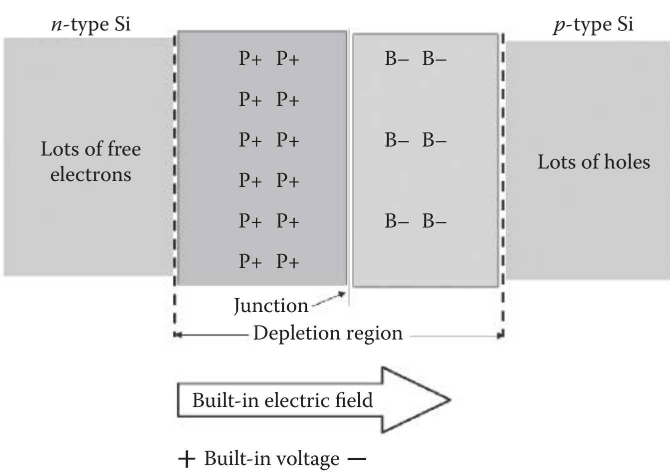


FIGURE 6.5 Enlarged view of the region near the p–n junction. A built-in electric field exists due to the fixed dopant ions.

6.2 Diodes

The simplest device that can be made with a p–n junction is a diode. This is a two-terminal circuit element that allows current to flow in only one direction. Various aspects of a diode are illustrated in Figure 6.6.

Henceforth, the voltage measured or applied from anode to cathode will be designated V_D. The current flowing through the diode in the direction from anode to cathode will be designated I_D.

Four modes, or operating regions, are defined for the diode. When V_D is zero, then I_D is also zero, and the diode is said to be in the zero-bias state. This is also called "equilibrium."

When V_D is positive by a few tenths of volts, the diode enters the forward-biased region. In this case, the current increases exponentially with voltage. A practical model for the diode in this region is the so-called constant voltage

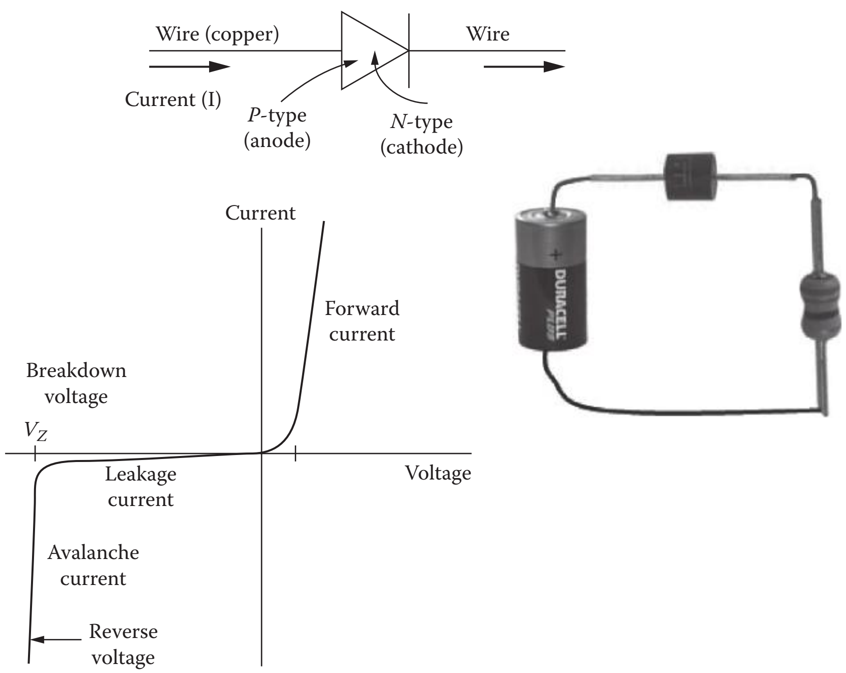


FIGURE 6.6 The p–n junction diode. Top: Circuit symbol and notation. Bottom: Current–voltage plot. Right: Series circuit containing a battery, a diode, and a resistor.

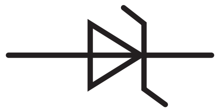


FIGURE 6.7 Circuit symbol for a zener diode.

drop model, in which $V_D = 0.7$ V (for a silicon p–n junction diode), and I_D depends only on the rest of the circuit.

When V_D is negative, the diode is said to be reverse biased. In this case, a very small leakage current flows "backward" through the diode. In most practical applications, this current can be assumed to be zero.

When V_D becomes sufficiently negative, the diode will enter the reverse breakdown region. In this region, the current becomes large and negative. The diode has stopped blocking. This type of operation does no harm to the diode, as long as the external circuit limits the current flow to within the diode's current and power rating. This mode of operation is often used for voltage regulation and voltage clamping for protection of circuit inputs. Diodes that are fabricated specifically for use in this region are called zener diodes, and they have their own circuit symbol as shown in Figure 6.7.

The mathematical model for diode current in terms of diode voltage, not including breakdown, is

$$I_D = I_S e^{\left(-1+\frac{V_D}{nV_T}\right)} \tag{6.1}$$

where:

- I_S is the saturation current, typically in the range of 1×10^{-6} to 1×10^{-14} A.
- n is the ideality factor, which ranges between 1 and 2 depending on the fabrication details. We will assume a value of 1 unless otherwise noted.
- V_T is called the thermal voltage. It is calculated as

$$V_T = \frac{kT}{q}$$

where

k = Boltzmann's constant = 1.3807×10^{-23} J/K = 8.617×10^{-5} *e*V/K.

q = magnitude of charge on an electron = 1.602×10^{-19} C.

T = absolute temperature, degrees Kelvin (K).

The round number value for V_T most commonly used in electronics is 25 mV at "room temperature."

Example 6.1: Calculating Diode Current Given Diode Voltage

A certain diode has $I_S = 1$ pA. Calculate the current at room temperature if $V_D = 0.700$ V.

Solution: We will assume $n = 1$ and $V_T = 25$ mV. The units prefix "p" stands for pico, which means 1×10^{-12}. Then:

$$I_D = 1 \times 10^{-12} \exp\left(-1 + \frac{0.700}{25 \times 10^{-3}}\right) = 0.532\,\text{A}$$

Example 6.2: Calculating Diode Voltage Given Diode Current

A certain diode has $I_S = 1$ nA. Calculate the voltage at 35°C if $I_D = 1$ mA.

Solution: We will assume $n = 1$. The units prefix "n" stands for nano, which means 1×10^{-9}.

We will need to calculate V_T. This requires first determining the temperature in K.

$$T\text{ (in K)} = T\text{ (in °C)} + 273.15$$

$$T = 35 + 273.15 = 308.15\text{ K}$$

$$V_T = (1.3807 \times 10^{-23}\text{ J/K}) \times (308.15\text{ K}) / (1.602 \times 10^{-19}\text{ C}) = 26.56\text{ mV}$$

Now we can invert the diode equation, Equation (6.1), to solve for V_D in terms of I_D:

$$V_D = nV_T \ln\left(1 + \frac{I_D}{I_S}\right) \tag{6.2}$$

Applying this to our problem:

$$V_D = 26.56 \times 10^{-3} \ln\left(1 + \frac{1 \times 10^{-3}}{1 \times 10^{-9}}\right) = 0.367\,\text{V}$$

6.2.1 Diode Rectifier Circuits and Simple Dc Power Supplies

Diode rectifier circuits employ the one-way conduction of diodes to convert ac to dc. With the addition of a filter capacitor, one obtains a simple dc power supply. There are several types of rectifier circuits that one may encounter in practice. Of these, the half-wave rectifier and the full-wave rectifier are the most common.

6.2.2 Half-Wave Rectifier Circuit

A half-wave rectifier circuit is shown in Figure 6.8. The circuit consists of an ac supply (assumed sinusoidal), a diode, and a resistor. The resistor models the load to which we want to supply dc current. For a basic understanding of circuit operation, the diode can be treated as a perfect conductor when V_1 is greater than V_2, and as an open circuit when V_1 is less than V_2.

When the ac voltage is positive, V_1 will be greater than V_2, and the diode will conduct. By Ohm's law, the current will be equal to the voltage divided by the load resistance. The current will be a positive value, although not constant since it is following the sinusoidal variation of the supply voltage.

When the ac voltage is negative, V_1 will be less than V_2, and the diode will not conduct. Therefore, the current will be zero.

Although this circuit provides all-positive current, which is technically dc, we usually want a constant value. We can smooth out the variation and provide nearly constant voltage to the load by placing a capacitor in parallel with the load (i.e., connected between the node labeled V_2 and the ground node). A capacitor used this way is called a filter capacitor. The resulting circuit is

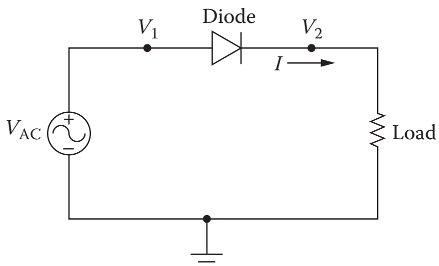


FIGURE 6.8 Half-wave rectifier circuit.

a simple but inexpensive and useful design known as a linear, unregulated dc supply. Such power supplies are used, for example, in microwave ovens to supply current to the magnetron tube. The design details are beyond the scope of this discussion, but this type of power supply was standard for many years until the advent of computer components that required more stable, more efficient, and smaller power supplies. The latter are provided by modern switching power supplies.

6.2.3 Full-Wave Bridge Rectifier Circuit

A number of improvements can be made to the half-wave rectifier at the cost of additional components. One of the most common circuits is the full-wave bridge (FWB) rectifier.

The circuit is shown in Figure 6.9. Careful consideration of this circuit using the same understanding of the diode as for the half-wave rectifier leads to the following two important conclusions: (1) The load current will flow during both phases of the ac, not just the positive phase; and (2) the current through the load, R_L, will always flow in the same direction: top to bottom in the diagram.

When the ac voltage is positive (the top terminal is at higher potential with respect to the bottom terminal), diodes D_1 and D_3 will be conducting, whereas D_2 and D_4 will be off (blocking). When the ac voltage is negative (the bottom terminal is at higher potential with respect to the top terminal), the diode roles will be reversed. D_2 and D_4 will be conducting, and D_1 and D_3 will be blocking.

Just as for the half-wave rectifier, a filter capacitor can be placed in parallel with the load terminals to smooth out the sinusoidal voltage variations

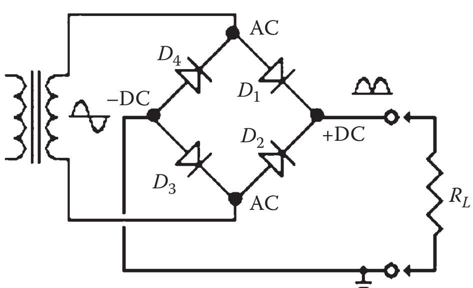


FIGURE 6.9 Full-wave bridge rectifier circuit.

and provide a more constant dc output. Because the full-wave rectifier provides output during both half-cycles of the ac, the filter capacitance required for a given ripple allowance is half of that required for a half-wave rectifier.

6.3 Transistors: An Overview

Transistors are the workhorse circuit elements in modern electronics. They are used as switches and as amplifiers. Switches serve to turn signals on and off, while amplifiers change the size of signals. It can be helpful to visualize the function of a transistor by analogy to a simple water system. A transistor functions like a valve which can control the flow of water through a pipe. It can be adjusted from fully on to fully off, or it can be varied slightly around a setpoint. In the usual operation of a water valve, the position is controlled by hand, but the setting of the transistor is controlled by an electrical signal. This control signal might come from a manually adjusted knob, switch, or dial, or it might come from a computer or microcontroller under the control of software.

There are two main types of transistors: bipolar junction transistors (BJTs) and field effect transistors (FETs). (The acronym FET can be pronounced by sounding out the letters ["ef ee tee"] or by rhyming it with "let." The former seems to be the preferred usage.)

The FET has overtaken the BJT as the most commonly used transistor in modern electronics, but both are important in their own right.

A simple view of transistor function is presented in Figure 6.10. This figure shows the transistor as a three-terminal device. The amount of the working current that flows is determined by a control current (for BJTs) or a control voltage (for FETs). For BJTs the control current, which is typically a hundred times smaller than the working current, merges into the working current, so that the amount exiting is slightly more than the amount entering. For FETs, no significant current flows into the control terminal; instead, the working current is controlled by the effect of the electric field set up inside the device by the control voltage.

Transistors are studied using their current–voltage characteristics, less formally referred to as the "I-V curves." The I-V curves for a transistor are a plot (or data table) of terminal current versus terminal voltage.

Since transistors have three terminals (four if one counts the substrate terminal of the metal–oxide–semiconductor field effect transistor, or MOSFET),

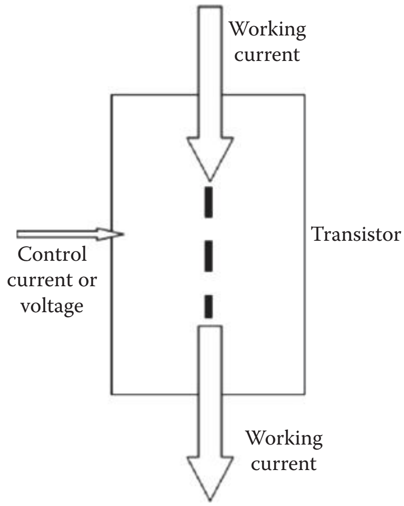


FIGURE 6.10 **Simplified overview of a transistor function.**

it is reasonable to ask which of the three terminal currents and three terminal voltages (e.g., a base emitter, base collector, or collector emitter) are of interest. The answer is provided by practical experience. There are two plots generally used in transistor work: the "transfer characteristics" and the "output characteristics." The transfer characteristics plot the output current versus the control (input) voltage, while the output characteristics plot the output current versus the output voltage, with the control voltage (MOSFET) or current (BJT) as a parameter. Figures 6.11, 6.12, 6.13, and 6.14 show a set of generalized input and output characteristics for a MOSFET and for a BJT.

These I-V plots are divided into regions of operation. Considering first the MOSFET output characteristics shown in Figure 6.12, the region to the right of the knees in which the curves are fairly flat is called the "saturation region." The region to the left of the knees is called the *linear* region, or sometimes the "triode region." When the gate–source voltage is below the threshold voltage, the device is said to be in the "cutoff region."

For the BJT output characteristics shown in Figure 6.14, similar regions exist. The flat region is called the forward active region. The region to the left of the knees is called the saturation region. When the collector current is very small, the BJT is cut off. There is also a reverse active region which is entered when the roles of the collector and emitter terminals are swapped. This region

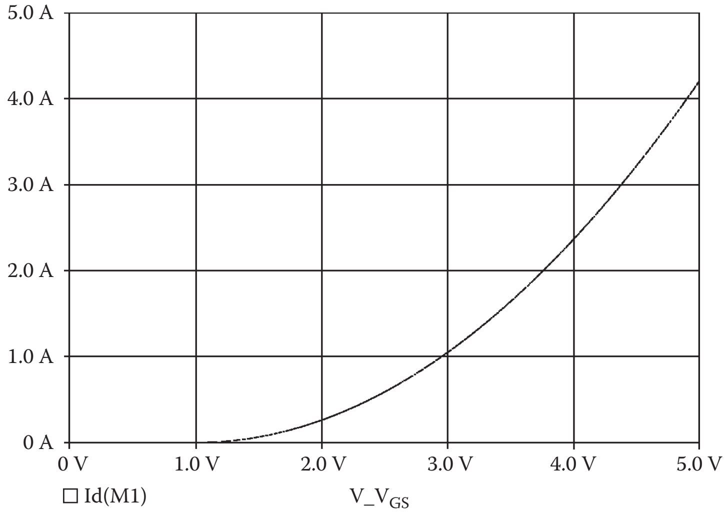


FIGURE 6.11 Transistor characteristics: MOSFET input (transfer) characteristics. Drain current plotted versus gate–source voltage in saturation.

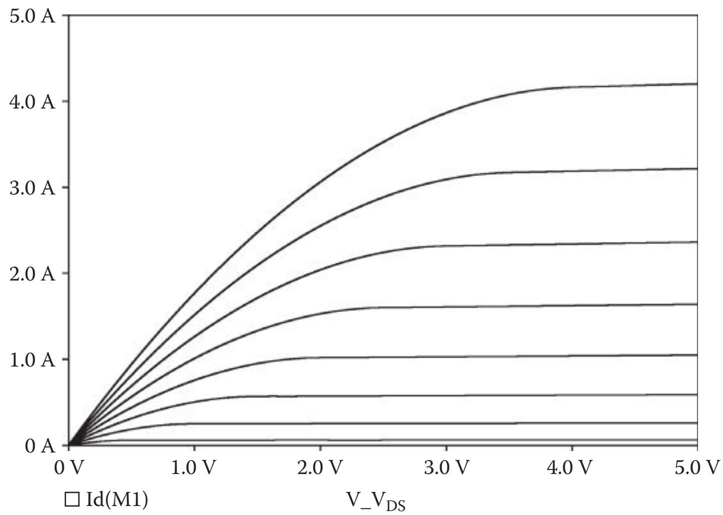


FIGURE 6.12 Transistor characteristics: MOSFET output characteristics. Drain current plotted versus drain–source voltage with gate–source voltage as a parameter.

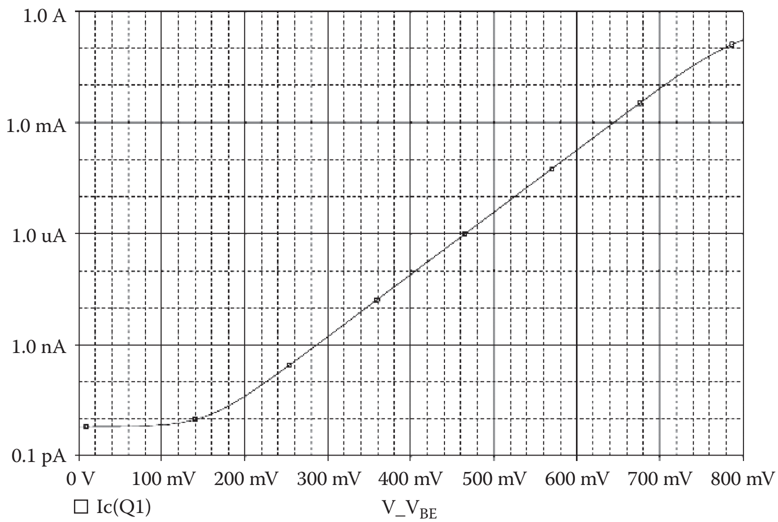


FIGURE 6.13 Transistor characteristics: BJT input (transfer) characteristics. This graph shows the collector current as a function of the base–emitter voltage with fixed collector–emitter voltage. The current scale (vertical axis) is logarithmic.

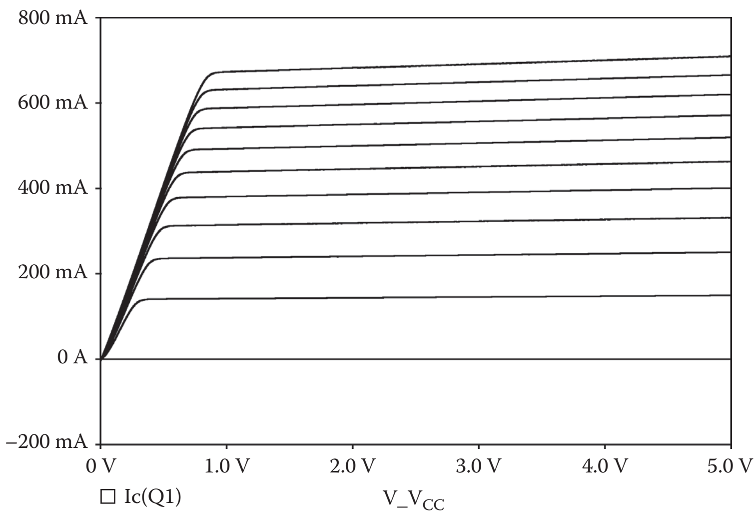


FIGURE 6.14 Transistor characteristics: BJT output characteristics. Collector current plotted versus collector–emitter voltage with base current as a parameter.

is rarely used intentionally, but is often observed when novice experimenters fail to read the data sheet carefully.

Unfortunately, as the careful reader might have noticed, the term "saturation" refers to a completely different region for the BJT than for the MOSFET. For the MOSFET, saturation is the flat region of the curves to the right of the knees. For the BJT, it is the steep sloped region to the left of the knees. There is no rational explanation for this—it is just one of the many accidents of history that we must deal with when learning a new subject. The people who developed BJT theory early on were a different group from those who developed MOSFET theory early on (and largely continue to be). They did not talk to each other much or read each others' papers, so the terminology became entrenched without coordination.

The next few paragraphs will be devoted to a more in-depth look at the MOSFET. The BJT will be treated following.

6.4 The Field Effect Transistor (FET)

The two primary types of FETs are junction FETs (JFETs) and MOSFETs. Only MOSFETs will be discussed here.

MOSFETs can be n-channel or p-channel, and they can also be depletion or enhancement type. These variations provide the circuit designer with the necessary options for practical work. They will be explained in more detail in the following paragraphs.

6.4.1 MOSFET Construction

Discrete (individually packaged) MOSFETs are made, like all modern electronics devices, in a batch process involving adding and removing layers of materials to a silicon wafer, and then cutting the wafer into individual dies (chips), which are packaged for sale. Integrated circuits are made the same way, but contain tens to millions of transistors on a single packaged chip. In either case, the structure of a MOSFET is essentially that shown in Figure 6.15.

The source and drain regions are formed by introducing dopants of opposite types to the substrate through a masking pattern to form bathtub-shaped wells. The dopant concentration in the source and drain regions is higher than that in the substrate.[2] By this means, a p–n junction is formed around

[2] The substrate is also called the *body* or the *bulk*. The corresponding device terminal is usually labeled with the letter "B" to distinguish it from the source terminal, which is labeled with "S."

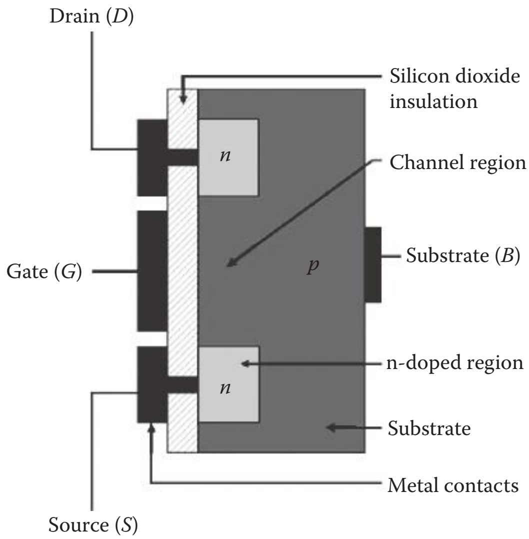


FIGURE 6.15 **MOSFET structure. The drawing is not to scale. A typical MOSFET incorporated in an integrated circuit would have a substrate thickness of several hundred micrometers, and the distance from source to drain would be a fraction of a micrometer. An n-channel MOSFET is shown here; a p-channel MOSFET would have an n-type substrate and p-type source and drain regions.**

the source and drain regions, which helps to constrain the flow of current. With suitable voltage polarity applied to the source, drain, gate, and substrate terminals, current can be caused to flow between the source and drain regions and thus through the external connected circuitry.

For simplicity, we will only consider MOSFETs in which the substrate and source are connected. These are so-called three-terminal devices. If this is not the case, so-called back-gating effects can occur. These must be taken into account in the most accurate analyses, but will be ignored here.

When sufficient voltage is applied to the gate relative to the source, a temporary channel forms between the source and drain, and current can flow. The amount of current that flows is controlled by the amount of voltage applied between the gate and source. From this, we see that the essential function of the MOSFET is a voltage-controlled resistance. This resistance is nonlinear (i.e., it does not follow Ohm's law).

Some discrete MOSFETs are symmetric in that the source and drain terminals are interchangeable, but many are not, so it is important to read the manufacturer's data sheet carefully to determine the right way to connect the device in a circuit.

Enhancement MOSFETs are fabricated without a channel. For an n-channel enhancement MOSFET, a positive voltage greater than the threshold voltage must be applied between the gate and source terminals in order to establish a channel of free electrons between the source and drain. The channel exists only as long as the gate-to-source voltage is maintained above threshold. While the channel exists, the resistance between the source and drain is low, and current can flow. The channel resistance is a function of the gate-to-source voltage, so therefore is the current.

When the gate–source voltage V_{GS} is below the threshold voltage (typically 1 V or less), there is no channel, so the source and drain regions remain isolated by the p–n junction that surrounds them, and no current flows even if a voltage is applied between the drain and source. In other words, the MOSFET acts like an infinite resistance, or open circuit.

When the gate–source voltage is raised above the threshold voltage, an electric field penetrates the silicon substrate, and a temporary channel forms composed of electrons drawn near to the silicon surface by the electric field. Since this channel contains the same type of charge carriers as the free carriers in the source and drain, it links the two regions. The resistance of the channel depends upon the channel dimensions (length, width, and thickness), and the concentration of electrons in the channel. The electron concentration and the channel thickness both depend upon the strength of the electric field from the gate–source voltage. Consequently, the channel resistance is a nonlinear function of the applied control voltage.

For p-channel enhancement MOSFETs, the same concept applies but with the polarities reversed. A gate-to-source voltage more negative than the (negative) threshold must be applied in order to create a channel of holes between the source and drain regions.

Depletion MOSFETs are fabricated with an extra step so that a channel exists even with no voltage applied. The applied gate-to-source voltage can either deplete or enhance the channel, depending upon the polarity. The basic rule is "Opposites attract." For example, for an n-channel depletion MOSFET, a negative applied gate-to-source voltage will repel the channel electrons, thus depleting the channel, increasing the channel resistance, and reducing the flow of current. If the applied voltage is

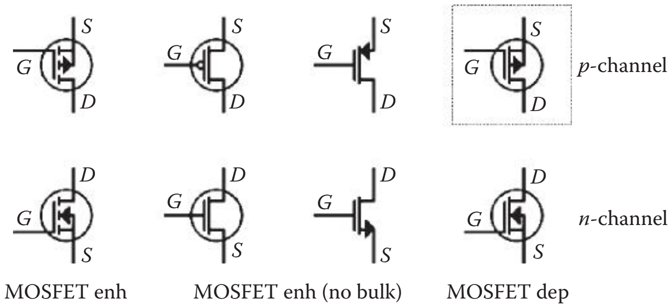


FIGURE 6.16 **Some commonly used circuit symbols for three-terminal MOSFETs.**

sufficiently negative, the channel will be completely depleted and the current will cease to flow.

6.4.2 MOSFET Circuit Symbols

A variety of circuit symbols are used to represent MOSFETs in schematic diagrams. Some standard symbols for three-terminal devices are shown in Figure 6.16.

6.4.3 MOSFET I-V Characteristics

We will begin by discussing the n-channel enhancement MOSFET, and then state the changes that need to be made for p-channel and depletion types.

The MOSFET characteristics shown in Figure 6.11 and Figure 6.12 are for an n-channel enhancement MOSFET. The input characteristics (Figure 6.11) show that the MOSFET has a threshold voltage followed by a quadratic control region. The corresponding equation in the saturation region is

$$I_D = \frac{K_n}{2}(V_{GS} - V_{Tn})^2(1 + \lambda V_{DS}) \tag{6.3}$$

for

$$V_{DS} \geq V_{SATn} \tag{6.4}$$

where:

I_D is the drain current (working current)—the current *entering* the drain terminal.

K_n is the transconductance parameter.

V_{GS} is the gate–source voltage.
V_{Tn} is the threshold voltage (positive for enhancement and negative for depletion).
λ is the channel-length modulation parameter.
V_{DS} is the drain–source voltage.
$V_{SATn} = V_{GS} - V_{Tn}$ is the saturation value of V_{DS}.

In the linear region, the MOSFET equation is

$$I_D = K_n\left(V_{GS} - V_{Tn} - \frac{V_{DS}}{2}\right)V_{DS} \tag{6.5}$$

for

$$0 \le V_{DS} < V_{SATn} \tag{6.6}$$

The linear region is thusly named because of the linear relation between I_D and V_{GS}.

When the MOSFET is used as a switch, the ON, or CLOSED, switch is approximated by the linear region. In this region, the working voltage (V_{DS}) and the resistance are both small. The OFF, or OPEN, switch is approximated by the cutoff region. The resistance is very large, and the working current is essentially zero.

Example 6.3: n-Channel MOSFET Drain Current Calculation

The NXP BSS83 MOSFET has $K_n = 14.3$ mA/V^2, $V_{Tn} = 1.5$ V, and $\lambda = 0.2$ V^{-1}. Calculate the drain current if $V_{GS} = 2$ V and $V_{DS} = 9$ V.

Solution:

First, we go online to verify that the given device is an n-channel MOSFET. (Since the threshold voltage V_{Tn} is positive, we know it is an enhancement device. This does not change the calculated result, but it is useful to know in general.)

We need to know whether the MOSFET is in the saturation or linear region in order to determine which equation to use. We can learn this by comparing V_{DS} to V_{SAT}.

$$V_{SAT} = V_{GS} - V_{Tn} = 2 - 1.5 = 0.5 \text{ V}$$

$$V_{DS} = 9 \text{ V (given)}$$

Since $V_{DS} > V_{SAT}$, the MOSFET is in saturation, and we should use Equation (6.3).

$$I_D = \frac{K_n}{2}(V_{GS} - V_{Tn})^2(1 + \lambda V_{DS})$$

$$= \frac{14.3\,\text{mA}}{2}(2 - 1.5)^2(1 + 0.2 \times 9)$$

$$= 5.01\,\text{mA}$$

The terminal currents and terminal-pair voltages are shown in Figure 6.17.

For a p-channel MOSFET, the equations are modified to reflect the changed polarities.

$$I_D = \frac{K_p}{2}(V_{SG} + V_{Tp})^2(1 + \lambda V_{SD}) \tag{6.7}$$

for

$$V_{SD} \geq V_{SATp} \tag{6.8}$$

where

I_D is the drain current (working current)—the current *leaving* the drain terminal.
K_p is the transconductance parameter.
V_{SG} is the source–gate voltage.
V_{Tp} is the threshold voltage (negative for enhancement and positive for depletion).
λ is the channel-length modulation parameter.
V_{SD} is the source–drain voltage.
$V_{SATp} = V_{SG} + V_{Tp}$ is the saturation value of V_{SD}.

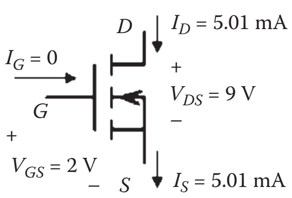


FIGURE 6.17 MOSFET voltages and currents for Example 6.3.

In the linear region, the p-channel MOSFET equation is

$$I_D = K_p\left(V_{SG} + V_{Tp} - \frac{V_{SD}}{2}\right)V_{SD} \tag{6.9}$$

for

$$0 \le V_{SD} < V_{SATp} \tag{6.10}$$

Figure 6.18 shows a practical circuit in which an n-channel enhancement MOSFET is used to switch a dc motor on and off.

In this circuit, the MOSFET is used as a switch. Typically V_{in} will be a logic-level voltage (on or off). The input signal is fed to the gate of the MOSFET. Usually R1 is small, typically 100 Ω or less. Its purpose is to limit the inrush current that can occur due to the gate capacitance for large MOSFETs switching large motor currents. In small or medium-sized applications, it can be replaced with a direct connection (R1 = 0). The purpose of R2 (typically 10 kΩ) is to pull down the gate voltage if V_{in} goes to its logic low value, or becomes undefined. The latter case might occur if V_{in} is produced by so-called tri-state logic. Tri-state logic has an extra input that can disable (internally disconnect) the outputs. Then, the gate voltage of the MOSFET would be floating (undefined) and the circuit would behave unpredictably.

The purpose of the diode is to suppress the large transient voltage spikes that can occur when the motor is switched off. These happen because the motor winding is inductive. This means that the current flowing when the motor is on sets up a magnetic field around the winding. The energy stored in

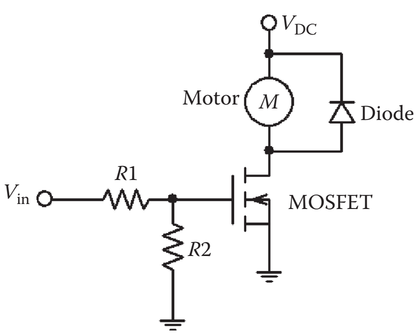


FIGURE 6.18 MOSFET circuit used to switch a dc motor.

this magnetic field must be dissipated before the current can go to zero. The result is that a large voltage spike appears across the motor terminals when the current is changed rapidly. The polarity of this voltage spike is opposed to the direction of current flow. In Figure 6.18 it is positive at the bottom of the motor with respect to the top of the motor. However, since the voltage at the top of the motor is fixed by the power supply VDC, the large spike will appear across the MOSFET, possibly destroying it. With the diode in place as shown, the diode will be forward-biased by the voltage spike. It will limit the size of the voltage spike to about one volt, and it will absorb the energy of the collapsing magnetic field from the motor winding.

One common method for speed control of a dc motor is pulse-width modulation (PWM). In this approach, a pulsed control signal is applied to the MOSFET at high frequency (much faster than the motor can respond). The motor speed is then determined by the average value of the pulses, which can be varied by changing the pulse width.

6.4.4 MOSFET Switches: Practical Details

When choosing which MOSFET to use in a particular switching application, one must consider the maximum voltage, the maximum current, and the maximum power that the switch will need to handle. In some cases, the switching speed is an additional consideration.

These considerations are best illustrated through an example.

Example 6.4: MOSFET Switch

A dc motor is to be switched. The motor is rated at 12 V and has a maximum (stall) current of 2 A. The control voltage to the MOSFET is to be supplied from a microcontroller output pin that has logic high = 3.3 V and logic low = 0 V.

A suitable circuit is shown in Figure 6.18.

We can see that the motor current is the same as the MOSFET working current, I_D. When V_G is low = 0 V, the MOSFET is off (cutoff region), and $I_D = 0$. In this case, there will be no voltage drop across the motor, so the voltage at the drain of the MOSFET is 12 V. This provides the worst-case (maximum) value required for V_{DS} for this MOSFET.

When the MOSFET is on, current will flow through the motor. The actual voltage drop across the motor terminals will depend upon the load applied to the motor. In the extreme case, if the load fully stalls the motor, the rating tells us that the current will be 2 A if the voltage is 12 V. However, V_{DS} of the MOSFET cannot go to zero, so the worst-case motor voltage drop will be slightly less than 12 V and the corresponding current will be slightly less than the full rated value. Nevertheless, we will look for a MOSFET with a current rating of at least 2 A.

There are many suppliers and sources of data for MOSFETs. Due to our familiarity and past experience, we somewhat arbitrarily decide to try to find a suitable part from NXP. A visit to the NXP Web site and a couple of menu selections later, we find the list of available MOSFETs. Using the parameter selection feature, we first seek all the n-channel MOSFETs with a maximum V_{DS} of 12 V. This suddenly reduces the selection to a single part, the PMN28UN. Hopefully, we look at the next critical value: I_D (A). We are pleased to find that this far exceeds our requirements, since the given value is 5.7 A. Next, we examine the channel resistance, R_{DS}(on), to estimate what the working voltage V_{DS} will be when the maximum current of 2 A is flowing through the MOSFET. Our value of V_{GS} will be 3.3 V. Unfortunately, this exact value is not provided, but we can interpolate between $R_{DS\,\mathrm{on}}$ [V_{GS} = 4.5 V] and $R_{DS\,on}$ [V_{GS} = 2.5 V]. The former value is 34 mΩ, and the latter value is 40 mΩ. Rough interpolation yields R_{DSon} = 37 mΩ for V_{GS} = 3.3 V. Then we can use Ohm's law to compute the on-state voltage across the MOSFET:

$$V_{DSon} = R_{DSon} \times I_{D\max} = 37 \times 10^{-3}\ \Omega \times 2\ \text{A} = 74\ \text{mV} = 0.074\ \text{V}$$

This MOSFET provides a very low on-state voltage drop, which is desirable.

The next thing to check is the power dissipation. In the off state, there is no power dissipation in the MOSFET because there is no current flow. In the on state, the power dissipation will be

$$P_D = V_{DSon} \times I_D = 74\ \text{mV} \times 2\ \text{A} = 148\ \text{mW}$$

Looking at the given specifications for this MOSFET, we find P_{tot} (W) = 1.75 W. Since our required power dissipation (roughly 0.15 W) is far less than this, we should be quite safe in using this part.

