Exploring Engineering

SECOND EDITION

An Introduction to Engineering and Design

Philip Kosky • Robert Balmer • William Keat • George Wise

Exploring Engineering

Exploring Engineering
An Introduction to Engineering and Design

Philip Kosky

George Wise

Robert Balmer

William Keat

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
525 B Street, Suite 1900, San Diego, California 92101-4495, USA
84 Theobald's Road, London WC1X 8RR, UK

Library of Congress Cataloging-in-Publication Data
Exploring engineering : an introduction to engineering and design / Philip Kosky ... [et al.].
p. cm.
Includes bibliographical references and index.
ISBN 978-0-12-374723-5 (hardcover)
1. Engineering--Textbooks. 2. Engineering design--Textbooks. I. Kosky, P. G. (Philip G.)
TA147.E97 2010
620--dc22
2009013583

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

For information on all Academic Press publications
visit our Web site at www.elsevierdirect.com

Printed in Canada
09 10 11 12 6 5 4 3 2 1

"... it is engineering that changes the world."

Isaac Asimov, *Isaac Asimov's Book of Science and Nature Quotations*, Simon and Schuster, 1970

"... (engineering) ... is the art of doing that well with one dollar that any bungler can do with two ..."

Arthur Wellington, *Economic Theory of the Location of Railways*, 2nd ed., Wiley, NY, 1887

"... the explosion of knowledge, the global economy, and the way engineers will work will reflect an ongoing evolution that began to gain momentum a decade ago."

Educating the Engineer of 2020, National Academy of Engineering, October, 2005

Table of Contents

Foreword

Engineers have made remarkable innovations during the twentieth century. The National Academy of Engineering (NAE) recently identified the top 20 engineering achievements of the twentieth century that "shaped a century and changed the world."

> NATIONAL ACADEMY OF ENGINEERING
>
> TOP 20 ENGINEERING ACHIEVEMENTS OF THE 20TH CENTURY
>
> 1. Electrification – to supply our homes and businesses with electricity
> 2. Automobile – for leisure and commercial transportation
> 3. Airplane – for rapidly moving people and goods around the world
> 4. Water Supply and Distribution – to supply clean, germ-free water to every home
> 5. Electronics – to provide electronic control of machines and consumer products
> 6. Radio and Television – for entertainment and commercial uses
> 7. Agricultural Mechanization – to increase the efficiency of food production
> 8. Computers – a revolution in the way people work and communicate
> 9. Telephone – for rapid personal and commercial communication
> 10. Air Conditioning and Refrigeration – to increase the quality of life
> 11. Highways – to speed transportation of people and goods across the land
> 12. Spacecraft – to begin our exploration of limitless space
> 13. Internet – a cultural evolution of the way people interact
> 14. Imaging – to improve healthcare
> 15. Household Appliances – to allow women to enter the workplace
> 16. Health Technologies – to improve the quality of life
> 17. Petroleum and Petrochemical Technologies – to power transportation systems
> 18. Laser and Fiber Optics – to improve measurement and communication systems
> 19. Nuclear Technologies – to tap a new natural energy source
> 20. High-performance Materials – to create safer, lighter, better products

However, engineering freshmen are less interested in what was or what is than they are in what will be. Young men and women exploring engineering as a career are excited about the future—their future—and about the engineering challenges 10 to 20 years from now when they are in the spring and summer of their careers. In the words of the four-time Stanley Cup winner and Hockey Hall of Fame member Wayne Gretzky,

I skate to where the puck is going to be, not where it's been.

The National Academy of Engineering also has proposed the following 14 Grand Challenges for Engineering in the 21st Century. In our second edition of this text, we have chosen to highlight material that engages these topics because they represent the future of engineering creativity.

NATIONAL ACADEMY OF ENGINEERING

ENGINEERING CHALLENGES FOR THE 21ST CENTURY

1. Make solar energy economical
2. Provide energy from fusion
3. Develop carbon sequestration methods
4. Manage the nitrogen cycle
5. Provide access to clean water
6. Restore and improve urban infrastructure
7. Advance health informatics
8. Engineer better medicines
9. Reverse-engineer the brain
10. Prevent nuclear terror
11. Secure cyberspace
12. Enhance virtual reality
13. Advance personalized learning
14. Engineer the tools of scientific discovery

The twenty-first century will be filled with many exciting challenges for engineers, architects, physicians, sociologists, and politicians. Figure 1 illustrates an enhanced set of future challenges as envisioned by Joseph Bordogna, Deputy Director and Chief Operating Officer of the National Science Foundation.[1]

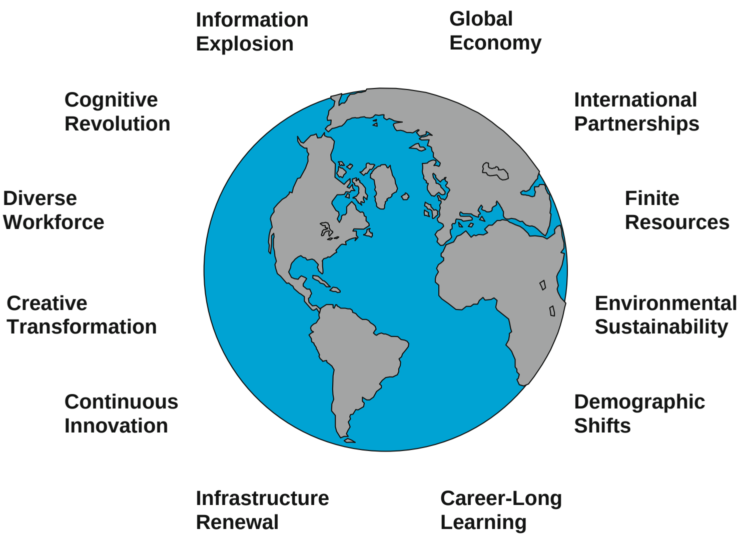


FIGURE 1 Future Trajectories in Science, Engineering, and Technology

[1] http://www.nsf.gov/news/speeches/bordogna/jb98_nrl/sld001.htm

THE STRUCTURE OF THIS TEXT

In this text we have tried to provide an exciting introduction to the engineering profession. Between its covers you will find material on classical engineering fields as well as introductory material leading to emerging twenty-first century engineering fields such as bioengineering, nanotechnology, and mechatronics.

This text is divided into two parts: **Part 1: Minds-on** and **Part 2: Hands-on**. Most chapters in Part 1 are organized around just one or two principles and have several worked examples and include exercises with an increasing level of complexity at the end of the chapter. Answers are given to selected exercises to encourage students to work toward self-proficiency.

Part 1 covers *introductory* material explicitly from the following engineering subdisciplines: bioengineering, chemical engineering, civil engineering, computer and electronic engineering, control systems engineering, electrical engineering, electrochemical engineering, materials engineering, manufacturing engineering and mechanical engineering and an introduction to engineering economics. The second edition of this text is organized around the theme of 21^{st} century engineering and provides a forward-looking entry into each of the engineering subdisciplines listed.

The topics covered are kept to a level compatible with the background of first year students. Some topics obviously are closer to the core material in one subdiscipline of engineering than to another, and some are generic to all. In order to cover such broad, and sometimes relatively advanced, subject matter we have taken some liberties in simplifying those topics. Instructors may expect to find shortcuts that will pain the purists; we have tried, nevertheless, to be accurate as to basic principles.

Part 2 provides the content for a Design Studio, and is associated with the design of engineering *systems*. This "Hands-on" section is just as essential and challenging as the minds-on aspects covered in Part 1. Also, for most students, *it is a lot more fun*. Few things are more satisfying than seeing a machine, an electronic device, or a computer program you have designed and built doing exactly what you intended it to do. Such initial successes may sound simple, but they provide the basis of a rigorous system that will enable an engineering graduate, as part of a team of engineers, to achieve the even greater satisfaction in designing a system that can provide new means of transportation, information access, medical care, energy supply, and such, and can change for the better the lives of people around the world.

We physically separated the two parts of this text to emphasize the different character of their content. Each chapter of the minds-on section has about the equivalent amount of new ideas and principles; our experience is that any chapter can be sufficiently covered in about two hours of lecture class time, and that the students can complete the rest of the chapter unaided. On the other hand, the Design Studio needs up to three contiguous laboratory hours per week to do it justice. It culminates in a team-orientated competition. Typically, student teams build a small model "device" that has wheels, or walks, or floats, that may be wireless or autonomous, and so forth. Students then compete head-to-head against other teams from the course with the same design goals plus an offensive and defensive strategy to overcome all the other teams in the competition. Our experience is that this is highly motivating for the students.

There is too much material, as well as too broad coverage, in this text provided for just one introductory course. Given the necessary breaks for testing and for a final examination, typically a class will cover several chapters of Part 1. Exactly which chapters will depend on the engineering disciplines offered at your institution. We certainly think the more fundamental chapters need to be included. Suggested Part I coverage should include the basics in Chapters 2, 3 and 4 plus several other chapters that can be selected for suitability for particular students' subdisciplines. Part 2 of this text can be thought of as independent of Part 1, but should be taught as an integral part of a first-year engineering course.

The approach taken in this first year text is unique, in part because of the atypical character of authorship. Two of the authors have *industrial* backgrounds, mostly at the GE Research Center in Niskayuna, NY, with one in engineering research and applied science and the other in industrial communications and bring a working knowledge of what is core to a practicing engineer. The other two authors have followed more traditional academic career paths and have the appropriate academic experience and credentials upon which to draw. We believe the synergy of the combined authorship provides a fresh perspective for first-year engineering education. Specifically, though elementary in coverage, this textbook parallels the combined authors' wide experience that engineering is not a "spectator sport" We therefore do not duck the introduction of relatively advanced topics in this otherwise elementary text. Here are some of the nonstandard approaches to familiar engineering topics.

1. We introduce spreadsheets early in the text, and almost every chapter of Part 1 has one or more spreadsheet exercises.
2. We try to rigorously enforce the use of appropriate significant figures throughout the text. For example, we always try to differentiate between 60. and 60 (notice the decimal point or its absence). We obviously recognize often it appears to be clumsy to write numbers such as 6.00×10^1 but we do so to discourage bad habits such as electronic calculator answers to undeserved significant figures.
3. We develop all[2] our exercise solutions in a rigorous format using a simple mnemonic Need–Know–How–Solve to discourage the student who thinks he or she knows the answer and writes the wrong one down (or even the correct one!). This too can appear to be clumsy in usage, but it is invaluable in training a young engineer to leave an audit trail of his or her methods, a good basic work habit of practicing engineers.
4. We recognize that the Engineering English unit system of lbf, lbm, and g_c will be used throughout the careers of many, if not most, of today's young engineers. A clear exposition is used to develop it and to use it so we can avoid the terrible results of a factor of 32.2 that should or shouldn't be there!
5. Conservation principles, particularly energy and mass, are introduced early in the text as well as emphasis on the use of control boundaries that focus on the essential problem at hand.
6. The use of tables is a powerful tool, both in the hands of students and of qualified engineers. We have developed a number of tabular methods for stoichiometric and for thermodynamic problems that should eliminate the problem of the wrong stoichiometric coefficients and of sign errors, respectively. Methods based on tables are also fundamental to design principles as taught in the Design Studio section of the book.
7. We have emphasized the power of electrical switches as vital elements of computer design and their mathematical logic analogues.
8. Since standard mathematical control theory is far too advanced for our intended audience, we have used spreadsheet methods that graphically show the effects of feedback gains, paralleling the results of the standard mathematical methods. Most students will still find this chapter to be very challenging.
9. We have developed a simple solution method for standard one-dimensional kinematics problems using a visual/geometric technique of speed-time graphs rather than applying the standard equations by rote. We believe this is a usefully visual way to deal with multi-element kinematics problems. Of course we

[2]Except for answers to ethics problems, which have their own formalism.

have also quoted, but not developed, the standard kinematics equations because they are derived in every introductory college textbook and their use does not increase basic understanding of kinematics *per se*.

10. The design methodology in the Design Studio is presented in a stepwise manner to help lead student and instructor through a hands-on design project.

11. Pacing of hands-on projects is accomplished through design milestones. These are general time-tested project assignments that we believe are the most powerful tool in getting a freshman design course to work well.

12. The many design examples were selected from past student projects, ranging from the freshman to the senior year, to appeal to and be readily grasped by the beginning engineering student. In one chapter we present a couple of typical first-year design projects and follow the evolution of one team's design from clarification of the task to detailed design.

13. The culmination of the hands-on Design Studio is a head-to-head team competition, and it is recommended that all first-year engineering courses based on this text should strive to include it.

14. The Accreditation Board for Engineering and Technology (ABET) sets curriculum criteria[3] that require students to have "an understanding of professional and ethical responsibility." In order to avoid creating this unintentional contrast between ethics and engineering, we have introduced a new pedagogical tool: **the engineering ethics decision matrix**. The rows of the matrix are the canons of engineering ethics and the columns are possible ways to resolve the problem. Each box of the matrix must be filled with a very brief answer to the question, "Does this one particular solution meet this one particular canon?" This is a structured approach that will bring discipline to this subject for first-year engineers. Each chapter in Part 1 has ethics problems pertinent to that particular chapter, and some with suggested answers given. We believe that it is more useful to infuse ethics continually during the term, than as a single arbitrarily inserted lecture.

PGK, GW, RTB and WDK
Union College, Schenectady, New York

[3]According to ABET, engineering programs must demonstrate that students attain an ability to (a) apply the knowledge of mathematics, science, and engineering; (b) design and conduct experiments and analyze data; (c) design a system, component, or process within economic, environmental, social, political, ethical, health-safety, manufacturability, and sustainability constraints; (d) function on multidisciplinary teams; (e) identify, formulate, and solve engineering problems; (f) understand professional and ethical responsibility; (g) communicate effectively; (h) understand engineering solutions in a global, economic, environmental, and societal context; (i) engage in life-long learning; (j) gain a knowledge of contemporary issues; (k) apply modern engineering tools to engineering practice.

A companion web site for this textbook is available at:

www.elsevierdirect.com/companions/9780123747235

It has resources including time management and study skills information, links to unit conversion programs, and practice exercises with some solutions.

For instructors, a solution manual, design contest material, and Power Point slides are available by registering at:

www.textbooks.elsevier.com

It contains worked solutions to every exercise using the Need–Know–How–Solve paradigm as developed in this text.

Acknowledgments

We wish to acknowledge help, suggestions, and advice from several Union colleagues and especially from co-teachers for the Union freshman engineering course: Dean Cherrice Traver, Professors Brad Bruno, James Hedrick, Thomas Jewell, John Spinelli, and Frank Wicks. Dr. John Rogers, Mechanical Engineering Division, West Point, and Dr. Andrew Wolfe, Civil Engineering Technology, State University of New York—Institute of Technology, Utica, NY were also of great assistance in developing this text.

In addition we have received advice, assistance, and most importantly individual chapter reviews, from Professors Nicholas Krouglicof (Memorial University, Newfoundland, Canada) and Thomas Jewell (Union College).

We would also like to thank the following instructors who provided feedback on the revision plan:

Aaron Budge, Minnesota State University, Mankato
Mauro Caputi, Hofstra University
Kelly Crittenden, Louisiana Tech University
Brian DeJong, Central Michigan University
Michael Gregg, Virginia Tech
Daniel Gulino, Ohio University
Jerry C. Hamann, University of Wyoming
Robert Krchnavek, Rowan University
Steven McIntosh, University of Virginia
Francelina Neto, California State Polytechnic University, Pomona
Jin Y. Park, Minnesota State University, Mankato
James Riddell, Baker College

The competition-based hands-on approach to teaching design was inspired by Professor Michael C. Larson, Mechanical Engineering Department, Tulane University, New Orleans, Louisiana, and by Mr. Daniel Retajczyk, then a graduate student at Clarkson University, New York.

Mr. Craig Ferguson (Computer Science/Mechanical Engineering, Union College) developed the student design for the "A Bridge Too Far" example in Part 2.

The graphic illustrations carried over from the first edition were produced by Ted Balmer at March Twenty Productions (http://www.marchtwenty.com)

PART

1

Minds-On

This page intentionally left blank

CHAPTER

What Engineers Do

1

Source: © iStockphoto.com/Antonis Papantoniou

1.1 INTRODUCTION

What is an engineer, and what does he or she do? You can get a good answer to this question by just looking at the word itself. The word *engine* comes from the Latin *ingenerare*, meaning "to create." About 2000 years ago, the Latin word *ingenium* ("the product of genius") was used to describe the design of a new machine. Soon after, the word *ingen* was used to describe all machines. In English, "ingen" was spelled "engine," and people who designed creative things were known as "engine-ers." In French, German, and Spanish today the word for *engineer* is *ingenieur*, and in Italian it is *ingegnere*.

So, again—

What is an engineer?
An engineer is a creative, ingenious person.
What does an engineer do?
Engineers create ingenious solutions to societal problems.

Thus engineering is creative design and analysis that uses energy, materials, motion, and information to serve human needs in innovative ways. Engineers express knowledge in the form of variables, numbers, and units. There are many kinds of engineers, but all share the ideas and methods introduced in this book.

1.2 WHAT DO ENGINEERS DO?

Isaac Asimov once said that "Science can amuse and fascinate us all but it is engineering that changes the world."[1] Almost everything you see around you has been touched by an engineer. Engineers are creative people that use mathematics, scientific principles, material properties, and computer methods to design new products and to solve human problems. Engineers do just about anything, including designing and building roads, bridges, cars, planes, space stations, cell phones, computers, medical equipment, and so forth.

[1]Isaac Asimov's *Book of Science and Nature Quotations*, 1970. (Simon & Schuster)

Engineers can be classified into at least a dozen types, and many subtypes, according to the kind of work they do—administration, construction, consulting, design, development, teaching, planning (also called applications), production, research, sales, service, and test engineers. Because engineering deals with the world around us, the number of engineering disciplines is very large, and includes areas such as aerospace, agricultural, architectural, automotive, biomedical, ceramic, chemical, civil, computer, ecological, electrical, engineering physics, environmental health and safety, geological, marine, mechanical, metallurgical and materials, mining, nuclear, ocean, petroleum, sanitary, systems, textile, and transportation.

Engineers work in industry and government, in laboratories and manufacturing plants, in universities, on construction sites, and as entrepreneurs. They work in an office most of the time, and occasionally travel around the world to manufacturing or construction or equipment test sites. Civil engineers often work outdoors part of the time.

Engineers usually work in teams. Sometimes the team has only two or three engineers, but in large companies, engineering teams have hundreds of people working on a single project (the design and manufacture of a large aircraft, for example). Engineers are responsible for communicating, planning, designing, manufacturing, and testing, among other duties.

Engineers are capable of designing the processes and equipment needed for a project, and sometimes that involves inventing new technologies. Engineers must also test their work carefully before it is used by trying to anticipate all the things that could go wrong, and make sure that their products perform safely and effectively.

More than 1.2 million engineers work in the United States today, making engineering the nation's second-largest profession. According to a survey by the National Association of Colleges and Employers, baccalaureate degree engineering majors have the highest starting salaries.

An engineering degree also opens doors to other careers. Engineering graduates can move into other professions such as medicine, law, and business, where their engineering problem-solving ability is a valuable asset. The list of famous engineers includes American presidents, Nobel Prize winners, astronauts, corporate presidents, entertainers, inventors, and scientists.[2]

Distinguished engineers may be elected to the National Academy of Engineering (NAE); it is the singular highest national honor for engineers.

You can determine what today's engineers do within their specialties by searching the Internet. Here are some of the societies that represent engineers with different subdisciplines: ASME (mechanical engineers), IEEE (electrical engineers), AIChE (chemical engineers), ASTM (materials and testing engineers), ASCE (civil engineers), BMES (biomedical engineers), ANS (nuclear engineers), AIAA (aeronautical engineers), and many others.[3]

A typical engineering society has several functions. They define the core disciplines needed for membership and advocate for them. They also define codes and standards for their discipline, provide further educational courses, and offer a code of engineering ethics customized for that particular profession.

Not surprisingly, you will discover that the basic college engineering courses have much in common with all engineering disciplines. They cover scientific principles, application of logical problem-solving processes, principles of design, value of teamwork, and engineering ethics. If you are considering an engineering career, we highly recommend you consult web resources to refine your understanding of the various fields of engineering.

[2]See http://www.sinc.sunysb.edu/Stu/hnaseer/interest.htm

[3]Canadian engineering societies basically follow a similar nomenclature as do others worldwide.

1.3 WHAT MAKES A "GOOD" ENGINEER?

This is actually a difficult question to answer because the knowledge and skills required to be an engineer (i.e., to create ingenious solutions) is a moving target. The factors that will lead to your career success are not the same as they were 20 years ago. In this book, we illustrate the key characteristics of a successful engineer by exploring the multidisciplinary creative engineering processes required to produce "good" competitive products for the twenty-first century.

So just what *does* the twenty-first century hold for the young engineer? It will be characterized by the convergence of many technologies and engineering systems. The products of today and of tomorrow will be "smarter," in which computers, sensors, controls, modern metal alloys, and plastics are as important as continuing expertise in the traditional engineering disciplines. This book is also intended to appeal to a number of aspects of modern engineering subdisciplines.

Obviously, in a beginning engineering text, we can discuss only a small segment of all the engineering disciplines. Some of the major engineering disciplines are as follows:

- **Bioengineers** deal with the engineering analysis of living systems.
- **Chemical engineers** deal with complex systems and processes including, for example, the way atoms and molecules link up and how those connections shape the properties of materials.
- **Civil engineers** design and analyze large-scale structures such as buildings, bridges, water treatment systems, and so forth.
- **Computer and electronic engineers** design embedded computers and electronic systems that are essential for the operation of modern technology.
- **Control system engineers** design and analyze systems that sense changes in the environment and provide responses to ensure that processes are kept within predetermined tolerances.
- **Electrochemical engineers**, essentially a sub-branch of chemical engineering, mechanical engineering and electrical engineering, work in fields that combine chemistry and electricity such as refining of metals, batteries and fuel cells, sensors, etching, separations, and corrosion.
- **Electrical engineers** design and analyze systems that apply electrical energy.
- **Manufacturing engineers** design manufacturing processes to make products better, faster, and cheaper.
- **Materials engineers** design and apply materials to enhance the performance of engineered systems.
- **Mechanical engineers** work in one of the most diverse of the engineering disciplines, and design and analyze many kinds of predominantly mechanical systems.

1.4 WHAT THIS BOOK COVERS

In your mind, what makes a good consumer product, say an automobile? If you were in the market to purchase one, you might want one that has high performance and good gas mileage and is roomy, safe, and stylish. Or you might describe it in categories like new or used; sedan, sports car, or SUV; two doors or four doors. Or maybe you would be interested only in the price tag.

As a consumer making a decision about purchasing a car, it is enough to use these words, categories, and questions to reach a decision. But engineers think differently. They design and analyze, and consequently they must have a different set of words, categories, and questions. In order to design and analyze, engineers ask precise questions that can be answered with **variables**, **numbers**, and **units**. They do it to accomplish a safe and reliable product. From this point of view, an automobile is an engineer's answer to the question, "What's a good way to move people safely and reliably?"

The purpose of this book is to introduce you to the engineering profession. It does so by introducing you to the way engineers think, ask, and answer questions like: What makes an automobile—or a computer, or an airplane, or a washing machine, or a bridge, or a prosthetic limb, or an oil refinery, or a space satellite—*good*?