Further design refinements can be made from this starting point. For example, a reasonable design choice would be to select a MOSFET with a higher V_{DS} rating to provide a safety margin. A choice of 16 V or 20 V would be reasonable. It might also be better to choose a MOSFET with a

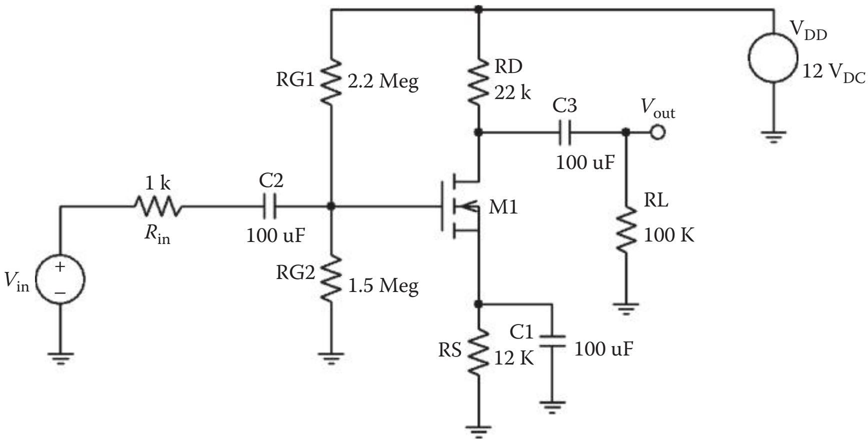


FIGURE 6.19 **Practical MOSFET amplifier.**

lower current and power rating if cost and/or size can be reduced, since this device exceeds our requirements by such a large margin.

The MOSFET can also be used to amplify electrical signals.

For a single-transistor amplifier, the MOSFET must be biased.[3] "Biasing" refers to causing a fixed reference value of current to flow through the device when no signal is applied. When an input signal voltage is applied, the output current is proportionally modulated, and a larger voltage can be developed across the output terminals. The voltage gain of the amplifier circuit is the ratio of the output voltage magnitude to the input voltage magnitude. A practical MOSFET amplifier circuit is shown in Figure 6.19.

The type of amplifier shown in Figure 6.19 is an ac-coupled amplifier suitable for amplifying signals in the audio-frequency range. The values of the capacitors C1, C2, and C3 largely determine the low-frequency response, whereas the choice of MOSFET determines the high-frequency performance. Using a typical n-channel MOSFET intended for small-signal amplifiers, such as the Infineon BSR202N, this circuit can provide a voltage gain of about 800 (or 58 dB).

The bias current in this circuit is about 320 μA. This is established by the values of R_D, R_S, R_{G1}, and R_{G2} together with the properties of the MOSFET and the power supply voltage V_{DD}.

[3] This is referred to as a Class-A amplifier configuration.

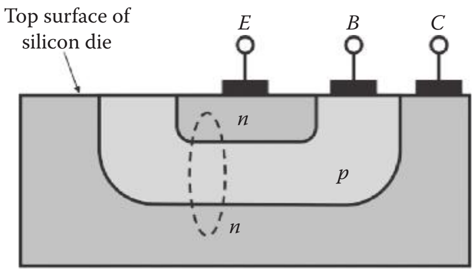


FIGURE 6.20 Structure of an npn bipolar junction transistor (BJT). The dashed oval shows the core region of transistor activity.

6.5 The Bipolar Junction Transistor (BJT)

The BJT is fabricated differently from the MOSFET, but it shares many of the same capabilities. The BJT structure is shown in Figure 6.20.

The core of the device is an npn (or pnp) sandwich formed by the emitter, base, and collector regions. The base–emitter voltage controls the flow of current through the base region. The underlying physics dictates that the control law is exponential, so small changes in base–emitter voltage can produce large changes in collector current. Properly biased, the BJT can produce amplification. The BJT can also be used as an on–off switch.

6.5.1 BJT Circuit Symbols

The common circuit symbols used for BJTs are shown in Figure 6.21. The emitter terminal is marked with an arrow. The direction of the arrow distinguishes npn from pnp. The base terminal is in the center, and the terminal opposite the emitter is the collector

6.5.2 BJT I-V Characteristics

Remarkably, the BJT characteristics are similar in appearance to those of the MOSFET, even though the underlying physics and equations are very different.

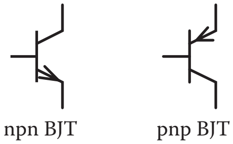


FIGURE 6.21 Common circuit symbols for BJTs.

Table 6.1 BJT Junction Conditions and Operating Regions

Junction Conditions	Operating Region
Both junctions forward biased	Saturation (closed switch)
Both junctions reverse biased	Cutoff (open switch)
EBJ forward biased, CBJ reverse biased	Forward active (amplifier)
EBJ reverse biased, CBJ forward biased	Reverse active (rarely used)

The BJT operating regions are distinguished by the condition of the p–n junctions between the base and emitter, and between the base and collector. Table 6.1 shows the relevant conditions.

A p–n junction is forward biased when the p-type side is at a higher voltage than the n-type side to such a degree that significant current can flow. The meaning of "significant current" is somewhat arbitrary and related to the actual application and specific devices under consideration. For a large power transistor that is expected to handle tens of amperes of currents, we might say it is forward biased when 1 A of current is flowing. For a small transistor in a computer interface circuit handling mA of current, forward bias might be defined at 100 microamps. In either case, the corresponding junction voltage for silicon near room temperature will typically be in the range of 0.5 to 0.8 V. Most practitioners use a somewhat arbitrary value of 0.7 V for the forward bias voltage.

When the junction voltage between the p-side and the n-side is less than required to establish forward bias, we say the junction is reverse biased. This is actually a misnomer when the voltage is not actually negative, but it is common usage.

There are several models which can be used to describe the underlying physics and the observed behavior of the BJT. One of these is the Ebers–Moll model, which is based on the behavior of the individual p–n junctions within the transistor. In practice, these models are simplified for everyday use in design, estimation, and hand calculations. In the following paragraphs, we will describe the simplified models for the npn BJT in each region. Similar models apply for the pnp BJT with appropriate reversal of polarities.

6.5.3 The BJT in Cutoff

Since both junctions are reverse biased, no significant current will flow into any device terminal. The voltages will be completely determined by the external circuit.

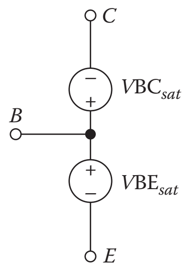


FIGURE 6.22 Model for the npn BJT in saturation.

In switching applications, the BJT in cutoff corresponds to the open switch position.

6.5.4 The BJT in Saturation

Both junctions are forward biased, so there will be a voltage drop of approximately 0.7 V across each junction. The model is shown in Figure 6.22. From this model, we can see that V_{CEsat} is equal to the difference between V_{BEsat} and V_{BCsat}, which will be small. Typical values of V_{CEsat} are in the range of 50 mV.

The terminal currents will be determined by the external circuit. In switching applications, the saturated BJT corresponds to the closed switch position.

6.5.5 The npn BJT in the Forward Active Region

In this case, the I-V equations are governed by the base–emitter junction. The relevant equations are as follows:

$$I_C = I_S e^{\frac{V_{BE}}{V_T}}\left(1+\frac{V_{CE}}{V_A}\right) \tag{6.11}$$

$$I_B = I_C/\beta_F \tag{6.12}$$

$$\beta_F = \beta_F\left(1+\frac{V_{CE}}{V_A}\right) \tag{6.13}$$

where

I_C is the current *entering* the collector from the external circuit.
I_B is the current *entering* the base from the external circuit.
I_S is the saturation current parameter.

V_{CE} and V_{BE} are the collector–emitter and base–emitter junction voltages, respectively.
V_T is the thermal voltage, previously defined for the diode in Equation 6.1.
V_A is the Early voltage, a parameter which accounts for the increase in collector current with increasing V_{CE}.
β_{Fo} is the current gain at the onset of the forward active region (knee of the output characteristic curves).
β_F is the current gain at any specific value of V_{CE} in the forward active region.

We can also write the emitter current since by Kirchhoff's voltage law (KVL) we know that

$$I_E + I_B + I_C = 0$$

where I_E is the current entering the emitter terminal.
Therefore:

$$I_E = -(I_B + \beta\, I_B) = -I_B(1+\beta) \tag{6.14}$$

With a little algebraic manipulation, we can also show that

$$I_C = -\alpha I_E \tag{6.15}$$

where

$$\alpha = \beta/(1+\beta) \tag{6.16}$$

The physical significance of β is the ratio of collector current to base current (i.e., the current gain or control ratio of the BJT). Typical values of β are 100 to 400. For $\beta = 100$, we need to apply 1 mA of base current in order to cause 100 mA of collector current to flow.

The physical significance of α is the fraction of charge carriers that survive the trip through the base region on their way from the emitter to the collector. If $\beta = 100$, then $\alpha = 0.99$. The remaining 1% of electrons (in an npn BJT) that don't make it will recombine in the base region with positively charged holes injected via the base current.

6.5.6 The BJT in the Reverse Active Region

In this case, the roles of emitter and collector are interchanged. The equations are the same if the subscripts E and C are swapped. However, the reverse

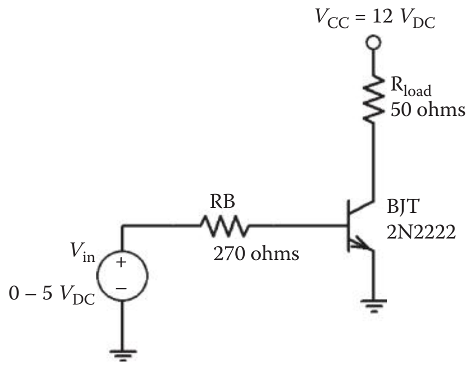


FIGURE 6.23 **Practical BJT switching circuit. In this circuit, approximately 15 mA of base current switches nearly 250 mA of collector current.**

current gain β_R is much smaller than the forward current gain, and is often less than one.

6.5.7 BJT Switching and Amplifying Circuits

Figure 6.23 shows a practical switching circuit. As the input switches between 0 V and 5 V, the BJT goes from its cutoff state to its saturated state. When the BJT is in cutoff, no current will flow through the load. When the BJT is saturated, the collector terminal (to which the bottom of the load is connected) will be nearly at a small voltage, and load current will flow.

BJT amplifiers can be built using a circuit similar to the MOSFET amplifier shown in Figure 6.19, although the biasing details are somewhat different. A BJT-inverting amplifier with a gain of 150 V/V is shown in Figure 6.24.

6.6 Integrated Circuits

Integrated circuits, commonly called ICs or chips, contain many transistors and other circuit components on a single piece of a silicon wafer, called a die.[4] The die is usually packaged in a housing that allows it to be handled and inserted into a printed circuit board or test equipment.

Some commonly encountered ICs are listed in Table 6.2. Integrated circuits may be classified according to several features; some common classifications are listed in Table 6.3.

[4] In IC terminology, *die* is singular. There are three commonly used plurals: *dice*, *dies*, and *die*.

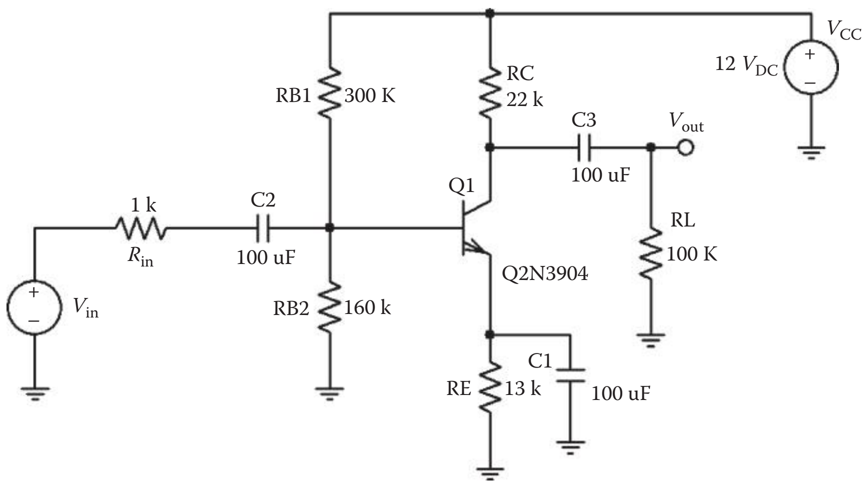


FIGURE 6.24 BJT audio-frequency amplifier.

6.7 Data Sheets

Nearly all manufacturers supply a data sheet for each part or family of parts they sell. Data sheets can range from a few pages for simple digital logic circuits to six hundred pages or more for a complex microcontroller or digital signal processor.

Table 6.2 Some Examples of Common Integrated Circuits

Common Name	Typical Manufacturer and Part Number	Description
555 Timer	National Semiconductor LM555	Circuit that can provide controlled pulses; usually used for timing and triggering other system functions
741 Op-Amp	National Semiconductor LM741	Linear operational amplifier
7400	Texas Instruments SN7400	TTL Quad NAND gate
HC12	Freescale Semiconductor HCS12	16-bit microcontroller

Table 6.3 Classifications of Integrated Circuits

Classifiers	Description
Linear, digital, mixed-signal, or interface	Type of function. Digital ICs perform logic operations. Linear ICs provide amplification. Mixed-signal circuits perform both logic and linear functions. Interface circuits act as the glue between different logic families, power levels, and/or external signals.
MSI, LSI, VLSI, and ULSI	Number of transistors per chip. MSI = medium-scale integration, L = large, VL = very large, and UL = ultra large.
Consumer, industrial, military, or medical grade	Ruggedness, qualification, and/or temperature range. Consumer grade is typically the least stringent, military grade demands the highest temperature range, and medical grade may have the most stringent fault tolerance requirements.
CMOS, TTL, ECL BiCMOS, etc.	Logic family. CMOS is constructed from MOSFETs, TTL and ECL from BJTs, and BiCMOS from both.

Reading a data sheet can be intimidating because of the density of information. Several data sheet examples will be discussed next.

6.7.1 Reading a Data Sheet for a Simple Digital Logic IC

The first several pages of the data sheet for a TTL NAND gate are shown in Appendix B. This particular part is the SN5400 / 7400 series TTL NAND gate supplied by Texas Instruments, Inc.

The overview of this data sheet is as follows:

Page 1: Title and overview, package layout including pin numbers, and Boolean logic function.
Page 2: Ordering information (part numbers), function table, and logic diagram.
Page 3: Internal circuit schematics.
Pages 4–6: Limits (absolute maximum ratings), recommended operating conditions, electrical characteristics, and switching characteristics for all variants.
Page 7–22: Packaging information.
Page 23: Legal disclaimers and company contact information.

More detail about each page is described in the following paragraphs.

6.7.2 Data Sheet Page 1

This particular data sheet covers a family of six electrically distinct parts; these are labeled SN5400, SN54LS00, SN54S00, SN7400, SN74LS00, and SN74S00.

The first challenge is to understand the syntax of the label. The first two letters, SN, are specific to this manufacturer, Texas Instruments (TI). The last two digits, 00, indicate the logic function NAND.

The two digits 74 are industry traditional for standard-grade TTL logic, which is fabricated using bipolar junction transistors (as distinct, e.g., from CMOS logic which uses MOSFETs). Texas Instruments changes the leading digit from 7 to 5 to indicate their higher grade parts, which have a wider temperature range of operation.

The intermediate letters, in this case nothing, LS, or S, indicate more detailed differences among the performance of the parts. This is discussed in more detail below.

In the data sheet, a shorthand is used whereby the parts SN54LS00 and SN74LS00 are referred to as ‘LS00; the parts SN54S00 and SN74S00 are referred to as ‘S00, and the others are referred to as ‘00. Also, we will use the notation 54xx and 74xx to distinguish between the parts with 5 in the label and those with 7 in the label.

It can sometimes be challenging to determine precisely what is different among the different device numbers. In this case, one clue comes from the temperature range given on page 2 in the “Ordering Information” table, where we notice that the 54xx parts have a much wider operating temperature range than the 74xx parts. The wider range of −55°C to +125°C is generally referred to as “mil spec,” short for “military and aerospace specification.” However, we note with some caution the disclaimer on the last page that actual “mil spec” parts have different packaging material as well as the enhanced temperature range. One can order the 54xx parts perhaps for breadboarding and prototype work targeted at military or aerospace finished products, which would ultimately be built with verified mil spec parts.

Other differences among the family members are indicated by the insertion of letters LS and S. Parts with S are Schottky-TTL designs. This design uses Schottky diode technology to reduce the extent to which the BJTs saturate, thereby speeding up the switching between logic states. The corresponding circuit diagram is shown on page 3 of the data sheet, labeled ‘S00. The reduction in switching delay can be seen by comparing the average of the typical propagation delays for the ‘00 parts (switching characteristics table at the top of page 5) and the ‘S00 parts (table at the bottom of page 6). The ‘00 parts average (1/2) (7 + 11) = 9 ns, whereas the ‘S00 parts average (1/2) (4.5 + 5) = 4.8 ns for the highest load listed.

Parts with LS in the part number are low-power Schottky designs. This design combines the use of Schottky diodes with larger value resistors. The larger resistors result in lower power consumption, but they slow down the switching between logic states. The Schottky diodes compensate for the switching speed reduction, so the overall design has about the same switching speed as the ‘00 parts, but with lower power consumption.

Besides the electrical performance differences, each of the six devices is available in a variety of package types. Each package has a unique combination of thermal and mechanical properties. A packaging matrix is provided on page 2 which shows the available packages for each part and the corresponding order number. The terms “tube” and “tape and reel” refer to the way the chips are shipped to the customer. For prototype work, tubes containing usually a small number of chips (typically 25) are ordered. For manufacturing, tape and reel are commonly used. The individual chips are stuck to a polymer “tape” with a light adhesive, and the tape is rolled up onto a reel which may contain several thousand chips. The reel is mounted onto a feeder system which is part of an automated assembly system.

6.7.3 Data Sheet Page 2

On page 2, we find the following items:

Ordering Information Table. This shows the available packages and the corresponding orderable part numbers, as well as the actual markings to be found on the packages.

Function Table (often called the Truth Table for simple combinational logic). This lists the logic output for every combination of inputs. The electrical meaning of High and Low logic levels are not provided here—those details are provided beginning on page 4.

Logic Diagram. This shows the function of the logic gate using standard logic symbols.

6.7.4 Data Sheet Page 3

On page 3 we find electrical schematics for each of the variants—standard ('00), Schottky ('S00), and low-power Schottky ('LS00). This can be useful for an engineer who needs to do detailed circuit simulation, or for someone who needs to design a nonstandard interface to the chip.

6.7.5 Data Sheet Page 4

On page 4, we start seeing the detailed electrical data for this part. Some of these data are contractual in nature. In other words, they are written as a guarantee of performance with various disclaimers.

The first item is "Absolute Maximum Ratings." The corresponding note gives a clear statement that the device is not guaranteed to perform as advertised if these limits are exceeded.

The next data table on page 4 is titled "Recommended Operating Conditions." These ROCs are the specified values for normal operation. Closer inspection reveals that this table is only for the '00 parts. (Tables for the 'S00 and 'LS00 parts are provided separately on the following pages of the data sheet.) Both parts, SN5400 and SN7400, are listed, with columns for "Min," "Nom," and "Max" giving the minimum, nominal, and maximum values of the various parameters, respectively.

For example, we read across the first line to learn that the supply voltage, V_{CC}, should be 5 V for both parts (this is the nominal value). However, we can still operate within specifications if the voltage dips as low as 4.5 V or rises as high as 5.5 V for the SN5400. The allowable range for the SN7400 is narrower: from 4.75 V to 5.25 V.

The last line of the table gives the allowable range of free-air temperature for operating the devices. The term "free-air" implies that no enclosure or fan is used which would affect heat convection from the chip. In this table, we note the significant difference between the 54xx military-grade parts and the 74xx standard parts.

The last table on page 4 shows "Electrical Characteristics over Recommended Free-Air Temperature Range." The table header states "unless otherwise noted." Indeed, the typical values (TYP column) are all noted to be valid only at $T_A = 25°C$. The term "T_A" stands for ambient

Table 6.4 Definition of Data Sheet Parameters

Parameter	Definition
V_{IK}	Input clamping voltage. All logic inputs should be positive voltage, but sometimes noise causes negative spikes. There are input clamping diodes that protect the chip against these spikes. This is the (negative) clamping voltage.
V_{OH}	Output voltage corresponding to a logic HIGH state.
V_{OL}	Output voltage corresponding to a logic LOW state.
I_I	Input current under worst-case conditions.
I_{IH}	Input current under nominal conditions, input logic HIGH.
I_{IL}	Input current under nominal conditions, input logic LOW. A negative sign means current is flowing out of the chip.
I_{OS}	Output current from an output pin that has been shorted to ground. A negative sign means current is flowing out of the chip.
I_{CCH}	Power supply current when the gate output is logic HIGH.
I_{CCL}	Power supply current when the gate output is logic LOW.

temperature, which for this purpose means essentially the same thing as free air.

Unfortunately, the data sheet does not clearly define the parameters listed. The notation used is fairly standard in the field, but by no means uniform. The parameters are defined in Table 6.4.

Use of this table is illustrated in Example 6.5.

Example 6.5: Reading a Data Sheet

(a) Determine the typical and extreme values of the output low voltage for the SN7400. State the conditions which apply.
(b) Determine the average typical power supply current for the SN5400 and SN7400. State the conditions which apply.

Solution: First, we observe the footnote in the data sheet table which states, "For conditions shown as MIN or MAX, use the appropriate value specified under recommended operating conditions."

For part (a), the data sheet values we need are highlighted in Figure 6.25. Reading the row for V_{OL} and the columns under the SN7400 heading, we

Recommended operating conditions (see Note 3)

		SN5400			SN7400			Unit
		Min	NOM	Max	Min	NOM	Max	
V_{CC}	Supply voltage	4.5	5	5.5	4.75	5	5.25	V
V_{IH}	High-level input voltage	2			2			V
V_{IL}	Low-level input voltage			0.8			0.8	V
I_{OH}	High-level output current			–0.4			–0.4	mA
I_{OL}	Low-level output current			16			16	mA
T_A	Operating free-air temperature	–55		125	0		70	°C

Note 3: All unused inputs of the device must be held at V_{CC} or *GND* to ensure proper device operation. Refer to the Tl application report, *implications of Slow or Floating CMOS inputs*, literature number SCBA004.

Electrical characteristics over recommended operating free-air temperature range (unless otherwise noted)

Parameter	Test Conditions‡			SN5400			SN7400			Unit
				Min	TYP§	Max	Min	TYP§	Max	
V_{IK}	V_{CC} = MIN.	I_I = –12 mA				–1.5			–1.5	V
V_{OH}	V_{CC} = MIN.	V_{IL} = 0.8 V.	I_{OH} = –0.4 mA	2.4	3.4		2.4	3.4		V
V_{OL}	V_{CC} = MIN.	V_{IH} = 2 V.	I_{OL} = 16 mA		0.2	0.4		0.2	0.4	V
I_I	V_{CC} = MAX.	V_I = 5.5 V				1			1	mA
I_{IH}	V_{CC} = MAX.	V_I = 2.4 V				40			40	μA
I_{IL}	V_{CC} = MAX.	V_I = 0.4 V				–16			–1.6	mA
I_{OS}¶	V_{CC} = MAX.			–20		–55	–18		–55	mA
I_{CCH}	V_{CC} = MAX.	V_I = 0 V			4	8		4	8	mA
I_{CCL}	V_{CC} = MAX.	V_I = 4.5 V			12	22		12	22	mA

‡ For conditions shown as MIN or MAX use the appropriate value specified under recommended operating conditions.
§ All typical values are at V_{CC} = 5 V, T_A = 25°C
¶ Not more than one output should be shorted at a time.

FIGURE 6.25 Section of data sheet showing the values used in the solution for part (a).

find the typical (TYP) value is 0.2 V and the maximum (MAX) value is 0.4 V. The conditions are as follows:

i. For the MAX value of $V_{OL} = 0.4$ V, the condition is V_{CC} = MIN which we find from the recommended operating condition table to be 4.75 V for the SN7400.
ii. For the TYP value of $V_{OL} = 0.2$ V, the condition is $V_{CC} = 5.0$ V and ambient temperature is $T_A = 25°C$ according to the footnote § below the "Electrical Characteristics" table.
iii. $V_{IH} = 2$ V
iv. $I_{OL} = 16$ mA
v. Ambient temperature range from 0°C to 70°C for the maximum value

For part (b), we will use the last two rows in the electrical characteristics table for I_{CCH} and I_{CCL}. We note that the typical (TYP) values are the same for both parts. The values for I_{CCH} and I_{CCL} are per gate, and this chip contains four gates. We need to devise a reasonable method of averaging over time. It seems most reasonable to assume that each gate will spend half of its life at logic HIGH and half at logic LOW. In this case, we can compute the average power supply current per gate:

$$I_{CC\text{-}AVG} = (1/2)\,(I_{CCL} + I_{CCH}) = (1/2)(4 + 12) \text{ mA} = 8 \text{ mA}$$

Since there are four gates, the average per chip will be

$$I_{CC\text{-}AVG\text{-}CHIP} = 4 \times 8 \text{ mA} = 32 \text{ mA}$$

The conditions are (i) $V_{CC} = 5.0$ V for both the SN5400 and the SN7400 since we are using typical values, (ii) $V_I = 0$ V if logic LOW and 4.5 V if logic HIGH, and (iii) temperature $T_A = 25°C$.

6.7.6 Data Sheet Pages 5–6

The next page has one last table pertaining to the '00 parts—the switching characteristics table. This tells us how fast the logic gate can switch between HIGH and LOW states. The parameter t_{PLH} is the propagation delay from LOW to HIGH on the output; t_{PHL} is the propagation delay from HIGH to LOW on the output. The definitions of these parameters are shown graphically on page 7 of the data sheet. They are essentially the time delays between the halfway points on the input and output waveforms. Usually, the values of t_{PHL} and t_{PLH} are averaged to get a single number called the propagation delay for a given device.

The next three tables on page 5 are repeats of the '00 tables, but now they are for the 'LS00 parts. The same three tables for the 'S00 parts are on page 6 of the data sheet.

6.7.7 Data Sheet Page 7

Page 7 of the data sheet provides details on how parameters are measured and also defines some of the parameters graphically. The measurement circuits can be helpful to a test engineer who needs to set up testing to qualify incoming parts. There are three test circuits shown, although only one of them pertains to the parts on this data sheet. All the parts on this data sheet have a two-state totem-pole output, so only the leftmost test circuit applies. Also, only the two diagrams on the left half of the page apply to these parts. This is an example of a common situation where the data sheet includes generic information and the engineer must determine what is relevant to his or her application. Also, as is common in data sheets, the footnotes contain very important information without which the rest of the page cannot be understood.

6.7.8 Data Sheet Pages 8 and Following

The remainder of the data sheet (which is not included in Appendix B) contains a wealth of information about the mechanical details of the packaged parts. This is intended for use by the engineers and technicians who perform printed circuit board layout and automated assembly equipment setup. Some of the details included in these pages are spacing between pins, chip dimensions, arrangement of the chips on tape and reel dispensers (used in automated assembly), and materials used in the packaging. Dimensions are given to very high precision (0.001 inch or 0.01 mm).

Problems

6.1. Use a reliable reference to identify the generally accepted resistivity range of insulators, semiconductors, and conductors. What are the units used?

6.2. List the metals aluminum, silver, gold, and copper in order from best electrical conductor to worst. What material is used for most electrical power transmission lines? Why is this material used?

6.3. What material is the best electrical insulator? What material is used to insulate most electrical power lines? Why is this material used?

6.4. Produce a table of semiconductors and their bandgap energy values at room temperature. Include at least six semiconductors, and list them from highest bandgap to lowest.

6.5. Write a simple explanation of the existence and behavior of holes in semiconductors.

6.6. Briefly describe the process by which purified silicon wafers are produced for electronics use from common sand.

6.7. Make a 3D sketch of a silicon atom bonded to its four nearest neighbors in its natural crystalline structure. What are the bond angles and bond lengths?

6.8. Draw a sketch similar to Figure 6.3 for p-type Si with boron doping.

6.9. As shown in Figure 6.5, a built-in voltage exists across the p–n junction. However, this voltage cannot be measured directly. For example, if one uses a voltmeter to measure the voltage across the terminals of an unconnected diode, it will read zero. Explain this.

6.10. Explain in simple terms why current can flow in only one direction through a p–n junction.

6.11. Use the diode equation, Equation (6.1), to calculate the current in a diode for which $V_D = 0.700$ V, $n = 1$, $I_S = 1$ pA, and $T = 300$ K.

6.12. Use the diode equation, Equation (6.1), to calculate the current in a diode for which $V_D = 0.600$ V, $n = 1$, $I_S = 1$ nA, and $T = 295$ K.

6.13. Make a table of values of I_D versus temperature for a diode in which $V_D = 0.700$ V, $n = 1$, $I_S = 1$ pA, and $T = 0, 10, 20, \ldots 100$°C. (b) Plot the values calculated in part (a) on a semilog plot (current on a log axis, and temperature on a linear axis).

6.14. Make a table of values of I_D versus V_D for a diode in which $T =$ 25°C, $n = 1$, $I_S = 1$ pA, and $V_D = 0.40, 0.45, \ldots 0.80$ V. (b) Plot the values calculated in part (a).

6.15. Make a table of values of I_D versus V_D for a diode in which $T =$ 25°C, $V_D = 0.700$ V, $I_S = 1$ pA, and $n = 1, 1.1, 1.2, \ldots 2.0$. (b) Plot the values calculated in part (a).

6.16. (a) On the same axes, plot I_D versus V_D for three different temperatures: $T = 15$°C, 25°C, and 35°C. Let $n = 1$ and $I_S = 1$ pA, and choose the values of V_D so that I_D ranges from 0 to 100 mA. (b)

Based on the results from part (a), how does the voltage change with temperature if the current is held constant?

6.17. On the same axes, plot I_D versus T for three different diode voltages: $V_D = 0.65$ V, 0.70 V, and 0.75 V. Let $n = 1$ and $I_S = 1$ pA, and let T range from 0°C to 100°C. Plot on a semilog plot (current on a log axis, and temperature on a linear axis).

6.18. Let $n = 1$, $I_S = 1$ pA, and $T = 300$ K. (a) Calculate the diode voltage if the current is 1 mA, 10 mA, and 100 mA. (b) Electronics engineers often use the approximate "60 mV per decade" rule. This rule states that the current through a forward-biased diode increases by a factor of 10 when the diode voltage increases by 60 mV. Compare your results from part (a) with this rule.

6.19. (a) Which operating region is *not* modeled by the diode equation, Equation (6.1)? (b) What value of V_D will yield $I_D = -I_S/2$? (c) What value of V_D will yield $I_D = -0.9I_S$? Assume $V_T = 25$ mV and $n = 1$ for (b) and (c).

6.20. A product designer wishes to use Equation (6.2) as the basis for an electronic medical thermometer. The product is intended to function over a range from 85°F to 115°F. (a) What is the range of diode voltages that will occur over this temperature range if $I_D = 1$ mA, $n = 1$, and $I_S = 1$ pA? (b) Plot V_D versus T.

6.21. Sketch the load voltage, V_2, that you would expect for the half-wave rectifier circuit in Figure 6.8 if the diode is ideal and V_{AC} is sinusoidal with a peak value of 170 V.

6.22. Sketch the load current that you would expect for the half-wave rectifier circuit in Figure 6.8 if the diode is ideal, V_{AC} is sinusoidal with a peak value of 170 V, and the load resistor has a value of 100 Ω.

6.23. Calculate the average power delivered to the load for the half-wave rectifier circuit in Figure 6.8 if the diode is ideal, V_{AC} is sinusoidal with a peak value of 170 V, and the load resistor has a value of 100 Ω.

6.24. Sketch the load voltage that you would expect for the full-wave bridge rectifier circuit in Figure 6.9 if the diode is ideal and V_{AC} is sinusoidal with a peak value of 170 V.

6.25. Sketch the load current that you would expect for the full-wave bridge rectifier circuit in Figure 6.9 if the diode is ideal, V_{AC} is sinusoidal with a peak value of 170 V, and the load resistor has a value of 100 Ω.

6.26. Calculate the average power delivered to the load for the full-wave bridge rectifier circuit in Figure 6.9 if the diode is ideal, V_{AC} is sinusoidal with a peak value of 170 V, and the load resistor has a value of 100 Ω.

6.27. An n-channel enhancement MOSFET has $K_n = 100$ µA, $\lambda = 0.01$, and $V_{Tn} = 1.00$ V. Calculate the drain current if $V_G = 5$ V, $V_S = 3$ V, and (a) $V_D = 6$ V and (b) $V_D = 3.5$ V.

6.28. An n-channel enhancement MOSFET has $K_n = 100$ µA, $\lambda = 0.01$, $V_{GS} = 2$ V, and $V_{Tn} = 1.00$ V. Plot I_D versus V_{DS} for $0 \le V_{DS} \le 5$ V.

6.29. An n-channel enhancement MOSFET has $K_n = 100$ µA, $\lambda = 0.01$, and $V_{Tn} = 1.00$ V. Plot the output characteristics: I_D versus V_{DS} for $0 \le V_{DS} \le 5$ V, and $V_{GS} = V_{Tn}$, $V_{Tn} + 1$ V, $V_{Tn} + 2$ V, … $V_{Tn} + 5$ V.

6.30. An n-channel enhancement MOSFET has $K_n = 100$ µA, $V_{GS} = 2$ V, and $V_{Tn} = 1.00$ V. Plot I_D versus V_{DS} for $0 \le V_{DS} \le 5$ V, with $\lambda =$ 0.01, 0.02, and .03 V^{-1}.

6.31. An n-channel depletion MOSFET has $K_n = 100$ µA, $\lambda = 0.01$, and $V_{Tn} = -1.00$ V. Calculate the drain current if $V_G = 5$ V, $V_S = 3$ V, and (a) $V_D = 6$ V and (b) $V_D = 3.5$ V.

6.32. An n-channel depletion MOSFET has $K_n = 100$ µA, $\lambda = 0.01$, $V_{GS} =$ 2 V, and $V_{Tn} = -1.00$ V. Plot I_D versus V_{DS} for $0 \le V_{DS} \le 5$ V.

6.33. An n-channel depletion MOSFET has $K_n = 100$ µA, $\lambda = 0.01$, and $V_{Tn} = -1.00$ V. Plot the output characteristics: I_D versus V_{DS} for $0 \le V_{DS} \le 5$ V, and $V_{GS} = V_{Tn}$, $V_{Tn} + 1$ V, $V_{Tn} + 2$ V, … $V_{Tn} + 5$ V.

6.34. An n-channel depletion MOSFET has $K_n = 100$ µA, $V_{GS} = 2$ V, and $V_{Tn} = -1.00$ V. Plot I_D versus V_{DS} for $0 \le V_{DS} \le 5$ V, with $\lambda = 0.01$, 0.02, and .03 V^{-1}.

6.35. A p-channel enhancement MOSFET has $K_p = 100$ µA, $\lambda = 0.01$, and $V_{Tp} = -1.00$ V. Calculate the drain current if $V_S = 5$ V, $V_G = 3$ V, and (a) $V_D = 3$ V and (b) $V_D = 4.5$ V.

6.36. A p-channel enhancement MOSFET has $K_p = 100$ µA, $\lambda = 0.01$, $V_{SG} = 2$ V, and $V_{Tp} = -1.00$ V. Plot I_D versus V_{SD} for $0 \le V_{SD} \le 5$ V.

6.37. A p-channel enhancement MOSFET has $K_p = 100$ µA, $\lambda = 0.01$, and $V_{Tp} = -1.00$ V. Plot the output characteristics: I_D versus V_{SD} for $0 \le V_{SD} \le 5$ V, and $V_{SG} = -V_{Tp}$, $-V_{Tp} + 1$ V, $-V_{Tp} + 2$ V, … $-V_{Tp} + 5$ V.

6.38. A p-channel enhancement MOSFET has $K_p = 100$ µA, $V_{GS} = 2$ V, and $V_{Tp} = -1.00$ V. Plot I_D versus V_{DS} for $0 \le V_{DS} \le 5$ V, with $\lambda =$ 0.01, 0.02, and .03 V^{-1}.

6.39. A p-channel depletion MOSFET has $K_p = 100\ \mu A$, $\lambda = 0.01$, and $V_{Tp} = 1.00$ V. Calculate the drain current if $V_S = 5$ V, $V_G = 3$ V, and (a) $V_D = 3$ V and (b) $V_D = 4.5$ V.

6.40. A p-channel depletion MOSFET has $K_p = 100\ \mu A$, $\lambda = 0.01$, $V_{SG} =$ 2 V, and $V_{Tp} = 1.00$ V. Plot I_D versus V_{SD} for $0 \le V_{SD} \le 5$ V.