We are using the automobile as an example at this point strictly for convenience. It no more and no less expresses the essence of engineering than would an example based on a computer, an airplane, a washing machine, a bridge, a prosthetic limb, an oil refinery, or a space satellite. In each case, the essence of the example would focus on the creative use of energy, materials, motion, and information to serve human needs, so a more detail-oriented engineer might answer our original question like this:

> *A good twenty-first-century automobile employs stored energy (on the order of 100 million joules), complex materials (on the order of 1000 kilograms (about one ton) of steel, aluminum, glass, and plastics), and information (on the order of millions of bits processed every second) to produce an automobile capable of high speed (on the order of 40 meters/second at approximately 90 mph), low cost (a few tens of cents per mile), low pollution (a few grams of pollutants per mile), and high safety.*

That's a long and multidimensional answer, but an engineer would be unapologetic about that. Engineering is *inherently* multidimensional and multidisciplinary. It needs to be multidimensional to create compromises among conflicting criteria, and it needs to be multidisciplinary to understand the technical impact of the compromises. Making a car heavier, for example, might make it safer, but it would also be less fuel efficient. Engineers often deal with such competing factors. They break down general issues into concrete questions. They then answer those questions with design variables, units, and numbers.

Engineering is not a spectator sport. It is a hands-on and minds-on activity. In this book, you will be asked to participate in a "Design Studio." This is the part of the book that is hands-on—and, it's *fun*! But you will still learn the principles of good design practice (regardless of your intended engineering major), and you will have to integrate skills learned in construction, electrical circuits, logic, and computers in building a device (the "device" could be a car, robot, boat, bridge, or anything else appropriate to the course) that will have to compete against similar devices built by other young engineers in your class whose motivation may be to stop your device from succeeding in achieving the same goals! You will learn how to organize data and the vital importance of good communication skills. You will present your ideas and your designs orally and in written format. In the design studio you will design and build increasingly complex engineering systems, starting with the tallest tower made from a single sheet of paper and ending with a controlled device combining many parts into a system aimed at achieving complex goals.

As a start to the minds-on portion of the book, can you mentally take apart and put back together an imaginary automobile or toaster, or computer or bicycle? Instead of using wrenches and screwdrivers, your tools will be mental and computerized tools for engineering thought.

Example 1.1

Figure 1.1 shows a generic car with numbered parts. Without cheating from the footnote, fill in the correct number corresponding to the object in each of the blanks.[4]

[4]Answer: 1-distributor, 2-transmission, 3-spare tire, 4-muffler, 5-gas tank, 6-starter motor, 7-exhaust manifold, 8-oil filter, 9-radiator, 10-alternator, 11-battery.

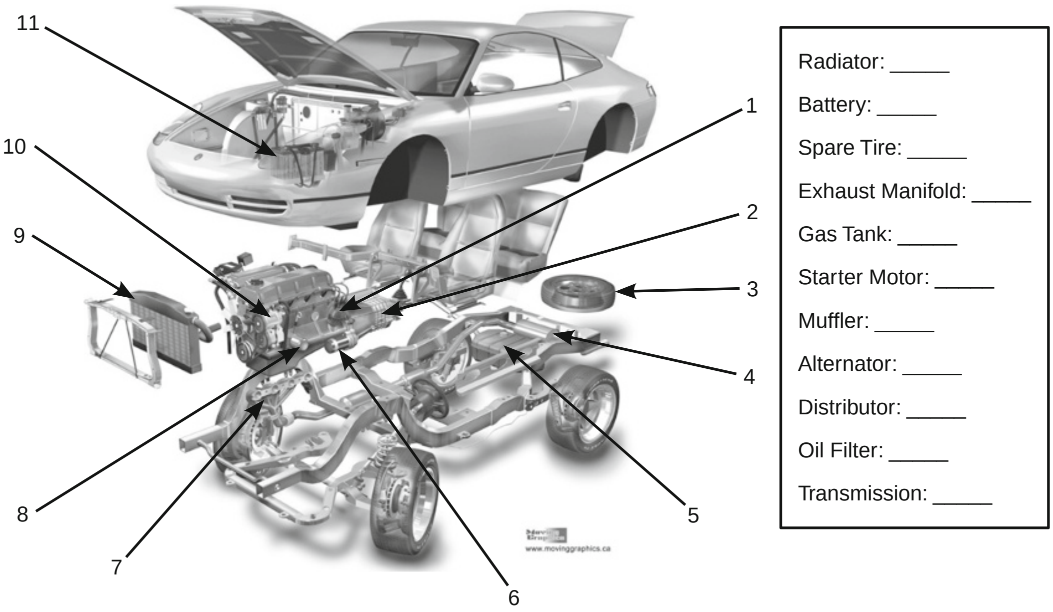


FIGURE 1.1 Exploded View of a Modern Automobile. © Moving Graphics

As visually appealing as this figure is, an engineer would consider it inadequate because it fails to express the functional connections among the various parts. Expressing in visual form the elements and relationships involved in a problem is a crucial tool of engineering, called a **conceptual sketch**. A first step in an engineer's approach to a problem is to draw a conceptual sketch of the problem. Artistic talent is not an issue, nor is graphic accuracy. The engineer's conceptual sketch will not look exactly like the thing it portrays. Rather, it is intended to (1) help the engineer identify the elements in a problem, (2) see how groups of elements are connected together to form subsystems, and (3) understand how all those subsystems work together to create a working system.

Example 1.2

On a piece of paper draw a conceptual sketch of what happens when you push on the pedal of a bicycle. Before you begin, think about these questions:

1. What are the key components that connect the pedal to the wheel?
2. Which ones are connected to each other?
3. How does doing something to one of the components affect the others?
4. What do those connections and changes have to do with accomplishing the task of accelerating the bicycle?

Solution

Here is what your sketch should contain (see Figure 1.2). The pedal is connected to a crank, and the crank is connected to a sprocket. A chain connects the sprocket to a smaller sprocket on the rear wheel. This sprocket is connected to some type of transmission with gears that turns the rear wheel.

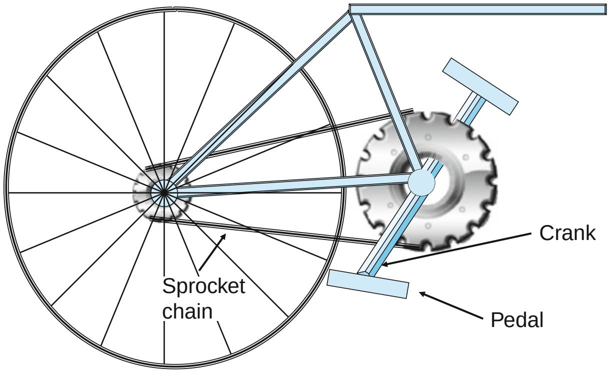


FIGURE 1.2 Bicycle Transmission

For any engineering concept, many different conceptual sketches are possible. You are encouraged to draw conceptual sketches of each of the key points in the learning sections in this book.

1.5 PERSONAL AND PROFESSIONAL ETHICS

What are *personal* ethics . . . and what do they have to do with engineering?

Personal ethics are the standards of human behavior that individuals of different cultures have constructed to make moral judgments about personal or group situations. Ethical principles have developed as people have reflected on the intentions and consequences of their acts. Naturally, they vary over time and from culture to culture, resulting in conflict when what is acceptable in one culture is not in another. For example, the notion of privacy in US culture is very strong, and a desk is considered an extension of that privacy; whereas in another culture, such as Japan, office space is open and a desk would be considered public domain.

Suppose you are a passenger in a car driven by a close friend. The friend is exceeding the speed limit and has an accident. There are no witnesses, and his lawyer tells you that if you testify that your friend was not exceeding the speed limit, it will save him from a jail sentence. What do you do?

Lying is more accepted in cultures that stress human relationships, but it is less accepted in cultures that stress laws. People in cultures that emphasize human relationships would most likely lie to protect the relationship, whereas people in cultures that put greater value on laws would lie less in order to obey the law.

How do you reconcile a belief in certain moral absolutes such as “I will not kill anyone” with the reality that in some circumstances (e.g., war) it might be necessary to endanger or kill innocent people for the greater good? This issue gets particularly difficult if one denies tolerance to other faiths, yet the prevailing morality that most of us would describe as “good” is to extend tolerance to others.

1.5.1 The Five Cornerstones of Ethical Behavior

Here are some examples of codes of personal ethics. At this point you might want to compare your own personal code of ethics with the ones listed here.[5]

- Do what you say you will do.
- Never divulge information given to you in confidence.
- Accept responsibility for your mistakes.
- Never become involved in a lie.
- Never accept gifts that compromise your ability to perform in the best interests of your organization.

1.5.2 Top Ten Questions You Should Ask Yourself When Making an Ethical Decision[6]

1. Could the decision become habit forming? *If so, don't do it.*
2. Is it legal? *If it isn't, don't do it.*
3. Is it safe? *If it isn't, don't do it.*
4. Is it the right thing to do? *If it isn't, don't do it.*
5. Will this stand the test of public scrutiny? *If it won't, don't do it.*
6. If something terrible happened, could I defend my actions? *If you can't, don't do it.*
7. Is it just, balanced, and fair? *If it isn't, don't do it.*
8. How will it make me feel about myself? *If it's lousy, don't do it.*
9. Does this choice lead to the greatest good for the greatest number? *If it doesn't, don't do it.*

And the #1 question you should ask yourself when making an ethical decision:

10. Would I do this in front of my mother? *If you wouldn't, don't do it.*

1.6 WHAT ARE *PROFESSIONAL* ETHICS?

A professional code of ethics has the goal of ensuring that a profession serves the legitimate goals of *all* its constituencies: self, employer, profession, and public. The code protects the members of the profession from some undesired consequences of competition (for example, the pressure to cut corners to save money) while leaving the members of the profession free to benefit from the desired consequences of competition (for example, invention and innovation).

Having a code of ethics enables an engineer to resist the pressure to produce substandard work by saying, "As a professional, I cannot ethically put business concerns ahead of professional ethics." It also enables the engineer to similarly resist pressures to allow concerns such as personal desires, greed, ideology, religion, or politics to override professional ethics.

1.6.1 National Society of Professional Engineers (NSPE) Code of Ethics for Engineers

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. Accordingly, the services provided by engineers require honesty, impartiality, fairness,

[5]Manske, F.A., Jr., *Secrets of Effective Leadership*, Leadership Education and Development, Inc., 1987.
[6]From http://www.cs.bgsu.edu/maner/heuristics/1990Taylor.htm

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.[7]

1.6.2 Fundamental Canons[8]

Engineers, in the fulfillment of their professional duties, shall:

- Hold paramount the safety, health, and welfare of the public.
- Perform services only in areas of their competence.
- Issue public statements only in an objective and truthful manner.
- Act for each employer or client as faithful agents or trustees.
- Avoid deceptive acts.
- Conduct themselves honorably, responsibly, ethically, and lawfully so as to enhance the honor, reputation, and usefulness of the profession.

Example 1.3: An Ethical Situation

The following scenario is a common situation faced by engineering students. Read it and discuss how you would respond. What are your ethical responsibilities?

You and your roommate are both enrolled in the same engineering class. Your roommate spent the weekend partying and did not do the homework that is due on Monday. You did the homework, and your roommate asks to see it. You are afraid he or she will just copy it and turn it in as his or her own work. What are you ethically obligated to do?

a. Show your roommate the homework.
b. Show the homework but ask your roommate not to copy it.
c. Show the homework and tell the roommate that if the homework is copied, you will tell the professor.
d. Refuse to show the homework.
e. Refuse to show the homework but offer to spend time tutoring the roommate.

Solution

For the purposes of this course, the answer to an ethics question will consist of appropriately *applying a code of ethics*. In this example, The Five Cornerstones of Ethical Behavior will be used since they are familiar to you in one form or another.

In subsequent chapters, the NSPE Code of Engineering Ethics will be used, but this does not constitute an endorsement of the code or any other particular code for personal ethics. Use of the NSPE Code of Engineering Ethics in subsequent answers, by contrast, *does* constitute a reminder that you must accept that code in your professional dealings if you want to be a professional engineer.

Let us see which of the Five Cornerstones apply here.

1. **Do what you say you will do.** If the teacher has made it clear that this is an individual assignment, then by participating in the assignment you have implicitly agreed to keep your individual effort private. Allowing your

---

[7] See http://www.nspe.org/ethics/eh1-code.asp
[8] Canons were originally church laws; the word has come to mean rules of acceptable behavior for specific groups.

homework to be copied means going back on this implicit promise. This implies that answer (d) or (e), "Refuse to show the homework," is at least part of the right answer.

2. **Never divulge information given to you in confidence.** Again, homework is implicitly a confidential communication between individual student and teacher. By solving the problem, you have created a confidential communication with the teacher. This is more support for choice (d) or (e).
3. **Accept responsibility for your mistakes.** Sharing your homework will enable your roommate to evade this standard. Being an accomplice in the violation of standards by others is itself an ethical violation. This is further support for choice (d) or (e).
4. **Never become involved in a lie.** Allowing your homework to be copied is participating in a lie: that the work the roommate turns in is his or her own work. This further supports choice (d) or (e).
5. **Never accept gifts that compromise your ability to perform in the best interests of your organization.** Since the roommate has not offered anything in exchange for the help, this standard appears not to apply in this case.

Four of the five cornerstones endorse choice (d) or (e), refuse to show the homework, and the fifth cornerstone is silent. These results indicate that your ethical obligation under this particular code of personal ethics is to refuse to show the homework.

Many people will find the Five Cornerstones to be incomplete because they lack a canon common to most of the world's ethical codes: the Golden Rule.[9] Including the Golden Rule would create the additional obligation to show some empathy for your roommate's plight, just as you would hope to receive such empathy if you were in a similar situation. This suggests the appropriateness of choice (e), offering to tutor the roommate in doing the homework. In much the same way, in subsequent exercises you may feel the need to supplement the Code of Engineering Ethics with elements from your own personal code of ethics. However, this must not take the form of *replacing* an element in the Code of Engineering Ethics with a personal preference.

1.7 ENGINEERING ETHICS DECISION MATRIX

In order to avoid creating an unintentional contrast between ethics and engineering, you will be asked to focus on a particular tool: **the engineering ethics decision matrix**. This tool presents a simple way of applying the canons of engineering ethics and further to see the spectrum of responses that might apply in a given situation. In particular it should give you pause not to accept the first simple do/do not response that comes to you.

In Table 1.1 the rows of the matrix are the canons of engineering ethics (the NSPE set) and the columns are possible ways to resolve the problem. (You can add additional columns as they occur to you.) Each box of the matrix must be filled with a very brief answer to the question: "Does this one particular solution meet this particular canon?" Like other engineering tools, the ethical decision matrix is a way to divide-and-conquer a problem, rather than trying to address all its dimensions simultaneously.

[9]There are many versions of the Golden Rule in the world's major religions. Here's one attributed to Confucius: "Do not do to others what you would not like yourself."

Table 1.1 The Engineering Ethics Decision Matrix

Options → NSPE Canons ↓	Go Along with the Decision	Appeal to Higher Management	Quit Your Job	Write Your State Representative	Call a Newspaper Reporter
Hold paramount the safety, health, and welfare of the public.					
Perform services only in the area of your competence.					
Issue public statements only in an objective and truthful manner.					
Act for each employer or client as faithful agents or trustees.					
Avoid deceptive acts.					
Conduct yourself honorably.					

Example 1.4

You are a civil engineer on a team designing a bridge for a state government. Your team submits what you believe to be the best design by all criteria, at a cost that is within the limits originally set. However, some months later the state undergoes a budget crisis. Your supervisor, also a qualified civil engineer, makes design changes to achieve cost reduction that he or she believes will not compromise the safety of the bridge. You are not so sure, though you cannot conclusively demonstrate a safety hazard. You request that a new safety analysis be done. Your supervisor denies your request on the grounds of time and limited budget. What do you do?

Solution

Table 1.2 shows a typical set of student responses. How would *you* fill out this table?

Notice the multidimensional character of the answers. Here's one way to make some sense of your answer. Total the yes's and the no's in each column (ignore maybe's). By this criterion, you should appeal to higher management, who of course might still ignore you. But that is the first action you should consider even though your boss may strenuously disagree with you. You have a powerful ally in the engineering ethics decision matrix to persuade others to your point of view. Some engineering ethics decision matrices will have just one overwhelming criterion that will negate all other ethical responses on your part—if so, you must follow that path—but usually the engineering ethics decision matrix has multiple conflicting factors. All you should expect from the matrix is that it will stimulate most or all of the relevant terms you should consider and not accept the first thought that entered your head.

Table 1.2 Student Responses to the Ethical Scenario

Options → Canons ↓	Go Along with the Decision	Appeal to Higher Management	Quit Your Job	Write Your State Representative	Call a Newspaper Reporter
Hold paramount the safety, health, and welfare of the public.	**No.** Total assent may put public at risk.	**Maybe.** Addresses risk, but boss may bury issue.	**No.** If you just quit, risks less likely to be addressed.	**Yes.** Potential risk will be put before public.	**Yes.** Potential risk will be put before public.
Perform services only in the area of your competence.	**Yes.** You are not a safety expert.	**Yes.** Though not a safety expert, you are competent to surface an issue.	**Maybe.**	**No.** You are not an expert in government relations.	**No.** You are not an expert in press relations.
Issue public statements only in an objective and truthful manner.	**No.** Silence may seem untruthful assent.	**Maybe.** You are publicly silent, but have registered dissent.	**No.** Quitting in order to avoid the issue is being untruthful.	**Maybe.** Your personal involvement may hurt your objectivity.	**No.** The press is likely to sensationalize what is as yet only a potential issue.
Act for each employer or client as faithful agents or trustees.	**Yes.** As an agent, you are expected to follow orders.	**Yes.** As an agent, you are expected to alert management to potential problems.	**Maybe.** Quitting a job is not bad faith.	**No.** As an agent or trustee, you may not make internal matters public without higher approval.	**No.** As an agent or trustee, you may not make internal matters public without higher approval.
Avoid deceptive acts.	**No.** Assent to something you disagree with is deceptive.	**Yes.** You honestly reveal your disagreement.	**No.** Quitting to avoid responsibility is deceptive.	**Yes.** You honestly reveal your disagreement.	**Yes.** You honestly reveal your disagreement.
Conduct themselves honorably.	**No.** Deceptive assent dishonors the profession.	**Yes.** Honorable dissent is in accord with obligations.	**Maybe.**	**Yes.** Honorable dissent is in accord with obligations.	**Maybe.** Might be publicity seeking, not honorable dissent.
Totals	Yes = 2 No = 4 Maybe = 0	Yes = 4 No = 0 Maybe = 2	Yes = 0 No = 3 Maybe = 3	Yes = 3 No = 2 Maybe = 1	Yes = 2 No = 3 Maybe = 1

1.8 WHAT YOU SHOULD EXPECT FROM THIS BOOK

The old joke goes something like this: "A year ago, I couldn't even spell *injuneer*, but now I are one!" Well, you will *not* be an engineer at the end of this course and, if anything, you will learn at least that much. On the other hand, if you pay attention, you will learn the following:

- Engineering is based on well-founded fundamental principles grounded in physics, chemistry, mathematics, and in logic, to name just a few skills.
- Its most general principles include (1) definition of a Newtonian force unit, (2) conservation of energy, (3) conservation of mass, and (4) the use of control boundaries.
- Engineering problems are multidisciplinary in approach, and the lines between each subdiscipline blur.
- Engineering success often is based on successful teamwork.
- The ability to carry out an introductory analysis in several engineering disciplines should be based on fundamental principles. It often depends on:
  a. Identifying the basic steps in the design process.
  b. Applying those basic steps to simple designs.
  c. Completing a successful team design project.
- You will require sound thinking skills as well as practical hands-on skills.
- The Design Studio will teach you that you will also need writing and oral presentation skills.
- No project is complete without reporting what you have accomplished. Therefore, you will need to demonstrate effective communication skills.
- Computer skills are essential to answer many kinds of practical engineering problems.
- Engineering skills can be intellectually rewarding as well as demanding.
- You should come away with some idea of what is meant by each subdiscipline of engineering and, for those who will continue to seek an engineering career, some idea of which of these subdisciplines most appeals to you.
- We offer a practical way to ask if your behavior is ethical according to well-established engineering ethical canons. If you always act in concordance, no matter the short-term temptations not to, you *will* come out ahead.

SUMMARY

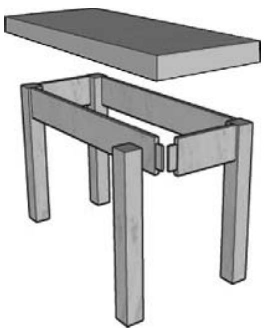
Engineering is about changing the world by creating new solutions to society's problems. This text covers introductions to bioengineering, chemical engineering, civil engineering, computer and electronic engineering, control systems engineering, electrical engineering, electrochemical engineering, materials engineering, and mechanical engineering.

What is common to the branches of engineering is their use of fundamental ideas involving variables, numbers, and units, and the creative use of energy, materials, motion, and information. Engineering is hands-on and minds-on. The hands-on activity for this book is the Design Studio, in which good design practices are used to construct a "device" to compete against similar devices built by other students. You will learn how to keep a log book and how to protect your designs. You will use conceptual sketches to advance your designs.

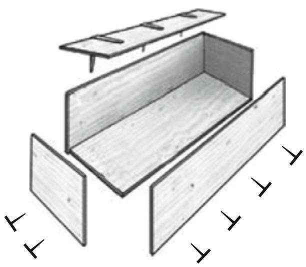
Finally, you will take your first steps to learn the need for professional ethics in your career; this is an ongoing activity you will need regardless of your specialization or job level in whatever direction your career takes you.

EXERCISES

1. Draw a conceptual sketch of your computer. Identify the keyboard, screen, power source, and information storage devices using arrows and labels.
2. Draw a conceptual sketch of an incandescent light bulb. Identify all the components using arrows and numbers as in Figure 1.1.
3. Draw a conceptual sketch of a ballpoint pen. Identify all the components with arrows and labels as in Figure 1.2.
4. The illustration below is an exploded view of a table. Identify and label all the components.



5. The following figure is an exploded view of a box. Identify and label all the components.



6. Repeat Example 1.3 using the NSPE Code of Engineering Ethics. Solve using the Engineering Ethics Matrix.
7. Repeat Example 1.4 using the Five Cornerstones of Ethical Behavior. Solve using the Engineering Ethics Matrix.

In exercises 8 through 12, use the Ethical Decision Matrix Table 1.1, which contains the six Fundamental Canons to respond to these ethical situations.

8. It is the last semester of your senior year and you are anxious to get an exciting electrical engineering position in a major company. You accept a position from Company A early in the recruiting process,

but continue to interview hoping for a better offer. Then your dream job offer comes along from Company B. More salary, better company, more options for advancement, it is just what you have been looking for. What should you do?

a. Just don't show up for work at Company **A**.
b. Send a letter to Company **A** retracting your job acceptance with them.
c. Ask Company **B** to contact Company **A** and tell them you won't be working for them.
d. Reject the offer from Company **B** and work for Company **A** anyway.

9. A company purchased an expensive computer program for your summer job with them. The license agreement states that you can make a backup copy, but you can use the program on only one computer at a time. Your senior design course professor would like you to use the program for your senior design project. What should you do?

a. Give the program to your professor and let him or her worry about the consequences.
b. Copy the program and use it because no one will know.
c. Ask your supervisor at the company that purchased the program if you can use it at school on your senior project.
d. Ask your professor to contact the company and ask for permission to use the program at school.

10. You are attending a student engineering organization regional conference along with five other students from your institution. The night before the group is scheduled to return to campus, one of the students is arrested for public intoxication and is jailed. Neither he nor the other students have enough cash for bail, and he doesn't want his parents to know. He asks you to lend him the organization's emergency cash so that he doesn't have to spend the night in jail; he'll repay you as soon as his parents send money. What should you do?

a. Lend him the money since his parents are wealthy and you know he can repay it quickly.
b. Tell him to contact his parents now and ask for help.
c. Give him the money, but ask him to write and sign a confessional note to repay it.
d. Tell him to call a lawyer since it's not your problem.

11. You are testing motorcycle helmets manufactured by a variety of your competitors. Your company has developed an inexpensive helmet with a liner that will withstand multiple impacts, but is less effective on the initial impact than your competitor's. The Vice President of Sales is anxious to get this new helmet on the market and is threatening to fire you if you do not release it to the manufacturing division. What should you do?

a. Follow the vice president's orders since he or she will ultimately be responsible for the decision.
b. Call a newspaper to "blow the whistle" on the unsafe company policies.
c. Refuse to release the product as unsafe and take your chances on being fired.
d. Stall the vice president while you look for a job at a different company.

12. Paul Ledbetter is employed at Bluestone Ltd. as a manufacturing engineer. He regularly meets with vendors who offer to supply Bluestone with needed services and parts. Paul discovers that one of the vendors, Duncan Mackey, like Paul, is an avid golfer. They begin comparing notes about their favorite golf courses. Paul says he's always wanted to play at the Cherry Orchard Country Club, but since it is a private club, he's never had the opportunity. Duncan says he's been a member there for several years and that he's sure he can arrange a guest visit for Paul. What should Paul do?[10]

a. Paul should accept the invitation since he has always wanted to play there.
b. Paul should reject the invitation since it might adversely affect his business relationship with Duncan.
c. Paul should ask Duncan to nominate him for membership in the club.
d. Paul should ask his supervisor if it's OK to accept Duncan's invitation.

Ethical Decision Matrix for Exercises 8–12				
Canons	**Option (a)**	**Option (b)**	**Option (c)**	**Option (d)**
1. Hold paramount the safety, health, and welfare of the public.				
2. Perform services only in the area of your competence.				
3. Issue public statements only in an objective and truthful manner.				
4. Act for each employer or client as faithful agents or trustees.				
5. Avoid deceptive acts.				
6. Conduct themselves honorably.				

In exercises 13 through 16, use the National Society of Professional Engineers (NSPE) Code of Ethics (see http://www.nspe.org/Ethics/CodeofEthics/index.html) to respond to these ethical situations.

13. Some American companies have refused to promote women into positions of high authority in their international operations in Asia, the Middle East, and South America. Their rationale is that business will be hurt because some foreign customers do not wish to deal with women. It might be contended that this practice is justified out of respect for the customs of countries that discourage women from entering business and the professions.

Some people feel that such practices are wrong and that gender should not to be used in formulating job qualification, and further, that customer preferences should not justify gender discrimination. Present and defend your views on whether or not this discrimination is justified.

14. Marvin Johnson is an Environmental Engineer for one of several local plants whose water discharges flow into a lake in a flourishing tourist area. Included in Marvin's responsibilities is the monitoring of water and air discharges at his plant and the periodic preparation of reports to be submitted to the Department of Natural Resources.

Marvin has just prepared a report that indicates that the level of pollution in the plant's water discharges slightly exceeds the legal limitations. However, there is little reason to believe that this excessive amount poses any danger to people in the area; at worst, it will endanger a small number of fish. On the other hand, solving the problem will cost the plant more than $200,000.

[10]Extracted from Teaching Engineering Ethics, A Case Study Approach, Michael S. Pritchard Editor, Center for the Study of Ethics in Society, Western Michigan University (http://ethics.tamu.edu/pritchar/golfing.htm).

Marvin's supervisor says the excess should be regarded as a mere "technicality," and he asks Marvin to "adjust" the data so that the plant appears to be in compliance. He explains: "We can't afford the $200,000. It would set us behind our competitors. Besides the bad publicity we'd get, it might scare off some of the tourist industry." How do you think Marvin should respond to Edgar's request?

15. Derek Evans used to work for a small computer firm that specializes in developing software for management tasks. Derek was a primary contributor in designing an innovative software system for customer services. This software system is essentially the "lifeblood" of the firm. The small computer firm never asked Derek to sign an agreement that software designed during his employment there becomes the property of the company.

Derek is now working for a much larger computer firm. Derek's job is in the customer service area, and he spends most of his time on the telephone talking with customers having systems problems. This requires him to cross reference large amounts of information. It now occurs to him that by making a few minor alterations in the innovative software system he helped design at the small computer firm, the task of cross referencing can be greatly simplified.

On Friday Derek decides he will come in early Monday morning to make the adaptation. However, on Saturday evening he attends a party with two of his old friends, Horace Jones and you. Since it has been some time since you have seen each other, you spend some time discussing what you have been doing recently. Derek mentions his plan to adapt the software system on Monday. Horace asks, "Isn't that unethical? That system is really the property of your previous employer." "But," Derek replies, "I'm just trying to make my work more efficient. I'm not selling the system to anyone, or anything like that. It's just for my use—and, after all, I did help design it. Besides, it's not exactly the same system—I've made a few changes." What should be done about this situation?[11]

16. Jan, a professional engineer on unpaid leave, is a part-time graduate student at a small private university and is enrolled in a research class for credit taught by Dimanro, a mechanical engineering professor at the university. Part of the research being performed by Jan involves the use of an innovative geothermal technology.

The university is in the process of enlarging its facilities, and Dimanro, a member of the university's building committee, has responsibility for developing a Request For Proposal (RFP) in order to solicit interested engineering firms. Dimanro plans to incorporate an application of the geothermal technology into the RFP. Dimanro asks Jan to serve as a paid consultant to the university's building committee in developing the RFP and reviewing proposals. Jan's employer will not be submitting a proposal and is not averse to having Jan work on the RFP and proposal reviews. Jan agrees to serve as a paid consultant.

Is it a conflict of interest for Jan to be enrolled in a class for credit at the university and at the same time serve as a consultant to the university?[12]

FINAL THOUGHTS[13]

A Calvin and Hobbes comic strip nicely illustrates the importance of thinking ahead in engineering and ethical issues. As they are cascading down a treacherous hill in Calvin's wagon they discuss their circumstance:

> **Calvin**: Ever notice how decisions make chain reactions?
> **Hobbes**: How so?

[11]Adapted from http://ethics.tamu.edu/pritchar/property.htm
[12]Adapted from NSPE Board of Ethical Review Case No. 91-5.
[13]This section is from Michael S. Pritchard, © 1992: Center for the Study of Ethics in Society, http://ethics.tamu.edu/pritchar/an-intro.htm

Calvin: Well, each decision we make determines the range of choices we'll face next. Take this fork in the road for instance. Which way should we go? Arbitrarily I choose left. Now, as a direct result of that decision, we're faced with another choice: Should we jump this ledge or ride along the side of it? If we hadn't turned left at the fork, this new choice would never have come up.

Hobbes: I note with some dismay, you've chosen to jump the ledge.

Calvin: Right. And that decision will give us new choices.

Hobbes: Like, should we bail out or die in the landing?

Calvin: Exactly. Our first decision created a chain reaction of decisions. Let's jump.

After crash-landing in a shallow pond,

Calvin philosophizes: See? If you don't make each decision carefully, you never know where you'll end up. That's an important lesson we should learn sometime.

Hobbes: I wish we could talk about these things without the visual aids.

Hobbes might prefer that they talk through a case study or two before venturing with Calvin into engineering practice.

CHAPTER

2 Key Elements of Engineering Analysis

Source: © iStockphoto.com/Mari Mansikka

2.1 ENGINEERING ANALYSIS

Many physical problems of interest to engineers are modeled by mathematical analysis. In the following chapters you will learn about a few such models. All of those models and the analysis methods used to construct them will share five key elements. One of them, **numerical value**, is familiar to you. Answering a numerical question requires coming up with the right number. But in engineering that is only one part of answering such a question. This chapter introduces other core elements of engineering analysis: **variables**, **dimensions**, **units**, and **significant figures**, as well as a fail-safe method of dealing with units and dimensions.

The essential idea to take away from this chapter is that arriving at the right numerical value in performing an analysis or solving a problem is only one step in the engineer's task. The result of an engineering calculation must involve the appropriate variables; it must be expressed in the appropriate units; it must express the numerical value (with the appropriate number of digits; or significant figures); and it must be accompanied by an explicit method so that others can understand and evaluate the merits and defects of your analysis or solution.

There is one variable introduced in this chapter that also has a strong claim to appear in another chapter that deals with energy and related subjects. That variable is **force**, and it is the scaffold on which much of modern engineering, as well as "classical" physics, relies. The definition of force and its associated units is crucial to what follows in much of this text. It is the strongest example of the notion of units and dimensions that appears in this text and thus has been placed in this chapter.

In addition to the preceding concepts, modern engineers have computerized tools at their fingertips that were unavailable just a generation ago. Because these tools pervasively enhance an engineer's productivity, it is necessary for the beginning engineer to learn them as soon as possible in his or her career. Today, all written reports and presentations are prepared on a computer. But there is another comprehensive computer tool that all engineers use: **spreadsheets**. This tool is another computer language that the engineer must master. We put its study in Chapter 3 so you can soon get some practice in its use.

2.1.1 Variables

Engineers typically seek answers to such questions as, "How hot will this get?", "How heavy will it be?", "What's the voltage?" Each of these questions involves a variable, a precisely defined quantity describing an aspect of nature. What an engineering calculation does is different from what a pure mathematical calculation might do; the latter usually focuses on the final numerical answer as the end product of an analysis.

For example, $\pi = 3.1415926 \ldots$ is a legitimate answer to the question, "What is the value of π?" The question, "How hot?" is answered using the variable "temperature." The question "How heavy?" uses the variable "weight." "What voltage?" uses the variable "electric potential."

For our purposes, variables will almost always be defined in terms of measurements made with familiar instruments such as thermometers, rulers, and clocks. Speed, for example, is defined as a ruler-measurement, distance, divided by a clock-measurement, time. This makes possible what a great engineer and scientist William Thomson, Lord Kelvin (1824–1907), described as the essence of scientific and engineering knowledge.

> *I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be.*

But expressing something in numbers is only the beginning of engineering knowledge. In addition to variables based on measurements and expressed as numbers, achieving Lord Kelvin's aspiration requires a second key element of engineering analysis: **units**.

2.1.2 Units

What if you are stopped by the highway patrol on a Canadian highway and get a ticket saying you were driving at "100"? You would probably guess that the variable involved is speed. But it would also be of interest to know if the claim was that you were traveling at 100 *miles* per hour (mph) or 100 *kilometers* per hour (kph), knowing that 100 kph is only 62 mph. Units can and do make a difference!

Although the fundamental laws of nature are independent of the **system of units** we use with them, in engineering and the sciences a calculated quantity always has two parts: the numerical value *and* its associated units, if any.[1] Therefore, the result of any engineering calculation must always be correct in two separate categories: **It must have the correct numerical value, and it must have the correct units.**

Units are a way of quantifying the underlying concept of **dimensions**. Dimensions are the fundamental quantities we perceive such as mass, length, and time. Units provide us with a numerical scale whereby we can carry out a measurement of a quantity in some dimension. On the other hand, units are established quite arbitrarily and are codified by civil law or cultural custom. How the dimension of length ends up being measured in units of feet or meters has nothing to do with any physical law. It is solely dependent on the creativity and ingenuity of people. Therefore, the basic tenets of units systems often are grounded in the complex roots of past civilizations and cultures.

2.2 THE SI UNIT SYSTEM

The international standard of units is the SI system or, officially, the International System of Units (*Le Système International d'Unités*), which has been abbreviated to SI in many languages. It is the standard of modern science and technology and is based on MKS units (meter, kilogram, second). The fundamental units in the SI system are:

- The meter (m), the fundamental unit of length
- The second (s), the fundamental unit of time
- The kilogram (kg), the fundamental unit of mass

[1]Some engineering quantities legitimately have no associated units—for instance, a ratio of like quantities.

- The degree kelvin (K), the fundamental unit of temperature
- The mole (mol), the fundamental unit of quantity of particles
- The ampere (A), the fundamental unit of electric current

Table 2.1 illustrates a variety of SI units that were all derived from proper names of scientists who made discoveries in each of the fields in which these units are used. Here are some rules regarding unit names.

- All unit names are written without capitalization (unless they appear at the beginning of a sentence), regardless of whether they were derived from proper names.
- When the unit is to be abbreviated, the abbreviation is capitalized if the unit was derived from a proper name.
- Unit abbreviations use two letters *only* when necessary to prevent them from being confused with other established unit abbreviations[2] (e.g., Wb for the magnetic field unit "weber" to distinguish it from the more common W, the watt unit of power), or to express prefixes (e.g., kW for kilowatt).
- A unit abbreviation is never pluralized, whereas the unit's name may be pluralized. For example, kilograms are abbreviated as kg, and *not* kgs, newtons as N and *not* Ns, and the correct abbreviation of seconds is s, not sec. nor secs.
- Unit name abbreviations are *never written with a terminal period* unless they appear at the end of a sentence.
- All other units whose names were not derived from the names of historically important people are both written and abbreviated with lowercase letters—for example, meter (m), kilogram (kg), second (s), and so forth.

Table 2.1 Some SI Units and Their Abbreviations

ampere (A)	henry (H)	pascal (Pa)
becquerel (Bq)	hertz (Hz)	siemens (S)
celsius (°C)	joule (J)	tesla (T)
coulomb (C)	kelvin (K)	volt (V)
farad (F)	newton (N)	watt (W)
gray (Gy)	ohm (Ω)	weber (Wb)

We assume you are familiar with **mass**, and you may indeed think you are, but it is not a trivial concept. As you will soon learn, mass is *not* weight. In a gravitational field, mass certainly produces weight, but the mass is present even where there is no gravity such as in outer space. Mass is best considered as the quantity of matter; it is a property of the substance.

In examples that follow in this text, we introduce a failsafe method that will always allow you to develop the correct units. We will follow the numerical part of the question with square brackets [. . .] enclosing the unit conversions that are needed. For example, knowing there are 12 inches in one foot would produce the units conversion factor [12 inch/ft], so if we wanted to covert 12.7 ft^2 to in^2, we would write 12.7 $[ft^2]$ $[12\ in/ft]^2 = 12.7 \times 144\ [ft^2][in/ft]^2 = 1830\ in^2$.

[2]Non-SI unit systems do not generally follow this simple rule. For example, the English length unit, foot, could be abbreviated "f" rather than "ft." However, the latter abbreviation is well established within society, and changing it at this time would only cause confusion.

Although this may seem ponderous in this example, in examples that are more complicated *it is essential* to follow this methodology. We will attempt to be consistent in presenting solutions and follow this technique throughout this text.

There are many SI units pertaining to different quantities being measured and their multiples thereof (Tables 2.2 and 2.3, respectively). Some of the rationale for their fundamental units will become clearer as we proceed.

Table 2.2 has value beyond merely listing these units: It relates the unit's name to the fundamental MKS units—that is, the fact that a frequency is expressed in "hertz" may not be as useful as the fact that a hertz is nothing but the name of an inverse second, s^{-1}. In Table 2.3, multiples of these quantities are arranged in factors of 1000 for convenience for very large and very small multiples thereof.[3]

Table 2.2 Some Derived SI Units

Quantity	Name	Symbol	Formula	Fundamental Units
Frequency	hertz	Hz	1/s	s^{-1}
Force	newton	N	$kg{\cdot}m/s^2$	$m{\cdot}kg{\cdot}s^{-2}$
Energy	joule	J	N·m	$m^2{\cdot}kg{\cdot}s^{-2}$
Power	watt	W	J/s	$m^2{\cdot}kg{\cdot}s^{-3}$
Electric charge	coulomb	C	A·s	A·s
Electric potential	volt	V	W/A	$m^2{\cdot}kg{\cdot}s^{-3}{\cdot}A^{-1}$
Electric resistance	ohm	Ω	V/A	$m^2{\cdot}kg{\cdot}s^{-3}{\cdot}A^{-2}$
Electric capacitance	farad	F	C/V	$m^{-2}{\cdot}kg^{-1}{\cdot}s^4{\cdot}A^2$

Table 2.3 SI Unit Prefixes

Multiples	Prefixes	Symbols	Submultiples	Prefixes	Symbols
10^{18}	exa	E	10^{-1}	deci	d
10^{15}	peta	P	10^{-2}	centi	c
10^{12}	tera	T	10^{-3}	milli	m
10^{9}	giga	G	10^{-6}	micro	μ
10^{6}	mega	M	10^{-9}	nano	n
10^{3}	kilo	k	10^{-12}	pico	p
10^{2}	hecto	h	10^{-15}	femto	f
10^{1}	deka	da	10^{-18}	atto	a

[3]In recent years, a new subcategory of materials and technology known as **Nanotechnology** has arisen; it is so called because it deals with materials whose size is in the nanometer or 10^{-9} m range.

2.3 FORCE, WEIGHT, AND MASS

Central to any scientific set of units is the definition of **force**. You probably have had some prior introduction to this concept in your high school physics classes. For example, for a constant mass system, Newton's Second Law of Motion is correctly stated as follows:

Force on a mass is proportional to the acceleration it produces.

Newton's Second Law can be written in an equation form:

$$F \propto ma \tag{2.1}$$

where a is the acceleration of mass m. To convert Newton's force law proportionality into an equality, we need to introduce a **constant of proportionality**.

Suppose there exists a set of units for which the force F_1 accelerates the mass m_1 by a_1. Then Newton's Second Law can be written as:

$$F_1 \propto m_1 a_1 \tag{2.2}$$

If we now eliminate the proportionality by dividing the general force-defining Equation (2.1) by Equation (2.2):

$$\frac{F}{F_1} = \frac{m}{m_1}\frac{a}{a_1} \quad \text{or} \quad F = \left(\frac{F_1}{m_1 a_1}\right) ma \tag{2.3}$$

Clearly, the proportionality constant, now explicitly the ratio of a specific force to a specific mass and to a specific acceleration, is very important to the calculations made with this equation. We must choose both the magnitude and the dimensions of this ratio. Any consistent set of units will satisfy this equation. With this degree of flexibility, it is easy to see how a large number of different force systems have evolved.

Example 2.1

Suppose you define a mass unit called the **slug**—a rather ugly word for a mass that accelerates by exactly 1 ft/s^2 when a force of exactly 1 lb is applied to it.

What is the **weight** of a mass of 5.00 slug? (Weight is the term for a force in the Earth's gravitational field, which produces an acceleration of 32.2 ft/s^2.)

Equation (2.3) is the applicable principle for this problem:

$$F = \left(\frac{F_1}{m_1 a_1}\right) ma \text{ in which } \left(\frac{F_1}{m_1 a_1}\right) = \frac{1}{1 \times 1} = 1\left[\frac{\text{lb force} \times \text{s}^2}{\text{slug} \times \text{ft}}\right]$$

hence

$$F = \left(\frac{F_1}{m_1 a_1}\right) mg = 1\left[\frac{\text{lb force} \times \text{s}^2}{\text{slug} \times \text{ft}}\right] \times 5.00[\text{slug}] \times 32.2\left[\frac{\text{ft}}{\text{s}^2}\right] = 161 \text{ lb force}$$

In fact, the slug system is still preferred in the United States by some engineering specialties. Its obvious advantage is that the constant of proportionality for Newton's law is exactly unity so Newton's law can be written in the familiar form of $F = ma$, with the stipulation that mass m is in units of slugs. Its disadvantage is that you probably have little feel for the size of a slug.

The SI system is somewhat similar to the slug system of units but with the choice for the constant of proportionality being both easy and logical: Pick the ratio F_1/m_1a_1 so that its numerical value is exactly 1, and pick it so that it *effectively* has no (i.e., [0]) dimensions.[4] The unit of force thus identified is called the **newton** when the unit of mass is the kilogram and the acceleration is exactly 1 m/s^2. It other words, the force F_1 of 1 N is defined to be that which causes the acceleration a_1 of exactly 1 m/s^2 when acting on a mass m_1 of exactly 1 kg. Thus, in the SI system, $F_1/m_1a_1 \equiv 1$ and Newton's Second Law is again written:

$$F = ma \tag{2.4}$$

This simplification leads to the familiar form of Newton's Law of Motion that most of you have seen before. Since the conversion factor F_1/m_1a_1,—that is, [N/(kg × m/s^2)]—is exactly unity, we have dropped it altogether.

Example 2.2

What is the force in newtons on a body of mass 102 g (0.102 kg) that is accelerated at 9.81 m/s^2?

Equation (2.4) is the principle we use:

$$F = ma = 0.102 \times 9.81\ [\text{kg}][\text{m/s}^2] = 1.00\ [\text{kg}\cdot\text{m/s}^2] = 1.00\ \text{N}$$

The last two examples use an acceleration of special interest, which is that caused by Earth's gravity: $g = 32.2\ \text{ft/s}^2 = 9.81\ \text{m/s}^2$. Looking at Equation (2.4), the SI force acting on 1 kg mass due to gravity is not 1 N but is 9.81 N—a fact that causes distress to newcomers to the subject. Weight is thus just a special force—that due to gravity. In this sense, Equation (2.4) can be modified for the acceleration of gravity to yield that special force we call **weight**, W, by writing it as:

$$W = mg \tag{2.5}$$

Example 2.3

What is the weight in newtons of a mass of 0.102 kg?

Equation (2.5) is the underlying principle; the arithmetic is identical to that of Example 2.2—the force on 0.102 kg of mass (which is about 4 oz in common English units) is just 1.00 N; in other words, **one newton is just about the weight of a small apple here on Earth!** Perhaps this will help you mentally imagine the magnitude of a force stated in newtons.

In addition to SI units, a North American engineer must master at least one of the other systems that relates mass and force, one whose persistence in the United States is due more to custom than logic: it is the **Engineering English** system of units. In this system, the conversion factor between force and mass × acceleration is *not* unity. Because of this we must carry an explicit proportionality constant every time we use this unit system. And because the proportionality factor is not unity, it requires the use of an explicit set of units as well.

[4]Read [0] to mean dimensionless.

This system has also evolved into a rather unfortunate convention regarding both the pound unit and the definition of force. It was decided that the name "pound" would be used both for mass *and* weight (force). Since mass and force are distinctly different quantities, a modifier had to be added to the pound unit to distinguish which (mass or weight) was being used. This was solved by simply using the phrase **pound mass** or **pound force** with the associated abbreviations lbm and lbf, respectively, to distinguish between them.

In the English Engineering system it was decided that a pound mass should weigh a pound force at standard gravity. (Standard gravity accelerates a mass by 32.174 ft/s^2.) This has the helpful convenience of allowing us the intuitive ability to understand immediately what is meant by, say, a force of 15 lbf. It would be the force you would experience if you picked up a rock of mass 15 lbm on the Earth's surface.

This convenience was accomplished by setting the ratio $(F_1/m_1a_1) = 1/32.174$ [lbf·s^2/lbm·ft]. In other words, 1 lbf is defined as the force that will accelerate exactly 1 lbm by exactly 32.174 ft/s^2. The designers of the Engineering English system cleverly decided to define the ***inverse*** of the proportionality constant as g_c defined as:

$$g_c \equiv 32.174 \frac{\text{lbm} \times \text{ft}}{\text{lbf} \times s^2} \tag{2.6}$$

The g_c symbolism originally was chosen because the *numerical value* (but not the dimensions) of g_c is the same as that of the acceleration in standard gravity in the English Engineering units system. However, this is awkward because it tends to make you think that g_c is the same as (i.e., equal to) the acceleration due to local gravity, *g, which it definitely is not*. The constant g_c is nothing more than a proportionality constant with dimensions of [mass × length/(force × time^2)]. Because the use of g_c is so widespread today in the United States and because it is important that you are able to recognize the meaning of g_c when you see it elsewhere, it will be used in the relevant equations in this course except when we are using the much more convenient (and universal) SI units. For example, in the English Engineering unit system, we will henceforth write Newton's Second Law as:

$$F = \frac{ma}{g_c} \tag{2.7}$$

The consequence of this choice for F_1, m_1, and a_1 as expressed by g_c is that you can easily calculate the force in lbf corresponding to an acceleration in ft/s^2 and a mass in lbm.