6.41. A p-channel depletion MOSFET has $K_p = 100\ \mu A$, $\lambda = 0.01$, and $V_{Tp} = 1.00$ V. Plot the output characteristics: I_D versus V_{SD} for $0 \le V_{SD} \le 5$ V, and $V_{SG} = -V_{Tp}, -V_{Tp} + 1$ V, $-V_{Tp} + 2$ V, … $-V_{Tp} + 5$ V.

6.42. A p-channel depletion MOSFET has $K_p = 100\ \mu A$, $V_{GS} = 2$ V, and $V_{Tp} = 1.00$ V. Plot I_D versus V_{DS} for $0 \le V_{DS} \le 5$ V, with $\lambda = 0.01$, 0.02, and 0.03 V^{-1}.

6.43. Follow the procedure in Example 6.4 to design a MOSFET switching circuit. The load is resistive with $V_{DC} = 15$ V and a maximum current of 3 A.

6.44. The Infineon BSR202N n-channel enhancement MOSFET has $K_n = 8.00\ A/V^2$, $\lambda = 0.170\ V^{-1}$, and $V_{Tn} = 1.00$ V. Computer simulation of the amplifier circuit in Figure 6.19 shows that the bias point values are $V_D = 4.930$ V, $V_S = 3.857$ V, and $V_G = 4.865$ V, and $I_D =$ 321 μA. (a) Determine the operating region of the MOSFET M_1. (b) Using the appropriate MOSFET equation, calculate I_D using the simulated terminal voltages. (c) Calculate V_G directly from the circuit using voltage division, and compare with the computer simulated result.

6.45. A certain npn BJT has $V_B = 0.7$ V, $V_E = 0$ V, and $V_C = 10$ V. What is the operating region?

6.46. A certain npn BJT has $V_B = 0.7$ V, $V_E = 0$ V, and $V_C = 0.1$ V. What is the operating region?

6.47. A pnp BJT has $V_B = 5$ V, $V_C = 2$ V, and $V_E = 3$ V. What is the operating region?

6.48. Calculate the terminal currents I_C, I_B, and I_E for an npn BJT with $I_S = 1.00$ nA, $T = 295$ K, $\beta = 100$, and $V_A = 75$ V. Let $V_{BE} = 0.700$ V and $V_{CE} = 10.0$ V.

6.49. Determine the typical and extreme values of the output low voltage for the SN54LS00.

6.50. Determine the average typical power supply current for the SN54LS00 and for the SN74LS00. Compare with those for the ‘5400 and ‘7400 calculated in the text.

This page intentionally left blank

Chapter 7

Sensors and Instrumentation

Most electronic circuits are designed to interface ultimately with the real world. This requires acquiring input signals related to physical phenomena and/or producing output signals resulting from calculations, filtering, or other types of signal processing.

Sensors as discussed here are devices that convert physical stimulation into electrical signals suitable for input into a circuit or computer system. Table 7.1 lists some common sensor types.

In almost all cases of interest, the raw sensor output will need to be *conditioned* before connecting to the inputs of another device such as an amplifier or computer. "Conditioning" refers to the application of physical and/or electrical preprocessing, often within the sensor housing, which provides functions such as scaling, linearization, impedance modification, data storage, and/or digitizing. For example, a digital video camera records a stream of numerical data which represents the scene impinging on the lens. The front-end electronics and optics interact with the CCD or CMOS imaging chip to generate

Table 7.1 Common Sensor Types

Sensor	Stimulation (Example Units)	Example Applications
Accelerometer	Linear acceleration (meters per second)	Seat-belt tensioners in automobiles, air-bag initiators
Gyroscope ("gyro")	Angular rate (degrees per second)	Flight control in aircraft, steering control for unmanned vehicles
Thermocouple, thermistor, or resistance temperature detector (RTD)	Temperature (degrees C)	Industrial oven temperature, consumer digital thermometers, household or vehicle thermostat
Photocell	Light intensity (watts per sq cm)	Dusk or dawn sensor for automated lighting
Sonar	Distance to solid object (cm)	Obstacle sensing for robots
Current transformer	AC current (amperes)	Measuring current in an industrial motor
Microphone	Sound pressure level (dB)	Picking up the voice of a person speaking or singing

a rendering of the external world in an analog format. This analog signal is converted to a digital bit stream through the process of analog-to-digital conversion (ADC), and then recorded onto a suitable medium such as DVD, flash memory card, hard drive, or tape.

Example 7.1 illustrates the problem of designing a simple circuit to interface a photocell to a digital logic chip.

Example 7.1: A Photocell Interface Circuit

A photocell is used to detect light intensity. This photocell is interfaced to a logic chip. If the light is bright enough, we want to create a logic-high output signal; otherwise, the output should be logic low.

This example will use a photocell with the following characteristics:[1]

$$\text{Resistance in maximum illumination} = R_{\text{light}} = 5\ \text{k}\Omega$$

$$\text{Resistance in dark} = R_{\text{dark}} = 500\ \text{k}\Omega$$

We will use a digital logic inverter to produce the output. One important consideration when interfacing sensors, especially for slowly changing inputs, is eliminating chatter. "Chatter" refers to the situation that occurs at the threshold due to noise. If the stimulation (light intensity in this example) is near the threshold where we desire to switch the digital output from low to high, the presence of noise (which is inevitable in any real system) will cause the output to rapidly jump back and forth between high and low. This can confuse or physically damage the next stage of the system which might be designed to take significant action (such as opening or closing a door) each time a transition is detected.

The usual solution to chatter is *hysteresis*. When a system has hysteresis, it actually has two thresholds—one (higher) for when the stimulation is increasing, and the other (lower) for when the stimulation is decreasing. If the light intensity is increasing and crosses the upward threshold, a small amount of noise will be insufficient to push it backward across the lower threshold. Likewise, when the light intensity is decreasing and crosses the lower threshold on the way down, a small amount of noise cannot push it back up across the upper threshold.

[1] An example of a photocell with these characteristics is the PHOTOCELL1 from http://www.futurlec.com/Photocells.shtml. It lists for 25 cents.

In this way, the probability of chatter can be greatly reduced by carefully choosing the amount of hysteresis (the difference between upper and lower thresholds) with regard to the amount of noise expected. Of course, there will always be a noise level that can exceed the hysteresis and cause unintended switching.

In response to this need, manufacturers have produced digital logic chips that have hysteresis built in. This function is provided by "Schmitt trigger inputs." This terminology refers to the specific circuitry that implements the hysteresis within the chip.

We will use such a chip for our photocell circuit. We select the NC7SZ14, which is a CMOS device capable of operating from a supply voltage as high as 5.5 V. We will choose a supply voltage for this example to be $V_{CC} = 3.0$ V. In this case, the data sheet tells us the typical negative threshold voltage is 1.00 V, and the typical positive threshold voltage is 1.75 V. Taking the average of these thresholds gives us roughly 1.4 V. In order to simplify our design somewhat, we will use 1.5 V as our target switching level. This is convenient since it is one-half of the power supply voltage. The resistance of the conventional photocell changes in a nonlinear way with light intensity. (One can purchase linearized devices if desired, although they cost more.)

One way to accomplish a design would be to measure the resistance that corresponds to the desired threshold light intensity, but often that is difficult to do experimentally. A more typical approach is to include a variable resistor (a potentiometer, usually just called a "pot") in the circuit which can be adjusted to vary the threshold.

A suitable circuit is shown in Figure 7.1. The potentiometer R_{thresh} is used to adjust the threshold. For the purpose of illustration, let us suppose that R_{thresh} is set to 100 kΩ. In the dark, the photocell will have resistance

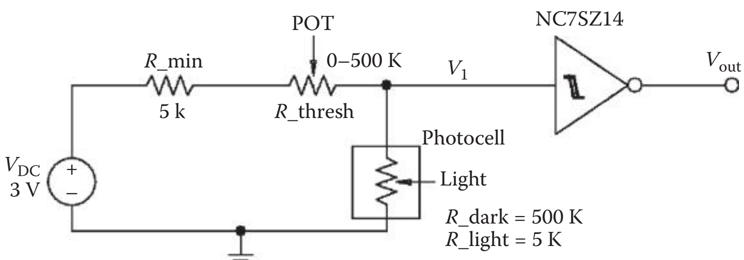


FIGURE 7.1 Photocell interface circuit.

of 500 kΩ, so by voltage division, the input to the logic inverter will be

$$V_1(\text{dark}) = \frac{R_{\text{dark}}}{R_{\text{thresh}} + R_{\text{min}} + R_{\text{dark}}} V_{DC} = \frac{500\, k\Omega}{105\,\text{k}\Omega + 500\,\text{k}\Omega} 3\text{V} = 2.48\ \text{V} \tag{7.1}$$

In the light, the photocell resistance will decrease to $R_{\text{light}} = 5$ kΩ resulting in $V_1 = 0.05$ V. When these values are provided as input to the digital logic inverter, the corresponding outputs will be logic low (in the dark) and logic high (in the light).

It is easy to ascertain that for any setting of the threshold potentiometer, there is always a light intensity that will cause the output to equal the switching threshold of 1.5 V. For greater light intensity, the output will become smaller, and for lesser light intensity, the output will become larger. Thus, the logic inverter output will behave as desired.

7.1 Mechanical Sensors: Accelerometers and Strain Gauges

Mechanical sensors such as accelerometers, strain gauges, gyroscopes, tilt sensors, and contact switches find wide use in industrial process control, navigation systems, and games, among many other applications. In this section, we will look at two types of devices—accelerometers and strain gauges.

7.1.1 Accelerometers

Accelerometers measure acceleration (rate of change of speed). The available types include single-axis and multi-axis (two- or three-axis) devices. Single-axis accelerometers respond, ideally, only to acceleration in the specified direction relative to the device package. Multi-axis accelerometers are made by mounting several single-axis devices in a single housing, or by using a single mechanical design that responds separately along each axis.

An accelerometer is typically attached to some object that is expected to undergo acceleration. The experienced acceleration may be due to a change in speed if the object is in motion. Another common application is to measure tilt with respect to the earth's gravitational field. This is used, for example, in mobile phones and handheld game controllers. In this latter case, a reading is obtained even when the device is not moving. Although any appropriate units can be used to characterize the response range of an accelerometer, one common unit is the *g*. One *g* is the standard value of acceleration for the earth's gravitational pull, 9.8 cm/s^2. Accelerometers are available which can measure from the micro-*g* range to hundreds of kilo-*g*'s. The latter values are experienced by projectiles launched from guns, while the former values can be used to track the motion of the moon, or detect miniscule vibrations of the earth's surface in response to geological events like distant earthquakes.

A typical device is the Analog Devices ADXL335. This is a three-axis accelerometer with a full-scale range of 3 *g*. Examination of the data sheet reveals a list of intended applications including mobile devices, gaming systems, disk drive protection, image stabilization, and sports and health devices. The manufacturer also specifies shock survival up to 10,000 *g*, a small size of 4 mm × 4 mm × 1.45 mm, and partially conditioned (amplified) outputs. Some signal conditioning is left up to the user, since some circuit component values governing the bandwidth and resolution are application dependent.

7.1.2 Strain Gauges[2]

Strain gauges are sensors which measure change in dimension for items that are subject to stress, such as bending beams. Usually, an engineer wants to know the relative (percentage) change along a particular dimension. Metal foil strain gauges are common. They are normally glued to the device under test (DUT), and wired to an instrument that can measure the change in resistance that occurs when the DUT is stressed. The ratio of relative resistance change to relative dimension change is called the gauge factor (GF) or, equivalently, the strain sensitivity (k). Most common metal foil strain gauges have a gauge factor of approximately 2.0.[3] The precise value is usually provided on the data sheet or on the shipped package for a particular batch. Example 7.6 in Section 7.5 provides an illustration of interfacing a strain gauge.

7.2 Acoustic Sensors: Microphones and Sonar

7.2.1 Microphones

Microphones convert sound waves into electrical signals. Besides the common applications of amplifying speech and music, microphones are widely used to monitor and diagnose mechanical equipment. Unusual vibrations in

[2] The spellings "gauge" and "gage" are equally accepted in U.S. English.

[3] http://www.efunda.com/designstandards/sensors/strain_gages/strain_gage_sensitivity.cfm.

equipment can produce sounds which can be traced to specific types of degradation or failure, such as shaft imbalance or gear wear.

The majority of microphones for consumer use are of the electret condenser type. In these, one plate of a capacitor is moved by the acoustic pressure wave. Fixed charge is supplied by a permanently polarized electret material. The resulting output is a voltage corresponding to the instantaneous sound pressure level.

Microphone selection parameters include element type, impedance levels (e.g., low or high), acoustic pattern (omni, cardioids, etc.), and connector types for specific applications.

7.2.2 Sonar (Ultrasound) Sensors

Sonar sensors transmit a pulse of high-frequency sound waves (ultrasound) and then listen for the reflection. The total round-trip time of the sound pulse is used to determine distance.

Sonar was originally an acronym for SOund Navigation And Ranging. One of the most common consumer-grade sonar sensors, widely used in robotics, is the PING))) sold by Parallax. This device can measure distance to an object at distances ranging from about 2 cm to 3 m. The resolution of the distance measurement is largely determined by how precisely the user can measure the duration of the echo pulse. Changing air temperature will change the apparent distance reported by a sonar, since the speed of sound in air changes with temperature.

Sonar sensors are typically used to determine range (distance) to nearby, large objects. They are used, for example, as obstacle indicators on automobiles. In the latter application, they are sometimes referred to as "parking assist" sensors.

Example 7.2: Sonar Sensor Used to Stop a Warehouse Cart

A *PING)))* sonar sensor is used to detect obstacles in front of an automated cart in a warehouse. Design an interface that will send a "STOP" signal when it detects an object closer than 1 meter in front.

Solution:

According to the PING))) data sheet, the sensor needs three wires: +5 VDC power, ground, and signal. Standard three-wire servo connectors, available in multiples of 6-inch lengths, can be used to connect directly from the PING))) pins to a header on the circuit board containing our interface circuit. The standard color code is red for power, black for signal, and yellow (or white) for signal.

Our interface circuit should provide +5 VDC and ground needed by the PING))). It must also generate a trigger pulse at regular intervals. Each trigger pulse will initiate a measurement. The more often a trigger is produced, the more likely we will detect objects moving quickly into the path of the cart. After each trigger pulse, our circuit must go into a timing mode to determine the length of the echo pulse.

One interesting thing about the PING))) is that the same pin is used by the PING))) to receive the trigger pulse and to send back the echo pulse. In other words, the signal line is bidirectional—it is used to both send and receive information.

Although one could design a mixed-signal (analog and digital) circuit from basic parts to accomplish this, this application just begs for us to use a microcontroller. In fact, the PING))) is intended for microcontroller interface, as is clear from reading the data sheet.

Considering the wide variety of microcontrollers available, the most beneficial starting point for a design is at a schematic and pseudo-code level. A system-level drawing is shown in Figure 7.2.

The pseudo-code to perform the desired action is shown below.

Assumptions:

- Microcontroller output pin $S = 1$ will send a stop signal to the cart drive system, and $S = 0$ will enable the cart to move.

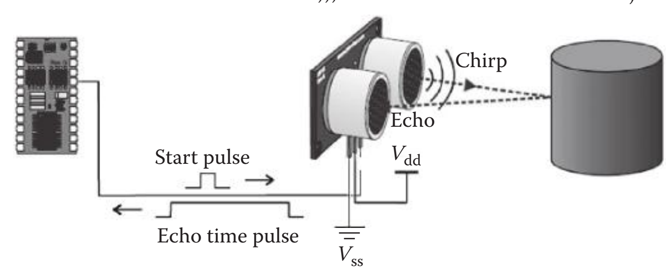


FIGURE 7.2 PING))) sonar sensor interface to a generic microcontroller.

- Microcontroller pin *P* is a bidirectional logic-level interface to the PING))).
- *scale_factor* is the multiplier for converting echo return time to distance. This will be determined from calibration or from a reference source.

Syntax:

`OUT` *`pin _ name`*`,` *`time`* produces a high output for *`time`* on pin `pin _ name`.
`IN` *`pin _ name`*`,` *`t _ in`* reads the duration of a high-level pulse appearing on `pin` *`pin _ name`*.
SET *`pin _ name`* forces pin *`pin _ name`* to logic 1.
`RESET` *`pin _ name`* forces pin *`pin _ name`* to logic 0.

Begin Loop Forever:	
RESET S	/*Initialize to no object detected—cart can move */
WAIT 200 us	/*force 200 µs delay between measurement pulses, per data sheet */
OUT P, 5us	/*send out a 5 µs trigger pulse on pin P to start measurement, per data sheet */
IN P, time_in	/* get duration of echo return pulse */
Distance = time_in * scale_factor	/*compute distance to object */
IF (dist ≤ 1 m)	/* compare with 1 meter */
SET S	/* if object is within 1 meter, put logic 1 on pin S to stop the cart */
ELSE	
RESET S	/* if no object within 1 meter, put logic 0 on pin S so cart can move */
End Loop Forever	

7.3 Optical Sensors: IR Sensors and Photoresistors

Many sensors respond passively to light, or use light actively to detect objects. Two of the more common types are infrared (IR) range sensors and photoresistors.

7.3.1 Infrared (IR) Range Sensors

Sonar sensors are not very good at detecting small objects, due to their fairly large beam width. Higher spatial resolution can often be obtained using IR range sensors. The trade-off is that IR sensors are typically restricted to shorter distances, and their output voltage is highly nonlinear with respect to distance. For example, the Sharp GP2Y0A21YK0F Analog Distance Sensor has an output voltage–distance relationship shown in Figure 7.3. The data sheet for this device states that it is designed for measuring distance in the range of 10 cm to 80 cm. This may be compared with the maximum distance of 3 m (300 cm) for the PING))) sonar sensor. For IR range sensors, best performance is obtained when the sensed object is optically reflective. Shiny, light-colored surfaces work better than dull or dark-colored surfaces. These sensors work by emitting a chopped (turned on and off at a special frequency) IR light pulse using an LED, and then detecting the reflection of that pulse

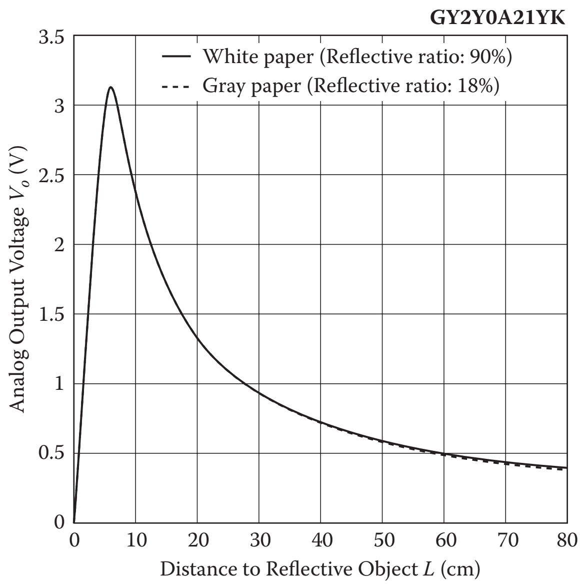


FIGURE 7.3 Response curve for an infrared range sensor.

using a special device called a "position-sensitive detector" (PSD). The PSD is an array of very small photodiodes. The distance of the reflecting surface from the sensor determines the angle of the reflected beam, and thus determines which photodiode is activated. This is converted to a voltage by the internal circuitry of the sensor. Because the emitted light beam is chopped at a known frequency, it is much easier for the detection circuit to lock in on the reflection even in the presence of varying background illumination. As a consequence, these sensors can work over a wide range of varying light levels.

7.3.2 Photocells

Photoresistors (sometimes called photocells) are two-terminal devices that are designed to have a wide variation of resistance as incident illumination is changed. An example is the Advanced Photonix, Inc. PDV-P8101. This device is specified to have a dark resistance of 150 kΩ and a minimum resistance of 4 kΩ under illumination. A photograph is shown in Figure 7.4, and a plot of resistance versus illumination is shown in Figure 7.5.

7.4 Thermal Sensors: Thermocouples, Thermistors, RTDs, and PTATs

Temperature measurement is important for two primary reasons. On the one hand, process control often depends critically on accurate temperature measurement. On the other hand, many other types of measurement need to be temperature compensated for accuracy. For example, conversion of raw sonar reading (echo return time) to distance depends on the speed of sound in air, which is temperature dependent.

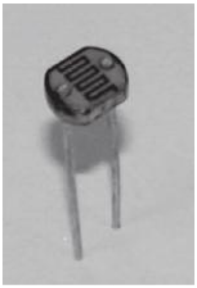


FIGURE 7.4 Photocell.

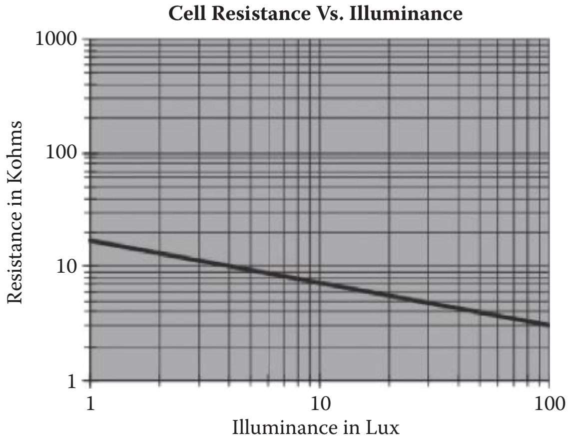


FIGURE 7.5 Resistance variation with light intensity.

Several common temperature measurement technologies are discussed in the following paragraphs.

7.4.1 Thermocouples

Thermocouple temperature sensors produce a temperature-dependent voltage. They are made by bonding together the ends of two different metal wires, often by spot welding. The principle of operation is the Seebeck effect, also known as the "thermoelectric effect," which relates the temperature difference along a wire to an induced voltage difference. The amount of voltage difference produced for a given temperature difference depends on the wire material. When two unlike materials are joined, the measurement junction ("hot end") is forced to be at a common potential, so a voltage difference will be measured between the two "cold" terminals.

Thermocouples are classified by the materials used. Table 7.2 lists some of the most common types. A brief discussion of each follows.

Type J (iron–constantan) has a more restricted range than type K, but a higher sensitivity of about 55 μV/°C. The Curie point of the iron (770°C) causes an abrupt change in the characteristic, which determines the upper temperature limit.

Type K (chromel [90% nickel and 10% chromium]–alumel [consisting of 95% nickel, 2% manganese, 2% aluminum, and 1% silicon]) is the most common general-purpose thermocouple with a sensitivity of approximately

Table 7.2 Common Thermocouples

Type	Temperature Range	Materials
J	0°C to 750°C 32°F to 1382°F	Iron–constantan
K	–200°C to 1250°C –328°F to 2282°F	Chromel (90% nickel and 10% chromium)–alumel (alumel consisting of 95% nickel, 2% manganese, 2% aluminum, and 1% silicon)
E	–200°C to 900°C –328°F to 1652°F	Chromel–constantan
T	–250°C to 350°C –328°F to 662°F	Copper–constantan

41 µV/°C, with chromel positive relative to alumel. It is inexpensive, and a wide variety of probes are available in its range. Type K was specified at a time when metallurgy was less advanced than it is today, and consequently characteristics vary considerably between samples. One of the constituent metals, nickel, is magnetic; a characteristic of thermocouples made with magnetic material is that they undergo a step change in output when the magnetic material reaches its Curie point (around 354°C for type K thermocouples).

Type E (chromel–constantan) has a high output (68 µV/°C) which makes it well suited to cryogenic use. Additionally, it is nonmagnetic.

Type T (copper–constantan) thermocouples are suited for measurements in the –200°C to 350°C range. They are often used as a differential measurement since only copper wire touches the probes. Since both conductors are nonmagnetic, there is no Curie point and thus no abrupt change in characteristics. Type T thermocouples have a sensitivity of about 43 µV/°C.

A comprehensive database of thermocouple calibrations is available online from the National Institute of Standards and Technology (NIST). The data tables can be viewed online or downloaded. The data are provided in tabular form, and curve–fit coefficients are also provided.

These tables (like most) provide the thermoelectric voltage that will be measured at a reference junction that is at 0°C. If the reference junction[4] is at a different temperature (such as room temperature), a correction must be applied. This process is demonstrated in Example 7.3.

[4] The reference junction is also called the "cold junction."

Example 7.3: Thermocouple Calculations

A type K thermocouple is used to measure the temperature inside an industrial furnace. The cold junction (reference junction) is at 27°C. (a) Determine the temperature corresponding to an output of 16 mV. (b) If the temperature is 700°C, what voltage output do you expect? Perform these calculations using (1) table lookup and (2) polynomial calculation.

Solution:

a. The steps required to calculate the thermocouple temperature are as follows:

Step 1. Determine the cold junction temperature.

Step 2. Determine the compensating voltage by calculation or table lookup.

Step 3. Add the compensating voltage to the measured (raw) Seebeck voltage to obtain the adjusted Seebeck voltage.

Step 4. Look up or calculate the temperature corresponding to the adjusted Seebeck voltage.

The NIST thermocouple database[5] provides voltage values at every 1°C, as well as coefficient tables for converting from temperature to voltage (used in Step 2) and from voltage to temperature (used in Step 4).

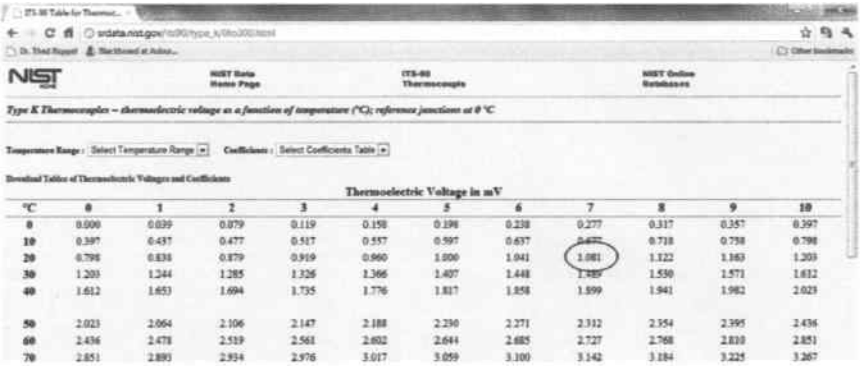
METHOD 1: TABLE LOOKUP

Step 1. The cold junction temperature is given in the problem statement. The value stated is 27°C. In practice, this might be determined using an electronic temperature sensor, such as a proportional-to-absolute temperature (PTAT) circuit, a resistance temperature detector (RTD), or a thermistor.

Step 2. A section of the NIST table for type K thermocouples is shown in Figure 7.6. This segment of the table covers the temperature range from 0°C to 80°C. Reading the table at 27°C (circled in the figure), we find $V_{comp} = 1.081$ mV.

Step 3. The Seebeck voltage is given as 16 mV. Since the NIST tables have microvolt resolution, we will write this as 16.000 mV for the purpose of table lookup.

[5] http://srdata.nist.gov/its90/type_k/0to300.html.


Type K Thermocouples – thermoelectric voltage as a function of temperature (°C); reference junctions at 0 °C

Temperature Range : Select Temperature Range
Coefficients : Select Coefficients Table

Download Tables of Thermoelectric Voltages and Coefficients

Thermoelectric Voltage in mV

°C	0	1	2	3	4	5	6	7	8	9	10
0	0.000	0.039	0.079	0.119	0.158	0.198	0.238	0.277	0.317	0.357	0.397
10	0.397	0.437	0.477	0.517	0.557	0.597	0.637	[illegible]	0.718	0.758	0.798
20	0.798	0.838	0.879	0.919	0.960	1.000	1.041	1.081	1.122	1.163	1.203
30	1.203	1.244	1.285	1.326	1.366	1.407	1.448	[illegible]	1.530	1.571	1.612
40	1.612	1.653	1.694	1.735	1.776	1.817	1.858	1.899	1.941	1.982	2.023
50	2.023	2.064	2.106	2.147	2.188	2.230	2.271	2.312	2.354	2.395	2.436
60	2.436	2.478	2.519	2.561	2.602	2.644	2.685	2.727	2.768	2.810	2.851
70	2.851	2.893	2.934	2.976	3.017	3.059	3.100	3.142	3.184	3.225	3.267

FIGURE 7.6 **Section of the NIST tables for type K thermocouples.**

Step 4. Adding the compensating voltage from Step 2, $V_{net} = V_{raw} + V_{comp} = 16.000$ mV + 1.081 mV = 17.081 mV.

Step 5. By browsing the NIST tables, we find the closest entry to this value is 17.074 mV, which corresponds to 416°C.

METHOD 2: POLYNOMIAL CALCULATION

In many cases, a microcomputer is used to interface a TC. Then it is often impractical to store the entire table of thermocouple voltages. Instead, one would use curve–fit polynomials to calculate the voltage given temperature, or temperature given voltage. The NIST database provides the needed polynomial coefficients. In the NIST table, the "inverse coefficients" are those used to find temperature given Seebeck voltage. The (noninverse) coefficients are those used to find Seebeck voltage given temperature.

For the present example, we find the following from the NIST tables:

$$T = d_0 + d_1 E + d_2 E^2 + \cdots + d_n E^n \tag{7.2}$$

where E is the thermoelectric (Seebeck) voltage in mV and T is the temperature in °C.

Furthermore, from the NIST tables we can find the inverse coefficients (the d_i values in Equation 7.2) for type K thermocouples. There are two different sets depending upon the range into which the thermoelectric voltage falls. In our case, the value of E falls into the range from 0 to 20.644 mV. The corresponding inverse coefficients are listed in Table 7.3.

Table 7.3 Inverse Coefficients for Type K Thermocouples for Thermoelectric Voltage in the Range from 0 to 20.644 mV

i	d_i
0	0.000000E+00
1	2.508355E+01
2	7.860106E-02
3	−2.503131E-01
4	8.315270E-02
5	−1.228034E-02
6	9.804036E-04
7	−4.413030E-05
8	1.057734E-06
9	−1.052755E-08

Placing the d_i values from Table 7.3 into Equation (7.2) and using $E = V_{net} = 17.081$ mV yields a temperature of 416°C, which is the same result obtained by table lookup.

(b) In this case, the steps are as follows:

Step 1. Determine the cold junction temperature.

Step 2. Calculate or look up the compensating voltage, V_{comp}.

Step 3. Calculate or look up the Seebeck voltage corresponding to the given temperature. This is actually the adjusted Seebeck voltage, V_{adj}.

Step 4. Subtract the compensating voltage from the Seebeck voltage to get the raw Seebeck voltage: $V_{raw} = V_{adj} - V_{comp}$. This is what we expect to measure.

Steps 1 and 2 were done in part (a), so we know $V_{comp} = 1.081$ mV.

Step 3: For method 1, we find by table lookup that $V_{adj} = 29.129$ mV at T = 700°C.

Step 4: $V_{raw} = 29.129 \text{ mV} - 1.081 \text{ mV} = 28.048$ mV.

If we perform Step 3 by method 2, polynomial calculation, we need the correct set of coefficients and the form of the equation. From the NIST

Table 7.4 Coefficients for Type K Thermocouples for Temperatures above 0°C

Coefficient	Value
a_0	0.118597600000E+00
a_1	−0.118343200000E-03
a_2	0.126968600000E+03
c_0	−0.176004136860E-01
c_1	0.389212049750E-01
c_2	0.185587700320E-04
c_3	−0.994575928740E-07
c_4	0.318409457190E-09
c_5	−0.560728448890E-12
c_6	0.560750590590E-15
c_7	−0.320207200030E-18
c_8	0.971511471520E-22
c_9	−0.121047212750E-25

database we find that the equation above 0°C is of the form:

$$E = a_o e^{a_1(T-a_2)^2} + \sum_{i=0}^{n} c_i T^i \tag{7.3}$$

where e is the natural logarithm constant, E is in mV, and T is in °C. The values of the coefficients a_0–a_2 and c_i are shown in Table 7.4.

Using $T = 700$°C and the coefficients from Table 7.4 in Equation (7.3) yields $E = 29.129$ mV, which is the same as that obtained by table lookup.

7.4.2 RTD

Another type of temperature sensor that finds widespread use is the resistance temperature detector, or RTD. The RTD is one of the most accurate temperature sensors. Not only does it provide good accuracy, but also it provides excellent stability and repeatability. RTDs are also relatively immune to electrical noise and therefore well suited for temperature measurement in industrial environments, especially around motors, generators, and other

Table 7.5 RTD Tolerance Classes for the European Calibration

Tolerance Class	Tolerance
European Class AA	±(0.1 + 0.0017*T)°C
European Class A	±(0.15 + 0.002*T)°C
European Class B	±(0.3 + 0.005*T)°C
European Class 1/10 B	±1/10 (0.3 + 0.005*T)°C

high-voltage equipment. Most RTD elements consist of a length of fine-coiled platinum wire wrapped around a ceramic or glass core, or alternatively a thin film of platinum patterned onto a ceramic substrate. The element is usually quite fragile, so it is often placed inside a sheathed probe to protect it.

The temperature range for commercially available thin-film RTDs is from –50°C to 500°C, whereas for wire-wound RTDs it is from –200°C to 850°C.

There are two calibration standards used in practice. The European standard, also known as the DIN or IEC standard, is considered the worldwide standard for platinum RTDs. This standard, DIN/IEC 60751 (or simply IEC751), requires the RTD to have an electrical resistance of 100.00 Ω at 0°C and a temperature coefficient of resistance (TCR) of 0.00385 Ω/Ω/°C between 0°C and 100°C. There are several standard tolerances available, depending on the precision required. These are listed in Table 7.5.

The American standard, used mostly in North America, has a resistance of 100.00 ±0.10 Ω at 0°C and a temperature coefficient of resistance (TCR) of 0.00392 Ω/Ω/°C nominal between 0°C and 100°C.

For high-accuracy work, the RTD resistance must be corrected for nonlinearity. This can be accomplished by using suitable equations and parameters, such as the Callendar–van Dusen equations and coefficients, or by table lookup. One source of RTD tables is Omega Engineering, Inc.[6]

7.4.3 Thermistor

A thermistor is a temperature-sensing element composed of sintered semiconductor material which exhibits a large change in resistance proportional to a small change in temperature. Thermistors are highly nonlinear and work over a limited temperature range of about 0°C to 100°C. However, their principal

[6] http://www.omega.com/temperature/z/pdf/z252-254.pdf.

advantages are large response (significant resistance change for small temperature change) and low cost. These features make them ideal for use in consumer devices such as inexpensive medical thermometers. Thermistors usually have negative temperature coefficients (NTC), which means the resistance decreases as the temperature increases, although positive temperature coefficient (PTC) thermistors are available.

As a consequence of their extreme nonlinearity, it is essential to use a set of calibration equations and parameters, or a table lookup, in order to convert resistance to temperature using a thermistor. A table of resistance values versus temperature for various thermistors can be found at the Omega Engineering, Inc. Web site.[7] A widely used mathematical model is the Steinhart–Hart equation:

$$\frac{1}{T} = a + b\ln(R) + c(\ln(R))^3 \tag{7.4}$$

where a, b, and c are called the Steinhart–Hart parameters, and must be specified for each device. Here T is the temperature in degrees Kelvin and R is the resistance in ohms. The Steinhart–Hart equation is unwieldy to calculate and somewhat difficult to invert (i.e., to solve for R given T). With some loss of accuracy, it can be simplified to the so-called B-parameter equation:

$$\frac{1}{T} = \frac{1}{T_o} + \frac{1}{B}\ln\left(\frac{R}{R_o}\right) \tag{7.5}$$

where the temperatures are in degrees Kelvin and R_o is the resistance at temperature T_o (usually 25°C = 298.15 K). This is a simplification of the Steinhart–Hart equation in which $a = (1/T_0) - (1/B)ln(R_0)$, where $b = 1/B$ and $c = 0$.

Equation (7.5) can be solved for R in terms of T:

$$R = R_o e^{B\left(\frac{1}{T} - \frac{1}{T_o}\right)} \tag{7.6}$$

Example 7.4: Thermistor Calculations

Typical values of the Steinhart–Hart equation parameters for a thermistor with a resistance of 3000 Ω at room temperature (25°C = 298.15 K) are

[7] http://www.omega.com/temperature/Z/pdf/z256-257.pdf.

as follows:

$$a = 1.40 \times 10^{-3}$$

$$b = 2.37 \times 10^{-4}$$

$$c = 9.90 \times 10^{-8}$$

a. Calculate the temperature if the resistance is 2000 Ω. (b) Calculate the resistance if the temperature is 100°C.