Example 2.4

What is the force necessary to accelerate a mass of 65.0 lbm at a rate of 15.0 ft/s^2? Since the problem is stated in English units, assume the answer is required in these units. Equation (2.7) is the principle used here:

$$F = \frac{ma}{g_c} = \frac{65.0 \times 15.0}{32.174}[\text{lbm}]\left[\frac{\text{ft}}{\text{s}^2}\right]\left[\frac{\text{lbf} \times \text{s}^2}{\text{lbm} \times \text{ft}}\right] = 30.3\ \text{lbf}$$

Notice how the **units** as well as the value of g_c enter the problem; without g_c our "force" would be in the nonsense units of lbm/ft/s^2 and our calculated numerical value in those nonsense units would be 975.

There are other consequences, too. In a subsequent chapter of this book you will be introduced to the quantity **kinetic energy**. An engineer using the SI system would define kinetic energy as ½ mv^2 (here the v stands for speed). However, an engineer using the Engineering English system would define kinetic energy as ½ $(m/g_c)v^2$. The convenient mnemonic for all applications of the Engineering English system is:

In the Engineering English system, when you see a mass m, divide it by g_c.

Of course, it is logically safer to argue the units using the [..] convention previously introduced. For example, if the grouping ½ mv^2 were expressed in English units, it would be [lbm] [ft/s]^2 if g_c were ignored. This is a meaningless collection of units, but using the g_c conversion factor, the definition of kinetic energy now would read dimensionally as [lbm] [lbf·s^2/lbm·ft][ft/s]^2 = [ft·lbf], a legitimate unit of energy in the English system (see Chapter 3).

Understand that the constant g_c has the same value everywhere in the universe—that is, the value of 32.174 lbm·ft/lbf·s^2, even if its domain of acceptance is confined to the United States! In this regard, as previously stated, it should not be confused with the physical quantity g, the acceleration due to gravity, which has different numerical values at different locations (as well as different dimensions from g_c).

The concept of "weight" always has the notion of the *local* gravity associated with it. Weight thus varies with location—indeed only slightly over the face of the Earth but significantly on nonterrestrial bodies.

To restate what has been learned: the weight of a body of mass m on the surface of the Earth is the force on it due to the acceleration "g" due to gravity—that is,

$$\text{In English Engineering units, } W = \frac{mg}{g_c} \tag{2.8a}$$

where $g = 32.174 \text{ ft/s}^2$ and $g_c = 32.174 \dfrac{\text{lbm}\cdot\text{ft}}{\text{lbf}\cdot\text{s}^2}$

or in SI units,

$$W = mg \tag{2.8b}$$

where $g = 9.81 \text{ m/s}^2$.

The weight of a body of mass m when the local gravity is g' is mg'/g_c in English units and mg' in SI. A person on the International Space Station experiences *micro*gravity, a much smaller g than we do on Earth, and a person on the Moon experiences about 1/6 of g compared to a person on the surface of the Earth. However, if that person is an engineer using the Engineering English system of units, he or she must use the same numerical value for g_c wherever in the universe he or she may be.

In summary, when you see Newton's Second Law written as $F = ma$ (in physics books, for example), you must use a unit system in which the proportionality constant between force and mass × acceleration is unity and is also effectively dimensionless, such as in the SI (MKS) system.[5]

Example 2.5

(a) What is the weight on Earth in Engineering English units[6] of a 10.0 lbm mass?
(b) What is the weight on Earth in SI units of a 10 kg mass?
(c) What is the mass of a 10.0 lbm object on the Moon (local g = 1/6.00 that of Earth)?
(d) What is the weight of that 10.0 lbm object on the Moon?

[5]Or in the cgs (centimeter-gram-second) system, the now outdated predecessor to SI.
[6]Generally, you should give your answers in the natural set of units that is suggested by the problem.

Solution

(a) $W = mg/g_c$ in English units. Thus, $W = 10.0 \times 32.2/32.2$ [lbm × ft/s^2]/[lbm·ft/lbf·s^2] = 10.0 lbf.

(b) $W = mg$ in SI units. Thus, $W = 10.0 \times 9.81$ [kg × m/s^2] = 98.1 [kg m/s^2] = 98.1 N. (See Table 2.2 for the definition of newtons in terms of MKS fundamental units.)

(c) Mass is a property of the material. Thus, the object still has a mass of 10.0 lbm on Earth, on the Moon, or anywhere in the cosmos.

(d) $W = mg/g_c$ in English units. On the Moon, $g = 32.2/6.00 = 5.37$ ft/s^2. Thus, on the Moon, $W = 10.0 \times (5.37)/32.2$ [lbm × ft/s^2]/[lbm·ft/lbf·s^2] = 1.67 lbf.

Until the mid-twentieth century, most English-speaking countries used one or more forms of the Engineering English units system. But because of world trade pressures and the worldwide acceptance of the SI system, many engineering textbooks today present examples and homework problems in both the Engineering English and the SI unit systems. The United States is *slowly* converting to common use of the SI system. However, it appears likely that this conversion will take at least a significant fraction of your lifetime. So to succeed as an engineer in the United States, you must learn the Engineering English system. Doing so will help you avoid future repetition of such disasters as NASA's embarrassing loss in 1999 of an expensive and scientifically important Mars Lander due to an improper conversion between Engineering English and SI units[7] (see Exercise 30 at the end of this chapter).

2.4 SIGNIFICANT FIGURES

Having defined your variable and specified its units, you now calculate its value. Your calculator will obediently spew out that value to as many digits as its display will hold. But how many of those digits really matter? How many of those digits actually contribute toward achieving the purpose of engineering, which is to design useful objects and systems and to understand, predict, and control their function in useful ways? This question introduces into engineering analysis a concept you have possibly seen in your high school science or mathematics courses: the concept of significant figures.

Even the greatest scientists have made the same kind of howling errors by quoting more significant figures than were justified. Newton wrote that "the mass of matter in the Moon will be to the mass of matter in the Earth as 1 to 39.788" (Principia, Book 3, proposition 37, problem 18). Since the ratio of the mass of the Earth to the mass of the Moon is actually $M_e/M_m = 81.300588$, it is clear that Newton had gone wrong somewhere. His value of M_e/M_m to five significant figures was completely unjustified.[8]

The use of the proper number of significant figures in experimental work is an important part of the experimentation process. Reporting a measurement of, say, 10 meters, or as 10. meters (notice the period after the zero), or as 10.0 meters, or as 10.00 meters, implies something about how accurately the measurement was made. The implication of 10 meters as written is that the accuracy of our measuring rule is of the order of ±10 meters. However, 10. meters implies the measurement was good to ±1 meter. Likewise, 10.0 meters implies accuracy to 0.1 meter and 10.00 meters to 0.01 meters and so on, a convention we will try to maintain in this book. Unless they are integers, numbers such as 1, 30, and 100 all have only one significant figure.

The concept of significant figures arises since arithmetic alone will not increase the accuracy of a measured quantity.[9] If arithmetic is applied indiscriminately it might actually decrease the accuracy of the

[7]See http://www.cnn.com/TECH/space/9909/30/mars.metric for more information.

[8]D. W. Hughes, "Measuring the Moon's Mass," from *Observatory*, Vol. 122, April 2002, p. 62.

[9]The averaging of a number of repeated measurements of the same quantity might seem to violate this statement, but it allows only increased confidence in the interval in which the averaged number will lie.

result. The use of significant figures is a method to avoid such blunders as 10/6 = 1.666666667 (as easily obtained on many electronic calculators), whereas the strict answer is 2! (since 6 and 10 are apparently known only to 1 *significant* figure here, and, if they represent real physical measurements, they apparently have not been measured to the implied accuracy of the arithmetical operation that produced 1.666666667.) In this sense, physical numbers differ from pure mathematical numbers.

"Exact" numbers (numbers such as in 1 foot = 12 inches, or numbers that come from counting, or in definitions such as diameter = 2 × radius) have no uncertainty and can be assumed to have an infinite number of significant figures. Thus they do not limit the number of significant figures in a calculation.

Definition

A significant figure[10] is any one of the digits 1, 2, 3, 4, 5, 6, 7, 8, 9, and 0. Note that zero is a significant figure *except* when it is used simply to fix the decimal point or to fill the places of unknown or discarded digits.

The number 234 has three significant figures, and the number 7305 has four significant figures since the zero within the number is a legitimate significant digit. But leading zeroes before a decimal point are not significant. Thus, the number 0.000452 has three significant figures (4, 5, and 2), the leading zeroes (including the first one before the decimal point) being place markers rather than significant figures.

How about trailing zeroes? For example, the number 12,300 is indeed twelve thousand, three hundred, but we can't tell without additional information whether the trailing zeroes represent the **precision**[11] of the number or merely its magnitude. If the number 12,300 was precise only to ±100, it has just three significant figures. If it were truly precise to ±1, then all five figures are significant. In order to convey unequivocally which ending zeroes of a number are significant, it should be written as 1.2×10^4 if it has only two significant figures, as 1.23×10^4 if it has three, as 1.230×10^4 if it has four, and as 1.2300×10^4 (or 12300.—note the decimal point) if it has five.

The identification of the number of significant figures associated with a measurement comes only through knowledge of how the measurement was carried out. For example, if we measured the diameter of a circular shaft with a ruler, the result might be 3.5 inches (two significant figures), but if it were measured with a digital micrometer (Figure 2.1), it might be 3.512 inches (four significant figures).

Engineering calculations often deal with numbers having unequal numbers of significant figures. A number of logically defensible rules have been developed for various computations. These rules are actually the result of strict mathematical understanding of the propagation of errors due to arithmetical operations such as addition, subtraction, multiplication, and division. Rule 1 is for addition and subtraction.

Rule 1

The sum or difference of two values should contain no significant figures farther to the right of the decimal place than occurs in the least precise number in the operation.

[10]There are many good web sites on the Internet that deal with this concept. The following site has a self-test that you can use to check yourself: http://science.widener.edu/svb/tutorial/sigfigures.html

[11]The word "precision" has a specific meaning: it refers to how many times using independent measurements we can *reproduce* the number. If we throw many darts at a dartboard and they all cluster in the double three ring, they are precise but, if we were aiming at the bull's eye, they were not accurate!

FIGURE 2.1 Digital Micrometer

For example, 113.2 + 1.43 = 114.63, which must now be rounded to 114.6. The less precise number in this operation is 113.2 (having only one place to the right of the decimal point), so the final result can have no more than one place to the right of the decimal point. Similarly, 113.2 − 1.43 = 111.77 must now be rounded to 111.8. This is vitally important when subtracting two numbers of similar magnitude since their difference may be much less significant than the two numbers that were subtracted. For example, 113.212 − 113.0 = 0.2 has only one significant figure even though the "measured" numbers each had four or more significant figures.

There is another rule for multiplication and division, as follows.

Rule 2

The rule for multiplication and division of figures is:

> *The product or quotient should contain no more significant figures than are contained by the term with the least number of significant figures used in the operation.*

For example, (113.2) × (1.43) = 161.876, which must now be rounded to 162, and 113.2/1.43 = 79.16, which must now be rounded to 79.2 because 1.43 contains the least number of significant figures (i.e., three) in each case.

Finally a rule for "rounding" is as follows.

Rule 3[12]

The rule for rounding numbers up or down is:

> *When the discarded part of the number is 0, 1, 2, 3, or 4, the next remaining digit should not be changed. When the discarded part of the number is 5, 6, 7, 8, or 9, then the next remaining digit should be increased by one.*

[12]There is another round-off rule corresponding to Rule 3: The so-called Bankers' Rule was used before computers to check long columns of numbers. When the discarded part of the number is exactly 5 followed only by zeros (or nothing), then the previous digit should be rounded up if it is an odd number, but it remains unchanged if it is an even number. It was meant to average out any rounding bias in adding the columns.

For example, if we were to round 113.2 to three significant figures, it would be 113. If we were to round it further to two significant figures, it would be 110, and if we were to round it to one significant figure, it would be 100 with the trailing zeroes representing placeholders only. As another example, 116.876 rounded to five significant figures is 116.88, which further rounded to four significant figures is 116.9, which further rounded to three significant figures is 117. As another example, 1.55 rounds to 1.6, but 1.54 rounds to 1.5.

SUMMARY

To summarize this chapter, the results of an engineering analysis must be correct in four ways. It must involve the appropriate variables. It must be expressed in the appropriate units. It must express the correct numerical value with the appropriate number of significant figures.

Newton's Second Law of Motion presents additional challenges in dimensional analysis, since so many possibilities are open to define its proportionality constant. In the MKS system, the proportionality constant is unity, so we can write it as $F = ma$; in the English Engineering system, the proportionality constant g_c is chosen to have a value of 32.174 lbm·ft/lbf·s^2 and Newton's law is written as $F = ma/g_c$. Dependent quantities such as kinetic energy are likewise modified in the English Engineering system of units by dividing quantities that contain lbm by g_c.

EXERCISES

To help get you in the habit of applying these elements, in the following exercises you will be graded on use of *all* the elements we have discussed in this chapter. In these problems, where necessary, assume that on the surface of the Earth, $g = 9.81$ m/s^2 $= 32.2$ ft/s^2 (i.e., each to three significant figures). Make sure you are reporting the solution to the proper number of significant figures.

1. If a US gallon[13] has a volume of 0.134 ft^3 and a human mouth has a volume of 0.900 in^3, then how many mouthfuls of water are required to fill a 5.00 US gallon can? (**A: 1.29 × 10^3 mouthfuls**)

2. Identify whether you would perform the following unit conversions by definition, by conversion factors, by geometry, or by scientific law.
   a. How many square miles in a square kilometer?
   b. How many microfarads in a farad?
   c. What is the weight on Earth in N of an object with a mass of 10.0 kg?
   d. How many square miles on the surface of the Earth?

3. The height of horses from the ground to their shoulder is still measured in the old unit of hands. There are 16 hands in a fathom and 6.0 feet in a fathom. How many feet high is a horse that is 13 hands tall? (**A: 4.9 ft**)

4. An acre originally was defined as the amount of land that an oxen team could plow in a day.[14] Suppose a team could plow 0.4 hectare per day, where a hectare is 104 m^2. There are 1609 meters in a mile. How many acres are there in a square mile?

[13]Distinguish US gallons from the old English measure of Imperial gallons; 1.00 Imp. gallon = 1.20 US gallons.
[14]The furlong or "furrow-long" was the distance of 220 yards that the oxen could plow in a day times a width of one "chain" or 22 yards. Multiplied together they defined the area of an acre.

5. There are 39 inches in a meter. What is the area in the SI system of the skin of a spherical orange that is 4.0 inches in diameter? (**A: 3.3** $\times$ **10**$^{-2}$ **m**^{2})

6. There are 39 inches in a meter. What is the volume in the Engineering English system of a spherical apple that is 10. cm (note the decimal point here) in diameter?

7. If the pressure in the tire on your car is 32.0 lbf/in^2 (or psi), what is its pressure in SI units?

8. Suppose the mass in Example 2.1 was 50.0 slugs. What would be its weight in lbf (pounds force)?

9. What would the 5.00 slug mass in Example 2.1 weigh on the Moon where the acceleration of gravity is only 1/6 of that on Earth?

10. What would be the force on the body in Example 2.2 if its mass were 856 grams?

11. What would be the weight of the body in Example 2.3 on the Moon where the acceleration of gravity is just $g_{Moon} = 1.64$ m/s^2?

12. What force would be necessary in Example 2.4 if the mass were 735 lbm?

13. What is the value and units of g_c in the Engineering English system on the Moon?

14. Acceleration is sometimes measured in g's, where 1.0 $g = 9.8$ m/s^2. How many g's correspond to the steady acceleration of a car doing "zero to sixty"[15] in 10.0 seconds? (**A: 0.27 *g***)

15. What is your mass in kilograms divided by your weight in pounds? Do you have to step onto a scale to answer this question? How did you answer the question?

16. If power (measured in W, or watts) is defined as work (measured in J, or joules) performed per unit time (measured in s), and work is defined as force (measured in N or newtons) $\times$ distance (measured in m) and speed is defined as distance per unit time (measured in m/s), what is the power being exerted by a force of 1000. N on a car traveling at 30. m/s? (Assume force and speed are in the same direction, and treat all numbers as positive.) (**A: 3.0** $\times$ **10**^{4} **W**)

17. A rocket sled exerts 3.00×10^4 N of thrust and has a mass of 2.00×10^3 kg. In how much time does it do "zero to sixty"? How many g's (see Exercise 14) does it achieve?

18. A person pushes a crate on a frictionless surface with a force of 100. lbf. The crate accelerates at a rate of 3.0 feet per second^2. What is the mass of the crate in lbm? (**A: 1.07** $\times$ **10**^{3} **lbm**)

19. The force of gravity on the Moon is one-sixth (i.e., 1/6.0) as strong as the force of gravity on Earth. An apple weighs 1.0 N on Earth. (a) What is the mass of the apple on the Moon, in lbm? (b) What is the weight of the apple on the Moon, in lbf? (Conversion factor: 1.00 kg = 2.20 lbm)

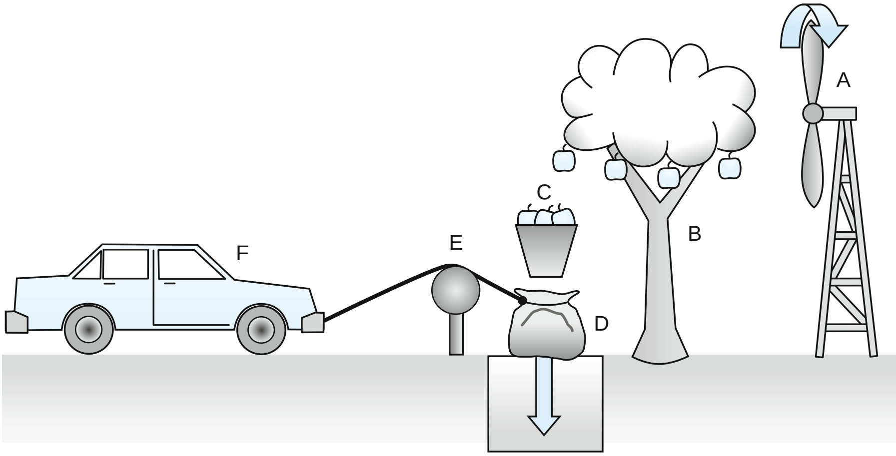
20. How many lbf does it take for a 4.0×10^3 lbm car to achieve 0 to 60 mph in 10. seconds? (**A: 1.1** $\times$ **10**^{3} **lbf**)

21. Suppose a planet exerted a gravitational force at its surface that was 0.6 the gravitational force exerted by Earth. What is g_c on that planet?

22. Suppose you were going to accelerate a 2000. kg car by the Rube Goldberg contraption shown in the following figure. The fan (A) blows apples (C) off the tree (B) into the funnel and thus into the bag (D). The bag is pulled downward by the force of gravity (equal to the weight of the apples in the bag), and that force

[15] In the language of the car enthusiast, a standard test to accelerate a car from a standing start to 60 mph is called "zero to sixty."

is transmitted via the pulley (E) to accelerate the car (F). About how many apples each weighing 1.00 N would have to fall into the bag in order to achieve 0 to 60.0 mph in 7.00 seconds? Assume the filled bag applies a constant force to the car, equal to the weight of the apples in the bag. (**A: 7.66** $\times$ **10^3 apples**)



23. Calculate with the correct significant figures: (a) $100/(2.0 \times 10^2)$, (b) $(1.0 \times 10^2)/(2.0 \times 10^2)$. (**A: 0.5, 0.50**)

24. Calculate with the correct significant figures: (a) 10/6, (b) 10.0/6, (c) 10/6.0, (d) 10./6.0, (e) 10.0/6.00.

25. What is 2.68×10^8 minus 2.33×10^3 to the correct significant figures? (**A: 2.68** $\times$ **10^8**)

26. A machinist has a sophisticated micrometer that can measure the diameter of a drill bit to 1/10,000 of an inch. What is the maximum number of significant figures that should be reported if the approximate diameter of the drill bit is: (a) 0.0001 inches, (b) 0.1 inches, (c) 1 inch?

27. Round off to three significant places: 1.53, 15.345, 16.67, 102.04, −124.7, and 0.00123456.

28. Suppose you were going to design a front door and doorway to fit snugly enough to keep out the drafts, yet to be easy to open. (You are not showing off precision carpentry here, but merely designing a convenient ordinary door by standard methods.) The dimensions are to be given in inches. To how many significant figures would you specify the length and width of the door and doorway?[16] Assume a standard door is 30.0″ by 81.0″.

29. You are browsing the Internet and find some units conversion software that may be useful in this course. You would like to download the software on your PC at school and use it in this course. What do you do?
a. Check with the Internet site to make sure this software is freeware for your use in this course.
b. Just download the software and use it because no one will know.
c. Download the software at home and bring it to school.
d. Never use software found on the Internet.

[16]The subject of tolerancing is important in mass manufacturing to ensure a proper fit with one part and another; since each part will not be exact their combined tolerance will determine how well, if at all, they fit together. It is the subject of significant statistical analysis. Applying the methods rigorously allowed the Japanese automotive industry to eclipse those of the rest of the world in terms of quality (see Chapter 15, Manufacturing Engineering).

Suggested method: Apply the Fundamental Canons and fill in an Engineering Ethics Matrix:

Options → Canons ↓	Check for Freeware	Just Download	Download at Home	Never Use Internet Software
Hold paramount the safety, health, and welfare of the public.				
Perform services only in the area of your competence.				
Issue public statements only in an objective and truthful manner.				
Act for each employer or client as faithful agents or trustees.				
Avoid deceptive acts.				
Conduct themselves honorably.				

30. On December 11, 1998, the Mars *Climate Orbiter* was launched on a 760 million mile journey to the Red Planet. On September 23, 1999, a final rocket firing was to put the spacecraft into orbit, but it disappeared. An investigation board concluded that NASA engineers failed to convert the rocket's thrust from pounds force to newtons (the unit used in the guidance software), causing the spacecraft to miss its intended 140–150 km altitude above Mars during orbit insertion, instead entering the Martian atmosphere at about 57 km. The spacecraft was then destroyed by atmospheric stresses and friction at this low altitude. As chief NASA engineer on this mission, how do you react to the national outcry for such a foolish mistake?

a. Take all the blame yourself and resign.
b. Find the person responsible, and fire, demote, or penalize that person.
c. Make sure it doesn't happen again by conducting a software audit for specification compliance on all data transferred between working groups.
d. Verify the consistent use of units throughout the spacecraft design and operations.

Suggested method: Apply the Fundamental Canons and fill in an Engineering Ethics Matrix.

CHAPTER

3 Solving Problems and Spreadsheet Analyses

Source: © iStockphoto.com/Johnny Lye

3.1 THE NEED–KNOW–HOW–SOLVE METHOD

We suggest the **need–know–how–solve** method for setting up and solving problems. It has the powerful benefit of enabling you to attack complicated problems in a systematic manner, thereby making success more likely. It may seem cumbersome in simple problems, but for those with an eye to their final course grades, it has an additional benefit of ensuring at least partial credit when a miscalculated number results in an incorrect answer! It also tracks an essential element of modern engineering practice by providing an "audit trail" by which past errors can be traced to their source, easing subsequent corrections. The need–know–how–solve method is a self-contained mnemonic device that should also help direct your thought processes.

Computers are essential for modern engineering; engineers must work with them, and need to learn a number of computer languages and high-level programs. Fortunately, one of the more intuitive and powerful tool is **spreadsheet analysis**. This tool is simply another computer language that the engineer has to master. We will put its study at the end of this chapter so that the student can get practice as soon as possible, and we will use spreadsheet techniques throughout this book in the relevant exercise sections.

We have previously emphasized the need to find the proper variables, to express them in a consistent set of units, and to express the answer to the appropriate number of significant figures. But what guarantees do we have that the answer, even if correctly expressed in this way, is actually the correct solution to a problem or a useful result that accurately reflects or predicts the performance of an actual object or system in the real world? The correctness of an answer can be no better than the correctness of the methods and assumptions used to obtain it. Typically, the more systematic your method, the more likely you will be to avoid errors. In addition, the best way to ensure the correctness of a method is to submit it to evaluation and criticism. The use of an explicit method will leave an audit trail for your colleagues and customers (and for your boss!).

In the need–know–how–solve method, the **need** is the variable for which you are solving. It is the very first thing you should write down. The **know** is the quantities that are known, either through explicit statement of the problem or through your background knowledge. Write these down next. They may be graphical sketches or schematic figures as well as numbers or principles. The **how** is the method you will use to solve the problem, typically expressed in an equation, although rough sketches or graphs may also be included. The "how" also includes the assumptions you make to solve the problem. Write it out in sentence or symbolic form before applying it. Only then, with need, know, and how explicitly laid out, should you proceed to **solve** the problem. Often it is also necessary to discuss the implications of your solution.

The pitfall this method guards against is the normal human temptation to look at a problem, *think* you know the answer, and simply write it down. But, rather than gamble with this hit-and-miss method, the need–know–how–solve scheme gives a logical development that shows your thought processes.

Many of the problems you will solve in this book will involve specific equations. But do not be deceived into thinking that this method involves the blind plugging of numbers into equations if you do not understand where those equations came from. The essence of the method is to realize that once you have correctly defined the variable you **need**, you will realize that you **know** a lot more about the problem than you thought you did. With the aid of that knowledge, there is a method (**how**) to **solve** the problem to an appropriate level of accuracy. Sometimes, this method will not require any equations at all.

In some sense, there is even a prior step before "need": Read the question and then reread the question. There is nothing more frustrating than solving the wrong problem!

It bears emphasizing: The need–know–how–solve process is guaranteed to get you a better grade, even if the final answer is the same one that got an "F" when all you bothered to write down was the final answer! This parallels what a practicing engineer does to document his or her work; by leaving an audit trail, prior mistakes can be spotted and rectified.

When applying the method, it may be especially helpful, as part of the "how" step, to sketch the situation described in the problem. Many people are discouraged from sketching by lack of artistic talent. Do not let this stop you. As the next example illustrates, even the crudest sketch can be illuminating.

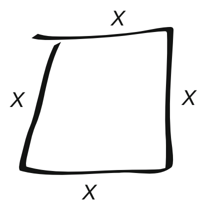
Example 3.1

A Texan wants to purchase the largest fenced-in square ranch she can afford. She has exactly $320,000. available for the purchase. Fencing costs exactly $10,000. a mile, and land costs exactly $100,000. a square mile. How large a ranch, as measured by the length of one side of a square, can she buy?

Need: The length of a side of the largest square of land the Texan can buy.

Know: Fencing costs $10,000. a mile, and land costs $100,000. a square mile. Our Texan has $320,000. to invest.

How: Let the unknown length = x miles. It may not be immediately obvious how to write an equation to find x. So sketch the ranch.



From the crude picture (and that is generally all you need), it is immediately obvious that the length of the fence surrounding the ranch is 4x, and the area of the ranch is x^2.

So the cost of the ranch is:

$$\text{Length of fence [miles]} \times 10{,}000.\ [\$/\text{mile}] + \text{area of ranch [square miles]} \times 100{,}000.\ [\$/\text{sq. mi}] = \text{total cost of ranch} = \$320{,}000.$$

or

$$4x\ [\text{miles}] \times 10{,}000.\ [\$/\text{mile}] + x^2\ [\text{sq. mile}] \times 100{,}000.\ [\$/\text{sq. mi}] = \$320{,}000.$$

Solve: $(4x) \times (10{,}000.) + x^2 \times (100{,}000.) = 320{,}000.$

Therefore, $10x^2 + 4x - 32 = 0$, which is a quadratic equation whose solutions are:

$$\mathbf{x = +1.6 \text{ and } x = -2.0}\ (\textbf{to just two significant figures, which is typical of land measurement})$$

So there are two solutions, x = 1.6 miles and x = −2 miles. Here we must apply a bit of knowledge so obvious that, although known, it was not listed in the **"Know"** section: that the length of the side of the ranch must be greater than zero (if for no other reason than to have room to put the cattle!). This yields the final answer: The ranch is 1.6 miles on a side. (It's always worth checking if this is correct: $1.6^2 \times \$1.0 \times 10^5 + 4 \times 1.6 \times \$1.0 \times 10^4 = \$3.2 \times 10^5$, which is arithmetically correct.)

Example 3.2

Consider the following problem: How many barbershops are there in the city of Schenectady (population about 60,000 people)? Your first reaction may be "I haven't got a clue." (You may want to try the need–know–how–solve method on this problem for yourself before looking at what follows.)

Need: The number of barbershops.

Know: There are about 60,000 people in the city of Schenectady, of whom about half are male. Assume the average male gets about 10 haircuts a year. A barber can probably do one haircut every half hour, or about 16 in the course of an eight-hour day. There are about three barbers in a typical barbershop.

How: The number of haircuts given by the barbers must be equal to the number of haircuts received by the customers. So if we calculate the number of haircuts per day received by all those 30,000 males, we can find the number of barbers needed to give those haircuts. Then we can calculate the number of shops needed to hold those barbers. Assume that barbershops are open 300 days per year.

Solve: 30,000 males require about 10 [haircuts/male-year] × 30,000 [males] = 300,000 haircuts/year.

On a per-day basis, this is about 300,000 [haircuts/year] × [one year/300 days] = 1000 haircuts/day.

This requires 1000 [haircuts/day] × [1 barber-day/16 haircuts] = 62.5 barbers. At three barbers per shop, this means 62.5 barbers × [1 shop/3 barbers] = 20.8 shops.

So the solution is **20 barbershops** (since surely not more than one of the digits in the calculation above is significant).

Looking in a recent yellow pages directory and counting the number of barbershops in Schenectady gives the result 23 barber shops. So we are within about 10 percent, which is a fortuitously good answer, given the roughness of our estimates and the many likely sources of error. Notice also in this problem how the method of carrying the units in square brackets [..] helps your analysis of the problem and directs your thinking.

Haircuts and barbershops and Texas ranches are not among the variables or units used by engineers. But, *applying common sense* as well as equations is a crucial component of engineering analysis as demonstrated in the next example.

Example 3.3

A 2.00 m steel wire is suspended from a hook in the ceiling with a mass of 10.0 kg that is tied to its lower end; the wire stretches by 15.0 mm under this load. If this same mass is used to stretch a 4.00 m piece of the same steel wire, how much will it stretch?

Need: Stretch = ____ mm for a 4.00 m piece of wire.

Know: 10.0 kg mass will stretch a 2.00 m wire by 15.0 mm.

How: We need to deduce a possible law for extending a wire under load. Without experimentation we cannot know if our "theoretical law" is correct, but a mixture of common sense and dimensional analysis can yield a plausible relationship.

Solve: A longer wire should stretch further than a shorter one if otherwise equivalent. A larger mass presumably will also stretch the wire further.

A *plausible* model is thus the extension x is proportional to the unstretched length of wire L (all other things such as the stretching force and the wire's cross section being equal).

Then $x \propto L$ and the ratio between the two cases is $\frac{x_2}{x_1} = \frac{L_2}{L_1}$.

Therefore the new extension $\mathbf{x_2} = 15.0 \times 4.00/2.00$ [mm][m/m] $=$ **30.0 mm.**

The proportionality law derived in this manner is not guaranteed to be correct. To guarantee that it is physically correct, we need to either understand the underlying principles of wire stretching or have good experimental data relating wire stretching to the wire's length. In fact it is correct and is part of a physical law named Hooke's Law.

In some problems, the need–know–how–solve method is decidedly clumsy in execution; nevertheless, it is recommended that you use the basic method and learn how to modify it to suit the peculiarities of the problem at hand. You will soon find there is ambiguity among what you *know*, and *how*, and even to *solve*. Never mind—the relevant information can go into one or the other baskets—just as long as it acts as a jog in the direction you need to solve the problem.

The five elements of engineering analysis introduced in this chapter will provide systematic ways of applying sense, both common and uncommon, to help meet engineering challenges.

3.2 SPREADSHEET ANALYSIS

Unfortunately, even though the **need–know–how–solve** method will get you to an understanding of the answer, the subsequent mathematics may be too difficult to solve on a piece of paper, or perhaps you will need multiple cases, or graphical solutions, or other complications that prevent you from immediately writing down the answer you are seeking. Today, computers come to your aid; an experienced engineer on his or her laptop computer can obtain solutions to problems that were daunting a generation ago and might have required a room full of people hand-cranking out small pieces of mathematical puzzles. In addition, spreadsheets do a superior job of displaying data—no small advantage if you want a busy person, perhaps your customer or your manager, to pay attention to the data.

Spreadsheets allow you to distinguish among three disparate concepts—**data, information**, and **knowledge**. First, what are data?[1]

What we call data are just a jumble of facts or numbers such 4, 3, 1, 1, 6; or an abstract set of colors such as red, violet, green, blue, orange, indigo, and yellow. They become information when you sort them out in some way such as 3, 1, 4, 1, 6, or red, orange, yellow, green, blue, indigo, and violet. They respectively become knowledge when the numerical sequence is interpreted as 3.1416 (i.e., π) and the color sequence is recognized as the visible spectrum arranged as dispersed by a prism (mnemonic ROYGBIV).

Spreadsheets were invented at the Harvard Business School in 1979 by a student[2] who was apparently bored with repetitive hand calculations. He and a fellow entrepreneur made the first spreadsheet tool, VisiCalc. It originally was intended simply to take the drudgery out of relatively mundane business ledgers. Not only do modern spreadsheets take the drudgery out of the computations, but they are also wonderful at displaying data in neat and tidy ways that can be real aids in understanding the displayed data.

The spreadsheet concept became one of the first so-called "killer-apps"—an application tool that no self-respecting personal computer could do without! Lotus 123 soon followed VisiCalc, and that was soon followed by Microsoft® Excel. In this book we will be explicitly using Excel, but the principles you learn here are readily transferred to IBM's Lotus 123® or to Corel's Quattro®. VisiCalc has faded from common use.

---

[1]Singular form *datum*!
[2]http://www.bricklin.com/visicalc.htm

Until VisiCalc, computers had tedious formatting requirements to display their results and typically as much time and energy was expended on getting results from the computer in an understandable way as was spent computing the results in the first place! What made spreadsheets interesting was the visual way they could handle large arrays of numbers; these numbers were identified by their position in these arrays by reading across their rows and down their columns. Interactions among these numbers were rapidly enhanced by providing sophisticated and complex mathematical functions that would manipulate the numbers in these arrays. Further, these arrays could be graphed in many ways so that an explicit crafted and individualized output was easily produced. Gone were most of the arcane skills in formatting that mainframe computers had demanded, and, since it was visual, silly mistakes instantly stood out like sore thumbs and could be expeditiously corrected. No wonder spreadsheets were one of the killer applications that drove the sales of PCs as indispensable tools for engineering, science, and for business.

Figure 3.1 is a reproduction of an Excel Spreadsheet with some guidelines as to its key parameters. It is navigated by the intersection of rows and columns. The first cell is thus A1. Active cells such as E10 in the example are boxed in a bold border. You can scroll across and down the spreadsheet using the horizontal and vertical scroll bars, respectively. Note there is even a provision to have multiple pages of spreadsheets that can be thought of as a 3D spreadsheet. Finally, the various icons and menu items that appear above and below this particular spreadsheet are idiosyncratic and can be changed at will to reflect your own personal preferences chosen among hundreds of different capabilities.

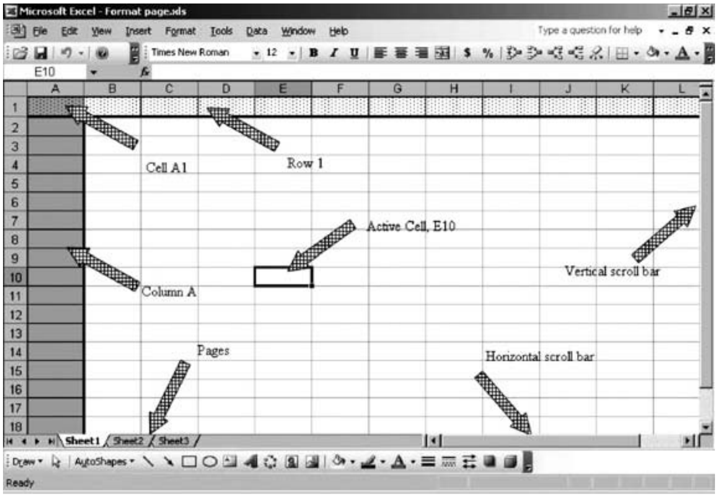


FIGURE 3.1 Typical Spreadsheet

3.2.1 Cell Addressing Modes

The window that you view is just a small fraction of the whole spreadsheet; the remaining virtual spreadsheet has thousands of rows and hundreds of columns of available space as well as many interactive sheets per workbook, thus enabling 3D addressing. Movement within a window is by the scroll bars either across or down. Movement to virtual positions can also be effected by moving the cursor to the boundary you wish to extend.

Types of data that can be in a cell

- Alphanumeric—text
  - Headings
  - Labels
- Numeric
  - Numbers: integers, floating point (meaning with decimals or exponents)
  - References to the contents of other cells (using the cell address)
- Equations
  - Formulae such as = B1 + B4/H13 add the contents of cell B1 to the result of dividing the contents of cell B4 by the contents of cell H13.
- Predefined functions
  - Sum
  - Average
  - Max
  - Min
- Many other functions (trigonometric, exponential, logical, statistical, etc.)

You will need to explicitly tell the computer what data type is to be stored in a cell:

- The default cell entry is text or plain numbers; text is left justified and numbers are right justified. (You can change these using the justifying pictographs in the heading rows.)
- If you put a single quotation mark (‘) at the beginning of your cell entry that means left-justified alphanumeric to follow.
- If you put an equal sign (=) at the beginning of your cell entry, that means equation or numeric to follow.
- If you put a plus (+) or minus (−) at the beginning of your cell entry, that means numeric to follow.[3]

Example 3.4

FleetsR'US owns a fleet of rental cars; it wants to offer its customers an inclusive all-in-one-contract that takes account of the cost of the gasoline used. To estimate the cost, and therefore, pricing to their customers, they do a survey of the typical renter's journey and determine how many miles are driven by a selection of typical drivers. The raw data are given in the following table.

Renter	City Miles	Suburban Miles	Highway Miles
Davis	16.0	28.0	79.0
Graham	10.0	31.0	112.
Washington	4.00	7.00	158.
Meyers	22.0	61.0	87.0
Richardson	12.0	56.0	198.
Thomas	5.00	22.0	124.
Williams	4.00	14.0	142.

[3]But "+ name" means a specially defined macro or independent program.

Using a spreadsheet, give the total miles in each category of driving by each driver and the average miles driven in each category. Show the algorithm used in the spreadsheet.

Need: Driver mileage and average miles driven in each segment of renter's journey.
Know: Mileage data in table.
How: Total miles for each driver is the sum of the individual miles driven and the average mileage in each category of driving is the total miles divided by the number of drivers.
Solve: Use a spreadsheet to input the data table and the indicated arithmetical operations.

Table 3.1 Output of Spreadsheet Corresponding to Example 3.4

	A	B	C	D	E
1	**Miles driven in various categories**				
2					
3					
4	**Renter**	**City Miles**	**Suburban Miles**	**Highway Miles**	**Total Miles**
5	Davis	16.0	28.0	79.0	123
6	Graham	10.0	31.0	112	153
7	Washington	4.00	7.00	158	169
8	Meyers	22.0	61.0	87.0	170.
9	Richardson	12.0	56.0	198	266
10	Thomas	5.00	22.0	124	151
11	Williams	4.00	14.0	142	160.
12					
13	Average	10.4	31.3	129.	172

The sheet title heading conveniently appears in cell A1 (even though the actual input sprawls across adjacent cells). It helps in organizing the spreadsheet to entitle it descriptively so that it can be read by you (or others) later when the purpose of the spreadsheet has faded from immediate memory. In addition to the preceding spreadsheet, descriptive labels for some columns are given in row 4 as A4:E4. (Read the colon to mean it includes all the cells in the rectangular area prescribed by the limits of the addresses, A4 and E4 in this case.) The labels may be simple such as these are, or they may be informative such as the units involved in the column (or row) that follows. Other descriptive labels that appear in this spreadsheet are A5:A11 and A13. All of these are **text** statements. The labels are the first steps in interpreting data as more than mere entries. Labels are there to codify what you want to arrange as information.

The numerical input data for this example are given in the block B5:D11. The results of some arithmetical operations appear as row E5:E11 and as B13:E13. These data are the beginnings of information about the journeys. Note the variables are formatted to the correct number of significant figures. (The formatting structure appears under the headings “Format, Cells, Number”; it is highly recommended you experiment with fonts, borders, alignment, and number formats.) A block of cells can be highlighted before each operation and the formatting applied *en masse*.

Table 3.2 shows the mathematical operations behind Table 3.1 and what’s in each cell. Toggle these from the spreadsheet of Table 3.1 by simultaneously pressing the control key (^) and a tilde (~). This is written in shorthand as ^~. This command also deletes decimal points and trailing zeros, which masks the number of significant figures present.

Table 3.2 Expanded Spreadsheet from Table 3.1 Using ^~ (control-tilde)

	A	B	C	D	E
1	**Miles driven**				
2					
3					
4	**Renter**	**City Miles**	**Suburban Miles**	**Highway Miles**	**Total Miles**
5	Davis	16	28	79	=SUM(B5:D5)
6	Graham	10	31	112	=SUM(B6:D6)
7	Washington	4	7	158	=SUM(B7:D7)
8	Meyers	22	61	87	=SUM(B8:D8)
9	Richardson	12	56	198	=SUM(B9:D9)
10	Thomas	5	22	124	=SUM(B10:D10)
11	Williams	4	14	142	=SUM(B11:D11)
12					
13	Average	=AVERAGE(B5:B11)	=AVERAGE(C5:C11)	=AVERAGE(D5:D11)	=AVERAGE(E5:E11)

Note the actual arithmetical operations now appear in cells E5:E11 and B13:E13. We declared them as such by inputting the mathematical operation after an = sign. The mathematical operations are the sum and the average, both with the specified ranges. Finally, notice in this view that the individual formatting for each cell has been removed and the actual input data as typed are revealed, while the long statement that appeared in cell A1 has now been truncated.

FleetsR’US now realizes that the gas mileage in each segment of the journey is different; the length of the journey is less important than the gallons used in each segment. It also realizes that gas mileage is a moving target; vehicles can get both more and less fuel efficiency (a large SUV vs. a compact hybrid vehicle, for example). It needs to keep its options open and to be able to update the spreadsheet when the market changes. Spreadsheets have a very powerful way of showing you how to do this. It introduces a powerful new tool: **absolute and relative cell addressing**. An absolute address is preceded by a $ sign for either or both rows and columns; relative addressing is the default mode.

Example 3.5

FleetsR'US wants to add the gas mileage to their spreadsheet in a flexible manner so it can be updated later if needed. It wants to now know the gallons used per trip and the average used by their typical drivers. Here are the miles per gallon (mpg) by journey segment:

	City	Suburbs	Highway
mpg	12	18	26

Need: Average gallons per journey segment and average per driver.
Know: Mileage data from Example 3.1 plus mpg just given.
How: Divide mileage by mpg to get gallons used in each segment of the journey.
Solve: Here are several ways we can reprogram our spreadsheet shown in Table 3.3. (We will just look at the spreadsheet mathematics using our ^~ method. Be assured that each cell *will* give the correct answer.)

What, if anything, is wrong with these solutions? To start with, cells B5:D5 do exactly what we have been asked, but they are *inflexible* should we later want to enter different mpg, since we would have to hand-input them to every affected cell (here just three, but possibly tens of thousands or more).

Table 3.3 shows that we can use a constant, such as the contents of cell B17, to divide each cell in column B; then, when we again wish to change, all we subsequently do is change the number in the cell B17, and it will be updated wherever it is used. This is a powerful improvement over entering the data in individual cells.

Table 3.3 Constants and Absolute Addressing

	A	B	C	D	E
4	**Renter**	**City Gallons**	**Suburban Gallons**	**Highway Gallons**	**Total Gallons**
5	Davis	=16/12	=28/18	=79/26	=SUM(B5:D5)
6	Graham	=10/B17	=31/$C$17	=112/D$17	=SUM(B6:D6)
7	Washington	=4/B17	=7/$C$17	=158/D$17	=SUM(B7:D7)
8	Meyers	=22/B17	=61/$C$17	=87/D$17	=SUM(B8:D8)
9	Richardson	=12/B17	=56/$C$17	=198/$D17	=SUM(B9:D9)
10	Thomas	=5/B17	=22/$C$17	=124/$D17	=SUM(B10:D10)
11	Williams	=4/B17	=14/$C$17	=142/$D17	=SUM(B11:D11)
12					
13	Average	=AVERAGE(B5:B11)	=AVERAGE(C5:C11)	=AVERAGE(D5:D11)	=AVERAGE(E5:E11)
14					
15					
16		**City**	**Suburbs**	**Highway**	
17	**mpg**	12	18	26	

We can also change the weightings of the cells to reflect mpg in columns C and D. We'll use a slightly different method. In fact, we will fix the contents of the cell C17 in a tricky way and indicate that as $C$17. Such an address, for reasons seen shortly, is called "absolute" addressing. It will be used here to modify cells C6:C11; we can also use as a constant either D$17 or $D17 (called a "mixed" mode of relative and absolute addressing) to modify the cells in D6:D11. What these do is *fix* the reference to the row or number operated on by the "$" sign. What does this mean and why do it? The importance of these operations is best explained by copying and pasting a section of the spreadsheet as in the next example.

Example 3.6

FleetsR'US wants to copy part of its spreadsheet to another area on the spreadsheet to make it clearer for someone to read. Assume the origin is no longer cell A1, but shift the whole spreadsheet to a new origin, H1. Analyze what you see and report whether the translocated spreadsheet is correct or not; if not, explain what is wrong.

Need: Copied spreadsheet
Know: You can copy and paste across the spreadsheet form.
How: Use Edit, Copy, Paste commands.
Solve: Suppose you want to copy the table to another part of the spreadsheet. All you do is highlight the block of information A1:E13 that you want to move and go to Edit, then Copy.[4] Move your cursor to the corner cell you want to move to, say cell H1, and then go to Edit, Paste. Observe the effects after this is done in Table 3.4.