Solution:

a. $\ln(R) = \ln(2000) = 7.6009$. Applying Equation (7.4):

$$\frac{1}{T} = 1.40 \times 10^{-3} + b(7.6009) + c(7.6009)^3 = 3.24489 \times 10^{-3}$$

$$T = 1/(3.24489 \times 10^{-3}) = 308.18 \text{ K}$$

$$T = 308.18 - 273.15 = 35.03°\text{C}$$

b. We will use Equation (7.6). First calculate $B = 1/b = 1/2.37 \times 10^{-4} = 4219.4$. Then convert the given temperature to degrees Kelvin:

$$T(K) = T(C) + 273.15 = 100 + 273.15 = 373.15 \text{ K}$$

$$R = 3000\, e^{4219.4\left(\frac{1}{373.15} - \frac{1}{298.15}\right)} = 174.5\,\Omega$$

NTC thermistors can be used as inrush-current-limiting devices in power supply circuits. They present a higher resistance initially which prevents large currents from flowing at turn-on, and then heat up and develop much lower resistance to allow higher current flow during normal operation. These thermistors are usually much larger than measuring-type thermistors, and are purposely designed for this application.

NTC thermistors are regularly used in automotive applications. For example, they monitor things like coolant temperature and/or oil temperature inside the engine, providing data to the ECU and, indirectly, to the dashboard. They are also commonly used to monitor the temperature of battery packs while charging.

PTC thermistors can be used as current-limiting devices for circuit protection, as replacements for fuses. Current through the device causes a small amount of resistive heating. If the current is large enough to generate more heat than the device can lose to its surroundings, the device heats up, causing its resistance to increase and therefore causing even more heating. This creates a self-reinforcing effect that drives the resistance upward, reducing the current and voltage available to the device.

7.4.4 Semiconductor (PTAT) Temperature Sensors

Thanks to semiconductor technology, it is possible to create a two-terminal device in which the current through the device is proportional to absolute temperature, that is, temperature in degrees Kelvin. This type of device is called a PTAT sensor. As an example of a PTAT, we will consider the Analog Devices AD590. This device functions like a current source: it passes a current of 1 μA/K for any applied voltage in the range from 4 V to 30 V. A simple circuit is shown in Figure 7.7 which will produce an output voltage proportional to temperature.

The current I_{temp} is the PTAT current. By Ohm's law, the voltage across R_1 is $V_{temp} = R_1 \times I_{temp} = 10\ \text{k}\Omega \times (1\ \mu\text{A}) \times T$ (deg. K).

Example 7.5: PTAT Calculations

Determine the output voltage of the PTAT circuit in Figure 7.7 at (a) room temperature and (b) freezing.

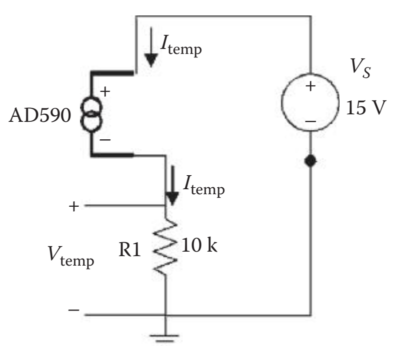


FIGURE 7.7 PTAT-based circuit to produce an output voltage proportional to temperature.

a. "Room temperature" is an imprecise quantity. However, a commonly used value is 300 K, which is equal to approximately 27°C or 80°F. At this temperature:

$$V_{temp} = 1 \times 10^4 \times 1 \times 10^{-6} \times 300 = 3.00 \text{ V}$$

b. Freezing is 0°C, which is approximately 273 K. The circuit output would be

$$V_{temp} = 1 \times 10^4 \times 1 \times 10^{-6} \times 273 = 2.73 \text{ V}$$

Another very useful device for use in digital systems is the integrated circuit temperature sensor with digital output. One of the most commonly used devices of this type is the Maxim DS18S20 Digital Thermometer. This device measures temperature from −55°C to +125°C (−67°F to +257°F) and offers ±0.5°C accuracy from −10°C to +85°C. It provides temperature conversion in a maximum of 750 ms (3/4 second). The chip reports its reading in digital format, which is convenient and far less prone to error than, for example, determining the resistance of a thermistor. The digital output format also provides the capability to communicate with the device over longer distances. The device requires a small number of connections to operate. It uses a single wire both to receive commands from a microcontroller and to reply with data. Because each device has a unique identification code, a microprocessor can interface to many sensors over the same single wire. A typical application might be temperature monitoring throughout an HVAC system in an office or factory.

7.5 Sensor Interfacing: Bridges, Op-Amps, ADC, and DAC

Many raw sensor signals tend to be small. A common practical problem is to amplify the raw sensor signal sufficiently to match the input range of an analog-to-digital converter, while suppressing noise. Special circuits for this purpose can be designed using bridges and operational amplifiers (op-amps).

7.5.1 Bridge Circuits

Bridge circuits are widely used as sensor interface circuits. The principle employed is to measure the offset that results from the deviation of a sensor element (e.g., resistance) from its nominal value. Bridges can form part of a null-seeking circuit, in which feedback is used to force the sensor element back to its nominal value. In this case, the output is measured as the amount of feedback required to null the bridge. Such null-seeking circuits are employed, for example, in accelerometers, where the feedback is used to keep the moving element near zero deflection which maximizes linearity while helping to prevent damage from overdriving the moving element.

The resistor Wheatstone bridge circuit consists of four circuit branches, each containing one resistor, as well as an ac or dc voltage bias. Each leg contains one resistor. If just one branch contains a variable resistor, the bridge is called ¼-active ("quarter-active"). If there are two branches with varying resistors, the bridge is called ½-active ("half-active"). Bridges with all four branches containing varying resistors are called full-active. Figure 7.8 shows a ¼-active resistive bridge powered by a dc bias source V_{DC}. The output of the bridge is the difference between the voltages V_A and V_B indicated in Figure 7.8. The resistor labeled R_X is the resistor whose value is desired to be measured. This is often a sensor element, for example a thermistor (a temperature-sensitive resistor), a photoresistor (resistance changes with light intensity), a piezoresistor (a force-sensitive resistor used in pressure sensors and accelerometers), or a strain gauge (in which resistance depends on the amount of stretching). The other three resistors labeled R_f are fixed resistors. The value of R_f is chosen

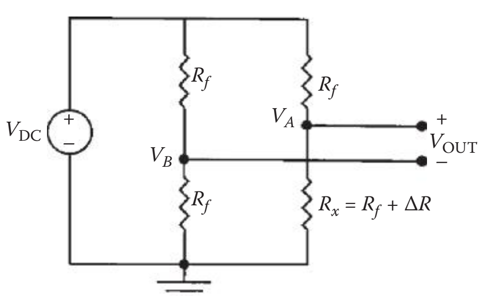


FIGURE 7.8 Wheatstone bridge circuit used to interface a resistive sensor.

to be equal to the nominal value of R_x. For sensors, this is often the value of resistance with zero or some established reference value of the measurand[8] applied. For example, a strain gauge may be specified to have a resistance of 120 Ω under zero strain, or a thermistor might be specified to have a value of 1000 Ω at 25°C. The difference between the value of R_x and its nominal value is designated ΔR. It is usually this value which we desire to determine, since it tells us how far the measurand has deviated from its reference value.

Using simple circuit theory, we can derive the equation for the output voltage in terms of the unknown resistance for the circuit shown in Figure 7.7.

$$V_A = V_{DC}\frac{R_x}{R_x + R_f} \tag{7.7}$$

$$V_B = V_{DC}\frac{R_f}{R_f + R_f} = \frac{V_{DC}}{2} \tag{7.8}$$

$$V_{OUT} = V_A - V_B = V_{DC}\left(\frac{R_x}{R_x + R_f} - \frac{1}{2}\right) \tag{7.9}$$

For sensor interfacing, we would like to have an expression for V_{OUT} that depends linearly on the change in R_x (i.e., V_{OUT} should be proportional to ΔR). Although Equation (7.9) is not linear in ΔR, if we assume ΔR is small compared with R_f we can use a Taylor-series expansion to derive the linear approximation shown in Equation (7.10).

$$V_{OUT} = \frac{V_{DC}}{4}\frac{\Delta R}{R_f} \tag{7.10}$$

In some applications, it is possible to design a measurement system with two or four suitably matched sensors. In such cases, one can use half-active or full-active bridge circuits, in which two or four branches, respectively, of the bridge contain changing resistances. One advantage of the half- and full-active bridge types is that they are inherently linear, so no approximation is required as is the case for the ¼-active bridge. Another advantage is that they can have larger output voltage than the ¼-active bridge for the same amount

[8] "Measurand" means whatever we are measuring with a particular sensor. For example, for thermistors, the measurand is temperature.

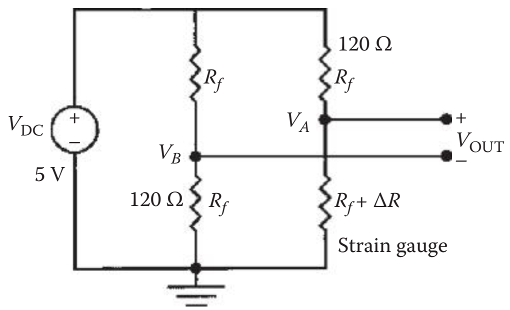


FIGURE 7.9 **Wheatstone bridge circuit used to interface a strain gauge sensor.**

of resistance change, twice as large for the half-active, and four times as large for the full-active.

Example 7.6: Interfacing a Strain Gauge Using a Wheatstone Bridge

In this example, we will design a ¼-active Wheatstone bridge circuit to provide an interface for a strain gauge.

A steel beam is being tested for bending strength in a certification laboratory. The test engineers have attached strain gauges to each side of the beam at various locations along its length. All the strain gauges are Omega Engineering Inc. Part # KFG-02-120-C1-11L1M2R from the same batch. The gauge factor provided with the batch is 2.05. The maximum expected strain is 1000 µStrain.[9] The dc supply voltage available for the bridge is 5 V. Our requirement is to construct a suitable Wheatstone bridge, and to develop the specifications for an amplifier that will provide an output signal of 5 V when the strain is at its maximum.

The circuit design is shown in Figure 7.9. The given part number is identified from the Omega Engineering Inc. Web site as a two-wire sensor with a nominal resistance of 120 ohms, temperature matched to steel. Since this is a ¼-active bridge, the three fixed resistors in the bridge should be selected

[9] This odd-sounding unit is used by convention, but strain is actually dimensionless, since it is the ratio of dimension change to a dimension reference value (e.g., 0.001 cm per cm).

to have a nominal resistance value of 120 ohms, to match the zero-strain value of the sensor. That way, the bridge output will be zero when the strain is zero.

The relationships between strain, sensor resistance, and bridge output can be determined from the following equations:

$$\frac{\Delta R}{R} = k\varepsilon \tag{7.11}$$

$$V_{OUT} = \frac{V_{DC}}{4}\frac{\Delta R}{R} = \frac{V_{DC}}{4}k\varepsilon \tag{7.12}$$

where

- ΔR is the resistance change of the strain gauge corresponding to the bending of the beam.
- R is the nominal resistance of the strain gauge.
- k is the gauge factor of the strain gauge.
- ε is the strain being measured.
- V_{OUT} is the output voltage from the Wheatstone bridge circuit.
- V_{DC} is the dc supply voltage for the Wheatstone bridge circuit.

For our design, $R = 120$ ohms, $k = 2.05$, $V_{DC} = 5$ V, and the maximum value of ε is 1000 μStrain, which is the same as 0.001.[10]

The maximum output of the bridge will be

$$V_{OUT}^{\max} = \frac{5V}{4}(2.05)(0.001) = 2.56\,\text{mV} \tag{7.13}$$

In order to obtain an output voltage of 5 V, we would need an amplifier with a gain of 5 V/2.56 mV, which is equal to 1951. This is a large amplifier gain, and consequently great care would have to be taken to minimize noise and dc offsets throughout the circuitry.

The same principle employed in the resistance bridge can be used for sensor elements that are capacitive or inductive. Such elements are sometimes used in position, proximity, and accelerometer sensors. In this case, each branch of the bridge contains capacitors or inductors which match the nominal value

[10] Recall that μ stands for 1×10^{-6}.

of the unknown element, and the bridge must be powered with an ac source of suitable frequency.

7.5.2 Linearization and Reduction of Error

Traditionally, linear response has been a crucial requirement for sensors. However, the widespread use of computerized data acquisition and signal conditioning is gradually eliminating this requirement by making it possible to correct for nonlinearity by using calibration curves generated by table lookup or calculation. Nevertheless, in any system that is well designed, the sources of nonlinearity must be considered. These sources may be divided into those which are mathematically predictable and those which are process and environment dependent. Mathematically predictable sources include the bridge circuit equations, the effects of self-heating, and the nonlinear relationship between ΔR and the measurand. Common process- and environment-dependent sources include resistor mismatches, lead and contact resistance, magnetic and galvanic potentials, leakage currents, and temperature, humidity, and aging effects.

One important consideration in any resistance-measuring circuit is "self-heating error." Self-heating occurs whenever current is passed through the resistance to be measured. If the current is I and the resistance is R_X, then an amount of heat is dissipated in the resistor equal to I^2R_X. This can raise the temperature and thereby change the value of resistance. Self-heating error can generally be kept small or negligible with careful design of the measuring circuit and attention to the temperature coefficient of resistance (TCR) of the device under Figure 7.10 shows a ¼-active bridge where the

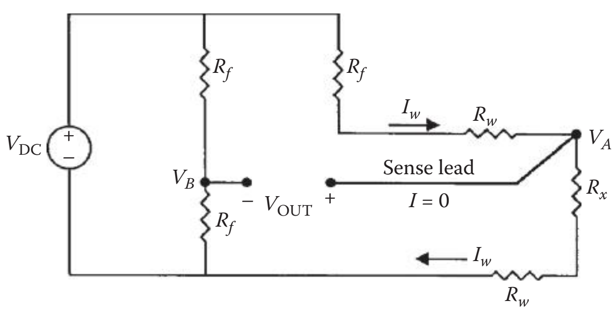


FIGURE 7.10 Three-wire Wheatstone bridge circuit used to compensate for long sensor lead wires.

active element is connected by means of long leads having nonnegligible resistance. A typical application would be interfacing a resistance temperature detector (RTD), where the RTD element must be placed at the temperature source, but the bridge is maintained at room temperature to avoid TCR-induced drifts in the fixed resistors. In this case, the lead wire resistance can be compensated by using a three-wire connection as shown in Figure 7.10.

The lead-wire voltage drops $I_W R_W$ cancel since they are in opposite legs of the bridge. The sense wire carries no current and thus $V_{OUT} = V_A - V_B$, as desired. This approach is effective in the usual case of equal lead lengths and wire types. When V_{OUT} is connected to an amplifier, a small amount of current will flow in the sense lead, but it can usually be made small enough to be negligible in overall effect.

The most accurate conventional method of measuring resistance is the four-wire Kelvin probe, illustrated in Figure 7.11. Although not really a bridge technique, it is included in this section for completeness.

A reference current is forced to flow through R_x, and the voltage is measured using two additional sense leads. If the sense leads are connected to a high-impedance input (e.g., a digital voltmeter [DVM], instrumentation amplifier, or analog-to-digital converter), they carry essentially zero current so no voltage drop is induced. The voltage V_x is thus directly proportional to R_x. This is true regardless of whether the leads are matched in length and type.

A modification of this approach that is especially suitable for multichannel analog-to-digital converter-based systems is shown in Figure 7.12.

Here, the need for a precision current source is eliminated. Instead, a precision resistor R_{REF} is used to sample the current. The ADC system must acquire both V_x and V_{REF}, one of which requires a differential measurement, for a total

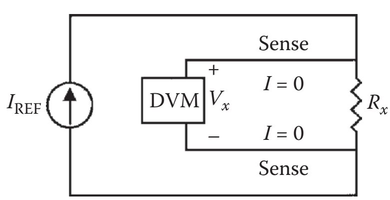


FIGURE 7.11 Four-wire Kelvin probe used for accurate resistance measurements.

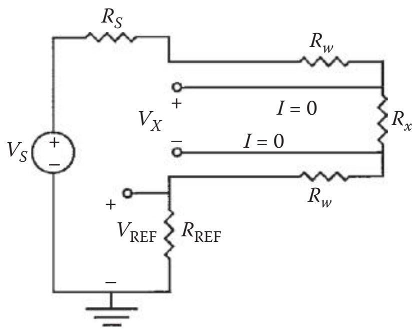


FIGURE 7.12 **Modified Kelvin probe using a voltage source instead of a current source.**

of three single-ended channels.[11] The resistance R_x can be calculated from the measured voltages and the known reference resistance:

$$R_X = \frac{R_{REF}}{V_{REF}} V_X \tag{7.14}$$

The choice of the series resistor *RS* involves a trade-off. As *RS* is made smaller, the power dissipation in R_x increases, thus contributing to self-heating error. As R_S is made larger, V_{REF} and V_x are reduced in magnitude, which reduces the achievable resolution and the signal-to-noise ratio.

7.5.3 Operational Amplifier (Op-Amp) Circuits

The integrated circuit operational amplifier (op-amp) is the most widely used interface component in sensor systems. Flexibility and low cost are key reasons. The op-amp circuit symbol is shown in Figure 7.13.

The input terminals are called the *noninverting terminal* (+) and the *inverting terminal* (−). In order to simplify circuit diagrams, the dc power supplies (V+) and (V−) are normally not shown. Models are also available that use a single supply voltage. Most of the selection process for an op-amp involves trade-offs among input impedance, bandwidth, slew rate, noise, and other parameters, versus cost. Examples of several applications that demand high-performance op-amp models will be discussed later in this section.

[11] A differential measurement is one in which neither input is ground. A single-ended measurement, also called "ground-referenced," is one in which one of the input connections is system ground.

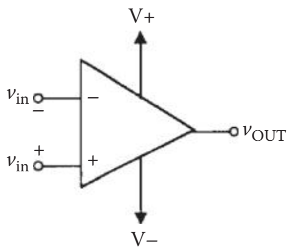


FIGURE 7.13 Operational amplifier circuit symbol.

Although op-amps are complicated circuits containing dozens of transistors and other components, they are often treated as a single circuit element with the following relationship between the inputs and the output:

$$v_{OUT} = A\left(v_{in}^{+} - v_{in}^{-}\right) \tag{7.15}$$

where A is called the "open-loop gain" of the op-amp. As we shall shortly see, the op-amp is almost always used in a feedback configuration, which yields another gain called the "closed-loop gain."

The primary usefulness of the op-amp stems from two design features. First, A is designed to be very large; typical values range from 100,000 to 1,000,000. Second, the input impedance is designed to be very large, so that almost zero current flows into the input terminals.

The effect of these two design features is best understood by considering a particular circuit—the unity gain buffer, shown in Figure 7.14. We will follow common practice and employ an idealized model of the op-amp in

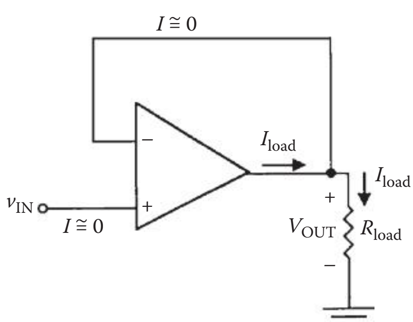


FIGURE 7.14 Unity gain buffer amplifier.

which the value of A approaches infinity, and the input terminal currents are zero.

By applying Equation (7.15) to the circuit in Figure 7.14, we can write

$$V_{OUT} = A(V_{IN} - V_{OUT}) \tag{7.16}$$

where we have used the fact that the feedback connection forces the voltage at the inverting terminal to be the same as the output voltage. Solving for V_{OUT} yields

$$V_{OUT} = V_{IN}\frac{A}{1+A} \tag{7.17}$$

If we let A approach infinity, then we have, in the limit:

$$V_{OUT} = V_{IN} \tag{7.18}$$

Thus we see that this circuit has a closed-loop gain:

$$A_V = \frac{V_{OUT}}{V_{IN}} = 1 \tag{7.19}$$

whence the name "unity gain."

Furthermore, we see that so far as the input source is concerned, it does not need to supply any appreciable current, since it is looking directly into the high-impedance op-amp terminal. All of the load current is supplied by the op-amp (ultimately this comes from the op-amp dc power supplies). This is why the circuit is called a "buffer."

In summary, this circuit provides an exact copy of the input signal to the load, but it draws negligible current from the signal source. This is especially useful for low-power sources, such as microphones, medical electrodes, radio receivers, and the like.

The circuit shown in Figure 7.15 also provides buffering, but in addition it can provide a closed-loop gain greater than one.

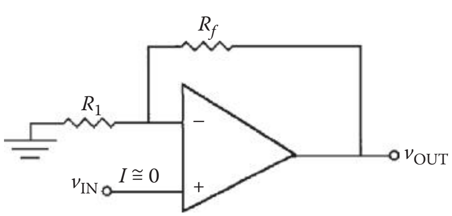


FIGURE 7.15 Noninverting amplifier.

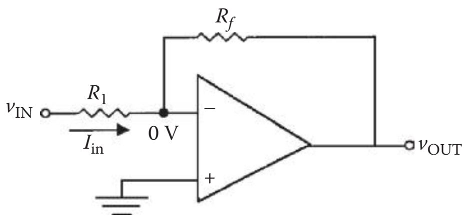


FIGURE 7.16 **Inverting amplifier.**

Applying Equation (7.15) together with the ideal specifications of infinite open-loop gain and zero input currents, we find

$$V_{IN}^{-} = \frac{R_1}{R_1 + R_f} V_{OUT} \tag{7.20}$$

$$V_{OUT} = A(V_{IN} - V_{IN}^{-}) \tag{7.21}$$

Combining Equation (7.20) and Equation (7.21), and letting A approach infinity, we find

$$V_{OUT} = \left(1 + \frac{R_f}{R_1}\right) V_{IN} \tag{7.22}$$

For this circuit, the closed-loop gain is

$$A_V = \frac{V_{OUT}}{V_{IN}} = 1 + \frac{R_f}{R_1} \tag{7.23}$$

This circuit can provide any gain greater than or equal to one, depending on the values of resistors used. The unity-gain buffer is a special case of this circuit in which $R_f = 0$ (an ideal wire is equivalent to a zero-ohm resistor)[12] and R_1 = infinity (an open circuit).

Another useful op-amp circuit is the inverting amplifier shown in Figure 7.16. In this circuit, the locations of the input and ground have been exchanged. Following the same procedure as for the noninverting circuit, we find

$$A_V = \frac{V_{OUT}}{V_{IN}} = -\frac{R_f}{R_1} \tag{7.24}$$

[12] On circuit diagrams, wires are usually assumed to be ideal. In the real world, wires have nonzero resistance. If they did not, there would be no limit on how much current they could carry.

The minus sign that appears in Equation (7.24) is the reason this is called an inverting amplifier. Unlike the noninverting amplifier, this is not a high-impedance buffer circuit since the signal source sees an impedance equal to R_1. Besides amplification, this circuit can provide *attenuation*, which is the term for a gain less than one in magnitude. If R_f is less than R_1, the output will be smaller in absolute value than the input.

Op-amp circuits can amplify ac as well as dc signals. The upper limit on the signal frequency that can be amplified is determined by a parameter called the gain-bandwidth product. This is normally provided on the op-amp data sheet, and is usually specified in MHz.

Other widely used op-amp circuits include instrumentation amplifiers, integrators, summing amplifiers, and filters. Designs for these can be found in many readily available reference books and internet sources.

7.5.4 Analog-to-Digital Conversion (ADC) and Digital-to-Analog Conversion (DAC)

The output of a sensor such as the accelerometer previously discussed is often in analog form. Most modern applications require a digital version of the signal in order to act upon it or record it. Circuits that convert analog signals to digital data are called "analog-to-digital converters" (ADCs). Circuits that convert in the other direction from digital to analog are called "digital-to-analog converters" (DACs). An electronic module or computer card that incorporates multiple ADCs and DACs and operates under software control is referred to as a "data acquisition system" (DAQ).

ADCs and DACs are rated according to the resolution and the sampling (or conversion) rate that they provide. For an ADC, the resolution is the number of data bits generated for each analog sample value. Typical resolution values for commercially available parts are even numbers in the range from 8 to 24 bits. Sample rate is the number of conversions produced per second. Typical values are in the range from thousands to millions of samples per second. A useful calculation is to determine the smallest signal change that can be detected for a given resolution. The calculation is

$$\Delta V = \frac{V_{FS}}{2^N}$$

where V is the smallest detectable voltage change, V_{FS} is the full-scale voltage of the ADC, and N is the number of bits of resolution.

Example 7.7: Design an Op-Amp Circuit to Interface a Sensor to an ADC

Design an op-amp circuit to interface a sensor to an ADC. The sensor outputs a voltage in the range from 0.1 V to 0.5 V over the range of stimulation that we are interested in. The ADC has a full range of 0 V to 5 V.

Solution:
The necessary gain is 5 V/0.5 V = 10. We should use a noninverting amplifier. We could choose R_2 and R_1 in Figure 7.15 to be 18 kΩ and 2 kΩ, respectively. Since only the ratio matters, there are an infinite number of choices mathematically. From a practical standpoint, resistors in op-amp circuits are usually chosen to be in the range from several kohms to several megohms. Larger values lead to more noise but less power drain. The designer must therefore make a trade-off suitable to his or her particular problem.

In this example, the lower end of the input (0.1 V) will yield an amplified value of 1 V. The ADC input range below this value will be "wasted" in the sense of never being used. A more sophisticated design would use a level shifter to subtract off 0.1 V before amplifying. Then the gain could be 5 V/0.4 V = 12.5, and maximal use would be made of the ADC. This would improve the signal-to-noise ratio.

Example 7.8: Smallest Detectable Voltage Change Using ADC

A 10-bit ADC is used with a 5 V power supply. Determine the smallest detectable voltage change, V, if the full-scale voltage is calibrated to be 5.00 V.

Solution:

$$\Delta V = \frac{V_{FS}}{2^N} = \frac{5.00\,\text{V}}{2^{10}} = \frac{5.00\,\text{V}}{1{,}024} = 4.88\,\text{mV}$$

This means that an analog voltage change of 4.88 mV will cause the converted binary value to change by 1 bit.

Example 7.9: Binary Output of an ADC

Suppose the 10-bit ADC in the previous example has an input of 3.000 V. What is the corresponding binary conversion?

Solution:
The number of 1-bit increments corresponding to 3.000 V is

$$NB = 3.000 / 4.88 \times 10^{-3} = 614.4$$

Thus we can obtain the binary value by rounding to the nearest integer, and converting to binary:

$$614D = 1001100110_B$$

Example 7.10: Voltage Change per Bit for an ADC

If the binary value changes to 1001100100, what is the corresponding voltage change, and what is the new voltage?
Solution:
We first determine the change in binary value:

$$\text{Delta BV} = 1001100100_B - 1001100110_B = -10_B$$

The corresponding change is –2 bits. Since each bit represents 4.88 mV, we find the voltage change is

$$\text{Delta V} = -2\ (4.88\ \text{mV}) = -9.76\ \text{mV}$$

The new voltage is therefore

$$V = 3.000 - 0.00976 = 2.990\ \text{V}$$

The calculations for a DAC are similar to those for the ADC.

One common application area for ADCs and DACs is in digital signal processing (DSP) for audio signals. Most music that we listen to is processed by a number of DSP chips or systems between the recording studio and our ears. A well-known theorem in communications theory, the Shannon sampling theorem, states that an analog signal must be sampled at a rate at least twice that of the highest frequency component in order to obtain a correct digital rendition. This rate is called the "Nyquist rate." For example, if the highest frequency in a certain audio signal is 10 kHz, it must be sampled at a rate of at least 20 kHz. For a variety of practical reasons, most sampling systems are designed to *oversample* at a rate which is several times the Nyquist rate. Four times (4×) oversampling is common, for instance. Commonly an audio signal is filtered before sampling so that the highest frequency is known. Failure to do so can result in a phenomenon known as aliasing, which is usually noticed as low-frequency rumbling or distortion.

Example 7.11: Anti-Alias Filter for FM Radio

A certain audio signal is to be digitally processed before transmitting over FM analog broadcast radio. What type of filter should be applied, and what should be the sampling rate?

Solution:

A little research shows that conventional FM radio supports a bandwidth of 15 kHz. Therefore, we want to use a low-pass filter that provides significant power reduction for frequencies at and above 15 kHz. Filter design per se is beyond the scope of this discussion, but a typical anti-aliasing filter might be an eighth-order low-pass one such as the National Instruments NI SCXI-1141. The sampling rate should be at least 2 × 15 kHz = 30 kHz (30,000 samples per second), but a good design would double or quadruple this. For 4× oversampling, the rate is

$$SR = 30{,}000 \times 4 = 120{,}000 \text{ samples per second}$$

The higher the resolution (sometimes called “bit depth”), the less noise the signal will incur. Typical values for higher end audio are 22 or 24 bits. A part such as the Analog Devices AD7766 could be selected. This chip has a conversion rate of 128,000 samples per second at 24-bit resolution, and costs around \$6.

Problems

7.1. Calculate the value of V_1(light) in Figure 7.1 if the photocell is exposed to light and has resistance $R_{light} = 5\ \text{k}\Omega$, and $R_{thresh} = 100\ \text{k}\Omega$.

7.2. (a) Plot the value of V_1 versus photocell resistance over the range from R_{light} to R_{dark}, for the parameters given in Figure 7.1, with R_{thresh} =

100 kΩ. (b) On the same axes, plot V_{OUT} for both increasing light intensity (decreasing photocell resistance) and decreasing light intensity. Indicate the thresholds and the hysteresis.

7.3. The Analog Devices ADXL335 is proposed for use in several systems. Briefly discuss the suitability of this part for each: (a) Monitoring acceleration on a roller coaster that produces 2 g of acceleration in each dimension; (b) recording acceleration during space shuttle lift-off, which is typically 8 g in one dimension; and (c) recording acceleration for a formula car that can go from 0 to 124 mph in 3.8 seconds.

7.4. The range of the PING)))™ sonar sensor is specified to be 2 cm to 3 m, and the operating temperature range is specified to be 0°C to 70°C. Make a table showing the round-trip echo return time at the extremes of temperature and distance. Use the Web or another resource to look up the speed of sound in dry air at the given temperatures. What is the overall longest time you might have to wait for an echo return?

7.5. Repeat Example 7.3 for (a) an output of 18 mV and (b) a temperature of 1000°C.

7.6. Repeat Example 7.3 for a type J thermocouple.

7.7. Calculate the nominal resistance of an American Standard RTD at 0°C, 50°C, and 100°C.

7.8. Calculate the nominal resistance, as well as the minimum and maximum possible, for an RTD that meets the requirements of European Class A at 0°C, 50°C, and 100°C.

7.9. A certain thermistor has resistance $R_o = 3000\ \Omega$ at $T_o = 25°C$. The Steinhart–Hart equation parameters are known to be $a = 1.40 \times 10^{-3}$, $b = 2.37 \times 10^{-4}$, and $c = 9.90 \times 10^{-8}$. (a) Calculate the temperature if $R = 1000\ \Omega$. (b) Calculate the resistance corresponding to a temperature of T = 20°C.

7.10. A certain thermistor has resistance $R_o = 3000\ \Omega$ at $T_o = 25°C$. The Steinhart–Hart equation parameters are known to be $a = 1.40 \times 10^{-3}$, $b = 2.37 \times 10^{-4}$, and $c = 9.90 \times 10^{-8}$. Plot the resistance versus temperature over the range from approximately 0°C to 100°C. Make sure your plot data have a resolution of at least 1°C. *Hint*: Select, by trial and error, the values of R that approximately yield the desired end-point values of T. Then choose a suitable increment of R to give good resolution in T. Finally, plot R versus T.

7.11. An AD590 PTAT sensor is used in the circuit shown in Figure 7.7. The power supply voltage is 10 V, and the resistor R_1 is 6.8 kΩ. What is the difference in V_{temp} as the temperature changes from 0°C to 100°C?

7.12. The Wheatstone bridge circuit in Figure 7.8 is used to interface a resistive sensor, R_X. The sensor resistance is nominally 1000 Ω, but it can range from 900 to 1100 Ω. Calculate the corresponding range of the output voltage.

7.13. Consider the Wheatstone bridge circuit in Figure 7.8 with R_f = 1000 Ω. (a) If ΔR is 100 Ω, what is the percent error in V_{OUT} that arises from using the linear approximation in Equation (7.10) versus the exact expression in Equation (7.9)? (b) Repeat for ΔR = 500 Ω.

7.14. For the strain gauge interface circuit in Example 7.6, use Equation (7.11) to calculate the maximum resistance change, ΔR. Use this result to determine the exact value of the maximum bridge output voltage based on Equation (7.9). What is the percent error that arises from using the approximation (Equation 7.12) compared with Equation (7.9)?

7.15. The Wheatstone bridge circuit in Figure 7.8 has R_f = 100 Ω and V_{DC} = 5 V. At the nominal value of $R_X = R_f$, determine the power dissipated in R_X. If it is known that the sensor temperature increases by 0.01°C/mW due to self-heating, what is the change in sensor temperature? If the temperature coefficient of resistance (TCR) is 1% per °C, what is the resistance change of the sensor due to self-heating? By how much does the output voltage change if none of the fixed resistors change?

7.16. Design a noninverting amplifier similar to Figure 7.15 for which the gain is (a) 10 V/V, (b) 100 V/V, and (c) 200 V/V. For each design, use resistor values that are no larger than 1 MΩ and no smaller than 1 kΩ.

7.17. A non-inverting amplifier is desired to have a gain of 100. Choose the values of resistors R_1 and R_f (as defined in Figure 7.15) from the industry-standard 5% tolerance values. Use resistor values that are no larger than 1 MΩ and no smaller than 1 kΩ. What is the actual gain that your design achieves? What is the percent error when compared with the ideal gain?

7.18. Design an inverting amplifier similar to Figure 7.16 for which the gain is (a) –10 V/V, (b) –20 V/V, and (c) –30 V/V. For each design, use resistor values that are no larger than 1 MΩ and no smaller than 1 kΩ.

7.19. An inverting amplifier is desired to have a gain of −44. (a) Choose the values of resistors R_1 and R_f (as defined in Figure 7.16) from the industry-standard 5% tolerance values. Use resistor values that are no larger than 1 MΩ and no smaller than 1 kΩ. What is the actual gain that your design achieves? What is the percent error when compared with the ideal gain? (b) Repeat if the resistors must be selected from industry-standard 10% tolerance values.

7.20. A 12-bit ADC has the full-scale voltage calibrated to be 3.30 V. Determine the smallest detectable voltage change.

This page intentionally left blank

Chapter
8

Digital Logic

This chapter covers a number of topics relevant to the use of digital logic for the control of equipment. We first introduce the basic ideas of binary arithmetic and Boolean algebra by which we can construct and calculate logical functions. Next, logic chips and circuits used for interfacing (often referred to as "glue logic") are presented. We then introduce programmable logic controllers (PLCs), which are widely used in industry for controlling manufacturing processes. The chapter concludes with a discussion of microcontrollers and digital signal processors.

8.1 Binary Arithmetic and Boolean Algebra

All modern digital systems use binary numbers. Whereas the conventional number system in most human cultures is decimal (base 10), with digits 0, 1, 2, … 9, the binary system used by computers is base 2. This means there are just two possible values: 0 and 1. These are sometimes referred to by various pairs of opposites, such as *high* and *low*, *true* and *false*, *up* and *down*, or the like.

8.1.1 Binary Numbers

One digit of a binary number is called a *bit* (compaction of "binary digit"). Eight bits constitute a *byte*. Two bytes form a 16-bit *word*. The leftmost bit in a written binary number is called the *most significant bit* (MSB), and the rightmost bit is called the *least significant bit* (LSB).

Consider the byte 10100101. The decimal value can be calculated easily using place-value arithmetic. The LSB has a place value of $2^0 = 1$, the next bit to the left has place value $2^1 = 2$, and the bits in sequence moving left have place values given by increasing powers of 2, that is, 4, 8, 16, 32, 64, and 128. The calculation is shown in Equation (8.1).

$$\begin{aligned}(10100101)_2 &= (1\times 2^7 + 0\times 2^6 + 1\times 2^5 + 0\times 2^4 + 0\times 2^3 + 1\times 2^2 + 0\times 2^1 + 1\times 2^0)_{10}\\ &= 128 + 32 + 4 + 1\\ &= 165_{10}\end{aligned} \tag{8.1}$$

Discovering the binary equivalent of a decimal value proceeds by a comparison process. We first find the largest power of 2 that is smaller than the given decimal value, and this determines the place value of the MSB. The rest of the process is best illustrated with an example.

Example 8.1: Converting a Decimal Value to Binary

Problem: Convert $(19)_{10}$ to binary.

Solution: Writing powers of 2 in sequence (1, 2, 4, 8, 16, 32), we see that the largest of these values that is smaller than the given decimal value is $2^4 = 16$. So our binary conversion so far is 1xxxx, where *x* indicates an as-yet-unknown bit value. The MSB (bit 5) is 1.