Table 3.4 Effect of Shifting the Spreadsheet of Table 3.3

	H	I	J	K	L
1	**Miles driven in various categories**				
2					
3					
4	**Renter**	**City Gallons**	**Suburban Gallons**	**Highway Gallons**	**Total Gallons**
5	Davis	1.33	1.56	3.04	5.93
6	Graham	#DIV/0!	1.72	#DIV/0!	#DIV/0!
7	Washington	#DIV/0!	0.39	#DIV/0!	#DIV/0!
8	Meyers	#DIV/0!	3.39	#DIV/0!	#DIV/0!
9	Richardson	#DIV/0!	3.11	7.62	#DIV/0!
10	Thomas	#DIV/0!	1.22	4.77	#DIV/0!
11	Williams	#DIV/0!	0.78	5.46	#DIV/0!
12					
13	Average	#DIV/0!	1.74	#DIV/0!	#DIV/0!

[4]There are alternate ways of copying and pasting using smart tags; however, the method suggested here is conceptually easier for a beginner in spreadsheet manipulations.

Notice we did not get all the expected results: Instead of modifying cells to the correct answers, cells I6:I11 and K6:K8 give the wrong answers. (Other divisions by zero also occur, since the spreadsheet's mathematical functions of sum and average also reference cells with one or more zero divisions.) What has happened? Use the toggle ^~ again and observe the results in Table 3.5.

Table 3.5 Contents of Cells after Moving Location

	H	I	J	K	L
4	**Renter**	**City Gallons**	**Suburban Gallons**	**Highway Gallons**	**Total Gallons**
5	Davis	=16/12	=28/18	=79/26	=SUM(I5:K5)
6	Graham	=10/117	=31/$C$17	=112/K$17	=SUM(I6:K6)
7	Washington	=4/117	=7/$C$17	=158/K$17	=SUM(I7:K7)
8	Meyers	=22/117	=61/$C$17	=87/K$17	=SUM(I8:K8)
9	Richardson	=12/117	=56/$C$17	=198/$D17	=SUM(I9:K9)
10	Thomas	=5/117	=22/$C$17	=124/$D17	=SUM(I10:K10)
11	Williams	=4/117	=14/$C$17	=142/$D17	=SUM(I11:K11)
12					
13	Average	=AVERAGE(15:111)	=AVERAGE(J5:J11)	=AVERAGE(K5:K11)	=AVERAGE(L5:L11)

Of course, the manually transplanted input cells I5:K5 are correct, since they call for no information outside of the particular cells, but the entries to I6:I11 are divided, not by B17 but by I17. In other words, when we translocated the cells, we also shifted B17 relative to the move (i.e., cell B17 is now displaced to I17 as is cell A1 is to H1). And cell I17 (not shown) contains a default empty cell value of zero—hence, the apparent disaster that the gallons used in city driving are now all infinite!

Notice that cells J6:J11 are multiplied by the contents of cell $C$17 as desired, since we used the $ designator for both row and column. Cell $C$17 has not translocated relative to the move; its address is thus absolute.

The sundry results of the translocated highway mileage column K is explained by the mixed mode of addressing used there. Cells K6:K8 now have been divided by K$17, since the column designator translocated from D to K (again, just as did column A to H). Since we fixed the row designator at $17, it did not translocate to a new row. But the cell K$17 contains a default value of zero and hence the cells K6:K8, all of which contain this as a divisor, are each infinite.

However, cells K9:K11 gave the correct answer because they were divided by $D17. The $ fixed the column designator at column $D. We did not choose to change the rows when we copied the block of cells from A1 to H1, and the divisor $D17 still refers to the original mpg in cell D17. We simply took advantage of our simple horizontal move from A1 to H1 in that it did not change the row locations. Had we also changed the row location—say, by pasting the block at A1:E13 to H2 instead of H1—the rest of highway gallons calculations would also have been in error due to an imputed mpg call to cell D18.

Sometimes we want to use relative, absolute, or mixed modes. It depends on what we want to do. With experience, your ability to do this will improve. One nice feature of spreadsheets is that your mistakes are usually embarrassingly and immediately clear. If, for example, you forget to fix an absolute cell address and add, multiply, divide, exponentiate, and so on, by some unintended cell, the worksheet will complain as the target cell picks up whatever cell values were in the inadvertent cells.

Example 3.7

Correct Example 3.6 using relative and absolute addressing modes.

Need: Corrected spreadsheet.
Know: How to use different cell addressing modes.
How: Make sure that no translocation cell addresses occur to undesired cells.
Solve: Had all the cells B5:D11 row by row been divided by their respective absolute constants, $B$17:$D$17, the final translocated result would have been as initially desired—see Table 3.6.

Table 3.6 Results of Spreadsheet That Used Absolute Cell Constants

	H	I	J	K	L
1	**Miles driven in various categories**				
2					
3					
4	**Renter**	**City Gallons**	**Suburban Gallons**	**Highway Gallons**	**Total Gallons**
5	Davis	1.33	1.56	3.04	5.93
6	Graham	0.833	1.72	4.31	6.86
7	Washington	0.333	0.389	6.08	6.80
8	Meyers	1.83	3.39	3.35	8.57
9	Richardson	1.00	3.11	7.62	11.7
10	Thomas	0.417	1.22	4.77	6.41
11	Williams	0.333	0.778	5.46	6.57
12					
13	Average	0.869	1.74	4.95	7.55

3.3 GRAPHING IN SPREADSHEETS

One of the big advances of spreadsheeting is that you can display your results in many ways that can visually convey a great deal of information. In a real sense most of your knowledge will be accessible only then, since the raw data, even if arranged as logical information in neat tables, may still be difficult to interpret.

We will confine ourselves for the moment to simple Cartesian graphing, but you will soon recognize the pattern to follow if you want to use other modes of display.

Example 3.8

SpeedsR'US wants to sell after-market booster kits for cars. They measure a booster-modified vehicle that has the following characteristics on a level road:

Time, s	Mph Actual	Time, s	Mph Actual
0.00	0.00	5.0	78.6
1.00	18.6	6.0	86.9
2.00	32.1	7.0	96.9
3.00	46.5	8.0	103
4.00	58.2		

The car manufacturer says that the *un*modified car speed characteristic on a level road obeys the following equation: mph $= 136.5(1 - e^{-0.158t})$ with t in seconds. Is the modified car faster to 100. mph than the standard car?

Need: To compare the tabular data to the equation given.
Know: A spreadsheet will provide multiple ways of looking at data.
How: Graph the tabular data and compare to the theoretical equation.
Solve: Set up the spreadsheet in Table 3.7.

Table 3.7 Set Up for Graphing Function

	A	B	C	D
1	**Compare measured car speed to manufacturer's specification**			
2				
3	Speed in mph $= 136.5(1 - e^{-0.158t})$ with t in seconds			
4				
5				
6	**Time, s**	**Mph Actual**	**Mph, Theory**	
7	0	0	=136.5 *(1-EXP(-0.158*A7))	
8	=A7+1	18.6	=136.5 *(1-EXP(-0.158*A8))	
9	=A8+1	32.1	=136.5 *(1-EXP(-0.158*A9))	
10	=A9+1	46.5	=136.5 *(1-EXP(-0.158*A10))	
11	=A10+1	58.2	=136.5 *(1-EXP(-0.158*A11))	
12	=A11+1	78.6	=136.5 *(1-EXP(-0.158*A12))	
13	=A12+1	86.9	=136.5 *(1-EXP(-0.158*A13))	
14	=A13+1	96.9	=136.5 *(1-EXP(-0.158*A14))	
15	=A14+1	103	=136.5 *(1-EXP(-0.158*A15))	

Start with a spreadsheet title as in cell A1. Make it descriptive. The cell labels in A6:C6 should be helpful as to what the immediate columns under them mean. Of course, the numbers generated as A7:C15 are just raw data—not easy to interpret by merely looking at them. In columns A, B, and C are the data you want to plot—that is, the value of the variable mph given as input data and calculated by the manufacturer's equation.

You don't have to enter too much to get a lot of mileage.[5] In cells A7:B7, enter 0, and in C7, enter 136.5*(1 − exp (−0.158*A7). The result in C7 is zero, since $1 - e^0 = 1 - 1$ is zero. Now in A8 write = A7 + 1. The value that appears in A8 will be 1, since A7 is zero. Now copy A8 to A9:A15, and then copy C7 to C8:C15, and you will have the manufacturer's function defined corresponding to each time of measurement.

You might want to graph your results, A7:A15 vs. B7:B15 and also A7:A15 vs. C7:C15. No sweat: Just highlight A7:C15 and then go to Insert, Chart. Then tell the spreadsheet what kind of graph you want. You want an x-y (Scatter) plot. ("Scatter" is just a regular Cartesian graph!) You will see two sets of points corresponding to each column of speed data. Then click Next and fill in your preferred titles. Optionally toggle off the Legend box. You can finish by pasting the graph on the current worksheet (i.e., the working page in the spreadsheet). If you want, you can highlight either series of data and you can format them to discrete points or to a continuous line.

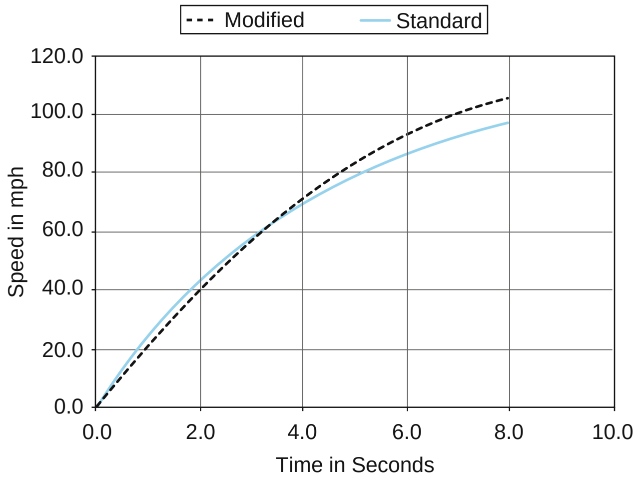


FIGURE 3.2 SpeedsR'US Modification

You should get the graph in Figure 3.2 for your troubles. Spreadsheets leave you plenty of flexibility if you use your imagination. On the basis of Figure 3.2, has the speed modification achieved its goal?

SUMMARY

A problem-solving method that guarantees you a path to the answer to engineering solutions is a valuable addition to an engineer's tool kit; we recommend the self-prompting mnemonic approach we call the **need–know–how–solve** method. It allows for a systematic approach to engineering problems. We suggest

[5]As previously noted, you can short-circuit much of this discussion by taking advantage of smart tags; these are pull-downs on the cells that will automatically copy or increment adjacent cells. See the Help menu for your particular spreadsheet to see if this use is supported.

you couple this method with the units method, [..], that we introduced in Chapter 2. Between these two methods you should be able to reason your way through even arcane problems that appear impossible at first glance. In addition, the systematic approach means you leave an auditing trail that can be used in your absence to check on, or to extend, or even to correct, your solution.

Many solutions to engineering problems require either or both complex analysis and repetitive solutions for many cases; still others require graphical representation. In other cases we may be confronted with jumbles of apparently uncorrelated **data**. Spreadsheeting allows you to organize these data and sort them so you have some understanding of gross **information**. Finally interpret the data to gain **knowledge** from the model you have developed. These situations are where **spreadsheet analysis** stands out. The heart of this method is to recognize that cells in a spreadsheet may be manipulated either in an **absolute** sense or in a **relative** sense. The former allows for facile repetition of algebraic and arithmetical operations using constants, and the latter allows us to easily extend tables of data, even while carrying their embedded mathematical operations.

Together, with an appreciation of the significance of engineering numbers and dimensional analysis, the contents of Chapters 2 and 3 are fundamental to what an engineer does and how successful he or she will be in pursuing an engineering career.

This lesson on spreadsheeting merely skims the surface of the very powerful tool of spreadsheet analysis. With time, both as a beginning engineer and then as a practicing engineer, you will discover many of the additional features of this tool. Used with skill and precision, this will become one of the most useful and generic skills that you will acquire. The sooner you make use of it, the sooner you will acquire these skills. They are urged upon you and you will find practice problems in virtually every chapter of this text that will be susceptible to spreadsheet analysis.

EXERCISES

Many of these exercises can be done using hand calculators; alternatively, you can also use a spreadsheet even for simple problems. *If you choose a spreadsheet, then for all your spreadsheet exercises for this course, go to File, then Print Setup, then Sheet, and click the boxes for Gridlines and for Row and Columns Headings so that your final cell locations will be printed when you submit the answers to these problems. You will lose points otherwise*, since it is nearly impossible to grade answers without the row and column indicators!

Use the need–know–how–solve method in setting up all these problems. *You may lose points otherwise*.

1. Suppose the ranch in Example 3.1 was a circle instead of a square. Using the same financial information ($32,000. available funds, $10,000. a mile for fence, and $100,000. per square mile land cost), what would be the diameter of the ranch?

2. Suppose the ranch in Example 3.1 was an equilateral triangle instead of a square. Using the same financial information ($32,000. available funds, $10,000. a mile for fence, and $100,000. per square mile land cost), what would be the length of one side of the ranch?

3. Suppose the ranch in Example 3.1 was a rectangle with the long side twice as long as the short side. Using the same financial information ($32,000. available funds, $10,000. a mile for fence, and $100,000. per square mile land cost), what would be the length of the short side of the ranch?

4. The great physicist Enrico Fermi used to test the problem-solving ability of his students at the University of Chicago by giving them the following problem: How many piano tuners are there in the city of Chicago? (Assume the population of Chicago is 5 million people.) (**Plausible answers: 50–250 tuners**[6])

[6]See http://www.grc.nasa.gov/WWW/K12/Numbers/Math/Mathematical_Thinking/fermis_piano_tuner.htmS

5. Of all the rectangles that have an area of one square meter, what are the dimensions (length and width) of the one that has the smallest perimeter? Solve by graphing on a spreadsheet.

6. Suppose you want to make a cylindrical can to hold 0.01 m^3 of soup. The sheet steel for the can costs \$0.01/m^2. It costs \$0.02/m to seal circular pieces to the top and bottom of the can and along the seam. What is the cost of the cylindrical can that is least expensive to make? (**A: 3 cents/can**)

7. Suppose the mass used in Example 3.3 was increased from 10.0 kg to 20.0 kg, and the wire stretched by twice as much. If the 20.0 kg mass was then used to stretch a 4.00 m piece of the same steel wire, how much will it stretch?

8. Use the spreadsheet analysis in Example 3.4 to determine the total miles driven by each driver and the average miles driven in each category for the following car renters:

Renter	City Miles	Suburban Miles	Highway Miles
Geske	35	57	93
Pollack	27	11	275
Loth	14	43	159
Sommerfeld	12	31	305
Thunes	22	16	132
Lu	5.0	21	417

9. Using the renter mileage information given in the previous exercise and the miles per gallon information given in Example 3.5, determine the average gallons per journey segment and average per driver for this set of drivers.

10. Using the technique introduced in Example 3.8, create a spreadsheet graph of the following data for the median annual salaries in dollars for engineers based on years of experience, supervisory responsibility, and level of education.[7]

	Number of Years After BS Degree					
	0	5	10	15	25	35
Nonsupervisory						
B.S.	55,341	63,649	73,162	80,207	85,116	92,748
M.S.	—	79,875	86,868	90,134	97,463	110,289
Ph.D.	—	—	91,352	98,053	108,747	122,886
Supervisory						
B.S.	—	—	72,632	80,739	92,029	107,844
M.S.	—	99,367	109,450	110,360	113,916	117,146
Ph.D.	—	—	—	110,877	132,800	147,517

Exercises 11–13 involve the following situation:

Suppose that the weight of the gasoline in lbf in a car's gas tank equaled the weight of the car in lbf to the 2/3 power (i.e., if G = gasoline weight in lbf, and W = car weight in lbf, then $G = W^{2/3}$). Assume that

[7]These data are from the 2007 report of the Engineering Workforce Commission of the American Association of Engineering Societies (AAES).

gasoline weighs 8.0 lbf/gal, and gas mileage, measured in miles/gallon or mpg, varies with weight according to the empirical formula:

$$\text{Gas mileage in mpg} = (84{,}500 \text{ mpg} \cdot \text{lbf}) \div (\text{car's weight in lbf}) - 3.0 \text{ mpg}$$

11. What is the fuel usage in miles per gallon of a 3.00×10^3 lbf car?

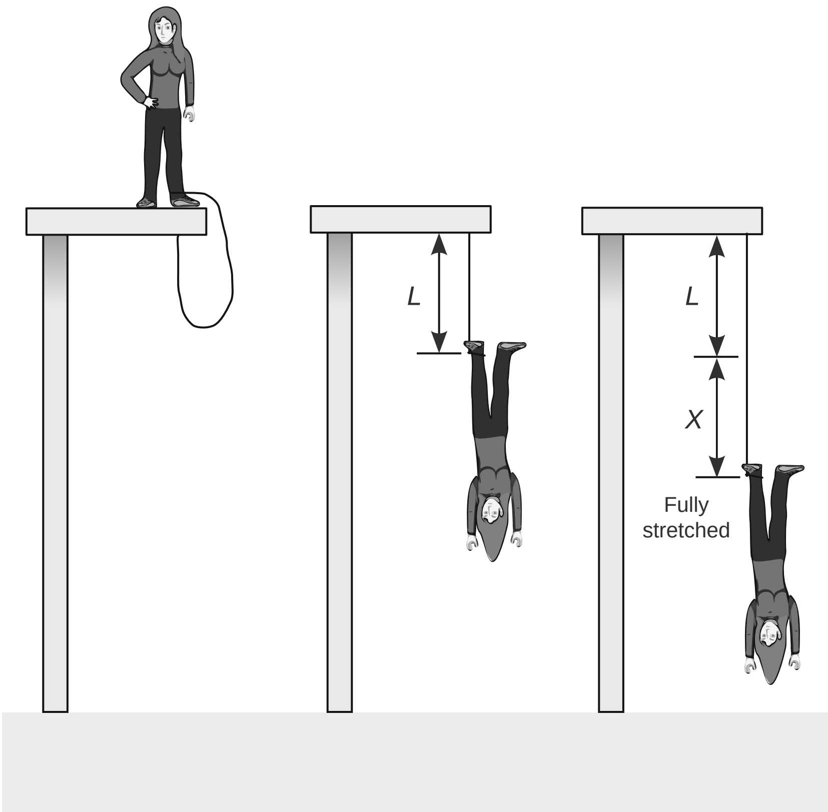
12. What is the heaviest car that can achieve a range of 600. miles? (**A: 3.69×10^3 lbf**)

13. Suppose the formula for weight of the gas was $G = W^b$, where b can be varied in the range 0.50–0.75. Graph the range of a 3.69×10^3 lbf car as a function of b.

Exercises 14 through 17 are concerned with bungee jumping as displayed in the figure. At full stretch, the elastic rope of original length L stretches to $L + x$. For a person whose weight is W lbf and a cord with a stiffness K lbf/ft, the extension x is given by this formula:

$$x = \frac{W}{K} + \sqrt{\frac{W^2}{K^2} + \frac{2W \times L}{K}}$$

and that can be written in spreadsheet script as $x = W/K + \text{sqrt}(W\text{^}2/K\text{^}2 + 2 * W * L/K)$.



14. If the height of the cliff is 150.0 ft, $K = 6.25$ lbf/ft, $L = 40.0$ ft, and the person's weight is 150.0 lbf, will the person be able to bungee jump safely? Support your answer by giving the final value for length $= L + x$. (**A: 114 ft < 150 ft, OK**)

15. Americans are getting heavier. What's the jumper's weight limit for a 40.0 ft unstretched bungee with stiffness of $K = 6.25$ lbf/ft? Graph final length $L + x$ vs. W for weights from 100. lbf to 300. lbf in increments of 25 lbf. Print a warning if the jumper is too heavy for a 150. ft initial height. (**Hint**: Look up the application of the IF statement in your spreadsheet program.)

16. If the height of the parapet is 200. ft, and the weight of the person is 150. lbf, and the unstretched length $L = 45.0$ ft, find a value of K that enables this person to stop exactly five feet above the ground. (**A: 2.60 lbf/ft**)

17. By copying and pasting your spreadsheet from exercise 15, find and plot the values of L needed (in ft) vs. W, weight of jumper (in lbf) for successful bungee jumps (coming to a stop 5.0 ft above the ground) for $K = 6.25$ lbf/ft and from a cliff of height 150. ft above the ground. The graph should cover weights from 100. lbf to 300. lbf in increments of 25 lbf. (**Hint**: The function Goal seek under Tools is one way to solve this exercise.)

For **Exercises 18 through 20**: The fixed costs per year of operating an automobile is approximately 20.% of the initial price of the car. Thus, the operating cost/mile = 0.20/yr × (purchase price of automobile)/(miles driven per year) + (price of gasoline/gallon) × (gallons used per mile). In the exercises that follow, assume that the automobile is driven 2.0×10^4 miles per year. Assume gasoline costs $5.00/gallon.

18. Estimate the operating cost per mile of an automobile with a price of $15,000 that gets 30.0 miles per gallon. (**A: $0.32/mile**)

19. If we were to double the price of the automobile in exercise 18, what would its gas mileage have to be in order to cost the same to operate per mile as the automobile in exercise 18?

20. Suppose that the purchase price of automobiles varies with weight according to the formula that weight in lbf × $8.00, and gas mileage varies according to mpg = (84,500 mile-lbf/gal)/W – 3.0 miles/gal. Graph the cost per mile of operating a car as a function of the car's weight, in increments of 500. lbf from 2000. lbf to 5000. lbf. (**Partial A: 29 cents/mile for 2000. lbf car and 76 cents/mile for 5000. lbf car**)

For **Exercises 21 through 23**: In visiting stores, we find the following prices for various things. Broccoli crowns cost $2.89 per pound. Soft drinks cost $2.00 per two-liter bottle. (A liter is 0.001 m^3.) A new automobile weighs 2.50×10^3 lbf and costs $$1.50 \times 10^4$. A dozen oranges, each of which is 0.06 m in diameter, costs $2.05. A 1.5 lb package of chicken thighs costs $5.35. A dictionary weighs 5.00 pounds and costs $20. A refrigerator weighs 200. lbf and costs $900. Assume that one cubic meter of any solid object or liquid weighs 1.00×10^4 N.

21. For the objects listed in the previous exercise, make a table and graph of the cost of objects in dollars as a function of their weight in newtons. It is suggested for this graph to use a *line* (not scatter) graph using lines with markers displayed at each data value. (Get rid of the unwanted line using the format series function.) The value of the line graph is that everything plotted is at the same horizontal displacement and not dependent on its value.

22. What (perhaps surprising) simple generalization about the cost of things might we make based on the table and graph of exercise 21? (**Hint**: Does the comparative lack of spread in price surprise you?)

23. Name a product or group of products that does *not* fit the generalization you made in exercise 22 and add and label the point on the graph in exercise 21. To get a better perspective use a *log scale* for the y-axis, \$/N, since the ordinate should be much larger than the coordinates of those points from exercise 21.

24. An unnamed country has the following population of passenger cars on its roads as determined by 250 kg mass differences. You have to make these data clear to the undersecretary to that country's transport minister. Plot these data by two methods: (1) as a pie chart and (2) as a histogram to show the distribution in an effective manner.

	B	C
1	Upper limit, kg	% all vehicles
2	1000.	12.1
3	1250.	13.1
4	1500.	15.4
5	1750.	18.6
6	2000.	14.8
7	2250.	9.20
8	2500.	7.50
9	2750.	6.30
10	4000.	3.00

Exercises 25 through 27 deal with Hubbert's Peak.[8] This is a model of supply and demand for oil. It looks at the amount of available oil and its rate of consumption to draw conclusions about continuing the current course of our oil-based economy.