Next, we subtract $(16)_{10}$ from $(19)_{10}$ to yield remainder R = 3. The comparison then proceeds as follows:

Test for bit 4: Is $2^3 = 8$ less than or equal to R? NO → bit 4 = 0
Test for bit 3: Is $2^2 = 4$ less than or equal to R? NO → bit 3 = 0
Test for bit 2: Is $2^1 = 2$ less than or equal to R? YES → bit 2 = 1
Since this bit was 1, subtract its value (2) from R to get a new value of R = 1.
Test for bit 1: Is $2^0 = 1$ less than or equal to R? YES → bit 1 (LSB) = 1.
Done, since LSB found.

Writing the bits left to right, or from MSB to LSB, we find $(19)_{10} = (10011)_2$. Many computer systems work with fixed-length data. For example, if we needed to represent $(19)_{10}$ in a full byte, we would pad from the left with zeros in order to make eight bits. The result would be $(00010011)_2$.

8.1.2 *Hexadecimal and Octal Notation*

It is often convenient to represent binary numbers in a more compact form for easier reading by humans. This can be accomplished by grouping the binary digits into groups of four (hexadecimal) or three (octal). Hexadecimal (hex for short) is base 16. The digits in hex are 0, 1, 2, … 9, A, B, C, D, E, and F, where the letters A–F are used to provide single-character representations for the decimal values 10, 11, … 15. Octal is base 8, and has the digits 0, … 7.

The binary representation of $(19)_{10}$ was found in Example 8.1 to be $(00010011)_2$. We can group the bits and write this in hex as follows:

$$(00010011)_2 = (0001\ 0011)_2 = (13)_{16}$$

The left-hand group of 4 bits, (0001), has value 1, and the right-hand group has value 2 + 1 = 3. This representation is generally more meaningful than a string of 1's and 0's. The decimal value can be verified readily. Using place-value arithmetic with each place now a power of 16 (1, 16, 256, …), we note that $(13)_{16} = 1 \times 16 + 3 \times 1 = (19)_{10}$.

The notation commonly used for hex varies depending in part on syntax defined by various computer languages. The most common variants for $(13)_{16}$ are 13H and 0 × 13. Others would include 13h, #0013, &#x0013, and some others which are less common.

Octal representation is not as common as hex in modern computer systems, but may still be encountered in special applications. In programming languages, octal values are typically identified with a variety of prefixes, including the digit 0, the letters o or q, or the digit–letter combination 0o. For example, $(45)_8$ might be represented as 045, o45, q45, or 0o45 in various languages.

8.1.3 Binary and Hex Arithmetic

We can perform arithmetic operations on binary and hex numbers just as we do for decimals. An example summation in binary is shown below:

$$\begin{array}{r} 1101 \\ +1100 \\ 11001 \end{array}$$

As usual, we start with the rightmost column and move left. The possible bit sums are $0 + 0 = 0$, $0 + 1 = 1 + 0 = 1$, $1 + 1 = (10)_2$, and $1 + 1 + 1 = (11)_2$. The latter occurs when there is a carry-over from the previous column and both bits are 1's. Subtraction is usually accomplished by adding the negative of the subtrahend (bottom value).

There are various methods of handling negative numbers in binary. The most common is called "two's complement" representation. The 8-bit two's complement of a binary number is defined as the value obtained by subtracting the number from $2^8 = 256_{10} = (100)_{16}$. The two's complement of the number behaves like the negative of the original number in most arithmetic, and it can coexist with positive numbers in a natural way. Example 8.2 illustrates the use of the two's complement representation to perform binary subtraction.

Example 8.2: Binary Subtraction Using Two's Complement Representation

Problem: Compute (1101) – (1100).

Solution: In decimal, this problem would be written 13 – 12, so we know the answer should be 1.

We first form the two's complement of (1100). This is most easily accomplished by padding to 8 bits yielding (00001100), followed by *complementing* each bit (that is, changing 1's to 0's and vice versa) to yield (11110011), and then adding 1. The result is $-(1100) = (11110100)_{2c} = -12_{10}$, where the subscript "2c" indicates two's complement representation.

The MSB in two's complement representation will be 1 for negative numbers, and 0 for positive ones. Therefore, it conveniently serves as a sign bit.

Now, adding the two's complement representation of the subtrahend, -12_{10}, to the padded minuend (0000 1101), we obtain

$$\begin{array}{r} 00001101 \\ +11110100 \\ \hline 100000001 \end{array}$$

By the rules of two's complement arithmetic, we discard the 1 bit on the far left, and then we are left with the correct answer of 1.

The range of decimal values that can be represented in 8-bit two's complement is -128 $(=10000000)_{2c}$ to $+127$ $(=011111111)_2$. However, in some computer applications, all numerical values are treated as unsigned, meaning that only values from 0 to 255 can be represented by one 8-bit byte. Subtraction might be internally performed using two's complement, but all results are treated as positive numbers. Serious performance errors can result if an arithmetic operation unexpectedly yields a negative value. For example, a robot arm could move to a completely unexpected position and cause damage.

8.1.4 Boolean Algebra

Boolean algebra was developed by George Boole in the 1840s. It is used in digital logic to combine logical statements to yield a result of TRUE or FALSE. In computers, the result is often stored in memory for further use.

In programmable logic controllers, the result often activates a switch, relay, or similar device.

Although the terms TRUE and FALSE are used in formal Boolean algebra, it is common to use other pairs of opposites in electronics and computers. The most common are HIGH and LOW, or 1 and 0. A Boolean function combines *predicates* into a result. The predicates and the result all have logical values, that is, they are either TRUE or FALSE. This is shown in Equation (8.2).

$$\text{Predicate} * \text{Predicate} * \text{Predicate} *\ldots * \text{Predicate} = \text{Result} \tag{8.2}$$

Here, the * symbol represents any possible logical operation. These include OR, AND, NOT, and others. Individual predicates can themselves be the result of logical calculations.

Typically, in a control system, a predicate is TRUE or FALSE depending on the state of an input. For example, if a switch is toggled ON, that condition could be read by a microcontroller and an internal memory bit would be set HIGH (=TRUE). One can imagine a switch panel interfaced to a computer system. The state of each switch could be represented by the predicates in Equation (8.2). Then we might wish the result to be TRUE for certain combinations of switch settings (that is, some action would occur, such as turning on a heating system) and FALSE for all other possible combinations.

8.1.5 Truth Tables

Data sheets for digital parts often contain *truth tables*. These are tables that define the output or response of the part to all possible combinations of inputs. For example, the truth table for a digital circuit that implements the OR logic function is shown in Table 8.1.

We can see that the output is TRUE when Input A is TRUE *or* when Input B is TRUE.

Example 8.3 illustrates the calculation of a Boolean logic function.

Table 8.1 OR Logic Truth Table

Input A	Input B	Output
FALSE (0)	FALSE (0)	FALSE (0)
FALSE (0)	TRUE (1)	TRUE (1)
TRUE (1)	FALSE (0)	TRUE (1)
TRUE (1)	TRUE (1)	TRUE (1)

Example 8.3: Calculation of a Boolean Logic Function

Problem: A switch panel has three switches labeled A, B, and C. Each switch can be ON or OFF. Determine the output for every possible combination of the three switch settings if they are connected to a circuit that performs the following logic function:

$$Y = (A + B) \bullet \overline{C} \tag{8.3}$$

where Y is the output.

Solution: In Boolean algebra notation, the + symbol stands for the logical OR operation, and the • symbol stands for the logical AND operation. Often the latter symbol is omitted and AND-ing is indicated simply by adjacency, much like multiplication in conventional algebra. Furthermore, the overbar stands for the COMPLEMENT operation (also called the NOT function).

In order to calculate all possible outputs in an orderly manner, we should create a truth table. Since there are three inputs, there are $2^3 = 8$ possible combinations, and the truth table will have eight rows. It will help to construct the truth table with an intermediate column for $A + B$ and for $\overline{C}$. The truth table is shown in Table 8.2.

Here we have used the fact that the AND function yields a 1 only when all of its inputs are 1, and the OR function yields a 1 when any of its inputs are 1.

Table 8.2 Truth Table for Example 8.3

Inputs			Intermediate Results		Output
A	B	C	$A+B$	$\overline{C}$	$Y = (A+B) \bullet \overline{C}$
0	0	0	0	1	0
0	0	1	0	0	0
0	1	0	1	1	1
0	1	1	1	0	0
1	0	0	1	1	1
1	0	1	1	0	0
1	1	0	1	1	1
1	1	1	1	0	0

As shown in the truth table, the output is 1 (or TRUE, HIGH) for the three cases indicated, and 0 (or FALSE, LOW) for the other five cases.

8.2 Logic Circuits

In modern digital systems, most logic is performed by a microcontroller or programmable logic controller (PLC). However, there is often also a need for simple interface logic to get the right signals to the microcontroller pins or the PLC inputs. Also, the practicing engineer or technician is likely to encounter a large installed base of digital logic implemented with fixed (nonprogrammable) digital logic chips from the time before microcontrollers were in widespread use. A number of integrated circuits are available to perform these functions. There are two classes of logic circuits—combinational logic and sequential logic. Combinational logic is not time dependent. The inputs are combined internally according to the chip's logic function to produce the outputs. Whenever an input changes, the output changes. These types of logic circuits are called *logic gates*. The OR function described in Section 8.1 could be implemented by an OR gate, or it might be implemented in software running on a computer, microcontroller, or PLC.

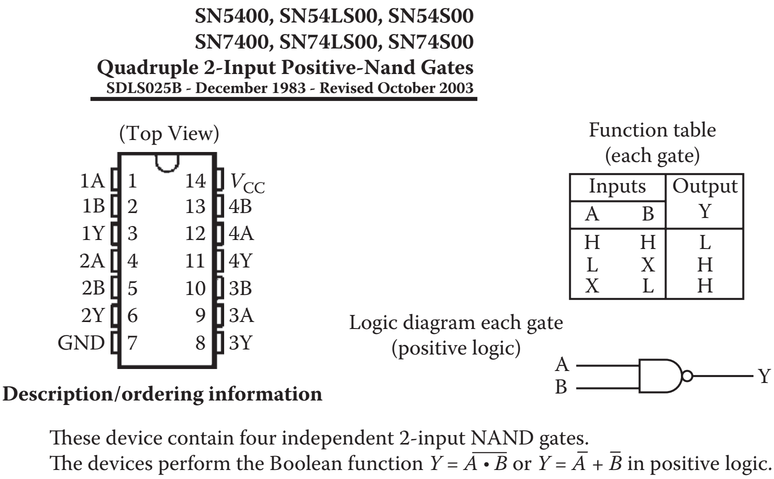
In contrast to combinational logic circuits, sequential logic circuits implement functions that are inherently time dependent. An example would be a *counter* or a *shift register*. A counter is a sequential circuit whose outputs change (usually as an increasing or decreasing binary value) each time an input pulse is detected. A shift register causes its input data bits to step through its internal registers as a *clock* signal is applied. Synchronous sequential logic circuits have clock inputs that synchronize their actions. Asynchronous sequential circuits operate without a clock.

Information from the data sheet of a typical logic gate is shown in Figure 8.1.[1]

This logic chip provides four independent two-input NOR gates.[2] The truth table is shown in Figure 8.1. We can see from the pinout that gate 1 has its inputs at pins 1 and 2, and its output at pin 3. Gate 2 has its inputs at pins 4 and 5, and its output at pin 6. The other two gates are connected on the other side of the chip to pins 8–13.

[1] A number of complete pages from this data sheet are included in Appendix B.

[2] The NOR function is the complement of the OR function.



SN5400, SN54LS00, SN54S00
SN7400, SN74LS00, SN74S00
Quadruple 2-Input Positive-Nand Gates
SDLS025B - December 1983 - Revised October 2003

Function table (each gate)

Inputs		Output
A	B	Y
H	H	L
L	X	H
X	L	H

Description/ordering information

These device contain four independent 2-input NAND gates.
The devices perform the Boolean function $Y = \overline{A \bullet B}$ or $Y = \bar{A} + \bar{B}$ in positive logic.

FIGURE 8.1 **Excerpts from the data sheet for a Texas Instruments 7400 series TTL NOR gate.**

Two other connections shown are very important—these are the power and ground pins. Pin 14, labeled V_{CC}, must be connected to a dc source which the data sheet specifies elsewhere to be 5 V. Pin 7, labeled GND, must be connected to the common ground of the circuitry in which this chip is used.

8.3 Programming Languages

Almost all digital systems with any complexity contain one or more programmable devices. These are typically microcontrollers, but even some sensors and peripheral devices require programming as well. Internal programmable devices that are part of a larger system or machine (such as a robot) are usually called "embedded systems." Sometimes the term embedded system is also used to refer to the entire system.

A variety of programming languages have evolved over the decades since the advent of the programmable digital computer in the mid-1960s. Today's microcontrollers are typically programmed in the C language and its derivatives, including C++. Many microcontrollers use a specialized or adapted variant of C. For example, the very popular Arduino microcontroller is programmed in a C-like language that is simplified and streamlined to make it easier to write working code. If maximum use is to be made of all microcontroller functions and code needs to run as fast as possible, one may write some or all of the code in assembly language. These are instructions that are as close to the machine as possible, in the sense that each command corresponds closely to some hardware operation. No matter what the user interface may be, all microcontrollers end up running machine language—instructions converted to pure binary form. The connection between the user's language (C, etc.) and the machine language is provided by a *compiler*. The compiler converts human-readable language into machine language code.

System-level software, particularly if intended to access the Internet, is sometimes written in higher level languages such as Java, and may include interpreted environments like PERL and Python. Visual Basic is another commonly used higher level language.

A typical setup for someone writing code for a microcontroller would include a desktop or laptop computer with the appropriate integrated development environment (IDE) software installed, as well as a microcontroller development board. IDE is an application, typically provided for free or purchased from the supplier of a particular microcontroller, which allows the user to write software, debug, download to the target (the microcontroller to be programmed), and perform other support functions. A development board usually contains the microcontroller chip and various peripherals to enable a user to conveniently interface with it. These would normally include a connector to communicate with the IDE host computer (often a USB port), power connectors, a manual reset button, and a few switches and LEDs that can be accessed by the microcontroller to aid in verifying the correct operation of a running program.

For writing code, the IDE can operate in a stand-alone mode, but when it is time to download the code it must be compiled, linked, and loaded. At this point, the target microcontroller must be connected to the IDE host computer in order to receive its copy of the program. After programming is successfully completed, the microcontroller can be disconnected from the host. There is usually a reset button on the microcontroller development board which allows the user to start the program running.

8.4 Programmable Logic Controllers

Programmable logic controllers (PLCs) are the most widely used control systems in factory automation today. A number of suppliers provide a wide variety of devices, but they have common features. PLCs are intended for real-time applications in which a set of input conditions (sensors and switches) control a set of outputs (motors, lights, etc.) according to some logic sequence.

While this seems like the type of behavior that can be implemented with conventional computers, the practical reality is that off-the-shelf computers cannot operate in harsh environments with sensitive timing requirements. PLCs are specifically enhanced for operation in challenging environments such as factory floors where they are routinely exposed to dust, dirt, vibration, power fluctuations, and extremes of temperature. Also, they are designed to operate in "hard" real time, meaning that the timing relationships between inputs and outputs are determined and predictable. Most computer systems using conventional operating systems are not able to provide this type of real-time functionality. Finally, PLCs are designed to handle real-world inputs and outputs without significant additional circuitry. This frees the user to concentrate on defining the logic for a specific application.

A typical PLC installation includes the PLC master unit, together with modules to connect inputs and outputs. The entire installation is often mounted in a specially designed housing such as a control cabinet or rack in a factory floor. The installation will provide a human interface through a programming terminal on a conventional PC, and/or a remote device or a control panel with pushbuttons, keypad, key switches, and a display. An example of a PLC installation is shown in Figure 8.2.

PLCs are most often programmed using ladder logic. This is a technique for describing a series of logical operations that, when diagrammed, appear in the form of a ladder. The general layout of a ladder logic program is shown in Figure 8.3.

Each rung (sometime called a "circuit") on the ladder is read sequentially by the PLC from top to bottom, left to right. Inputs are shown on the left, and outputs are shown on the right. Inputs can be true or false, and can be combined with simple logic. When the combination of inputs is true, the output becomes true. The left and right rails are sometimes called "power rails," but the ladder logic diagram is meant to help visualize the flow of "truth" rather than actual electrical current.

In the simplest case, an input is a physical input device such as a toggle switch, a momentary pushbutton, or a level sensor. However, as far as the PLC is concerned, the input is a logical construct. The "truth" of any input is determined by the PLC by looking at its internal state table. The state table is a section of memory that contains a bit (1 or 0) for the state of each input and output.

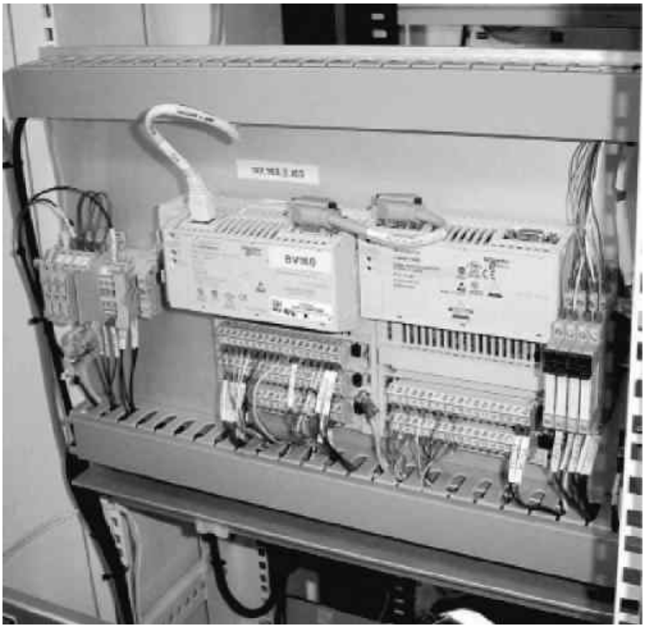


FIGURE 8.2 Example PLC installation.

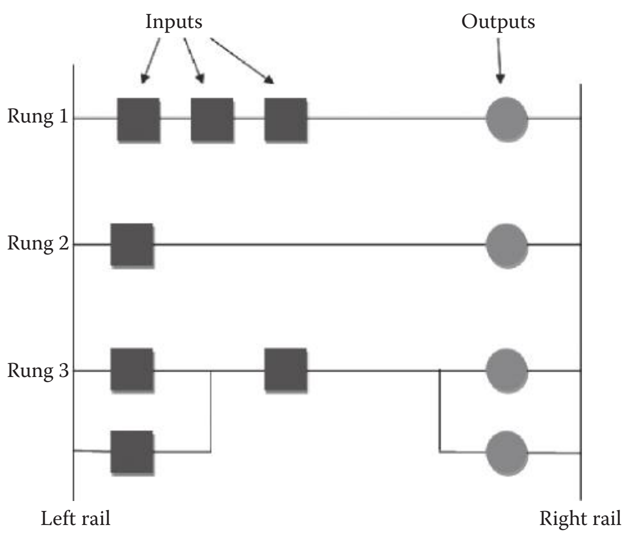


FIGURE 8.3 Anatomy of a ladder logic diagram.

Therefore, in the ladder logic diagram, an input can actually be the state of an output. For example, if a relay coil is energized so that the relay contacts are closed, the corresponding bit in the state table will be a "1." The state of the relay can appear on the left side of the ladder logic diagram as an input. Example 8.5 demonstrates a use for this by latching a momentary switch.

In order to reduce the possibility of logic conflicts, the PLC performs a first-pass reading through the entire ladder, and sets internal memory bits to indicate actions to perform on the outputs. During this pass, the outputs are not actually turned on or off. After this pass is completed, the outputs are turned on or off as indicated by the stored memory bit.

A brief digression to discuss switch notation is in order here. Although a seemingly simple topic, in fact there is a vast world of switch types and contact arrangements. Switches are often classified by the number of *throws* and the number of *poles*. The number of "poles" is the number of separate circuits which are controlled by a switch, while the number of "throws" is the number of separate positions that the switch can adopt. The abbreviations N.O. (or NO) and N.C. (NC) stand for normally open and normally closed, respectively. These are commonly encountered in ladder logic diagrams, and in circuit schematics in general. Figure 8.4 shows some of the most common switches.

Figure 8.5 shows typical notation for relay contacts. The movable contact connected to the COM (common) terminal is shown in its resting state (maintained by a spring or other mechanism), in which it is connected to the NC (normally closed) contact. When the coil is energized, the COM contact will

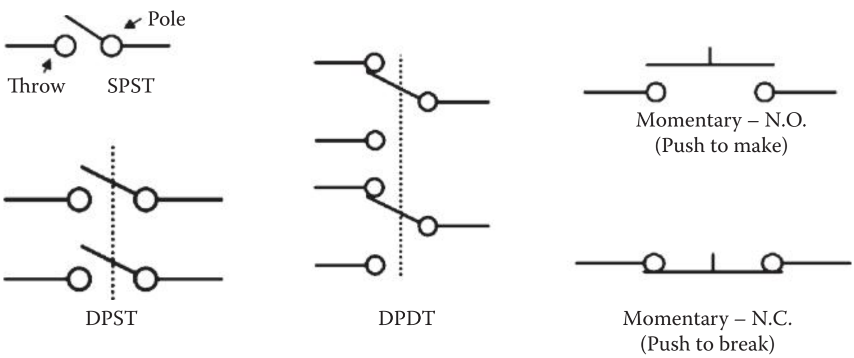


FIGURE 8.4 Types of switches. SPST: single-pole single-throw; DPST: double-pole single-throw; DPDT: double-pole double throw; N.O.: normally open; and N.C.: normally closed.

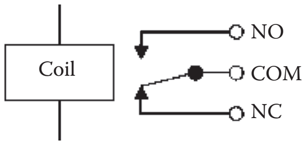


FIGURE 8.5 Relay contact notation.

be disconnected from NC and connected to NO for as long as the coil remains energized. When the coil is deenergized, the COM contact will return to NC. Relays often have more than one set of contacts actuated by the same coil.

Example 8.4: Using a Switch to Turn on a Fan

A very simple example of using a switch to turn on a fan is described. A schematic representation is shown in Figure 8.6. The ladder logic diagram is shown in Figure 8.7.

In this example, the ladder diagram corresponds closely to the electrical wiring diagram. When the Start switch is closed, the corresponding PLC internal memory bit will be set, the PLC output to which the fan is

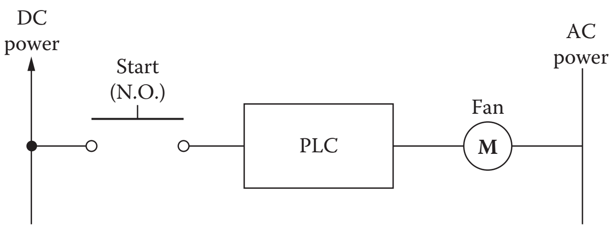


FIGURE 8.6 Schematic for turning on a fan using a momentary switch connected to a PLC.

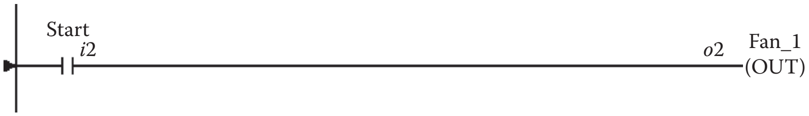


FIGURE 8.7 Ladder logic diagram corresponding to the circuit shown in Figure 8.6.

connected will operate to complete the electrical circuit, and the fan will be turned on. When the switch is opened, the PLC will read that event on a subsequent scan and turn off the fan. In order to turn on the fan, the PLC energizes the coil of a relay (or the equivalent solid-state circuitry) to which the fan is connected.

Example 8.5: Using Two Switches to Control a Fan

In this example, one switch is used to start a fan, and another switch is used to turn it off. Furthermore, a delay timer relay is used, which causes the fan to turn on 2 seconds after the Start button is pressed.

The wiring and ladder logic diagrams are illustrated in Figure 8.8 and Figure 8.9.

Operation is as follows:

- When the Start switch (N.O.) is pressed momentarily, the ensuing electrical continuity causes the corresponding PLC input table bit to become TRUE (1).
- As long as the Stop switch remains unpressed, it retains electrical continuity since it is N.C. The corresponding input table bit remains TRUE (1).
- With both inputs TRUE, the entire rung becomes true. The delay timer relay coil is energized, so its contacts will close in

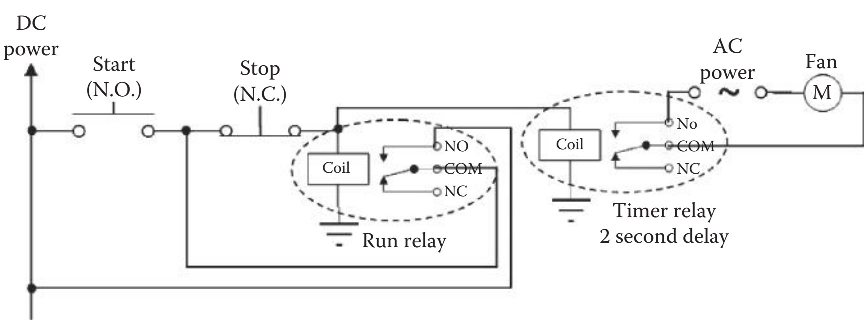


FIGURE 8.8 Equivalent circuit schematic. The relays are implemented by the PLC.

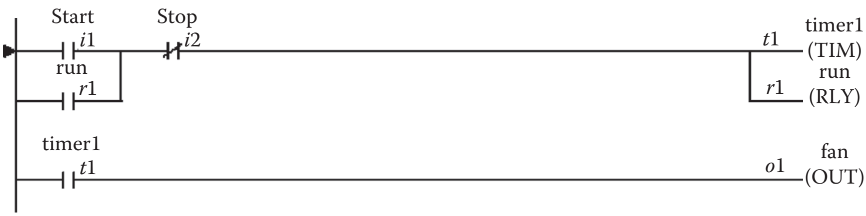


FIGURE 8.9 Using two switches to control a fan–ladder logic diagram.

2 seconds. At the same time, the RUN relay coil is energized and its contacts close immediately. This forms a closed circuit around the Start switch, so even if the start switch is released, the rung will remain energized. This is called *latching*.

- On the next rung, we see that when the timer contacts close, the fan will turn on.
- If the Stop switch is pressed at any time, its electrical continuity ceases; the internal state bit will become FALSE (0), and the top rung will become false, deenergizing both relay coils (timer1 and RUN). The fan will turn off, and the timer will reset.

Figure 8.10 shows the initial state of the simulation. Active elements (inputs, outputs, relays, etc.) are highlighted on the ladder logic diagram and in the state table. Initially, nothing is active.

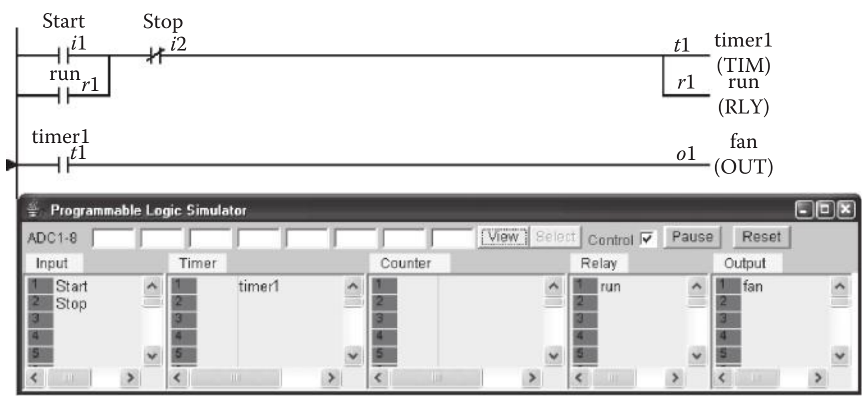


FIGURE 8.10 Initial state of the simulation.

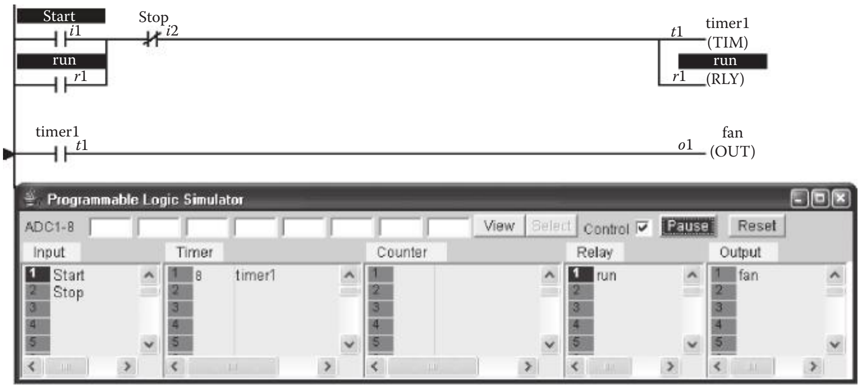


FIGURE 8.11 Start button pressed, delay timer running.

The simulation has been paused to capture the display shown in Figure 8.11. We can see the following from the ladder logic diagram and the state table:

- The *Start* button has been pressed since it is highlighted both on the ladder logic diagram (left side of top rung) and in the state table (item 1 under the heading "Input").
- The *run* relay coil is energized since it is highlighted both on the ladder logic diagram (right side of top rung) and in the state table (item 1 under the heading "Relay").
- The *run* relay contacts are closed—they are shown parallel to the Start switch input on the top rung, and they are highlighted.
- The delay timer (named "timer1") is running. The timer is programmed in 0.1 second intervals. For this example, the delay value was entered as 20, which corresponds to 2 seconds. The display shows that the timer has counted down from 20 to 8. When it gets to 0, the contacts will close and the fan will turn on.

Figure 8.12 shows the situation after the delay timer finishes counting down to 0. The timer contacts close, and the fan turns on. The *fan* output is highlighted on the ladder logic diagram (right side of lower rung) and in the state table (item 1 under heading "Output").

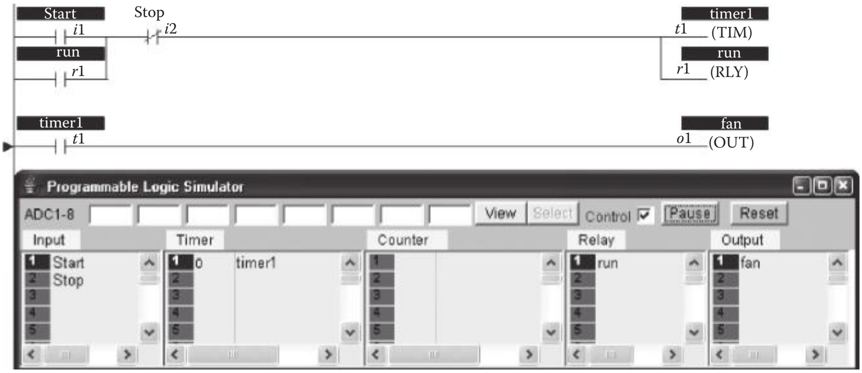


FIGURE 8.12 Delay finished, fan on.

In this state, momentarily pressing the Stop button will open the top rung, causing the PLC to deenergize the timer1, run relay coils, and turn off the fan. The system will not restart until the Start button is pressed.

The type of PLC environment described thus far is of the form known as a distributed control system. These are currently evolving into systems described as process automation systems (PASs) by the inclusion of additional functionality beyond basic control, including extensive communications networking and advanced human–machine interfaces. An example of this is the Rockwell Automation PlantPAx Process Automation System. It is proposed that the further evolution of PAS into collaborative PAS (CPAS) will add even more capability. In the next phase of their evolution, process automation systems will continue to facilitate process control but will also become the primary source of manufacturing data and information for collaborative manufacturing management (CMM) applications. The idea is for people at each level of the manufacturing environment with various responsibilities to tap into the database for information presented at a level and in a format that is useful to them.

8.5 Microcontrollers

Microcontrollers are the workhorses of modern digital technology. It is hard to think of an electronic system or device that does not contain at least one microcontroller-enabled function.

In many functional ways, a microcontroller is similar to a PLC. The main difference is that the microcontroller is designed to be embedded into a larger system rather than being self-sufficient. This means that the user must consider how to provide power, physical protection, and communications at a more detailed level than for a PLC.

A microcontroller is a single-chip product that contains a microprocessor together with input–output interfacing and memory. The microprocessor performs the core function of reading and executing program code, computing, and performing logic operations. The other onboard components provide access to the outside world.

Microcontrollers are the inevitable result of miniaturization that has taken place in the computer world since the 1960s. Early digital computers were made of racks of bread-boarded components. Integrated circuit technology led to the release of the first commercial microprocessor, the INTEL 4004, in 1971 (see Figure 8.13).

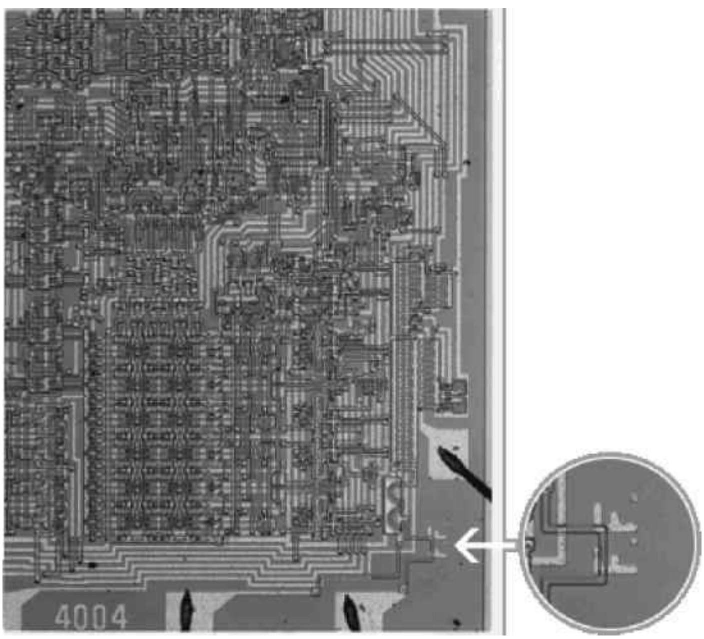


FIGURE 8.13 Initials F.F. (Federico Faggin) on the 4004 design (1971). The 4004 bears the initials F.F. of its designer, Federico Faggin, etched on one corner of the chip. Signing the chip was an original idea imitated after him by many Intel designers.

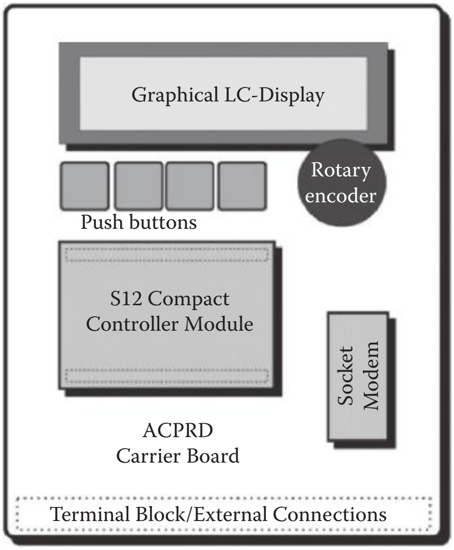


FIGURE 8.14 Alarm Control Panel Reference Design (ACPRD).

One general category to describe uCs is the bus width in bits. This gives the number of memory and address bits that flow in parallel through the device wiring. The 4004 was a 4-bit device. Today's standard uCs are 16 bits, though 32-bit devices are widely available and 64-bit units are available.

It is beyond the scope of this text to provide detailed instruction on the use of any one microcontroller, but several examples are provided to illustrate the procedures one can follow to select a device for a particular application, and then to design program code to enable a desired function. Some sections of code are shown.

Example 8.6: Alarm Panel Using Freescale HCS12

This example is based on the Alarm Control Panel Reference Design (ACPRD) by Freescale.[3] An alarm panel is designed as shown in Figure 8.14. The design uses a Freescale HCS12 microcontroller, specifically the MC9S12DP256 MCU. The Alarm Control Panel consists of two main units: the Carrier Board and the S12 compact Controller Module. The main components of the Alarm Control Panel are shown in Figure 8.15.

[3] http://www.freescale.com.

FIGURE 8.15 ACPRD board.

The S12 compact Controller Module is a small printed circuit board that holds the microcontroller unit (MCU) as well as standard circuitry for clock and reset generation, power supply and decoupling, plus a number of additional input–output devices. The Carrier Board provides a number of peripheral devices used to perform the specific functions of the Alarm Control Panel application. The Carrier Board is equipped with connections for alarm sensors, a flashlight, a siren, and a power supply. It also has sockets for a graphical LC display, a modem, and the S12 compact Controller Module.