25. Suppose the world originally had three trillion ($3. \times 10^{12}$) barrels of oil and its exploration began in 1850. Suppose 10.0% of the remaining *undiscovered* oil has been found in every quarter century since 1850. Call the *discovered* but not yet consumed oil, *reserves*. Suppose oil consumption was 1.0×10^{8} barrels in 1850, and further suppose oil consumption has grown by a factor of 5 in every quarter-century since 1850. When will the oil start to run out? (That is, when will the reserves become negative?) Give your answer to the nearest 25 years and provide a spreadsheet showing reserves and consumption as a factor. (**A: 2025**)

26. Suppose the world originally had 10 trillion (10.0×10^{12}) barrels of oil. Use the data of exercise 25 to predict again when the oil will start to run out.

27. Repeat exercise 26, but instead of assuming the exponential growth in consumption continuing unabated by a factor of 5 in every quarter-century since 1850, curtail growth since 2000. and assume consumption has stayed constant since then. Again predict when the oil will start to run out.

28. Your friend tells you that the need–know–how–solve problem-solving method seems overly complicated. He or she just wants to find the answer to the problem in the quickest possible way—say, by finding some formula in the text and plugging numbers into it. What do you tell him or her?

[8]M. King Hubbert was a geologist with Shell Oil who, in the 1950s, pointed out that the US supply of oil was going to fall short of demand by the 1970s, as it did. His methods have since been applied to world oil production and, based on demand exceeding production, predict an ongoing oil supply crisis. See http://www.hubbertpeak.com/hubbert/

a. Go ahead and do whatever you want, then you'll flunk out, and I'll survive.
b. Talk to the instructor and have him or her explain why this methodology works.
c. Find someone who has used this method and ask to copy his or her homework.
d. Explain why this technique will lead to a fail safe method of getting the correct answer.

29. You e-mail a classmate in this course for some information about a spreadsheet homework problem. In addition to answering your question, your classmate also attaches a spreadsheet solution to the homework. What do you do?
a. Delete the spreadsheet without looking at it.
b. Look at the spreadsheet to make sure she did it correctly.
c. Copy the spreadsheet into your homework and change the formatting so that it doesn't look like the original.
d. E-mail the spreadsheet to all your friends so that they can have the solution, too.

30. Stephanie knew Adam, the Environmental Manager, would not be pleased with her report on the chemical spill. The data clearly indicated that the spill was large enough that regulations required it to be reported to the state. When Stephanie presented her report to Adam, he lost his temper. "A few gallons over the limit isn't worth the time it's going to take to fill out those damned forms. Go back to your desk and rework those numbers until it comes out right." What should Stephanie do?[9]
a. Tell Adam that she will not knowingly violate state law, and threaten to quit.
b. Comply with Adam's request since he is in charge and will suffer any consequences.
c. Send an anonymous report to the state documenting the violation.
d. Go over Adam's head and speak to his supervisor about the problem.

[9] Abstracted from Engineering Ethics: Concepts and Cases at http://wadsworth.com/philosophy_d/templates/student_resources/0534605796_harris/cases/Cases.htm

CHAPTER

Energy: Kinds, Conversion, and Conservation

4

Source: http://www.beckersbakeryanddeli.com/picture.htm

4.1 USING ENERGY

Is the world running out of energy? Where does the energy come from that provides the light by which you are reading these words? Does a car possess more energy when it is sitting in your driveway, or 15 minutes later when it is traveling down the highway at 60 miles per hour? In this chapter, you will learn how to answer these questions, as well as how to address the more quantitative ones involving energy that engineers of all types encounter in their work.

Energy is *the capability to do work*. An important point is that energy comes in several discrete kinds. These kinds are capable of conversion from one kind to another. In the course of these conversions, energy is neither created nor destroyed, but rather conserved. That is, the total amount of energy in the universe remains constant. This is one of the most important principles you will have to master as an engineer. You will have to construct a useful system boundary known as a **control boundary**, to monitor the flows of energy.

Using energy as a variable, engineers create models that help with a wide range of applications, such as deciding what fuel to use in an automobile, designing methods of protecting buildings from earthquakes or lightning strikes, or advising citizens as to the technical capability of a society to replace fossil fuels with renewable resources such as solar power, wind power, and biomass, or with controversial but abundant energy sources such as uranium. But in order to achieve these wide-ranging applications, engineers must be very specific in the way they understand and use the concept of energy.

The way engineers use the word *energy* is examined in this chapter. As already stated, energy is the capability to do work. Work is narrowly defined based on Newton's Second Law of Motion because it introduced the notion of a force. In the simplest mathematical terms, work (symbol W) is defined as:

$$Work = W = F \times d \tag{4.1}$$

in which a force F moves through a distance d. The concept is broader than it appears from this straightforward definition and can be applied to a number of situations apparently far removed from this simple statement, while still inherently dependent on it. Energy, E, has the same units as work, W. In an imaginary system we might convert energy into work and then recover an equivalent amount to the original energy.

As noted, the different kinds of energy are capable of conversion from one kind to another and, as such, are a core component of the responsibility of several kinds of engineers. In the course of these conversions, energy is neither created nor destroyed but conserved. That is, the total amount of energy in the universe remains constant.[1]

Figure 4.1 represents one of the most important principles you will have to master as an engineer. The engineer uses something that we can generically call a **system**. It might be an engine, a sailboat, a lawn mower, an electric kettle, a crane, or anything. It is that which is contained within the inner dotted boundary. It connects to the rest of the universe through whatever elements are interposed between the system and the universe. You will have to use engineering skills and know-how to construct a useful system boundary, known as a **control boundary**, such that you can monitor the flows of energy (perhaps of several kinds) across it. When you do this, the **Law of Conservation of Energy** will enable you to determine the flows of energy necessary to maintain the balance.

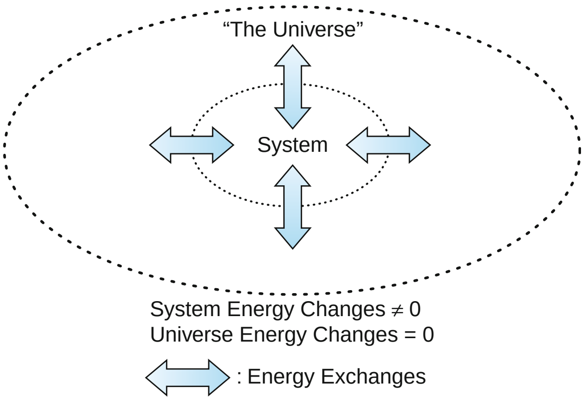


FIGURE 4.1 Conservation of Energy

When engineers use the term *energy conservation* they are not offering exhortations to turn off the lights, nor are they opposing the use of gas-guzzling sport utility vehicles. Rather, they are expressing a scientific principle that is fundamental to engineering and that will remain in effect regardless of the energy policies that people and nations choose to follow. If you remember only three words from this chapter remember this: *Energy is conserved*.

As a concrete example of what the engineers do in their analyses, we can divide the universe into at least two discrete regions: one of direct interest (the system, for example, a car's engine) and everything else. This is a very powerful tool in the hands of a capable engineer, since it allows the engineer to concentrate on the localized system of interest.

4.2 ENERGY IS THE CAPABILITY TO DO WORK

Understanding the variable energy begins by defining variables for the concepts of force and work. Force is a variable that may be thought of as a push or a pull, as measured with a spring balance or by the weight of an object in a gravitational field. As we have already seen, force has the units of newtons in the SI system and pounds force (lbf) in the Engineering English system. For example, at sea level, to lift a book of mass 1.0 kilogram requires a force of 9.8 newtons. In the Engineering English system, the force to lift this 1.0 kg book is 2.2 lbf (since 1.0 kg = 2.2 lbm, and 2.2 lbm "weighs" 2.2 lbf at standard gravity).

[1]There is one caveat that cosmological physicists might add: the radiant energy (which moves at the speed of light) in the universe may leak to beyond the observable edge of the expanding universe—something that is unlikely to disturb any engineers in the foreseeable future! For all practical purposes, the energy in the universe is indeed constant.

Work as defined earlier is the product of force and the distance over which the force is applied, where the distance is measured in the same direction as the application of the force. The units of work are **joules** in the SI system, and **foot pounds force (ft·lbf)** in the Engineering English system. A joule is exactly 1 N·m or equivalently 1 kg·m^2/s^2. For example, to raise a book weighing one kilogram from the surface of the Earth to a level one meter above the surface of the Earth requires an amount of work that equals 9.8 [N] × 1.0 [m] = 9.8 [N·m] = 9.8 joules. In the Engineering English system we have to account for the pesky g_c to calculate the amount of work as $(F \times d)/g_c$ = (mg × d)/g_c = 2.2 × 32.2 × 3.28/32.2 [lbm] [ft/s^2] [ft] [lbf·s^2/lbm·ft] = 7.2 ft·lbf (since 1.00 m is 3.28 ft).

Again, energy is the capability to do work. It may be stored in many objects, such as liquid gasoline, solid uranium, or a speeding train. Equally, the stored energy in a body can be released, such as water running through a water turbine, heat from burning gasoline, or electrical current from a generator. More precisely, any object that does an amount of work consisting of some number of joules will see its energy content decrease by that same number of joules. Any object that has an amount of work done upon it by its surroundings will see its energy similarly increase.

Often, an engineer wants to know how fast that work was done or how rapidly the amount of energy possessed by an object changed. This is determined using the variable **power**, which is defined as the time rate of doing work or, equivalently, as the time rate of change of energy. It is measured in watts in SI, where 1 watt is equal to 1 joule per second. For example, if a person takes two seconds to lift a 1 kg book a height of 1 meter above the surface of the Earth, that person is expending (9.8 joules)/(2.0 seconds) = 4.9 watts of power during those two seconds.

In the Engineering English system, power normally is expressed in units of ft·lbf/s or in horsepower,[2] where one horsepower is 550 ft·lbf/s. It is indeed roughly the power exerted by a working horse and roughly 5 to 10 times the power exerted by a person doing sustained physical labor such as shoveling or carrying.

A useful mnemonic connecting the SI and Engineering English systems is that 1 kilowatt and 1 horsepower are the same order of magnitude (1 kW = 1.34 hp). Some examples of quantities of energy in everyday life are shown in Table 4.1.

Table 4.1 Typical Energy Magnitudes

Number of Joules	Approximate Equivalent
100	Lighted match
10,000	Speeding bullet
5×10^5	Kinetic energy of a small car at 65 mph
10^6	A small meal or the kinetic energy of a SUV at 65 mph
3×10^7	Lightning stroke
10^8	One gallon of gasoline
10^{11}	One gram of uranium in fission
4.2×10^{12}	One kilo ton of TNT
60×10^{12}	Hiroshima size atomic bomb
10^{14}	Annihilation of one gram of any matter

[2]The Scottish engineer James Watt measured the rate of work that a good brewery horse could sustain as a way to sell his early steam engines.

In the past three centuries, a vast increase has occurred in the amount of power at the command of the ordinary person in the United States and other developed nations. In 1776, when Thomas Jefferson was writing the Declaration of Independence and James Watt was perfecting his steam engine, the average adult in the United States or in Britain could call on considerably less than one kilowatt of power for some fraction of a day. Today, the average person (adults and children included) in the United States and other industrialized nations typically has at his or her command more than 10. kW (15 hp) every day of the year, every second, day and night. This includes several kilowatts of electric power per person to say nothing of over 100 kilowatts of automotive power for selected periods of the day. Other kilowatts per person serve us each day—for example, in the form of the trucks that haul goods for us, the oil or gas that heats us in winter, the airplanes that transport us and our overnight deliveries, and the machinery that manufactures all these energy-using artifacts. So each citizen of an industrialized nation continually has on call the power of hundreds of horses at his or her service for a full 24/7. At the same time, vast numbers of people elsewhere in the world are still in that eighteenth century less-than-one-horsepower state or, worse, a one-person-power state. Resolving this global power discrepancy in an environmentally acceptable manner is another one of the great engineering and social challenges of the twenty-first century.

Let us now turn from these general concepts of work, energy, and power to the specific kind of energy that scientists and engineers have defined and to the phenomena in nature and technology that these kinds of energy help engineers to quantify and to use in engineering models.

4.3 KINDS OF ENERGY

Energy comes in various forms. Energy due to motion is called **kinetic energy** and includes the first two major types of energy we will discuss, **translational kinetic energy**[3], **TKE**, and, less obviously, **thermal energy**. Energy due to position is called **potential energy** and includes **gravitational potential energy, GPE**, and, less obviously, **chemical energy**[4] and **electromagnetic energy**.[5] There are other important types such as Einstein's discovery that mass is a form of energy, as embodied in his famous equation $E = mc^2$. We will not deal with it in this course, but note that this insight is the basis of nuclear energy in which about 1% or 2% of matter in uranium is converted to thermal energy, then to electrical energy.

Translational kinetic energy (TKE) is the energy of mass in straight-line motion. It is calculated by the formulae:

$$\text{TKE} = (^1/_2)\ mv^2\ \text{(SI) or } (^1/_2)\ mv^2/g_c\ \text{(English)} \tag{4.2}$$

where m is mass and v is the speed. TKE often is assumed to be the only form of kinetic energy; then it is called simply kinetic energy, and abbreviated **KE**.

[3]In some cases matter is not moving in a straight line but rather rotating around a central axis. In this case, the kinetic energy is called **rotational** kinetic energy, or RKE. It is no different in principle from TKE. It does, however, require more complicated mathematics, since many pieces of matter are moving in different directions and at different speeds. Due to these mathematical complications, RKE will not be required knowledge in this book.

[4]Unburned fuel such as a can of gasoline clearly has potential energy locked inside of it. Obviously this stretches the simple definition that potential energy is mass × height × g (see the section on GPE). Some part of this picture can be retained if you think of the electrons that surround the atoms of the fuel being in high energy states and endeavoring to reach lower or more stable states when being burnt in air.

[5]Electromagnetic energy depends on the potential of electrons to do work; hence voltage often is called "potential."

Note first that the formula results in the correct SI units of joules, since speed is m/s, and mass is in kilograms, so ½ mv^2 has units [kg][m^2/s^2], which are the same units that make a joule. In Engineering English units, we have to divide by g_c to get the proper units of [ft·lbf] from [lbm·ft^2/s^2] [lbf·s^2/lbm·ft].

Example 4.1

What is the translational kinetic energy of an automobile with a mass of 1.00×10^3 kg traveling at a speed of 65 miles per hour (29 m/s)?

Need: TKE of vehicle.
Know: Mass is 1.00×10^3 kg, speed is 29 m/s.
How: Apply Equation (4.2), TKE = ½ mv^2.
Solve: **TKE** = ½ × (1.00×10^3) [kg] × 29^2 [m/s]^2 = 420,500 kg m/s^2 = 4.2050×10^5 J.

Only two of these digits are significant (since the speed was stated only to two digits), so the answer is 4.2×10^5 J = **4.2×10^2 kJ**.

Anything that has mass and is moving in a straight line has TKE. Prominent examples of TKE in nature include the winds and the tides.

Example 4.2

Estimate the total kinetic energy of the wind on the Earth. As an introduction to this problem, consider that the wind is a movement of air that is produced by temperature differences in the atmosphere. Since hot air is less dense than cold air, air heated by the Sun at the equator rises until it reaches an altitude of about 6.0 miles (9.0 km) and then it spreads north and south. If the Earth did not rotate, this air would simply travel to the North and South Poles, cool down, and return to the equator along the surface of the Earth as wind. However, because the Earth rotates, the prevailing winds most of us see travel in a west–east rather than a south–north direction (in the northern hemisphere). In places where winds are strong and steady, it may make economic sense to install a windmill or wind turbine to capture that translational kinetic energy.

Need: Total TKE of the wind in joules.
Know: The atmosphere is about 9.0 kilometers thick (i.e., the height of Mt. Everest). The radius of the Earth is about 6.4 million meters. The surface area of a sphere is $4\pi R^2$, so the volume of an annular "shell" of thickness T around the Earth[6] is about $4\pi R^2 T$, since the thickness T is very small compared to the radius R. Air has a density of about 0.75 kg per cubic meter (averaging from sea level to the top of Everest). This air typically is moving at about 10. m/sec.
How: Find the mass of air in that 9.0 km thick shell around the Earth, and apply Equation (4.2), TKE = ½ mv^2.
Solve: Volume or air around the Earth is about $4\pi(6.4 \times 10^6)^2 \times 9.0 \times 10^3$ [m^2][m] = 4.6×10^{18} m^3. The mass of air around the Earth is therefore $0.75 \times 4.6 \times 10^{18}$ [kg/m^3][m^3] = 3.5×10^{18} kg.[7]

Therefore, TKE = ½ × $(3.5 \times 10^{18}) \times 10^2$ [kg] [m/s]^2 = **1.7×10^{20} J**.

[6]Since the volume of a sphere is $(4/3)\pi R^3$, then the volume of the atmosphere of thickness T is $(4/3)\pi(R+T)^3 - (4/3)\pi R^3$, which is approximately $(4/3)\pi R^2 T$ if R is much greater than T. You can show this yourself by expanding the equation and neglecting all terms containing T^2 and T^3.
[7]A better approximation is 5.1×10^{18} kg.

This is a number sufficiently large as to be meaningless to most people. But to an engineer it should inspire such questions as: Where does this energy come from? Where does it go? In fact, the total Sun's energy reaching the Earth is about 1.4 kW for every m^2 of the Earth's surface. Of this, about 31 percent is reflected back into space. We thus receive about 1 kW/m^2 net solar radiation. Over the Earth this amounts to about 1.7×10^{17} kW.[8]

Since we receive this on the average for 12 hours every day, our beneficent Sun delivers 7.3×10^{24} joules/day; apparently, our estimate of the winds accounts for only $[1.7 \times 10^{20} / 7.3 \times 10^{24}] \times 100 \approx 0.002\%$ of the daily received solar energy.

In addition, the Sun is the ultimate source of the energy in the fossil fuels that power virtually all of our automobiles and about two-thirds of our electric power stations, as well as the source of the biomass that, in the form of food, powers us. Moreover, it drives the cycles of flowing or falling water that power hydroelectric generating plants.

The role that the winds or other solar energy sources might play in meeting the needs of those billions of people who still live at an eighteenth-century energy standard is still debated. The major technical problems are that this source of energy is diffuse, usually unpredictable, and fluctuates daily and seasonally.

Thermal energy often is referred to as heat, and is a very special form of kinetic energy because it is the *random* motion of trillions and trillions of atoms and molecules that leads to the perception of temperature. Heat is simply the motion of things too small to see, an insight captured by the nineteenth-century German physicist Rudolf Clausius when he defined thermal energy as "the kind of motion we call heat."

There is a macroscopic analogue to thermal energy that may be familiar to many of you. The "mosh pit" at a rock concert is an example of the jostling motion of many bodies expending kinetic energy but not actually moving anywhere. It is the molecular amplification of this picture with trillions and trillions of atoms or molecules bouncing together that produces the net effect of what we call thermal energy.

Our analysis of the meaning of temperature will begin with the motion of a single particle of gas (either an atom or a molecule) in a box. We will point out the theoretical underpinnings of the famous **Ideal Gas Law** that you probably have seen before. The Ideal Gas Law for a fixed amount of gas is:

$$pV = N R T \tag{4.3}$$

in which p is the pressure, V is the volume of the gas, N is the amount of the gas in *moles*, T is its absolute temperature, and R is the gas constant per *mole* of gas.[9] This law is accurate for many real gases over a wide range of conditions.

We can use the Ideal Gas Law to analyze pressure from the point of view of the motion of a single gas atom in a box. The pressure on the walls of the box is the result of the atom repeatedly hitting the walls of the box. If there are few collisions, the internal pressure on the box walls is small, but if there are many collisions, the pressure is larger. If the temperature rises, the atom has more speed and thus more momentum and consequently the pressure on the walls increases. Now we need to calculate the rate of change of momentum as our single atom hits a wall of the box.

[8]And even this huge amount is only about 4×10^{-10} of the total energy generated by the Sun.

[9]A *mole* of gas is its mass/molecular mass. For example, helium has a molecular mass of 4 kg/kg mole so that one kg mole of it has a mass of 4 kg, and 2 kg moles = 8 kg, etc. For oxygen gas (a molecule with two oxygen atoms each of molecular mass 16 kg/kg mole), it has a molecular mass of 32, meaning one mole of it has a mass of 32 kg. These concepts are explained more fully in Chapter 5.

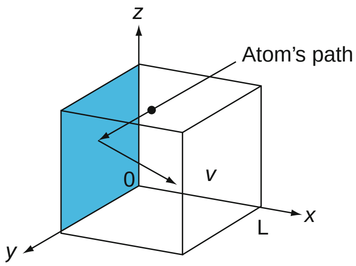


FIGURE 4.2 Cubical Box of Side L, Volume V Containing Just One Atom

Assume that the atom is moving with speed v (see Figure 4.2). On impact with the wall the atom's momentum will be transferred to the wall of the box, and the resulting force on the wall due to this single atom is proportional to its rate of change of momentum, or:

$$F \propto (mv)\, v = mv^2$$

But there are trillions of atoms (or molecules) in just one cm^3 of gas. Let this number of atoms be n. Thus the total force, F on each of the box's walls is proportional to nmv^2, or:

$$F \propto nmv^2$$

The speed of the atoms or molecules in this equation is that of the *fluctuations* as they interact with each other. It is *not* a steady wind.

It's hard to comprehend just how large n really is. The next example will show you why.

Example 4.3

If the number of atoms in one kg mole of a gas is 6.022×10^{26}, how many atoms of this gas occupy just 1.00 cm^3 at a pressure of 1.00×10^5 N/m^2 and a temperature of 0°C (273 K)?

Need: $n =$ ________ atoms in one kg of a gas at $p = 1.00 \times 10^5$ N/m^2 and $T = 273$ K. $R = 8314$ J/(kg mole·K).
Know: One kg mole of the gas contains 6.022×10^{26} molecules.
How: Use $pV = NRT$ to find N (the number of moles), and then convert N, into n, the number of atoms.
Solve:

$$N = \frac{pV}{RT} = \frac{1.00 \times 10^5 \times 1.00}{8314 \times 273} \times \left[10^{-2}\right]^3 \left[\frac{N}{m^2}\frac{cm^3}{}\left(\frac{m}{cm}\right)^3 \frac{kg\ mole\ K}{J}\frac{1}{K}\right] = 4.41 \times 10^{-8}\ kg\ moles\ in\ 1.00\ cm^3$$

Finally, convert N into the number of atoms in one cm^3:

$$n = 4.41 \times 10^{-8} \times 6.022 \times 10^{26}\ [\text{kg moles/cm}^3][\text{molecules/kg mole}] = 2.65 \times 10^{19}\ \text{atoms}$$

This is a very large number and there is no way of monitoring the motion of all these gas atoms or molecules as they collide with each other and the walls in random directions and at random speeds. Thus we use *averaging* techniques to describe the behavior of large numbers of atoms or molecules.

Returning to our calculation of the pressure on the walls of a container, it is just force per unit area, (F/A), and we can show that:

$$pV \propto nmv^2$$

Note that nm is the *total mass* of all the gas in the box so that our two expressions for pV are $pV = NRT$ and $pV \propto nmv^2$, from which we can conclude that:

$$T \propto \frac{1}{R}\frac{n}{N}(mv^2) = \frac{2}{R}\frac{n}{N}\left(\frac{mv^2}{2}\right)$$

In this equation, n, N, m, and R are constants, so the absolute temperature of the gas is proportional to the kinetic energy of the atoms or molecules, ½ mv^2 or:

$$T \propto {}^1\!/_2\, mv^2 \tag{4.4}$$

So when you feel hot it's because the kinetic energy of the molecules of air striking your body has increased.

Example 4.4

If you feel hot at 35.0°C and cold at 5.00°C, what is the percent of change in the speed of molecules contacting your skin?