The primary user interface is through the four pushbuttons and the jog dial shown below the display in Figure 8.15. The user can set operating modes and adjust parameters with these controls. Several inputs will be considered next by way of illustrating the connection between MPU architecture, system design, and software.

8.5.1 Input Voltage Sensing

A section of the ACPRD showing the ATD inputs is shown in Figure 8.16. There are two sources of input voltage: main input and battery backup. The nominal value is 12–16 VDC for the main power input at X1 and 10–12 VDC

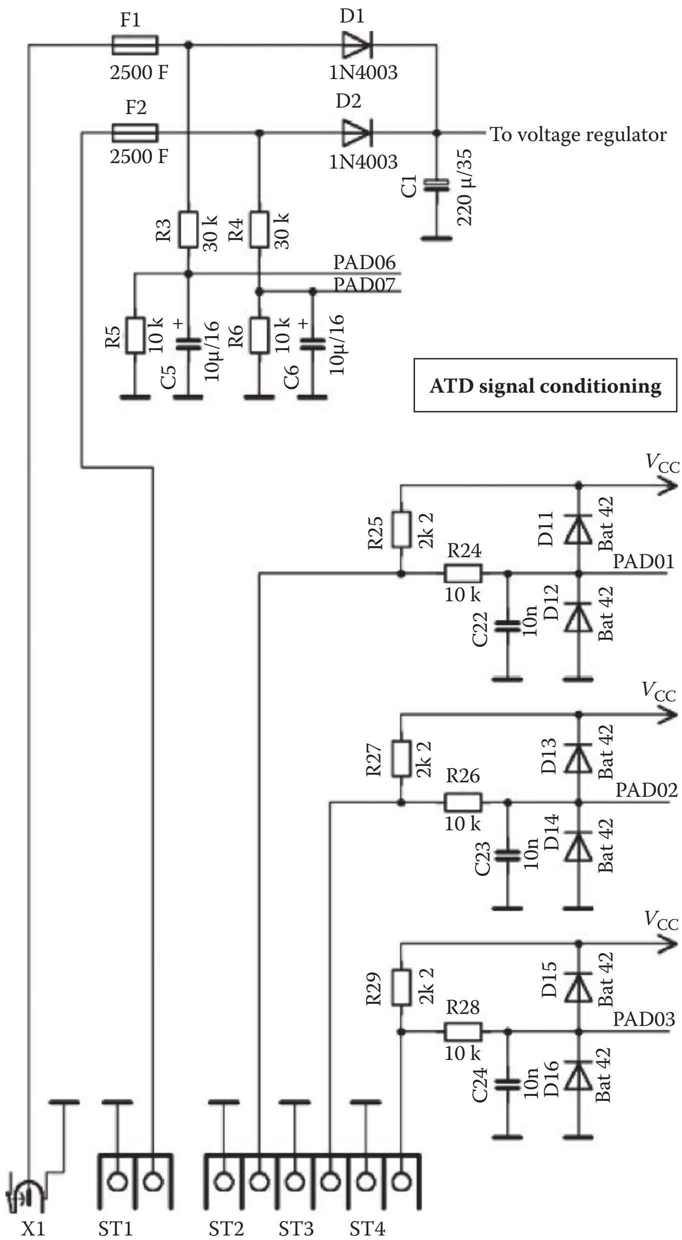


FIGURE 8.16 Circuit section showing ATD inputs.

for the backup battery input at ST1. (Physically, the input labeled X1 is a receptacle; the inputs labeled ST are screw terminal pairs. One of the screws connects to ground; the other one is the input.)

The voltage level of main power input and battery backup input is measured by PAD06 and PAD07, respectively. These two signals are inputs to the HCS12 on-chip analog-to-digital (ATD) converter. The input voltage levels are decreased by a factor of four using the resistor divider R3/R5 and R4/R6 in order to fit the input voltage range of the ATD, which is 0–5 V. Components C5 and C6 are filter capacitors to remove ripple from the unregulated input.

Next we look at the description of the software module by which the programmer can interface to the ATD, and we examine a segment of code which reads the power supply inputs PAD06 and PAD07.

8.5.2 Description of the S12 Analog-to-Digital Converter Module (S12_ATD)

This software module contains functions for driving the analog-to-digital converter of the HCS12. The MC9S12DP256 contains two such blocks, but this software module only deals with the first block (ATD0), which covers the input signals PAD00–PAD07.

The function *initATD0()* is called once in the initialization phase of the application. It enables the ATD hardware module and selects a 10-bit resolution mode.

The function *getATD0()* performs a single ATD conversion on a single channel. The channel number argument must be in the range of 0–7. The function will wait until the conversion is finished, which will cause an additional delay of about 10 microseconds.

8.5.3 Code Segment to Read Power Supply Voltages

A section of the code which instructs the MPU to read the input voltages is shown in Table 8.3. Lines beginning with a double forward slash // are comments. Lines 1–16 are required initialization. Line 17 calls the *initATD0* function which is required to enable the ATD module before a sample can be read. Line 19 sets the resolution to 10 bits. Line 29 calls the *getATD0* function which performs the actual read. The particular channel number to be read is

Table 8.3 Code Segment to Read Power Supply Voltages (Line Numbers Are for Text Reference)

1.	`//=============================================================================`
2.	`// File: S12_ATD.C - V1.00`
3.	`// Rem.: The ACPRD Project Page on the Web -> http://hc12web.de/acprd`
4.	`//=============================================================================`
5.	`//-- Includes ----------------------------------------------------------`
6.	`#include “datatypes.h”`
7.	`#include “hcs12dp256.h”`
8.	`#include “s12_atd.h”`
9.	`//-- Static Data -------------------------------------------------------`
10.	`UINT16 atd_results[ATD_AVERAGE_COUNT][ATD_MAX_CHANNELS];`
11.	`UINT16 atd_aidx;`
12.	`//-- Code --------------------------------------------------------------`
13.	`// Func: Initialize ATD module`
14.	`// Args: -`
15.	`// Retn: -`
16.	`//`
17.	`void initATD0(void) {`
18.	`// enable ATD module`
19.	`ATD0CTL2 = BM_ADPU;`
20.	`// 10 bit resolution, clock divider=12 (allows ECLK=6..24MHz)`
21.	`// 2nd sample time = 2 ATD clocks`
22.	`ATD0CTL4 = BM_PRS2 | BM_PRS0;`
23.	`}`

Table 8.3 Code Segment to Read Power Supply Voltages (Line Numbers Are for Text Reference) (Continued)

```
24. //-----------------------------------------------------
    -------------------
25. // Func: Perform single channel ATD conversion
26. // Args: channel = 0..7
27. // Retn: unsigned, left justified 10 bit result
28. //
29. UINT16 getATD0(UINT8 channel) {
30. // select one conversion per sequence
31. ATD0CTL3 = BM_S1C;
32. // right justified unsigned data mode
33. // perform single sequence, one out of 8 channels
34. ATD0CTL5 = BM_DJM | (channel & 0x07);
35. // wait until Sequence Complete Flag set
36. // CAUTION: no loop time limit implemented!
37. while((ATD0STAT0 & BM_SCF) == 0);
38. // read result register
39. return ATD0DR0;
40. }
41. //-----------------------------------------------------
    -------------------
```

specified in line 34 in the "channel" argument. As we learned above, the main power voltage input will be read on channel 6, and the battery backup input voltage will be read on channel 7.

After proper timing is completed, the result is returned at line 39.

8.5.4 Code Segment to Display Power Supply Voltages

A section of the code which instructs the MPU to display the input voltages is shown in Table 8.4. Line 4 calls the *drawVolts* function, which displays the readings on the LCD. Lines 8 and 9 display the readings obtained from ATD input channels 6 and 7, respectively, which correspond to the main power and battery backup voltages.

Table 8.4 Code Segment to Display Power Supply Voltages (Line Numbers Are for Text Reference)

```
//-----------------------------------------------------------
----------------------

UINT8 dobj_volts_x;
UINT8 dobj_volts_y;
BOOL dobj_volts_upd;

void drawVolts(void){

UINT16 sym, vpwr, vbat;

if(dobj_volts_upd == FALSE) return;
gotoxyDisp(dobj_volts_x, dobj_volts_y);
vpwr = shadow_atd0[6];
vbat = shadow_atd0[7];

sym = 0x93;
if((vpwr < ACPRD_VPWR_MIN) || (vpwr <= vbat))
sym += TA_BLINK;
_drawVolt(sym, vpwr);

sym = 0x94;
if(vbat < ACPRD_VBAT_MIN)
sym += TA_BLINK;
_drawVolt(sym, vbat);

// draw line below
gotoxyDisp(dobj_volts_x, dobj_volts_y + 1);
prDispMult(14, '\x8e');
}

//-----------------------------------------------------------
----------------------
```

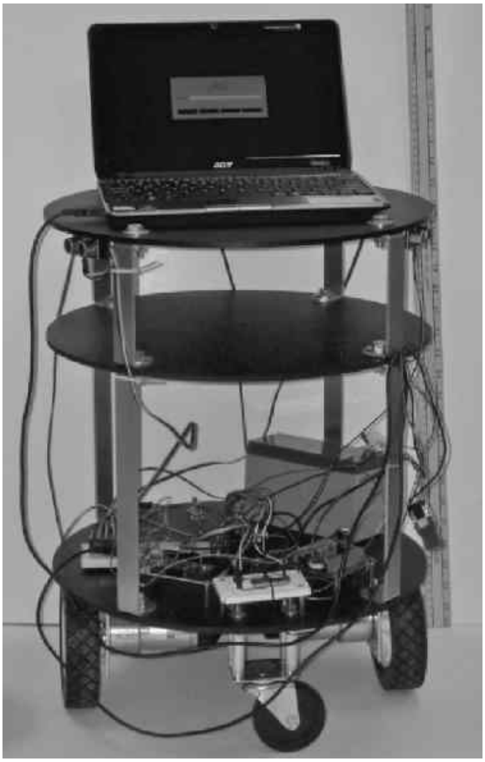


FIGURE 8.17 Mobile robot.

Referring to Figure 8.16, we can also see how the alarm inputs are read. They are connected to screw terminals ST2, ST3, and ST4 and read using ATD channels 1, 2, and 3 respectively.

Example 8.7: Controlling a Mobile Robot Using an Arduino Microcontroller

This example illustrates control of a mobile robot using an Arduino microcontroller. Information about Arduino can be found at the Arduino Home Page.[4] The mobile robot, which is used for research experiments in the Cooperative Robotics Research Laboratory at Auburn University, is pictured in Figure 8.17.

[4] http://www.arduino.cc.

This robot consists of the following components that are used for basic operation:

- Motors: two center-mounted differential drive 12 VDC motors
- Motor controller: a dual-channel pulse-width modulator (PWM) motor driver
- Motor shaft encoders
- Sonar sensors
- Servos to position sonar sensors

In addition, the high-level control such as path planning is performed on an onboard laptop, so the Arduino must communicate with the laptop.

This design uses the Arduino Pro board shown in Figure 8.18. The Arduino Pro used in this example is based on the ATMEGA328 microcontroller. A block diagram of the ATMEGA328 is shown in Figure 8.19.

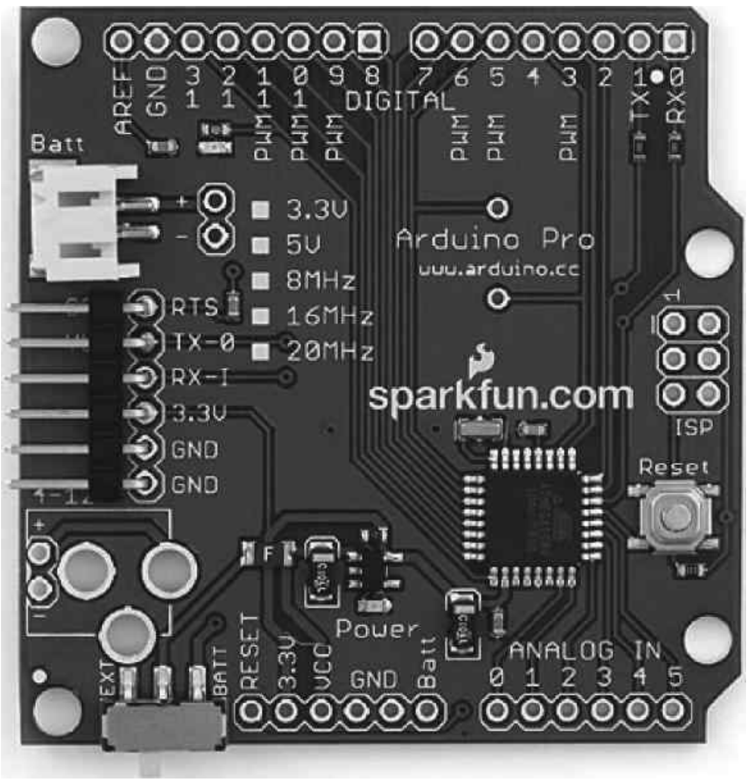


FIGURE 8.18 Arduino Pro.

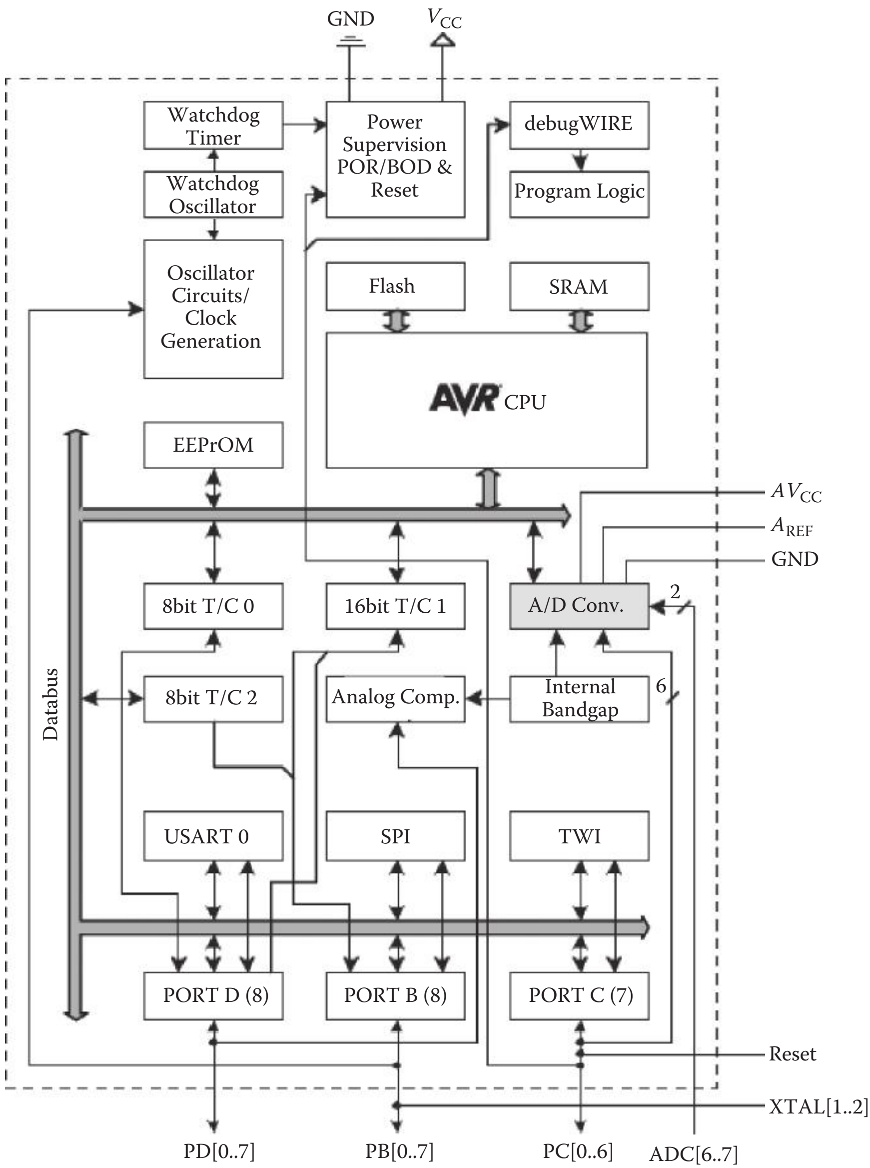


FIGURE 8.19 Block diagram of the ATMEGA328 microcontroller.

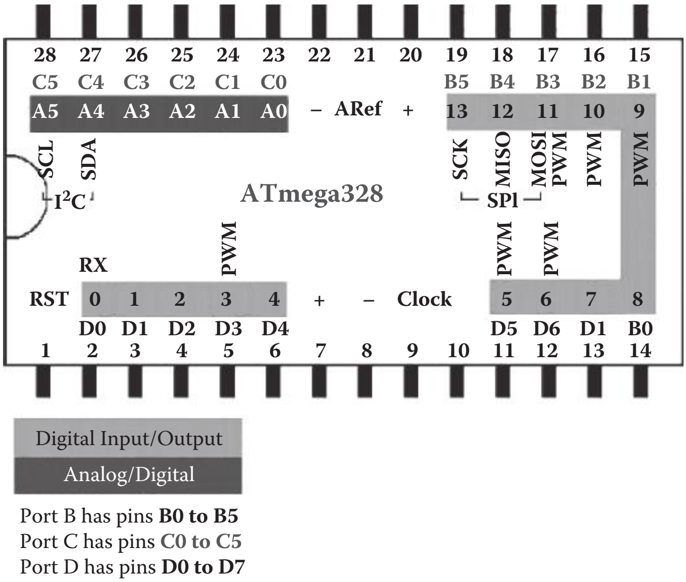


FIGURE 8.20 Pin assignments are for the 28-pin ATMEGA328 chip. Pins have shared functions.

The section labeled AVR CPU is the central processing unit. This is the core of the device, which performs the key functions of program execution, arithmetic, and logic. The sections labeled T/C are timers and counters, some of which also serve as pulse-width modulation (PWM) registers. The blocks labeled as ports along the bottom of the diagram are groups of inputs and outputs that serve various functions under program control, such as PWM, digital input–output, and analog input–output. The ports do not necessarily correspond one to one to physical pins on the chip; rather, they are logical constructs that can be addressed through software. The physical pins may have different functions depending on the current configuration defined by the software. This is illustrated in Figure 8.20.

The Flash and SRAM (static random access memory) blocks are on-chip memory registers that can be accessed directly by the CPU.

In order to provide a useful example, we consider a particular application of our mobile robot which is to patrol a particular floor in a building. The robot will traverse the building by following the walls at a fixed distance. It will thereby provide roving "eyes and ears" while connected to the building intranet.

8.5.5 Motor Operation

This robot uses a very common drive system—two center-mounted wheels attached to the shafts of dc gearmotors. The gearmotors are driven by a pulse-width modulated (PWM) signal. The PWM signal is a voltage that alternates rapidly between full battery voltage (high) and zero (low). The inertia of the motor–robot system causes the motor to respond only to the average value. The average value is changed by changing the duty cycle—that is the percentage of time when the voltage is high relative to the total period (high plus low time). Increasing the PWM duty cycle causes the motor to turn faster; reducing the duty cycle slows it down. By adjusting the duty cycles on the left and right motors, the robot can be steered. With a suitable controller capable of reversing the motor direction, the robot can be rendered highly maneuverable and can even turn in place.

The Arduino has a hardware and software environment specifically for enabling PWM operation on up to six channels. Only two are used for this example—one for the right motor and one for the left motor.

The PWM signal from the Arduino is a low-power digital output. It is not suited for driving a motor directly. The PWM signal is connected to a motor controller which interfaces the PWM signal, the battery, and the motors. In this case, the L298-based Compact Motor Driver kit from Solarbotics was used. The assembled circuit is shown in Figure 8.21.

The motor driver receives input control signals for each motor from the Arduino, including the PWM signals. It is also supplied raw battery voltage. The motor controller then applies the battery voltage to the motor according to the PWM and control signals.

A sample of code that drives the motors is shown in Table 8.5.

Lines 4–11 send the correct logic states to the Motor 1 Control and Motor 2 Control (refer to Figure 8.21) to enable the motors for PWM control. Lines 13 and 19 calculate the value to be sent to the motor controller. The PWM value can be from 0 to 255. Lines 14 and 20 cause the calculated values to be sent to the motor controller. Lines 25–28 send the control signals that will disable the motors in order to stop the robot.

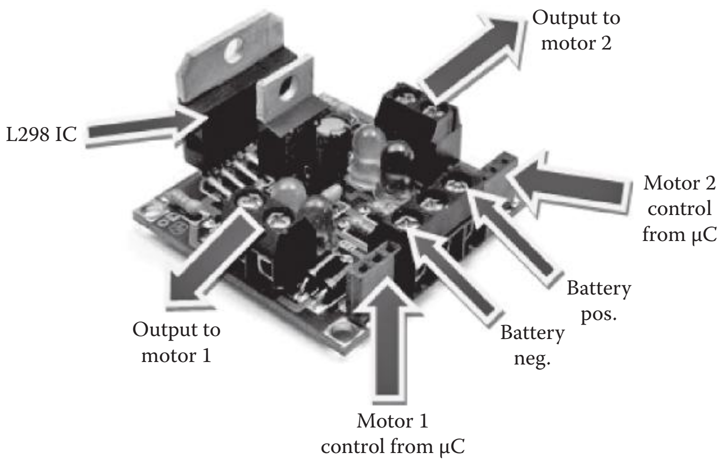


FIGURE 8.21 L298-based Compact Motor Driver kit, assembled.

Table 8.5 PWM Motor Control Code (Line Numbers Are for Text Reference Only)

1.	`// Here is the function we call to enable to PWMs for the motors. It is the forward() function in our case. This tells the controller to run the motors at whatever PWM percentage we told it last.`
2.	`void Motor::forward()`
3.	`{`
4.	`digitalWrite(_rightI1pin, HIGH);`
5.	`digitalWrite(_rightI2pin, LOW);`
6.	`digitalWrite(_leftI1pin, HIGH);`
7.	`digitalWrite(_leftI2pin, LOW);`
8.	`}`
9.	`// To update the percentages we use the right_mspeed(int percent) and left_mspeed(int percent) functions that take in a int between 0 and 100 and then updates the controller that percentage for the left and right motors respectively.`
10.	`void Motor::right_mspeed(int percent)`

Table 8.5 PWM Motor Control Code (Line Numbers Are for Text Reference Only) (Continued)

```
{
_rightPercent = percent;
_rightpwm = (_rightPercent*255)/100;
analogWrite(_rightEnablePin,_rightpwm);
}
void Motor::left_mspeed(int percent)
{
_leftPercent = percent;
_leftpwm = (_leftPercent*255)/100;
analogWrite(_leftEnablePin, _leftpwm);
}
// If we wish for the robot to stop, we use the
mstop() function
void Motor::mstop()
{
digitalWrite(_rightI1pin, HIGH);
digitalWrite(_rightI2pin, HIGH);
digitalWrite(_leftI1pin, HIGH);
digitalWrite(_leftI2pin, HIGH);
}
```

8.5.6 *Position Sensing*

The robot uses two types of inputs to determine its location. The motors have built-in encoders which provide a high-resolution digital indication of shaft position. The robot has also been equipped with several sonar sensors (Parallax PING) which are used to measure distance to the wall.

The shaft encoders output a *quadrature-encoded* signal. This type of signal allows determination of both shaft position and direction of rotation.

The encoder output is a series of pulses. These are interfaced to the Arduino using interrupts. An interrupt is a special type of input that triggers the microcontroller to stop what it is currently doing and perform some other action

FIGURE 8.22 **PING))) Sonar Sensor.**

instead. The other action is called an interrupt service routine (ISR). When the ISR is finished, the uC returns to where it left off.

Since the encoder calibration is known, each time the ISR is executed the microcontroller can accurately update the shaft position. This value can be used to estimate the location and heading of the robot using the appropriate equations of motion. In the absence of slippage and other errors, this approach, known as "dead reckoning," could be used exclusively to drive the robot. However, in practice a real robot cannot drive blindly very far without incurring significant error. We must have some other sensors to provide closed-loop control. Just as the driver of a car cannot close his eyes for long, we must have some sensor input from the external world in order to drive the robot successfully.

One could use a variety of sensors including cameras, infrared, and sonar, among others. For simplicity and low cost, we use sonar (ultrasound) sensors in this example. We use the PING))) sensor shown in Figure 8.22. Example 7.2 provides some further insight into using a PING))) with a microcontroller.

The PING))) emits a burst of ultrasound, and then provides a signal indicating the time required to hear the echo. The signal is digital in amplitude, but the Arduino must measure the time that it stays HIGH. A section of code that performs this is shown in Table 8.6.

Lines 11–13 send out the trigger signal required to start the PING)))'s ultrasound output burst. Line 11 sets the physical pin to which the PING))) signal line is connected to a logic HIGH state. Line 12 causes a 5-μs delay, and line 13 returns the PING))) connection pin to a logic LOW state. The net result is a 5-μs trigger pulse to the PING))). The PING))) will set the same pin HIGH while it waits for the ultrasound pulse to return. Line 17 sets the pin to input so the Arduino is now "listening" for a pulse. Line 18 reads the duration of the

Table 8.6 Code to Read a PING))) Sensor

```
int SONAR_Sensor::sonar_ping()
{
// establish variables for duration of the ping,
// and the distance result in inches and centimeters:
long duration, inches, cm;
// The PING))) is triggered by a HIGH pulse of 2 or more microseconds.
// Give a short LOW pulse beforehand to ensure a clean HIGH pulse:
pinMode(_SONAR_Pin, OUTPUT);
digitalWrite(_SONAR_Pin, LOW);
delayMicroseconds(2);
digitalWrite(_SONAR_Pin, HIGH);
delayMicroseconds(5);
digitalWrite(_SONAR_Pin, LOW);
// The same pin is used to read the signal from the PING))): a HIGH
// pulse whose duration is the time (in microseconds) from the sending
// of the ping to the reception of its echo off of an object.
pinMode(_SONAR_Pin, INPUT);
duration = pulseIn(_SONAR_Pin, HIGH);
// According to Parallax's datasheet for the PING))), there are
// 73.746 microseconds per inch (i.e. sound travels at 1130 feet per
// second). This gives the distance travelled by the ping, outbound
// and return, so we divide by 2 to get the distance of the obstacle.
// See: http://www.parallax.com/dl/docs/prod/acc/28015-PING-v1.3.pdf
//Uncomment this if inches are desired and comment out CM.
```

(continued)

Table 8.6 Code to Read a PING))) Sensor (Continued)

25.	`// inches = duration / 74 / 2;`
26.	`// The speed of sound is 340 m/s or 29 microseconds per centimeter.`
27.	`// The ping travels out and back, so to find the distance of the`
28.	`// object we take half of the distance travelled.`
29.	`cm = duration / 29 / 2;`
30.	`delay(100);`
31.	`return cm;`
32.	`}`

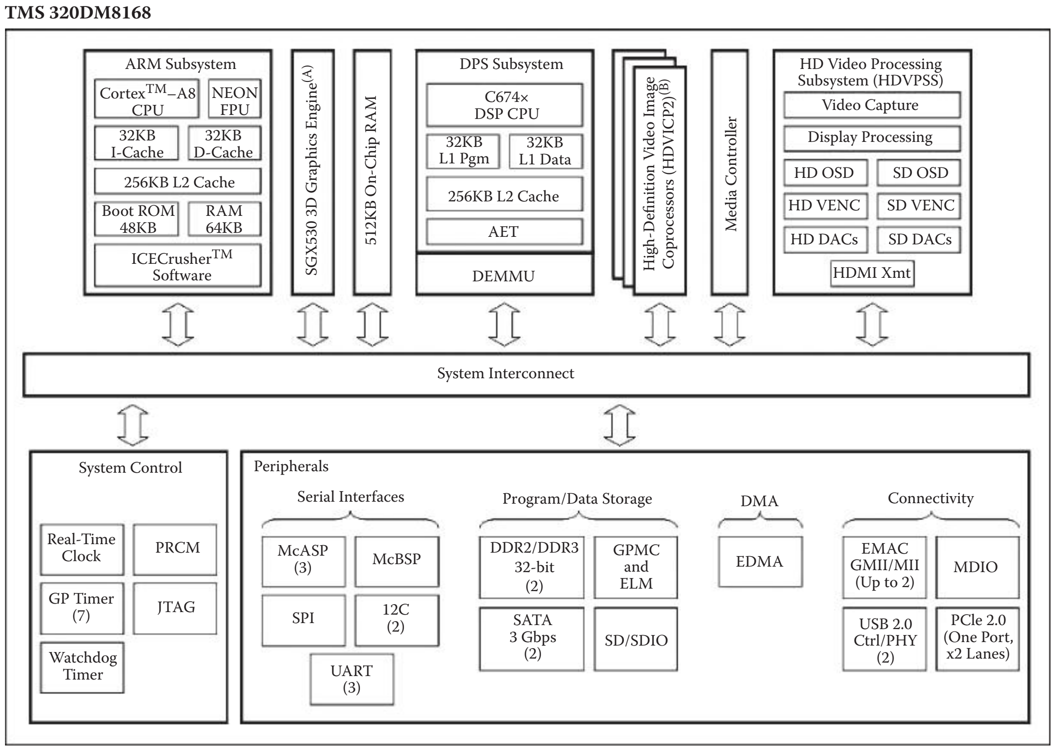
pulse—that is, the time that it remains HIGH. Line 29 performs the conversion from time to distance in centimeters. If distance is preferred in inches, line 25 can be used instead.

By aiming two PING))) sensors toward the wall, it is possible to determine the distance of the robot's center from the wall and its heading relative to the wall. For example, if both sensors indicate the same value, we can assume the robot is moving parallel to the wall. If the front sensor reads longer, we would know the robot is turning away from the wall, and we can correct accordingly. Using a simple proportional correction scheme, the robot can be made to follow the wall. By introducing several more sonar sensors, we can handle corners, intersections, and obstacles and cause the robot to move around the floorplan reliably.

8.6 Digital Signal Processors (DSPs) and Digital Media Processors (DMPs)

Microcontrollers that have evolved to handle arrays of signals simultaneously and to provide sophisticated signal-processing capability are known as "digital signal processors" (DSPs). "Processing" usually means filtering, encoding, and decoding. Conventional applications for DSP include real-time audio and radar signal processing.

Further evolution of these devices has led to a class of devices with even more capability, largely focused on video applications, known as digital media processors (DMPs). The block diagram for one such device is shown in Figure 8.23. The block in the upper left labeled "Cortex CPU" is the central



A. SGX350 is available only on the TMS320DM8168 and TMS320DM8166 devices.
B. Three HD Video Image Coprocessors (HDVICP2) are available on the TMS320DM8168 and TMS320DM8167 devices; two are available on the TMS320DM8166 and TMS320DM8165 devices.

FIGURE 8.23 Functional block diagram for the Texas Instruments TMS320DM8168 Digital Media Processor.

processing unit. To the right of that is the "NEON FPU" which is the floating-point unit. This handles most of the intensive floating-point arithmetic calculations. We can observe a module near the center top labeled "DSP subsystem." Thus we see that the previous generation of capability is embedded within this evolutionary step. There are several sections of various types of memory. There are also several sections dedicated to video capture and processing.

The functions performed by this chip would have taken racks of special-purpose equipment in past generations of electronics, costing many tens of thousands of dollars. Thus we see an illustration of Moore's Law, which predicts the geometric increase in functionality and the reduction in cost for electronics as time advances.

Problems

8.1. List three or more example applications where a PLC installation would be more suitable than a conventional computer, and explain why. List three or more example applications where a conventional computer would be preferred, and explain why.

8.2. In a certain PLC installation, a fan can be turned on by toggling either (or both) of two switches to the on position. Draw the corresponding ladder logic diagram.

8.3. In a certain PLC installation, a fan can be turned on only if both of two switches are toggled on. Draw the corresponding ladder logic diagram.

8.4. In a certain PLC installation, it is desired to use a momentary switch (on button) to turn on a motor. It is desired for the motor to stay on even when the button is released. The motor should turn off when another momentary switch (off button) is pressed. Draw the corresponding ladder logic diagram.

8.5. In a certain PLC installation, a toggle switch will be used to turn a motor on and off. It is desired to have a 5-second delay between when the switch is toggled on and when the motor starts. Also, as soon as the switch is toggled on, an alarm bell should be activated and a light should come on. When the motor starts, the alarm bell should turn off, but the light should remain on as long as the motor is running. Draw the corresponding ladder logic diagram.

8.6. Two 12 V automotive batteries are connected in series to operate a warehouse cart that uses 24 V motors. However, they need to be connected in parallel for charging. The operator will have a single manual switch with two positions labeled "charge" and "run" to accomplish this. Using one manual switch and a minimum number of relays, each of no greater complexity than DPDT, design a suitable circuit. Assume the relay coils are 12 VDC.

8.7. What is the most advanced microcontroller that you can find at the time you work this problem? How do you define "advanced"? What is the approximate cost?

8.8. What is the voltage divider ratio provided by resistors R3 and R5 in Figure 8.16? What is the largest voltage that can appear at main battery terminal X1 if PAD06 is not to exceed 5 V?

8.9. Consider the circuit schematic in Figure 8.16. If the battery backup voltage at ST1 is 12 V, what is the voltage at PAD07?

8.10. Give a mathematical proof that the average value of a pulse-width modulated signal is equal to the duty cycle (fraction of time the pulse is high relative to the period).

8.11. Consider a differential drive robot like the one discussed in Example 8.7. Assume the motors are identical. (a) If the right motor PWM duty cycle is 100% and the left motor duty cycle is 0%, describe the motion of the robot. Draw a set of simple sketches illustrating the motion. (b) Repeat if the right motor PWM duty cycle is 100% and the left motor duty cycle is 50%.

8.12. Consider a differential drive robot like the one discussed in Example 8.7. Assuming the motors are identical, if the right motor PWM duty cycle is 100% forward and the left motor duty cycle is 100% reverse, describe the motion of the robot. Draw a set of simple sketches illustrating the motion.

8.13. Which is more correct: a single pin on a microprocessor typically has a single function, or a single pin has multiple functions? Give two or more specific examples of pins with multiple functions on the ATMEGA328.

8.14. What is the primary reason for using a motor control board such as the L298-based kit described in the text? Why is it that a microcontroller output cannot be connected directly to a motor?

8.15. Make an accurate sketch of the waveform generated by lines 8–13 of the PING))) code shown in Table 8.6. Assume logic HIGH = 5 V and logic LOW = 0 V.

This page intentionally left blank

Chapter 9

Robots

This chapter introduces concepts underlying fixed-base industrial robots, as well as mobile robots such as those used for search-and-rescue operations. The inverse kinematics problem for industrial robots is introduced, and safety issues are discussed. Concepts of path planning, collaboration, and communication are introduced for mobile robots.

9.1 Industrial Robots: Classification and Terminology

9.1.1 Overview

Industrial robots are used throughout the world in nearly every type of industry for fabricating, packaging, and inspecting products. Typical applications of robots include welding, painting, assembly, packaging, product inspection, and testing. Most robots are powered by dc or ac electric motors. Motors are typically combined with control electronics to form what are commonly called *drives*. Electric drives are cheaper, cleaner, and quieter than other types, and can be very precisely controlled. Hydraulic actuators, or rarely pneumatic actuators, are also used to power industrial robots. Hydraulic drives have quick responses and can produce high torque. They excel at lifting heavy loads, such as automobile engines. Drawbacks of hydraulics include leaking hydraulic fluid (therefore, they are unsuitable for food processing and the handling of many types of materials), noise, and considerable peripheral equipment including pumps and hoses which require maintenance. Pneumatic drives are inexpensive and simple, but cannot be precisely controlled.

Industrial robots are composed of links and joints. The links are the rigid members; the joints allow the links to move relative to one another. The system of links and joints is called a "kinematic chain." Robots are broadly classified by their kinematic linkages: the number and types of joints that are used. Some of the most commonly used configurations are articulated robots, SCARA robots, and Cartesian coordinate robots (aka gantry robots or x–y–z robots).

An articulated robot is a robot with all rotary joints. Most commonly there are three joints, often designated "waist," "shoulder," and "elbow" by analogy with the human body. The linkage extending from the last joint is called the "forearm." The tooling is attached to the forearm by a wrist joint. The wrist

FIGURE 9.1 Articulated robot.

may have several (commonly three) axis motions as well. An articulated robot is shown in Figure 9.1.

SCARA stands for “selective compliant articulated robot arm” (or “selective compliant articulated robot for assembly”). The SCARA linkage has two rotary joints with parallel axes, and one linear joint. This particular configuration was originally developed to enable certain assembly tasks, such as parts insertion. In particular, it reduces the potential for binding when inserting parts into tight-fitting receptacles. A SCARA robot is shown in Figure 9.2.

Cartesian robots are attached to a framework (gantry) and move in Cartesian (x–y–z) coordinates. A large gantry robot is shown in Figure 9.3.

FIGURE 9.2 SCARA robot.

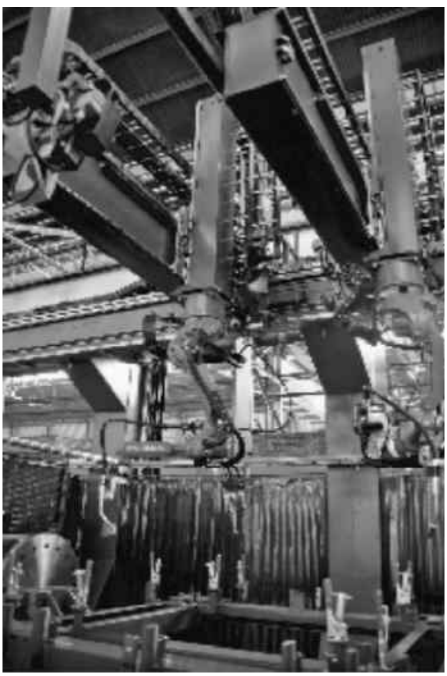


FIGURE 9.3 Gantry (Cartesian) robot.

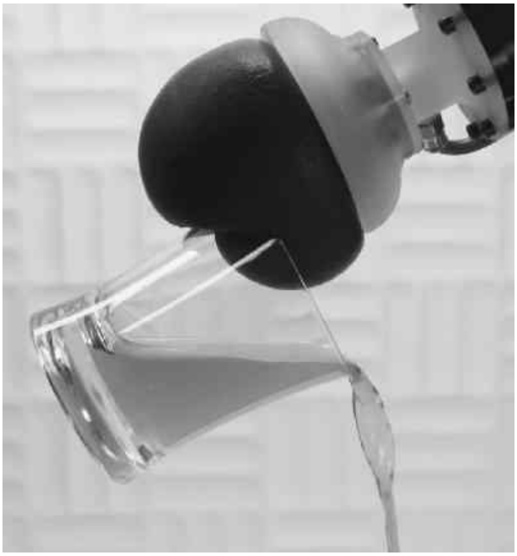


FIGURE 9.4 Highly specialized end effector.

A robot joint provides one axis of motion, or degree of freedom (DOF). A joint may provide linear (also called "translational" or "prismatic") motion, or rotary (also called "revolute") motion. Robot configurations are often designated by the kinematic linkage of their main joints. For example, an articulated robot is designated RRR to indicate that all three main joints are rotary. A SCARA robot would be designated RRP, having two rotary joints and a linear (prismatic) joint.

Generally, the purpose of a robot is to move an end effector (gripper, tool, etc.) in a particular path or to a particular point, and to perform a desired operation with it, such as welding or moving material. A specialized end effector used to lift and pour a glass is shown in Figure 9.4. A more typical end effector would be a paint sprayer, welder, polisher, or gripper. A deburring tool is shown in Figure 9.5.

A comparison between the five most widely used robot types is provided in Table 9.1.

The diagram of the joint structure for a spherical robot is shown in Figure 9.6. From this diagram we can see that the position of the end effector can be easily defined in spherical coordinates (azimuth angle, elevation angle, and radial travel).

FIGURE 9.5 Robotic end effector—deburring tool (ATI Industrial Automation Model RC-151).

Table 9.1 Comparison among Robot Types

Robot	Joints	Applications	Comments
Articulated	RRR	Welding, painting, grinding	Large freedom of movement in a compact space.
Spherical	RRP (mutually perpendicular axes)	Welding, painting, grinding	Relatively easy to control using a polar (spherical) coordinate system.
SCARA	RRP (mutually parallel axes)	Assembly of smaller parts, pick and place	Selective compliance helps with parts placement.
Cylindrical	RPP	Material transfer	Relatively easy to control using a cylindrical coordinate system.
Cartesian	PPP	Pick and place, heavy parts assembly	High structural rigidity, hence higher precision.

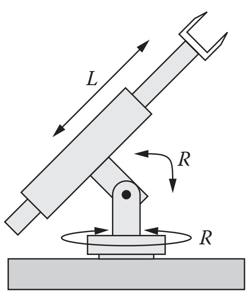


FIGURE 9.6 Diagram of a spherical robot joint structure. The kinematic arrangement is RRP. The end effector shown is a gripper. In general, this might be any type of tool.

9.1.2 Degrees of Freedom

One key descriptor of an industrial robot is the number of DOFs it offers, which is equal to the number of axis motions allowed. This is in turn determined by the number of joints. More DOF can provide more functionality, but also requires more sophisticated control. The majority of industrial robots are six-DOF machines. Six DOFs provide complete position and orientation in three-dimensional space.

For comparison, a human arm is considered to have seven DOFs. A shoulder gives pitch, yaw, and roll; an elbow allows for pitch and roll; and a wrist allows for pitch and yaw. Only three of those movements would be necessary to move the hand to any point in space, but people would lack the ability to grasp things from different angles or directions.

A robot (or object) that has mechanisms to control all six physical DOFs is said to be holonomic. An object with fewer controllable DOFs than its total DOFs is said to be nonholonomic, and an object with more controllable DOFs than its total DOFs (such as the human arm) is said to be redundant. An automobile is an example of a nonholonomic system, since the driver cannot move the vehicle in an arbitrary direction. If this were possible, parallel parking would be much simpler!

9.1.3 Workspace

The workspace of a robot is the region of space swept out by the end effector as it executes all possible motions. This will be determined by the geometry of the linkages, as well as by the inherent mechanical constraints on the joints. In particular, linear joints have travel limits and rotary joints are usually limited to less than a full 360° of rotation due to cabling and physical construction. Figure 9.7 shows the workspace for a cylindrical robot.

9.1.4 Control

The control problem is to move the end effector to a desired set of points or along a prescribed path, often while maintaining a required level of contact force with a workpiece. In principle, this is accomplished by controlling the actuators of each joint independently to achieve a known, desired configuration at each instant of time. There are several complications, however. First, the inverse kinematics must be solved to get each joint position as a function of the end effector position. This is not necessarily a unique solution. For example, think about how many ways you might arrange your joints while

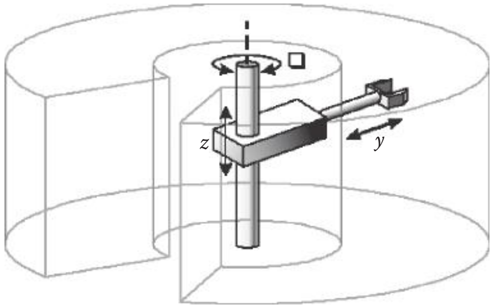


FIGURE 9.7 Cylindrical robot workspace.

grasping a pencil. Second, the joints are not fully independent since the links are not perfectly rigid. Third, operation is load dependent. A link carrying a heavy load will respond differently from one carrying a light load. Finally, all mechanical assemblies have variations such as backlash and slip, and are therefore neither perfectly accurate nor repeatable.

Considering these realities, engineers have developed an extensive body of knowledge involving a great deal of mathematical rigor tied to practical applications. Typical textbooks and references in the field have chapters on geometric transformations, forward and inverse kinematics, path and trajectory planning, dynamics, and other advanced topics.

By way of introducing the student to this topic, we will here consider a simple two-link robot with two rotating joints. The robot is illustrated schematically in Figure 9.8, and the geometry is shown in Figure 9.9.

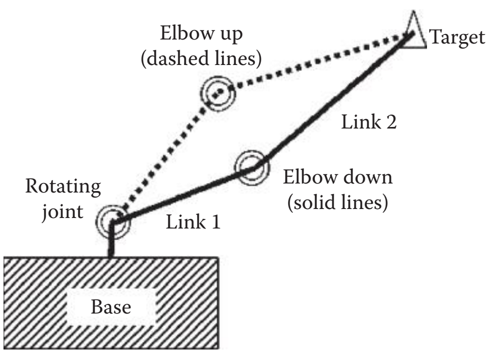


FIGURE 9.8 Two-link robot showing two possible configurations to reach the target.

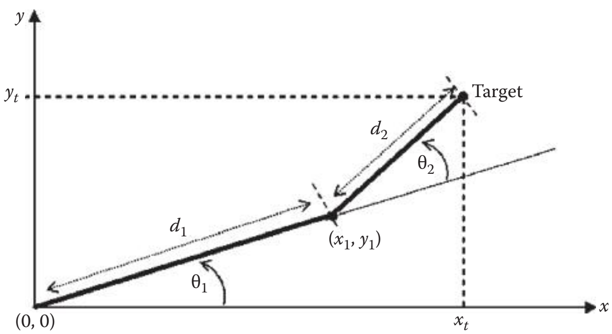


FIGURE 9.9 Geometry and definitions for the two-link robot problem.

The basic problem to be solved can be stated simply—given the target coordinates (x_t, y_t) and the link lengths d_1 and d_2, find the angles θ_1 and θ_2 that will place the end of link 2 at the target, if it is reachable. If it is not reachable, discover that and so indicate it. This problem is called the "inverse kinematics problem." It should be fairly clear that if the target is reachable, there are usually two equally valid solutions. These are referred to as "elbow up" and "elbow down" here. Other variants are commonly used, such as "lefty," "righty," and so forth.

We should reiterate that this two-joint, two-link problem is discussed for illustration only. Most industrial robots have more joints, together with numerous constraints on possible motions, making the real-world problem much more difficult. However, the student can glean an appreciation for the harder problem by studying this simple problem.

We begin the discussion by reminding the reader of the law of cosines given in Equation (9.1). This equation from geometry pertains to any triangle:

$$c^2 = a^2 + b^2 - 2ab\cos(\gamma) \tag{9.1}$$

where a, b, and c are the lengths of each side, and γ is the angle opposite to the side with length c.

We shall apply this to our problem with $\gamma = 180° - \theta_2$, where c = length of a straight line drawn from the origin (0, 0) to the target (x_t, y_t), $a = d_1$, and $b = d_2$. Referring to Figure 9.9, we note that γ is the interior angle formed by link 1 and link 2. This is illustrated for clarity in Figure 9.10.

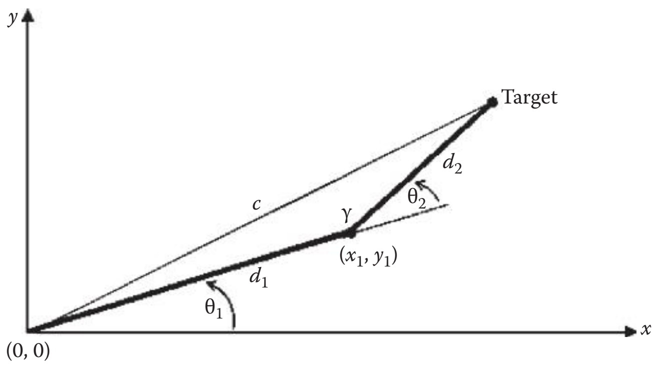


FIGURE 9.10 Application of the law of cosines to the two-link robot problem.

Observing that $c^2 = x_t^2 + y_t^2$, we can use Equation (9.1) to write

$$x_t^2 + y_t^2 = d_1^2 + d_2^2 - 2d_1d_2\cos(\gamma) \tag{9.2}$$

Using the fact that $\gamma = 180 - \theta_2$, we note

$$\cos(\gamma) = \cos(180 - \theta_2) = -\cos(\theta_2) \tag{9.3}$$

Substituting Equation (9.3) into Equation (9.2) and solving for $\cos(\theta_2)$ yields

$$\cos(\theta_2) = \frac{x_t^2 + y_t^2 - d_1^2 - d_2^2}{2d_1d_2} \triangleq D \tag{9.4}$$

in which we define D for convenience.

We can solve for θ_2 by taking the arccosine of D. At this point, we can further observe that the two possible solutions, elbow up and elbow down, are related by having the same magnitude of the angle between link 1 and link 2, but opposite signs. In particular, we can see that a positive sign on θ_2 in Figure 9.10 yields the elbow down configuration. We will adopt this convention in the following discussion and examples.

$$\theta_2 = \cos^{-1}(D) \quad \text{... elbow down configuration}$$

$$\theta_2 = -\cos^{-1}(D) \quad \text{... elbow up configuration} \tag{9.5}$$

Now that θ_2 has been found, we need to calculate θ_1. It can be shown after some work that θ_1 is given by

$$\theta_1 = \tan^{-1}\left(\frac{y_t}{x_t}\right) - \tan^{-1}\left(\frac{d_2 \sin(\theta_2)}{d_1 + d_2 \cos(\theta_2)}\right) \tag{9.6}$$

If the arctan function in Equation (9.6) is calculated carefully with regard to the sign of the argument, the correct results can be obtained in all quadrants. In particular, if one uses Microsoft Excel™, the ATAN2 function should be used, rather than ATAN. Example 9.1 illustrates the use of the foregoing equations to solve the inverse kinematics problem for a two-link robot.

Example 9.1: Inverse Kinematics for a Two-Link Robot

Microsoft Excel™ is used to solve the inverse kinematics problem for a two-link robot. The robot geometry is shown in Figure 9.8, Figure 9.9, and Figure 9.10. The joint angles are calculated using Equation (9.5) and Equation (9.6). The spreadsheet with formulas displayed is shown in Figure 9.11.

	A	B	C	D	E	F
1	This sheet uses the ATAN2 function which works properly in all 4 quadrants					
2	LINK LENGTHS					
3	1	d1 = length of link 1		<---user input		
4	2	d2 = length of link 2		<---user input		
5	TARGET COORDINATES					
6	2	x_targ	<---user input			
7	2	y_targ	<---user input			
8						
9	CALCULATE "D"					
10	=(A6^2+A7^2-A3^2-A4^2)/(2*A3*A4)	D = cos(theta2)				
11	ELBOW DOWN RESULTS					
12	=ACOS(A10)	theta2 down (radians)				
13	=180*A12/3.14159	theta2 down (degrees)				
14	=ATAN2(A$6, A$7)-ATAN2((A$3+A$4*COS(A12)),A$4*SIN(A12))	theta1 down (radians)				
15	=180*A14/3.14159	theta1 down (degrees)				
16	ELBOW UP RESULTS					
17	=-A12	theta2 up (radians)				
18	=180*A17/3.14159	theta2 up (degrees)				
19	=ATAN2(A$6, A$7)-ATAN2((A$3+A$4*COS(A17)),A$4*SIN(A17))	theta1 up (radians)				
20	=180*A19/3.14159	theta1 up (degrees)				
21						
22	CALCULATE JOINT COORD'S FOR PLOTTING					
23	x_down	y_down	x_up	y_up	x_targ	y_targ
24	0	0	0	0	=A6	=A7
25	=$A$3*COS(A14)	=$A$3*SIN(A14)	=$A$3*COS(A19)	=$A$3*SIN(A19)		
26	=A6	=A7	=A6	=A7		

FIGURE 9.11 Spreadsheet for calculating the two-link inverse kinematics problem.

	A	B	C	D	E	F	G	H
1	This sheet uses the ATAN2 function which works properly in all 4 quadrants.							
2	LINK LENGTHS							
3	1	d1 = length of link 1		<---user input				
4	2	d2 = length of link 2		<---user input				
5	TARGET COORDINATES							
6	2	x_targ	<---user input					
7	2	y_targ	<---user input					
8								
9	CALCULATE "D"							
10	0.75	D = cos(theta2)						
11	ELBOW DOWN RESULTS							
12	0.7227342	theta2 down (radians)						
13	41.409657	theta2 down (degrees)						
14	0.2987032	theta1 down (radians)						
15	17.114448	theta1 down (degrees)						
16	ELBOW UP RESULTS							
17	-0.722734	theta2 up (radians)						
18	-41.40966	theta2 up (degrees)						
19	1.2720931	theta1 up (radians)						
20	72.885628	theta1 up (degrees)						
21								
22	CALCULATE JOINT COORD'S FOR PLOTTING							
23	x_down	y_down	x_up	y_up	x_targ	y_targ		
24	0	0	0	0	2	2		
25	0.9557189	0.29428109	0.294281	0.9557				
26	2	2	2	2				

FIGURE 9.12 Numerical results for the two-link inverse kinematics problem.

For the simulation shown in Figure 9.11 and Figure 9.12, the link lengths are 1 and 2 for link 1 and link 2, respectively. The target (desired location of the end effector) is at (2, 2).

The numerical calculations for these particular inputs are shown in Figure 9.12. We can see that θ_1 (the angle between the first link and the base) is 17.1° for the elbow down configuration, and 72.9° for the elbow up configuration. The θ_2 values (the angle between link 1 and link 2) are equal and opposite, with a magnitude of 41.4°.

The resulting plot showing the elbow up (solid lines) and elbow down (dashed lines) configurations is shown in Figure 9.13.

Figure 9.14 shows the calculations and plot if the target is changed to position (–1, –2). In this case, with the target in the third quadrant, the elbow up calculation (corresponding to a negative sign on cos[θ_2]) actually generates a physical configuration in which the elbow is down. However, it is simpler to still call this “elbow up” rather than try to modify the nomenclature depending on the target quadrant.

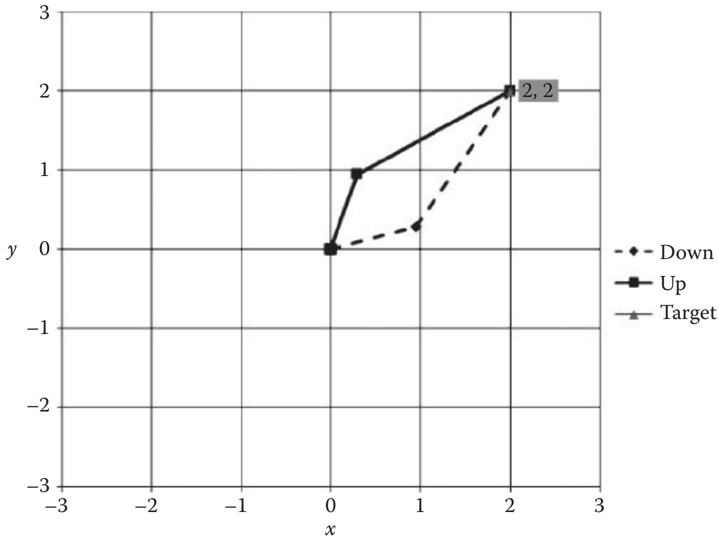
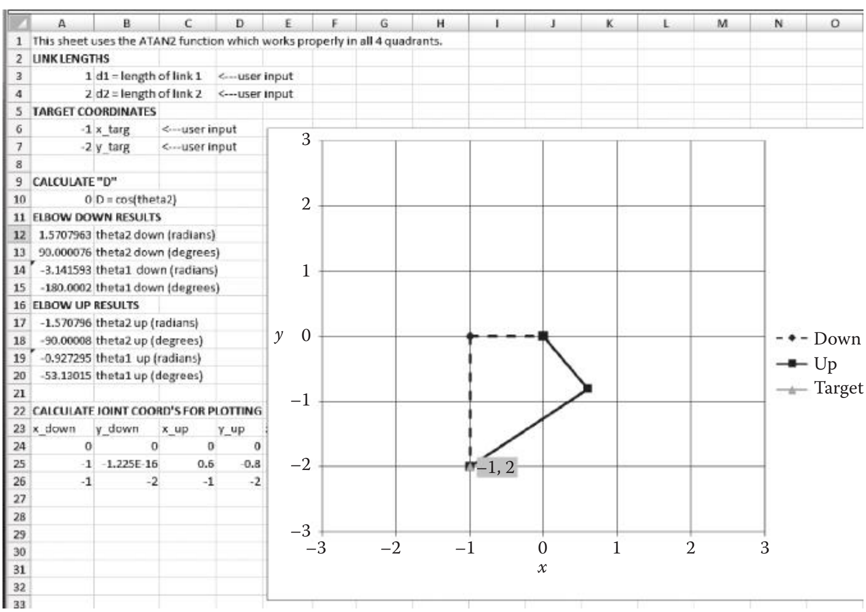


FIGURE 9.13 Plot of link positions. Solid lines are the elbow up configuration; dashed lines are the elbow down configuration. Link 1 has length 1; link 2 has length 2. The target is at (2, 2).



	A	B	C	D
1	This sheet uses the ATAN2 function which works properly in all 4 quadrants.			
2	LINK LENGTHS			
3	1	d1 = length of link 1		<---user input
4	2	d2 = length of link 2		<---user input
5	TARGET COORDINATES			
6	-1	x_targ	<---user input	
7	-2	y_targ	<---user input	
8				
9	CALCULATE "D"			
10	0	D = cos(theta2)		
11	ELBOW DOWN RESULTS			
12	1.5707963	theta2 down (radians)		
13	90.000076	theta2 down (degrees)		
14	-3.141593	theta1 down (radians)		
15	-180.0002	theta1 down (degrees)		
16	ELBOW UP RESULTS			
17	-1.570796	theta2 up (radians)		
18	-90.00008	theta2 up (degrees)		
19	-0.927295	theta1 up (radians)		
20	-53.13015	theta1 up (degrees)		
21				
22	CALCULATE JOINT COORD'S FOR PLOTTING			
23	x_down	y_down	x_up	y_up
24	0	0	0	0
25	-1	-1.225E-16	0.6	-0.8
26	-1	-2	-1	-2
27				
28				
29				
30				
31				
32				
33				

FIGURE 9.14 Configuration when the target is at (–1, –2).

9.2 Industrial Robots: Safety

A key consideration for industrial robots is safety. The American National Standards Institute (ANSI) publishes the relevant standard ANSI/RIA R15.06-1999, *American National Standard for Industrial Robots and Robot Systems: Safety Requirements*.[1] From the foreword of the *Standard*:

> The objective of this standard is to enhance the safety of personnel using industrial robot systems by establishing requirements for the manufacture (including remanufacture and overhaul), installation, safeguarding methods, maintenance and repair of manipulating industrial robots.

The standard provides these definitions:

Industrial robot: An automatically controlled, reprogrammable, multipurpose manipulator programmable in three or more axes which may be either fixed in place or mobile for use in industrial automation applications.

Industrial robot system: Equipment that includes the robot(s) (hardware and software) consisting of the manipulator power supply and control system, the end effector(s), and any other associated machinery and equipment within the safeguarded space.

A typical workcell protection environment specified by the *Standard* is shown in Figure 9.15.

Various aspects of protection include a barrier fence, emergency-stop (E-Stop) buttons, safety mats, light curtains, awareness signal(s), and other devices. The devices include a combination of active and passive, visual, and audible warnings, cutoffs, and enabling mechanisms.

Specific requirements for each are included in the standard. For example, the requirements for a barrier fence are detailed in Figure 9.16.

We see, for example, that for a barrier fence 36 inches from a hazard, the opening can be no more than 5 inches. The allowed opening size becomes smaller the closer the fence is to the hazard area.

[1] American National Standards Institute (ANSI). *American National Standard for Industrial Robots and Robot Systems: Safety Requirements* (ANSI/RIA R15.06-1999). Washington, DC: ANSI, 1999.

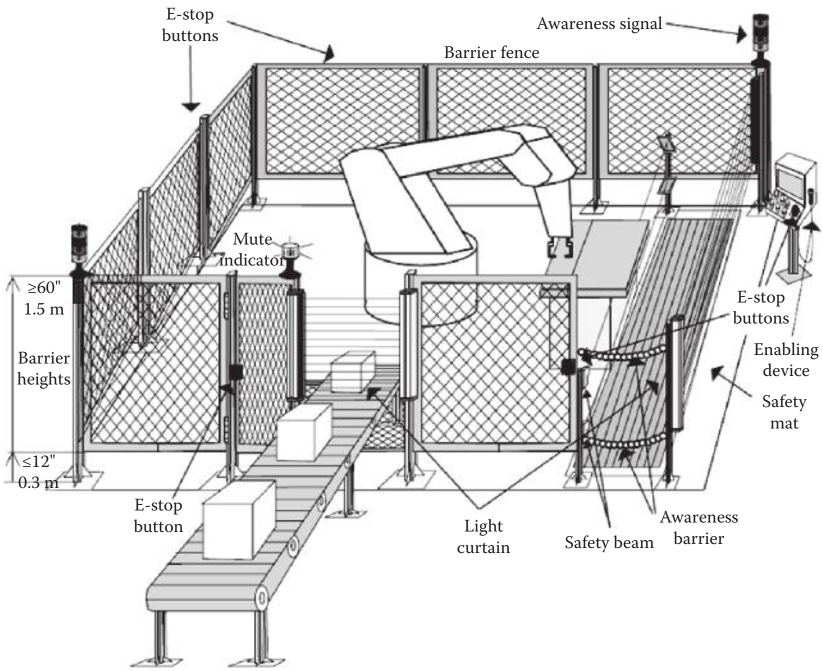


FIGURE 9.15 Typical workcell protection. Source: Adapted from American National Standards Institute (ANSI), *American National Standard for Industrial Robots and Robot Systems: Safety Requirements* (ANSI/RIA R15.06-1999) (Washington, DC: ANSI, 1999).

9.3 Industrial Robots: Programming

A typical robot program provides instructions on how each axis should move. Several means are used to program the robots:

Teach pendant: In this method, the programmer uses a remote control device, called a "teach pendant," to move the robot joints according to a desired trajectory. Once the teaching is finished, the program is stored and the robot executes the taught motions repetitively.

Manual guidance: In this method, the programmer moves the robot joints and tooling by hand. The robot control system is unpowered but able to store the motions.

Remote graphical user interface: The programmer uses a remote system to program motions and waypoints.

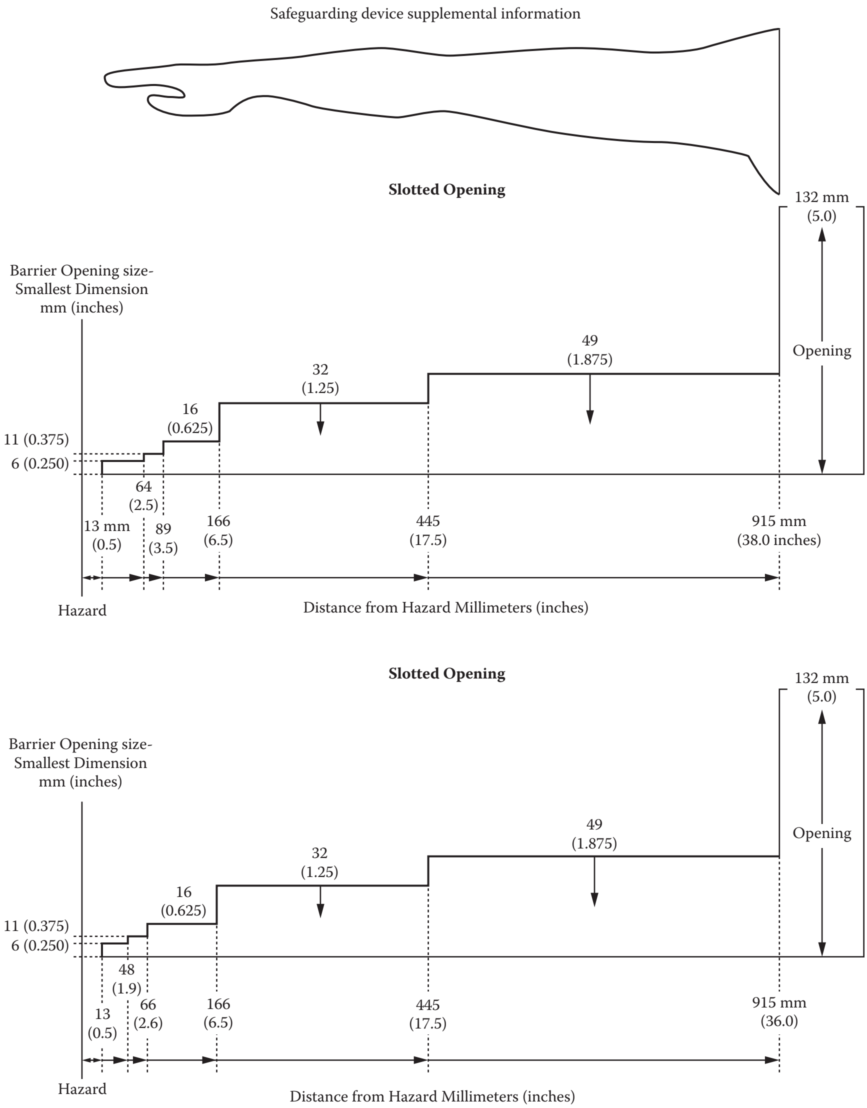


FIGURE 9.16 Barrier fence requirements. Source: Adapted from American National Standards Institute (ANSI), *American National Standard for Industrial Robots and Robot Systems: Safety Requirements* (ANSI/RIA R15.06-1999) (Washington, DC: ANSI, 1999).

9.4 Mobile Robots

Robots that can move on their own are called "mobile robots." Ground-based robots typically move using wheels or legs powered by electric motors. Underwater robots usually use electrically driven jets. Aerial robots use electric or fossil fuel drives.

Mobile robots have thus far found greatest application in hazardous applications such as military operations, search and rescue, nuclear power plant accidents, mine shaft recovery, undersea oil spills, and the like. Future applications are likely to include a much larger percentage of services, such as aid to the elderly and disabled, household chores, property maintenance, observation of children and pets, and other uses. Present-day machines are usually high-cost, special-purpose devices produced in small quantities.

The next era of mobile robotics will lean toward producing larger volumes of lower cost, versatile machines. Required capabilities will include natural language, vision, and other senses compatible with the human-centric environment. Behaviors such as climbing stairs and running will be essential.

9.4.1 Mobile Robots: Path Planning and Navigation

A basic function for mobile robots is moving from one point to another. Two aspects of this are path planning and navigation. Path planning tries to determine the needed motion control actions ahead of time (before actually moving). Navigation performs the control of energy (motors, jets, etc.) to follow the planned path, and also accommodates obstacles and opportunities that were unknown at the outset. In practice, path planning and navigation are intertwined. A suitable path should avoid obstacles that are known ahead of time, and accommodate obstacles that arise during travel.

The most basic algorithms are the BUG algorithms. These are similar to what a tourist without a map might use in an unfamiliar city. Suppose you find yourself in New York City for the first time, and you want to get to the Empire State Building. As long as you can see it, you can take local actions to get closer. You might not discover the best (i.e., shortest or quickest) path, but you can get there if a path exists.

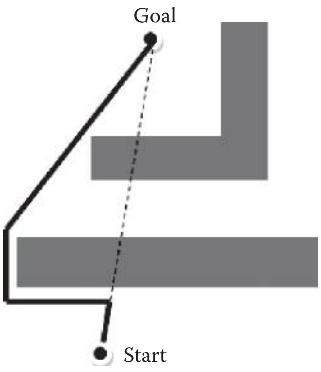


FIGURE 9.17 BUG 0 algorithm.

9.4.1.1 BUG 0 The BUG 0 algorithm assumes the robot can determine the direction to the goal, detect a wall when encountered, and turn to follow a wall as desired. It follows the simple strategy illustrated in Figure 9.17, in which the robot will

- head toward a goal,
- follow obstacles until it can head toward the goal again, and
- continue.

9.4.1.2 BUG 1 The BUG 1 algorithm assumes the robot can perform BUG 0 and also remember the closest approach to the goal while circumnavigating an obstacle. It follows the strategy illustrated in Figure 9.18, in which

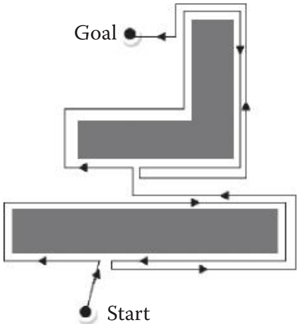


FIGURE 9.18 BUG 1 algorithm.

the robot will

- head toward a goal;
- if an obstacle is encountered, circumnavigate it and remember the closest point to the goal; and
- return to that closest point and head toward the goal.

BUG 0 is a "greedy" algorithm because it takes the first opportunity to move toward the goal without considering other possibilities. BUG 1 is an exhaustive search algorithm. It can also be shown to be complete, which means if a path exists, it will find it.

Other algorithms, including BUG 2 and Tangent BUG, use more elaborate variations of these basic ideas to enable motion in incompletely known environments.

9.4.2 Mobile Robots: Localization

For mobile robots, the question "Where am I?" is fundamentally important and often difficult to answer with precision.

For robots operating outdoors in favorable conditions, GPS can provide an approximate location. Indoor positioning systems generally similar to GPS do exist, but they require installation of equipment in the environment and are generally far more expensive than GPS. Further position refinement both indoors and outdoors can be obtained using additional sensors such as vision, sonar, infrared (IR), or tactile (touch).

One practical approach to localization follows the general scheme used by humans. If we begin at a known starting position and then use landmarks and beacons to navigate, we can build and refine our internal map while keeping track of our location within it. This process, known as "simultaneous location and mapping" (SLAM), has been formalized in the field of robotics.

Although the formal mathematics is beyond the scope of this book, an example is provided.

Example 9.2: Mobile Robot Collaboration

A team of mobile robots was built and operated in a laboratory-scale environment shown in Figure 9.19. Six robots were built in all. A series of experiments was conducted to examine the effects of coordinated activity. The robots are able to communicate with each other wirelessly. They are able to communicate their current maps and their best estimate of their

FIGURE 9.19 **Mobile robot environment.**

own position within the map. A map consists of a grid of cells. Each cell is approximately the size of a robot. Each cell is labeled with a value from 0 to 1 indicating the robot's belief about whether the cell is occupied by an obstacle (wall or other robot) or not. In an ideal world, only 0 or 1 values would be possible, indicating empty or occupied. However, due to imperfect knowledge obtained from the sensors, all values in the range of 0–1 are actually possible. For example, cells that lie outside the current range of the robot's sensors and that have not yet been sensed by any of the robots are assigned a value of 0.5, to indicate a completely unknown condition.

As the robots move, they gather sensor data and compute occupancy probabilities for each cell in their local (personal) map. The sensors used are dual-infrared (dual-IR) range sensors mounted on a turret that can swivel to cover 240°, as shown in Figure 9.20. The sensors provide an analog voltage related to the distance to the nearest reflective object. This voltage is read by an analog-to-digital converter in the robot's microcontroller, and then converted to actual distance using a programmed calibration routine. After updating their local maps, the robots broadcast their maps so that each robot has a copy of its own map as well as those of all the others.

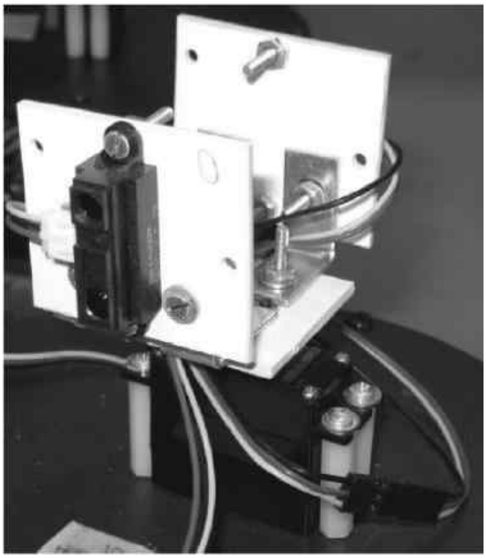


FIGURE 9.20 **Dual-infrared (dual-IR) sensor turret assembly. One IR range sensor (Sharp GP2D12) is visible in the photograph; the other is mounted opposite. The dual-IR sensor assembly is moved by the servo (underneath) under program control in order to scan the robot's surroundings. The IR sensors have a range from 10 cm to 80 cm.**

The robots then update their global map ("worldview") by combining all the local maps using standard rules of probability. Since this same procedure is performed locally by each robot, the system is highly redundant and is able to tolerate the failure of one or more robots. In order to measure motion (odometry), one needs to measure distance traveled with respect to a fixed object, such as the floor.

The robots described here use a pair of optical mice to measure motion. The mice are mounted so that they graze the floor while the robot is in motion. Optical mice have a camera, lens, and processor built in so that they report delta-x and delta-y through their serial output. This is easily read by the robot's microcontroller. Using two mice allows the robot software to resolve rotation and translation independently. (See Figure 9.21.)

Raw data from the IR sensors need to be refined in order to become useful. Several steps in the processing are shown in Figure 9.22. In Figure 9.22(a), the raw data are shown. Each point represents a reading from one of the two IR sensors. Figure 9.22(b) shows the result of an

FIGURE 9.21 **Optical mice used for odometry.**

algorithm that attempts to identify continuous line segments from the raw data. Figure 9.22(c) shows the best match of the line segment pattern in Figure 9.22(b) to the stored global map. Figure 9.22(d) shows the robot field of view (sensor scan region) superimposed on the matched wall lines.

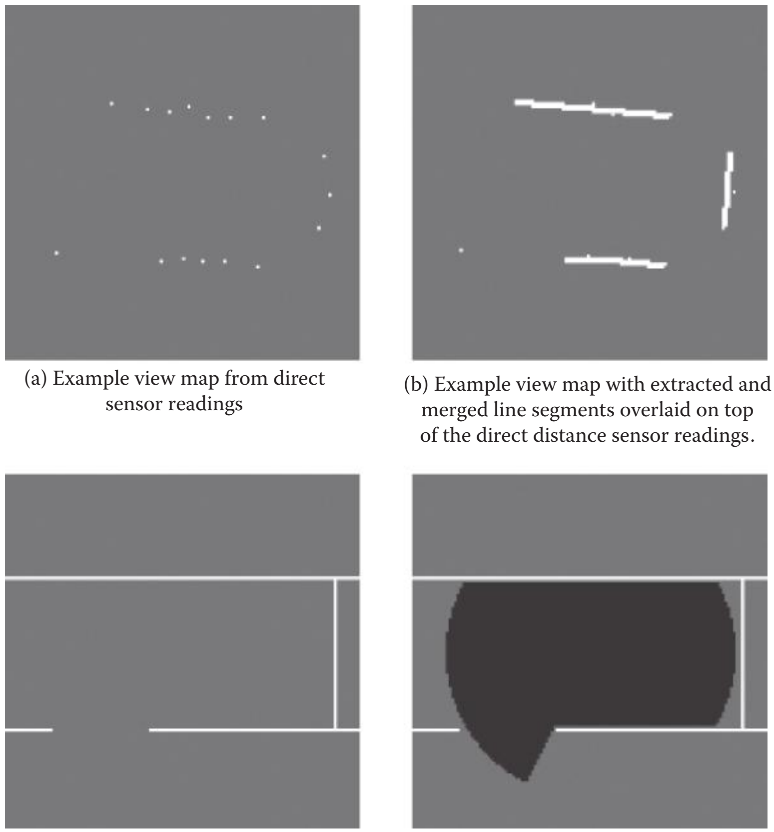
Using this procedure, it is possible to perform a probabilistic best match to any given set of sensor readings and to therefore identify the robot's location.

For search-and-rescue operations, robots are programmed to explore the environment as efficiently as possible, avoiding overlap with already explored areas. Regions to be explored are prioritized by how unknown they are and how difficult they are to reach. Cost maps are constructed, such as those shown in Figure 9.23.

Using a shared marketplace scheme, robots bid on their next destination. The winning bids are assigned to available robots, and each one proceeds to its next destination.

9.4.3 Mobile Robots: Communication

In the previous section, we said that the question "Where am I?" is fundamentally important. So too is the question "Who are you?" Increasingly, robots



(a) Example view map from direct sensor readings

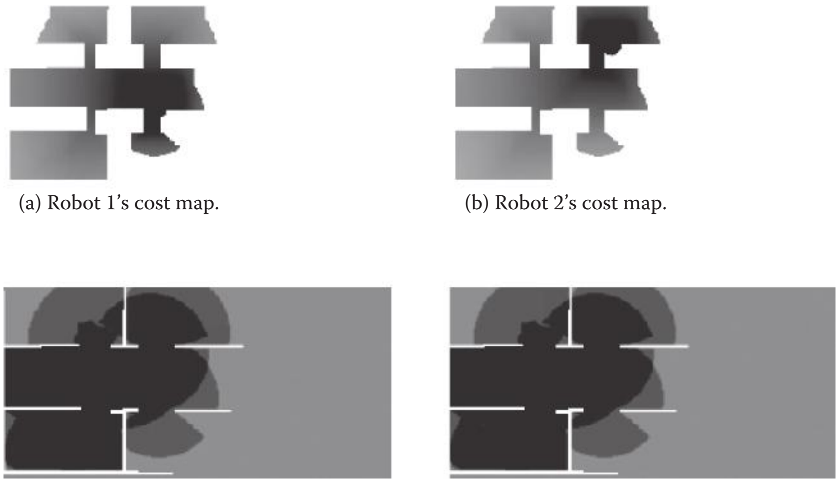
(b) Example view map with extracted and merged line segments overlaid on top of the direct distance sensor readings.

(c) Matched known wall lines

(d) Known wall lines with field for view

FIGURE 9.22 View maps at various stages of processing. The robot is located at the center of the view map.

are communicating in order to accomplish shared tasks with each other and with humans. In the summer of 2005, the Language for Intelligent Machines Workshop was held in the Mathematics Department at the U.S. Military Academy at West Point, New York. Both authors of this textbook were privileged to be in attendance. Linguists, philosophers, mathematicians, and engineers were brought together to think about how to define a language, spoken plus gestural, that could enable machines and humans to work together. A technology leader in a large automotive-manufacturing company has said that the goal of that company is to have a service robot in almost every U.S.



(a) Robot 1's cost map. (b) Robot 2's cost map.

(c) Robot 1's local map with exclusivity. (d) Robot 2's local map with exclusivity.

FIGURE 9.23 Cost maps and corresponding local maps of two robots in a three-robot test. Robot position can be found in the darkest areas of the cost map, which correspond to the least cost.

household by the year 2020. Clearly, such machines will need to communicate in natural ways with humans, and they will need to share data with each other in order to be maximally useful.

As people and robots interact, there will naturally arise emotional and philosophical issues. One only need observe how children imbue stuffed animals with personality to see that when robots become part of a household and participate in a family's everyday activities, there will be attachment issues. As robots become better at mimicking human behavior and looks, the issues will go deeper, and they certainly promise to be more interesting.

Problems

9.1. Make a simple sketch for each type of robot listed in Table 9.1, indicating the kinematic chain. Label each joint "R" for rotating or "P" for prismatic (linear).
9.2. Give one example each of physical systems that have one, two, and three degrees of freedom.

9.3. Is a jet airplane holonomic? Why or why not? Is a child floating on an inflated ring in a swimming pool holonomic? Why or why not?

9.4. Think of three things that would be hard to do if your arm had less than seven degrees of freedom.

9.5. Sketch the (two-dimensional) workspace for a planar kinematic chain having two links and two rotating joints. The end effector is at the end of the second joint. (a) Assume both links are the same length. (b) Assume the second joint is half as long as the first. (c) Assume the second link is twice as long as the first.

9.6. Calculate the angles and sketch the configuration for the robot in Example 9.1 for the following target (x, y) positions: (1, 2), (–1, 2), (–1, –2), and (1, –2).

9.7. Calculate the angles and sketch the configuration for the robot in Example 9.1 for the following target (x, y) positions: (2, 1), (–2, 1), (–2, –1), and (2, –1).

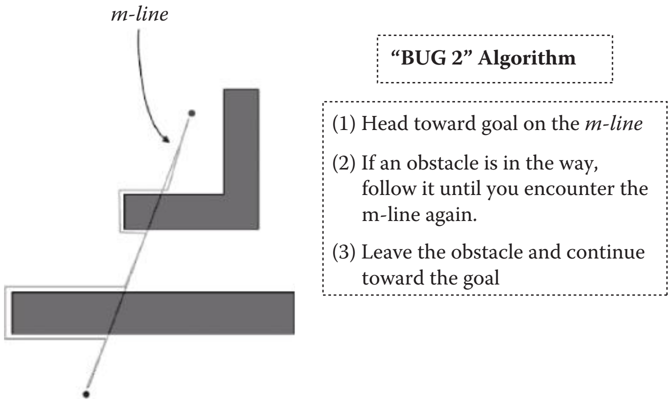
9.8. For the two-link robot discussed in Example 9.1, identify and sketch all the target positions that result in the elbow up and elbow down solutions being identical.

9.9. A square mesh barrier fence is being designed to enclose the workspace of an industrial robot. A certain portion of the fence will be 16 inches away from the hazard. What is the maximum allowed size of the mesh according to ANSI/RIA R15.06-1999?

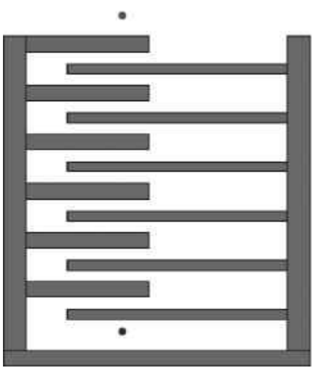
9.10. Show that the map below will foil the BUG 0 algorithm. Will BUG 1 be successful?

9.11. Consider the BUG 2 algorithm shown below. The m-line in the figure is a line connecting the start (lower dot) and goal (upper dot).

Compare it with BUG 1 for navigating the map shown in the figure. Which one performs best?



9.12. Compare BUG 1 and BUG 2 (defined in Problem 9.11) on the map shown below. Which one performs best?



Appendix A: Units and Conversion Factors

This appendix is provided as a convenient reference for units used in this book. For a comprehensive investigation of the subject, refer to the references cited at the end of this appendix. Contemporary practice encourages usage of SI units (SI is the accepted abbreviation for the International System of Units, or, more formally Le Système International d'Unités).

Quantity	Typical Symbol(s)	SI Unit (Abbreviation)
Mechanical		
Length	x,y,z,l,d	meter, m
Mass	m	kilogram, kg
Time	t	second, s
Velocity	v, u	meter/second, m/s
Force	f	newton, n
Torque	t	newton-meter, nm
Pressure	p	pascal, pa
Energy	w	joule, j
Power	p,p	watt, w
Angle	α,β,γ,θ	radian, rad
Angular velocity	ω	radian/second, rad/s
Mass polar moment of inertia	J	kg-m^2
Electrical		
Charge	q, Q	coulomb, C

(Electric) current	i,I	ampere, A
Voltage	v,V,e,E	volt, V
Frequency	f	hertz, Hz
Radian frequency	ω	radian per second, rad/s
Apparent power	S	voltampere, VA
Average power	P	watt, W
Reactive power	Q	voltampere-reactive, var
Resistance	R	ohm, Ω
Inductance	L	henry, H
Capacitance	C	farad, F
Impedance, reactance	Z , X	ohm, Ω
Conductance	G	siemens, S
Admittance, susceptance	Y, B	siemens, S
Thermal		
Temperature	θ, T	kelvin, K
Thermal power	P, Q	watt, W
Heat energy	W	joule, J
Thermal capacitance	C	joule/kelvin, J/K
Thermal conductance	G	watt/kelvin, W/K
Specific heat	c_p	joule/(kelvin-kilogram), J/(K-kg)
Magnetic		
Flux	ϕ	weber, Wb
Flux density	B	tesla, T
Magnetic field intensity	H	ampere/meter, A/m
Reluctance	$\Re$	1/henry, H^{-1}
Permeance	$\wp$	henry, H
Magnetomotive force	$\Im$	ampere-turns, A

SI Prefixes					
yotta	Y	10^{24}	yocto	z	10^{-24}
zetta	Z	10^{21}	zepto	z	10^{-21}
exa	E	10^{18}	atto	a	10^{-18}
peta	P	10^{15}	femto	f	10^{-15}
tera	T	10^{12}	pico	p	10^{-12}
giga	G	10^{9}	nano	n	10^{-9}
mega	M	10^{6}	micro	μ	10^{-6}
kilo	k	10^{3}	milli	m	10^{-3}
hecto	h	10^{2}	centi	c	10^{-2}
deka	da	10^{1}	deci	d	10^{-1}
Example usage: I = 115 A = 0.115 kA = 115000 mA.					

Selected Constants

Permeability of free space = 0.4 π μH/m
Permittivity of free space = 8.8542 pf/m
Absolute zero (temperature) = 0 K = –273.2°C = –459.7°F = 0°R
Freezing point of water = 273.2 K = 0°C = 32°F = 491.7°R
g = acceleration of gravity = 9.807 m/s^2 = 32.17 ft/s^2
G = gravitational constant = 66.72 pN (m^2) (kg $^{-2}$)
c = speed of light in vacuum = 0.2998 m/ns = 299.8 km/s
Avogadro's constant = 1.602×10^{23} = molecules per mole
e = charge on electron = 0.1602 aC
m_0 = electron rest mass = 9.11×10^{-28} gram
1 revolution = 2π radians = 360 degrees
1 standard atmosphere (atm) = 760 mm Hg = 101.3 kPa = 29.92 in Hg = 14.696 psi
Air density = 1.225 kg/m^3 @ 1 atmosphere, 15°C, 1 u = 1 amu = 1 atomic mass unit = 1.661×10^{-27} kg
1 toe (1 tonne oil equivalent) = 41.868 GJ = 1270 m^3 natural gas (STP) = 2.3 tonnes coal
1 ton TNT = 4.184 GJ

Selected Conversion Factors

1 inch = 2.54 cm; 1 foot = 12 inches; 1 meter = 3.28084 feet
1 pound force (lbf) = 4.4482 newtons; 1 kilogram = 2.188 pound mass (lbm)
J in kg-m^2 = 0.0.04246 × J in lbm-ft^2
1 atomic mass unit = u = 1.661 × 10^{-27} kg
Electron/positron mass = 0.000549 u
Proton mass = 1.007276 u
Neutron mass = 1.008665 u
1 slug = 14.59 kilogram
1 foot-pound = 1.356 newton-meter
1 horsepower = 746 watts
1 British Thermal Unit (BTU) = 1055 joule
1 calorie = 4.19 joule
Temperature in Kelvin (K) = T_K
Temperature in degrees Celsius (°C) = $T_C = T_K - 273.2$
Temperature in degrees Fahrenheit (°F) = $T_F = 1.8T_C + 32$
Temperature in degrees Rankine (°R) = $T_R = T_F + 459.7$
1 milligauss = 1 mG = 0.1 microtesla = 0.1 μT
1 ton (refrigeration) = 3517 watts = 12 kBTU/hr
1 tonne = 1 metric ton = 1000 kg
Speed in rad/s = (2π/60) × speed in rpm (rpm = revolutions per minute)
1 day = 24 hours; 1 hour = 60 minutes; 1 minute = 60 seconds
1 nautical mile = 6000 ft = 1.1364 mile
1 knot = 1 nautical mile per hour = 1.1364 mph = 0.5080 m/s

Temperature Scales

Fahrenheit	Rankine	Centigrade	Kelvin
212°F	672°R	100°C	373 K
100°F	560°R	38°C	311 K
32°F	492°R	0°C	273 K
0°F	460°R	−18°C	255 K
−460°F	0°R	−273°C	0 K

Notes

One unit C = one unit K; one unit F = one unit R; one unit K = I.8 F units.
Absolute zero (0 K) = −459.67°F = = −273.15°C.
Normal human body temperature is 98.2°F (36.8°C).

References

American National Standards Institute (ANSI). *Metric Practice* (ANSI Z210.1-1976). Washington, DC: ANSI, 1976.
American Society for Testing and Materials (ASTM). *Metric Practice* (ASTM E 380-76). West Conshohocken, PA: ASTM, 1980.
Dorf, Richard C. *Handbook of Engineering Tables*. Boca Raton, FL: CRC Press, 2004.
IEEE. *Standard 268-76*. Piscataway, NJ: IEEE, 2004.
Inland Steel Company. *Reference Manual for SI Units*. Chicago: Inland Steel Company, 1976.
Nelson, Robert A. *Guide for Metric Practice*. *Physics Today*, 1996. http://www.physicstoday.org/guide/metric.pdf

Unit Conversion Websites

http://www.unit-conversion.info/energy.html
http://www.unc.edu/~rowlett/units/

This page intentionally left blank

Appendix B: Data Sheet for SN7400 Series TTL NAND Gate

Nearly all manufacturers supply a data sheet for each part or family of parts they sell. Data sheets can range from a few pages for simple digital logic circuits to six hundred pages or more for a complex microcontroller or digital signal processor.

The first several pages of the data sheet for a TTL NAND gate are shown in Figures B-1 through B-7. This particular part is the SN5400 / 7400 series TTL NAND gate supplied by Texas Instruments, Inc. The complete data sheet is twenty-three pages in length. Only the pages related to the electrical performance are included here; the remaining pages contain mechanical drawings and related data needed for preparing board layouts and setting up automated assembly equipment.

Reference

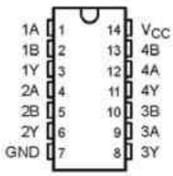
Texas Instruments. *SN7400 Series Data Sheet*. Rev. October 2003. http://focus.ti.com/lit/ds/symlink/sn74ls00.pdf.

SN5400, SN54LS00, SN54S00
SN7400, SN74LS00, SN74S00
QUADRUPLE 2-INPUT POSITIVE-NAND GATES

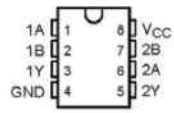
SDLS025B – DECEMBER 1983 – REVISED OCTOBER 2003

- **Package Options Include Plastic Small-Outline (D, NS, PS), Shrink Small-Outline (DB), and Ceramic Flat (W) Packages, Ceramic Chip Carriers (FK), and Standard Plastic (N) and Ceramic (J) DIPs**
- **Also Available as Dual 2-Input Positive-NAND Gate in Small-Outline (PS) Package**

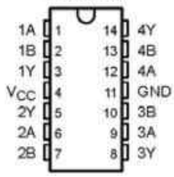
SN5400 . . . J PACKAGE
SN54LS00, SN54S00 . . . J OR W PACKAGE
SN7400, SN74S00 . . . D, N, OR NS PACKAGE
SN74LS00 . . . D, DB, N, OR NS PACKAGE
(TOP VIEW)



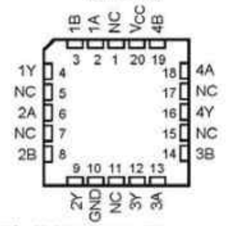
SN74LS00, SN74S00 . . . PS PACKAGE
(TOP VIEW)



SN5400 . . . W PACKAGE
(TOP VIEW)



SN54LS00, SN54S00 . . . FK PACKAGE
(TOP VIEW)



NC – No internal connection

description/ordering information

These devices contain four independent 2-input NAND gates. The devices perform the Boolean function $Y = \overline{A \bullet B}$ or $Y = \overline{A} + \overline{B}$ in positive logic.

Please be aware that an important notice concerning availability, standard warranty, and use in critical applications of Texas Instruments semiconductor products and disclaimers thereto appears at the end of this data sheet.

PRODUCTION DATA information is current as of publication date. Products conform to specifications per the terms of Texas Instruments standard warranty. Production processing does not necessarily include testing of all parameters.

TEXAS INSTRUMENTS
POST OFFICE BOX 655303 • DALLAS, TEXAS 75265

On products compliant to MIL-PRF-38535, all parameters are tested unless otherwise noted. On all other products, production processing does not necessarily include testing of all parameters.

FIGURE B.1 SN7400 series data sheet – Page 1.

SN5400, SN54LS00, SN54S00
SN7400, SN74LS00, SN74S00
QUADRUPLE 2-INPUT POSITIVE-NAND GATES

SDLS025B – DECEMBER 1983 – REVISED OCTOBER 2003

description/ordering information (continued)

ORDERING INFORMATION

T_A	PACKAGE†		ORDERABLE PART NUMBER	TOP-SIDE MARKING
0°C to 70°C	PDIP – N	Tube	SN7400N	SN7400N
			SN74LS00N	SN74LS00N
			SN74S00N	SN74S00N
	SOIC – D	Tube	SN7400D	7400
		Tape and reel	SN7400DR	
		Tube	SN74LS00D	LS00
		Tape and reel	SN74LS00DR	
		Tube	SN74S00D	S00
		Tape and reel	SN74S00DR	
	SOP – NS	Tape and reel	SN7400NSR	SN7400
			SN74LS00NSR	74LS00
			SN74S00NSR	74S00
	SOP – PS	Tape and reel	SN74LS00PSR	LS00
			SN74S00PSR	S00
	SSOP – DB	Tape and reel	SN74LS00DBR	LS00
−55°C to 125°C	CDIP – J	Tube	SNJ5400J	SNJ5400J
			SNJ54LS00J	SNJ54LS00J
			SNJ54S00J	SNJ54S00J
	CFP – W	Tube	SNJ5400W	SNJ5400W
			SNJ54LS00W	SNJ54LS00W
			SNJ54S00W	SNJ54S00W
	LCCC – FK	Tube	SNJ54LS00FK	SNJ54LS00FK
			SNJ54S00FK	SNJ54S00FK

† Package drawings, standard packing quantities, thermal data, symbolization, and PCB design guidelines are available at www.ti.com/sc/package.

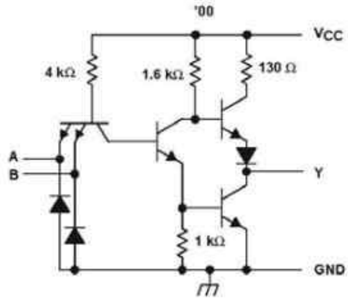
FUNCTION TABLE
(each gate)

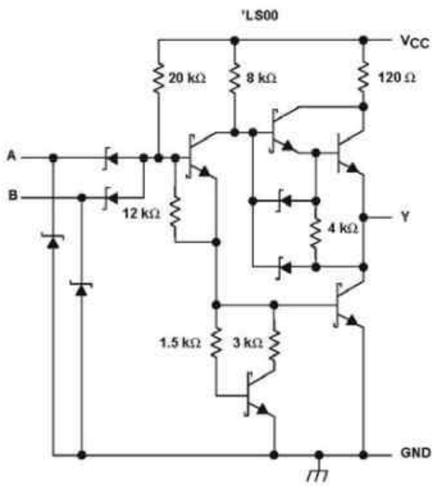
INPUTS		OUTPUT
A	B	Y
H	H	L
L	X	H
X	L	H

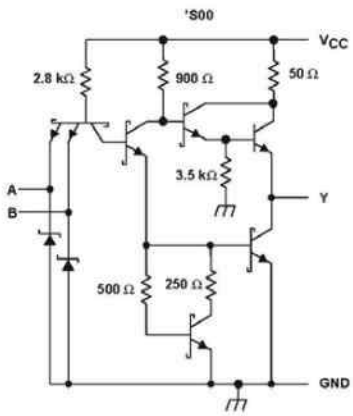
logic diagram, each gate (positive logic)

FIGURE B.2 SN7400 series data sheet – Page 2.

schematic







Resistor values shown are nominal.

FIGURE B.3 SN7400 series data sheet – Page 3

SN5400, SN54LS00, SN54S00
SN7400, SN74LS00, SN74S00
QUADRUPLE 2-INPUT POSITIVE-NAND GATES
SDLS025B – DECEMBER 1983 – REVISED OCTOBER 2003

absolute maximum ratings over operating free-air temperature (unless otherwise noted)†

Supply voltage, V_{CC} (see Note 1) .... 7 V
Input voltage: '00, 'S00 .... 5.5 V
'LS00 .... 7 V
Package thermal impedance, θ_{JA} (see Note 2): D package .... 86°C/W
DB package .... 96°C/W
N package .... 80°C/W
NS package .... 76°C/W
PS package .... 95°C/W
Storage temperature range, T_{stg} .... −65°C to 150°C

† Stresses beyond those listed under "absolute maximum ratings" may cause permanent damage to the device. These are stress ratings only, and functional operation of the device at these or any other conditions beyond those indicated under "recommended operating conditions" is not implied. Exposure to absolute-maximum-rated conditions for extended periods may affect device reliability.

NOTES: 1. Voltage values are with respect to network ground terminal.
2. The package termal impedance is calculated in accordance with JESD 51-7.

recommended operating conditions (see Note 3)

		SN5400			SN7400			UNIT
		MIN	NOM	MAX	MIN	NOM	MAX	
V_{CC}	Supply voltage	4.5	5	5.5	4.75	5	5.25	V
V_{IH}	High-level input voltage	2			2			V
V_{IL}	Low-level input voltage			0.8			0.8	V
I_{OH}	High-level output current			−0.4			−0.4	mA
I_{OL}	Low-level output current			16			16	mA
T_A	Operating free-air temperature	−55		125	0		70	°C

NOTE 3: All unused inputs of the device must be held at V_{CC} or GND to ensure proper device operation. Refer to the TI application report, *Implications of Slow or Floating CMOS Inputs*, literature number SCBA004.

electrical characteristics over recommended operating free-air temperature range (unless otherwise noted)

PARAMETER	TEST CONDITIONS‡			SN5400			SN7400			UNIT
				MIN	TYP§	MAX	MIN	TYP§	MAX	
V_{IK}	V_{CC} = MIN,	I_I = −12 mA				−1.5			−1.5	V
V_{OH}	V_{CC} = MIN,	V_{IL} = 0.8 V,	I_{OH} = −0.4 mA	2.4	3.4		2.4	3.4		V
V_{OL}	V_{CC} = MIN,	V_{IH} = 2 V,	I_{OL} = 16 mA		0.2	0.4		0.2	0.4	V
I_I	V_{CC} = MAX,	V_I = 5.5 V				1			1	mA
I_{IH}	V_{CC} = MAX,	V_I = 2.4 V				40			40	µA
I_{IL}	V_{CC} = MAX,	V_I = 0.4 V				−1.6			−1.6	mA
I_{OS}¶	V_{CC} = MAX			−20		−55	−18		−55	mA
I_{CCH}	V_{CC} = MAX,	V_I = 0 V			4	8		4	8	mA
I_{CCL}	V_{CC} = MAX,	V_I = 4.5 V			12	22		12	22	mA

‡ For conditions shown as MIN or MAX, use the appropriate value specified under recommended operating conditions.
§ All typical values are at V_{CC} = 5 V, T_A = 25°C.
¶ Not more than one output should be shorted at a time.

FIGURE B.4 SN7400 series data sheet – Page 4.

SN5400, SN54LS00, SN54S00
SN7400, SN74LS00, SN74S00
QUADRUPLE 2-INPUT POSITIVE-NAND GATES
SDLS025B – DECEMBER 1983 – REVISED OCTOBER 2003

switching characteristics, V_{CC} = 5 V, T_A = 25°C (see Figure 1)

PARAMETER	FROM (INPUT)	TO (OUTPUT)	TEST CONDITIONS		SN5400 SN7400		UNIT
				MIN	TYP	MAX	
t_{PLH}	A or B	Y	R_L = 400 Ω, C_L = 15 pF		11	22	ns
t_{PHL}					7	15	

recommended operating conditions (see Note 4)

		SN54LS00			SN74LS00			UNIT
		MIN	NOM	MAX	MIN	NOM	MAX	
V_{CC}	Supply voltage	4.5	5	5.5	4.75	5	5.25	V
V_{IH}	High-level input voltage	2			2			V
V_{IL}	Low-level input voltage			0.7			0.8	V
I_{OH}	High-level output current			−0.4			−0.4	mA
I_{OL}	Low-level output current			4			8	mA
T_A	Operating free-air temperature	−55		125	0		70	°C

NOTE 4: All unused inputs of the device must be held at V_{CC} or GND to ensure proper device operation. Refer to the TI application report, *Implications of Slow or Floating CMOS Inputs*, literature number SCBA004.

electrical characteristics over recommended operating free-air temperature range (unless otherwise noted)

PARAMETER	TEST CONDITIONS†			SN54LS00			SN74LS00			UNIT
				MIN	TYP‡	MAX	MIN	TYP‡	MAX	
V_{IK}	V_{CC} = MIN,	I_I = −18 mA				−1.5			−1.5	V
V_{OH}	V_{CC} = MIN,	V_{IL} = MAX,	I_{OH} = −0.4 mA	2.5	3.4		2.7	3.4		V
V_{OL}	V_{CC} = MIN,	V_{IH} = 2 V	I_{OL} = 4 mA		0.25	0.4		0.25	0.4	V
			I_{OL} = 8mA					0.35	0.5	
I_I	V_{CC} = MAX,	V_I = 7 V				0.1			0.1	mA
I_{IH}	V_{CC} = MAX,	V_I = 2.7V				20			20	µA
I_{IL}	V_{CC} = MAX,	V_I = 0.4 V				−0.4			−0.4	mA
I_{OS}§	V_{CC} = MAX			−20		−100	−20		−100	mA
I_{CCH}	V_{CC} = MAX,	V_I = 0 V			0.8	1.6		0.8	1.6	mA
I_{CCL}	V_{CC} = MAX,	V_I = 4.5 V			2.4	4.4		2.4	4.4	mA

† For conditions shown as MIN or MAX, use the appropriate value specified under recommended operating conditions.
‡ All typical values are at V_{CC} = 5 V, T_A = 25°C.
§ Not more than one output should be shorted at a time.

switching characteristics, V_{CC} = 5 V, T_A = 25°C (see Figure 1)

PARAMETER	FROM (INPUT)	TO (OUTPUT)	TEST CONDITIONS		SN54LS00 SN74LS00		UNIT
				MIN	TYP	MAX	
t_{PLH}	A or B	Y	R_L = 2 kΩ, C_L = 15 pF		9	15	ns
t_{PHL}					10	15	

FIGURE B.5 SN7400 series data sheet – Page 5.

SN5400, SN54LS00, SN54S00
SN7400, SN74LS00, SN74S00
QUADRUPLE 2-INPUT POSITIVE-NAND GATES
SDLS025B – DECEMBER 1983 – REVISED OCTOBER 2003

recommended operating conditions (see Note 5)

		SN54S00			SN74S00			UNIT
		MIN	NOM	MAX	MIN	NOM	MAX	
V_{CC}	Supply voltage	4.5	5	5.5	4.75	5	5.25	V
V_{IH}	High-level input voltage	2			2			V
V_{IL}	Low-level input voltage			0.8			0.8	V
I_{OH}	High-level output current			−1			−1	mA
I_{OL}	Low-level output current			20			20	mA
T_A	Operating free-air temperature	−55		125	0		70	°C

NOTE 5: All unused inputs of the device must be held at V_{CC} or GND to ensure proper device operation. Refer to the TI application report, *Implications of Slow or Floating CMOS Inputs*, literature number SCBA004.

electrical characteristics over recommended operating free-air temperature range (unless otherwise noted)

PARAMETER	TEST CONDITIONS†			SN54S00			SN74S00			UNIT
				MIN	TYP‡	MAX	MIN	TYP‡	MAX	
V_{IK}	V_{CC} = MIN,	$I_I = -18$ mA				−1.2			−1.2	V
V_{OH}	V_{CC} = MIN,	V_{IL} = 0.8 V,	I_{OH} = −1 mA	2.5	3.4		2.7	3.4		V
V_{OL}	V_{CC} = MIN,	V_{IH} = 2 V,	I_{OL} = 20 mA			0.5			0.5	V
I_I	V_{CC} = MAX,	V_I = 5.5 V				1			1	mA
I_{IH}	V_{CC} = MAX,	V_I = 2.7 V				50			50	µA
I_{IL}	V_{CC} = MAX,	V_I = 0.5V				−2			−2	mA
I_{OS}§	V_{CC} = MAX			−40		−100	−40		−100	mA
I_{CCH}	V_{CC} = MAX,	V_I = 0 V			10	16		10	16	mA
I_{CCL}	V_{CC} = MAX,	V_I = 4.5 V			20	36		20	36	mA

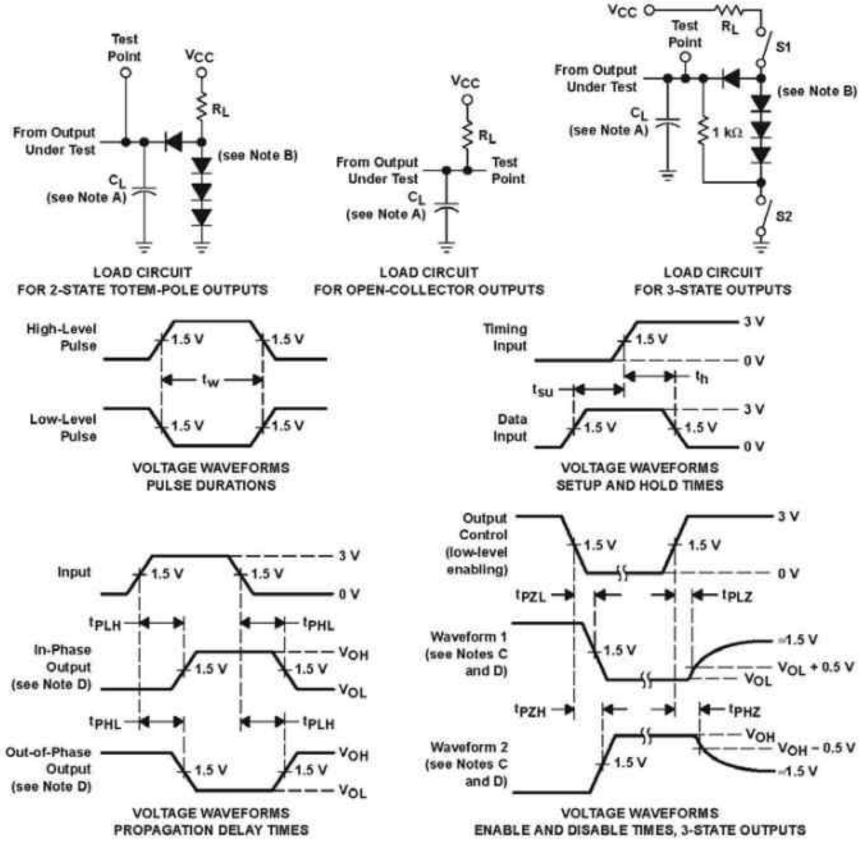
† For conditions shown as MIN or MAX, use the appropriate value specified under recommended operating conditions.
‡ All typical values are at V_{CC} = 5 V, T_A = 25°C.
§ Not more than one output should be shorted at a time.

switching characteristics, V_{CC} = 5 V, T_A = 25°C (see Figure 1)

PARAMETER	FROM (INPUT)	TO (OUTPUT)	TEST CONDITIONS		SN54S00 SN74S00			UNIT
					MIN	TYP	MAX	
t_{PLH}	A or B	Y	R_L = 280 Ω	C_L = 15 pF		3	4.5	ns
t_{PHL}						3	5	
t_{PLH}	A or B	Y	R_L = 280 Ω	C_L = 50 pF		4.5		ns
t_{PHL}						5		

FIGURE B.6 SN7400 series data sheet – Page 6.

PARAMETER MEASUREMENT INFORMATION
SERIES 54/74 DEVICES



NOTES: A. C_L includes probe and jig capacitance.
B. All diodes are 1N3064 or equivalent.
C. Waveform 1 is for an output with internal conditions such that the output is low except when disabled by the output control. Waveform 2 is for an output with internal conditions such that the output is high except when disabled by the output control.
D. S1 and S2 are closed for t_{PLH}, t_{PHL}, t_{PHZ}, and t_{PLZ}; S1 is open and S2 is closed for t_{PZH}; S1 is closed and S2 is open for t_{PZL}.
E. All input pulses are supplied by generators having the following characteristics: PRR ≤ 1 MHz, Z_O = 50 Ω; t_r and t_f ≤ 7 ns for Series 54/74 devices and t_r and t_f ≤ 2.5 ns for Series 54S/74S devices.
F. The outputs are measured one at a time with one input transition per measurement.

Figure 1. Load Circuits and Voltage Waveforms

FIGURE B.7 SN7400 series data sheet – Page 7.

This page intentionally left blank

Electrical Engineering

FUNDAMENTALS OF

Electrical Engineering

Real-world engineering problems are rarely, if ever, neatly divided into mechanical, electrical, chemical, civil, and other categories. Engineers from all disciplines will eventually encounter computer and electronic controls and instrumentation, which is going to require at least a basic knowledge of electrical and other engineering specialties, as well as associated economics, and environmental, political, and social issues.

Co-authored by Charles Gross—one of the most well-known and respected professors in the field of electric machines and power engineering—and his world-renowned colleague Thad Roppel, **Fundamentals of Electrical Engineering** provides an overview of the profession for engineering professionals and students whose specialization lies in areas other than electrical. For instance, civil engineers must contend with commercial electrical service and lighting design issues. Mechanical engineers have to deal with motors in HVAC applications, and chemical engineers are forced to handle problems involving process control.

Simple and easy to use, yet more than sufficient in rigor and coverage of fundamental concepts, this resource teaches EE fundamentals but omits the typical analytical methods that hold little relevance for the audience. The authors provide many examples to illustrate concepts, as well as homework problems to help readers understand and apply presented material.

In many cases, courses for non-electrical engineers, or non-EEs, have presented watered-down classical EE material, resulting in unpopular courses that students hate and senior faculty members understandingly avoid teaching. To remedy this situation—and create more well-rounded practitioners—the authors focus on the true EE needs of non-EEs, as determined through their own teaching experience, as well as significant input from non-EE faculty. The book provides several important contemporary interdisciplinary examples to support this approach. The result is a full-color modern narrative that bridges the various EE and non-EE curricula and serves as a truly relevant course that students and faculty can both enjoy.

CRC Press
Taylor & Francis Group
an informa business

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
711 Third Avenue
New York, NY 10017
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK

K11681
ISBN: 978-1-4398-3714-6
90000
9 781439 837146
www.crcpress.com