Need: Percent of change in molecular speed between a cold gas (at 5.00°C = 278K) and a hot gas (at 35.0°C = 308K).

Know-How: From Equation (4.4) we have that $T \propto v^2$.

Solve: $\frac{v_H}{v_C} = \sqrt{\frac{T_H}{T_C}}$ in which the subscripts refer to hot and to cold.

Therefore, $\frac{v_H}{v_C} = \sqrt{\frac{308}{278}} = 1.05$ or the difference in the average air molecule's speed is only 5%.

The key principle here is that temperature is a result of the kinetic energy of atoms and molecules. If we want the temperature of a roast in the oven, or a piece of steel being welded, or a block of ice, we cannot measure all the speeds of all the molecules. In practice, we use thermometers and other kinds of devices that automatically average the speed of the atoms or molecules.

Thermometry is the technology of temperature measurement. People have always been able to experience the sensations of hot and cold, but the development of an accurate temperature measurement technology did not occur until the seventeenth century. Galileo is credited with constructing the first practical thermometer in about 1592. By the eighteenth century, more than 30 different temperature scales were in use. These scales usually were based on the use of two fixed standard temperatures with the distance between them divided into equally spaced degrees. Some of these early scales are shown in Table 4.2.

The 100-division Celsius temperature scale became very popular during the eighteenth and nineteenth centuries, and was commonly known as the centigrade (from the Latin *centi* for 100 and *gradua* for step) scale until 1948, when Celsius' name formally was attached to it and the name centigrade officially was dropped.

In 1848 William Thomson (Lord Kelvin) developed an absolute scale based on the Celsius degree size that now bears his name. Soon thereafter an absolute temperature scale based on the Fahrenheit degree size was

Table 4.2 Early Temperature Scales

Inventor and Date	Fixed Points
Isaac Newton (1701)	Freezing water of (0°N) and human body heat (12°N)
Daniel Fahrenheit (1724)	Freezing water (0°F) and human body heat (96°F)[10]
René Réaumur (1730)	Freezing water (0°Re) and boiling water (80°Re)
Anders Celsius (1742)	Freezing water (0°C) and boiling water (100°C)

developed and named after the Scottish engineer William Rankine. The relationship between the modern temperature scales is shown as follows.

$T(°F) = (9/5)T(°C) + 32.0 = T(R) - 460.$
$T(°C) = (5/9)[T(°F) - 32.0] = T(K) - 273$
$T(R) = (9/5)T(K) = T(°F) + 460.$
$T(K) = (5/9)T(R) = T(°C) + 273$

Gravitational potential energy (GPE) is the energy acquired by an object by virtue of its position in a gravitational field—typically by being raised above the surface of the Earth. In **SI units**, it is calculated by the equation:

$$GPE = mgh \tag{4.5}$$

in which h is the height above some datum, usually the local ground level. In **Engineering English units**, this definition is modified by g_c to ensure that it comes out in ft·lbf units:

$$GPE = \frac{mgh}{g_c} \tag{4.6}$$

Note that gravitational potential energy only has meaning relative to a reference datum level. The choice of such a level is arbitrary but must be applied consistently throughout a problem or analysis.

Example 4.5

Wile E. Coyote holds an anvil of mass 100. lbm at the edge of a cliff, directly above the Road Runner who is standing 1000. feet below. Relative to the position of the Road Runner, what is the gravitational potential energy of the anvil?

Need: GPE.

Know: Anvil has mass 100. lbm. The reference datum level, where the Road Runner is standing can be chosen as zero. The height h of anvil is 1000. feet referenced to Road Runner's datum. Because we are calculating in Engineering English units, we will also have to use g_c.

How: Equation (4.6) is in Engineering English units, $GPE = \frac{mgh}{g_c}$.

Solve: $\text{GPE} = \frac{100.\,[\text{lbm}]\;1000.\,[\text{ft}]\;32.17\,[\text{ft/s}^2]}{32.17\,[\text{lbm}\cdot\text{ft/lbf}\cdot\text{s}^2]} = 1.00 \times 10^5\ \text{ft}\cdot\text{lbf}$ correct to three significant figures (as is the least known variable, the mass of the anvil).

[10] 0°F was the lowest temperature you could reach with a mixture of ice and salt. The modern Fahrenheit scale uses the freezing point of water (32°F) and the boiling point of water (212°F) as its fixed points. This change to more stable fixed points resulted in changing the average body temperature reading from 96°F on the old Fahrenheit scale to 98.6°F on the new Fahrenheit scale.

Electromagnetic energy (often merely called electricity) is a form of energy that is typically carried by electric charges[11] moving through wires, or electromagnetic waves (or particles) moving through space. It will be the subject of Chapter 7, so we just touch on it briefly here.

Like all forms of energy and power, electromagnetic energy can be measured in joules, and electromagnetic power can be measured in watts. Consider the electromagnetic power used by an electrical device that is connected by wires to an electrical battery to create an electrical **circuit**,[12] as shown in Figure 4.3.

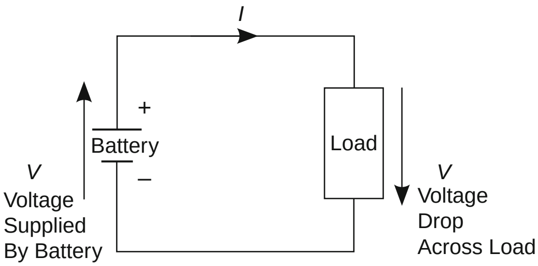


FIGURE 4.3 An Electrical Circuit

An electrical current, represented by the letter I, flows into and out of the device. Electrical current is measured in units called amperes (symbol A), which is simply a measure of the number of electrons passing through any cross section of the wire every second. An electrical potential, represented by the letter V, can be measured at any point on the circuit. Electrical potential is measured in units called volts (V). It might be thought of as an electrical pressure, originating in the battery that keeps the current flowing through the wire. The voltage has its most positive value at the terminal on the battery marked plus (+), and drops throughout the circuit, reaching its minimum value at the battery terminal marked minus (−). By measuring the voltages at any two points along the wire and subtracting to determine the difference between those two measured voltages, we can determine the voltage drop between those two points on a circuit. We determine the overall voltage drop by subtracting the voltage at the point nearer the minus terminal of the battery from the voltage at the point nearer the plus terminal of the battery. Again, that voltage drop might be thought of as a pressure pushing the current from the point of higher voltage to the point of lower voltage.

In mechanical systems, power, the rate of doing work, can be computed as the product of force × velocity. From electrostatics, the *force* on a charge Q coulombs[13] in a voltage gradient of V/d is $Q \times V/d$ (d is the distance over which the voltage changes from 0 to V)—hence, the work in moving the charge is $d \times QV/d = QV$. If the charge is moved along the voltage gradient in time t, the power to move the charge is VQ/t. We define the rate of movement of charge as the electric current, $I = Q/t$. If Q is in coulombs, t in seconds, then I is in amperes. Further, if V is in volts, then the power is in watts. Simply put, what you need to remember is that the electrical power is expressed as:

$$\textit{Electric Power},\ P = I \times V \tag{4.7}$$

[11]Specifically, negatively charged electrons.

[12]The word *circuit* has the idea of circle built in. Thus, an electrical circuit must start and finish in an unbroken loop. The physical reason is that the electrons as charge carriers are not consumed in the circuit.

[13]The charge on a single electron is 1.6×10^{-19} coulombs.

Example 4.6

A battery sustains a voltage drop of 3.0 V across a small lightbulb and produces a current of 0.1 A through the lightbulb. What is the power required by the lightbulb?

Need: Power of lightbulb in watts (W).
Know: Voltage across bulb is 3.0 V. Current through lightbulb $I = 0.1$ A.
How: Use Equation (4.7), $P = I \times V$, that is, power = voltage × current.
Solve: $P = 3.0$ [V] × 0.1 [A] = 0.3 W.

Although the power into the lightbulb departs from the electric circuit, it does not disappear. Instead, some of it is radiated away from the bulb in the form of electromagnetic power (or, equivalently, in the form of massless particles called photons) as visible light. Some of this electromagnetic energy is in the form of radiant heat. Yet, more energy is lost in lower-grade heat by heating the lightbulb's local surroundings. Calculating the value of the power carried by these mechanisms will not be covered in this book. However, once again, that power can be measured in watts.

Chemical energy is another form of potential energy in that it is determined by the relative distribution of electrons in the atoms that make up the structure of molecules. It is so important to our theme of twenty-first century engineering that we will devote an entire chapter to chemical energy. Suffice it here to say that it too is most conveniently measured in joules.

4.4 ENERGY CONVERSION

Sunlight drives the winds. An anvil hoisted to a cliff top can be used to deliver kinetic energy to an unwary Road Runner below (though it is more likely to end up falling on the head of Wile E. Coyote himself!). Burning fossil fuel (a form of stored-up solar electromagnetic energy) results in the rotational kinetic energy (known there as "shaft work") of a turbine that is then converted back into electromagnetic energy in a generator to light up a city. All these occurrences suggest a second key fact about energy: Its various kinds can be converted from one form to another.

Because all types of energy can be expressed in the same units, joules, this conversion can be expressed quantitatively in simple models. Even in Engineering English units, the number of conversion factors is relatively few.[14]

We mentioned earlier that the Sun delivers more than 10^{24} J of energy daily to the Earth in the form of the electromagnetic solar energy we call sunlight. What happens to that electromagnetic energy? A discussion provides the clues that can be made into a model. Part of the energy heats the atmosphere and drives the winds as previously noted as well as driving the water cycle. Part of the solar energy that reaches the Earth is converted into chemical energy via photosynthesis in trees, grass, and agricultural crops. Part of it simply heats the Earth and the ocean. Much of it is reflected away back into deep space. One thing the model must answer is the following question: If all that energy is arriving from the Sun every day, and much of it is converted to thermal energy, why isn't the Earth accumulating more and more thermal energy and continuously getting hotter? (Try to answer that question yourself before reading the next paragraph.)

[14]One useful conversion factor in the Engineering English system shows the relative magnitude of mechanical and thermal units: 1 Btu (British thermal unit) is the heat required to raise the temperature of 1 lbm of water by 1°F; it is equivalent to 778 ft·lbf of energy. In other words, if you dropped 1 lbm of water through 778 ft and all the original GPE were converted to thermal energy, you would heat that water only by 1°F. Another useful conversion factor is that 1 Btu is approximately equal to 1 kJ (actually, 1 Btu = 1.055 kJ).

The fact that the temperature of the Earth is staying roughly constant requires another element in our simple model. Roughly speaking, every day the Earth must reflect or radiate away into space the same amount of energy that it receives from the Sun. Should anything happen, either due to natural or human causes, to interfere with this energy conversion balance between absorption and radiation of solar energy, we will be in big trouble! The Earth might either cool off (as it did in the various ice ages) or heat up (as it appears to be doing right now according to many scientists who believe that global warming is taking place). So understanding energy conversion is crucial to projecting the future of life on our planet.

We can perform a similar analysis of energy conversion on a technological system, such as an automobile, a house, a toaster, or anything we choose. In the case of an automobile, we can regard the initial source of the energy as the chemical energy pumped into the vehicle at the gas station. In the automobile's engine, that chemical energy is converted to the translational kinetic energy of a piston in a cylinder. That translational kinetic energy is then converted into rotational energy in the crankshaft, transmission, axles, and wheels. That rotational kinetic energy in turn provides the translational kinetic energy of the automobile in its motion down the road. In this process, part of that energy—indeed, a substantial majority of the initial chemical energy—is transferred to the environment in the form of heat, either through the car's radiator and exhaust pipe or through friction and air resistance.

Example 4.7

A gallon of gasoline can provide about 1.30×10^5 kJ of chemical energy. Based on Example 4.1, if all the chemical energy of a gallon of gasoline could be converted into the translational kinetic energy of an automobile, how many gallons of gasoline would be equivalent to the vehicle's TKE if the automobile is traveling at 65. miles per hour on a level highway?

Need: Gallons of gasoline to propel the automobile at 65. mph (29.0 m/s).

Know: TKE of vehicle $= 4.2 \times 10^2$ kJ to 2 sig. figs. at 65. mph (as per our previous calculation in Example 4.1). Energy content of gasoline is 1.30×10^5 kJ/gallon.

How: Set TKE of vehicle equal to chemical energy in fuel. Let x = number of gallons needed to accomplish this.

Solve: $x \times 1.30 \times 10^5$ [gallons][kJ/gallon] $= 4.2 \times 10^2$ [kJ]. Therefore **x = 0.0032 gallons**.

This is a tiny amount of gasoline; as we will see later, it is a misleadingly small amount of gasoline. In fact, most of the gasoline is *not* being converted into useful KE but is mostly wasted elsewhere. The misleading nature of this calculation will become obvious as we proceed. The error is in the conditional statement "*if all* the chemical energy of a gallon of gasoline could be converted into the translational kinetic energy of an automobile." The lesson to be learned here is to beware of your assumptions!

4.5 CONSERVATION OF ENERGY

It is possible in principle (though impossibly difficult in practice) to add up all the energy existing[15] in the universe at any moment and determine a grand total. Scientists and engineers express this fact by saying that *energy is conserved*. This may sound like a theoretical claim of little practical use. However, engineers have

[15]In this case, we would also have to add in all the mass in the universe because cosmic events freely convert mass into energy and energy into mass via $E = mc^2$!

developed a method of applying the fact that energy is conserved to practical problems while avoiding the inconvenience of trying to account for all the energy in the universe. This method is called **control boundary** analysis and was schematically illustrated in Figure 4.1.

It consists simply of isolating the particular object or system under consideration, and making a simple model of the way that object or system exchanges energy with the rest of the universe. Making that model begins with a conceptual sketch of the object or system. You draw a dotted line representing the control boundary around the sketch to contain the item under analysis. Then draw arrows across the boundary representing specific types of exchanges of energy between the object or system and the rest of the universe. By limiting the energy exchange in this way, the principle of conservation of energy can be imported into the system under consideration to determine the results of various energy conversion processes that occur within the dotted line. Losses or gains are simply one form or another of energy that crosses that boundary.

As a simple example, consider your classroom as a closed system and draw an imaginary control boundary around it. Assume this boundary is impervious to energy flows so that what's inside in the classroom stays there. Now imagine you have a 1.0 kg book on your desk that is 1.0 meter high from the floor. The book has $GPE = mgh = 1.0 \times 9.8 \times 1.0$ [kg][m/s^2][m] $= 9.8$ J with respect to the floor of your classroom. Suppose that book now falls to the floor, thus losing all of its GPE. Since the classroom boundaries are impervious to energy exchanges, where did that GPE go? Have we violated conservation of energy?

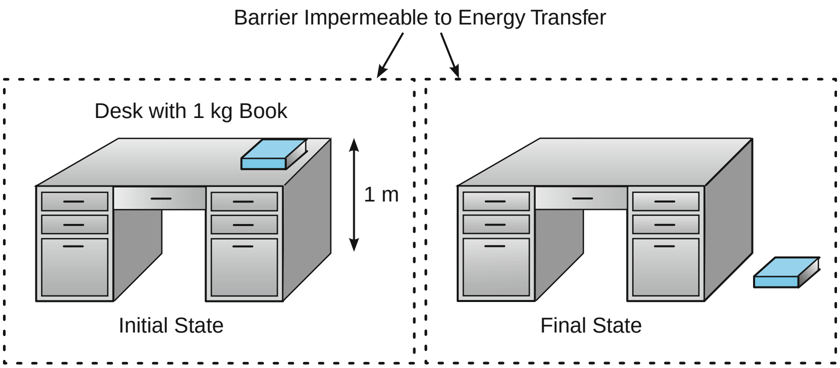


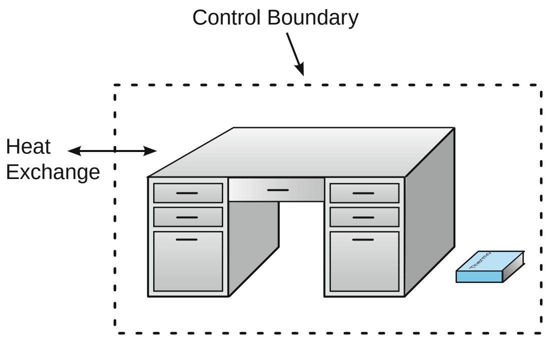
FIGURE 4.4 Where Did the Book's Potential Energy Go?

The fact is that GPE is still trapped within the room but not in the same form. What physically happens is: (1) the GPE of the falling book is converted to TKE; (2) when the book hit the floor, it sets up some sound waves in the floor material and in the surrounding air; and (3) eventually when all the transients died out, the forms of energy in (1) and (2) end up heating the room and its contents by exactly 9.8 J of energy (see Figure 4.4).

The principle of conservation of energy says energy is never lost, merely transformed into another form. This is one statement of energy conservation also called the **First Law of Thermodynamics**. Ultimately all forms of energy degrade to heat; this is one statement of the **Second Law of Thermodynamics**. Control boundary analysis is a very useful way to account for energy flows.

Example 4.8

Remove the impervious boundary in Figure 4.4 and replace it with a control boundary that allows energy flows only in the form of heat transfer from the room to the rest of the universe, keeping the room temperature constant. There is thermal equilibrium when the book is on the desk (i.e., no net heat flowing).



Suppose the initial energy with the classroom was 100.0 J when the room was isolated. What is the final energy of the classroom after the exchange of thermal energy?

Need: Energy of classroom, Q_{Final} = __ J.

Know–How: 9.8 J of potential energy was converted when the textbook fell to the floor. It cannot vanish, so what happened to it? It was eventually all converted to heat. When the book was on the table, the total energy in the room was 100.0 J.

Solve: After the textbook falls, 9.8 J of thermal energy was created from its original GPE. This eventually flows across the control boundary to the rest of the universe in the form of heat:

$$\text{The initial total energy} = \text{final total energy or } 100. = Q_{Final} + 9.8$$

Therefore $\mathbf{Q_{Final}}$ = 100.0 – 9.8 = **90. J**. (to 2 sig. figs.) In other words, the classroom's energy falls by 9.8 J.

This kind of model captures enough of reality that it can be highly useful in engineering analysis and design. Figure 4.5 shows a control boundary analysis model as applied to an automobile. In words, the picture illustrates the following "energy accounting" of a typical automobile trip. As indicated by the arrows crossing the dotted line, an automobile traveling at 65. miles per hour (29. m/s) transfers about 40.0 kilowatts to the atmosphere in the form of thermal energy from the radiator, about 20.0 kilowatts to the atmosphere in the form of heat and chemical energy out the exhaust, about 7.00 kilowatts of mechanical power[16] to the atmosphere in overcoming air resistance, about 11.0 kilowatts in frictional work (mainly the result of elastic compression and expansion of the rubber in the tires and applying the brakes), and about 2.00 kilowatts of work for other purposes such as pumping and operating such accessories as heat, lights, and air conditioning. All of these contributions total about 80.0 kW.

[16] In the atmosphere, due to friction, the displaced air will eventually slow, and the energy imparted to it will dissipate as an equivalent amount of thermal energy.

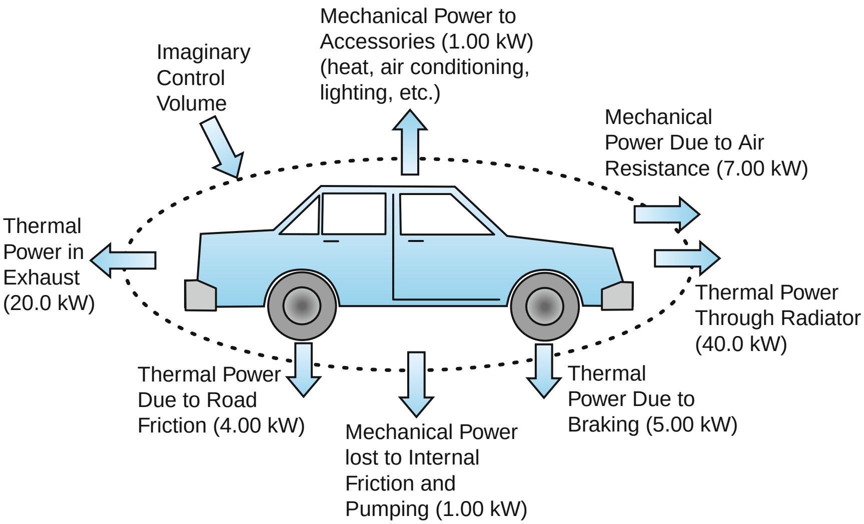


FIGURE 4.5 Control Boundary Analysis Model of Automobile

Example 4.9

If gasoline contains 1.3×10^5 kJ/gallon, how many gallons of gasoline must be used per second to provide the energy needed to sustain travel at 65. miles per hour on a level road? From this, estimate the fuel economy in mpg (miles per gallon).

Need: Amount of gasoline consumed per second in gallons.

Know: An automobile traveling at 65. miles per hour transfers about 80. kW or 80. kJ/s (or 110 hp) into various forms of energy.

How: Apply the principle of the conservation of energy. The energy lost by the car through the boundary of the control surface must come from somewhere. The only place it can come from within the car is by the decrease in the chemical energy of some of the gasoline. By the principle of conservation of energy, the decrease in chemical energy must equal the transfer out of the dotted lines of all the other types of energy.

Solve: Chemical energy needed per second = 80. kW or 80. kJ/s. Since gasoline contains about 1.3×10^5 kJ/gallon of chemical (i.e., potential) energy, this means $80./1.3 \times 10^5$ [kJ/s][gallon/kJ] = 6.2×10^{-4} gallons/s of gasoline are consumed.

Is this a reasonable answer? Note that a car going at 65. miles per hour travels 65./3600. miles/s. So converting to miles per gallon as follows:

$$(65./3600) \times 1/(6.2 \times 10^{-4})[\text{miles/s}][\text{s/gallon}] = 29.\ \text{mpg}$$

This is a reasonable estimate given the roughness of our calculations.

Note a couple of implications of this example. First, from an energy standpoint, we might view an automobile as a device for converting the chemical energy of fuel into thermal energy in the atmosphere. Second, operating an automobile requires a process for liberating 80. kilo-joules of energy every second from liquid fuel. This process, called combustion, will be the subject of the next chapter.

SUMMARY

The principle of conservation of energy can help us answer those questions that began this chapter. Is the world running out of energy? Drawing a dotted line around the Earth and applying control boundary analysis indicates that the amount of energy here on Earth either is remaining constant, or if you believe that global warming is happening, is increasing slightly (although significantly). So the answer is no—the world certainly is not running out of the total energy, although useful primary energy sources might be in decline.

So where does the energy come from that provides the light by which you are reading these words? This gets to the real point that people are making when they assert we are in an "energy crisis." Though the amount of energy around us is constant, we are rapidly converting reserves of easily exploited chemical energy (which is actually a form of stored-up solar electromagnetic energy derived from past sunlight) into much less useful thermal energy. The light by which you are reading this probably came from a lightbulb powered by a fossil-fuel burning electric power plant. As you read these words, that light is being converted from electromagnetic energy into heat at room temperature—a form of energy of very little use for anything beyond keeping you warm.

Originally we asked, "Does a car possess more energy when it is sitting in your driveway or 15 minutes later when it is traveling down the highway at 65 miles per hour?" Our control boundary analysis model of the automobile tells us that the automobile is continually transferring energy from its gas tank to the atmosphere. So it possesses less energy on the highway than it did 15 minutes earlier in the driveway, and until you pull into a gas station and fill up the gas tank, it will steadily decrease in the energy content within its dotted control boundary.

To sum up, an understanding of the concept of energy is one of the indispensable analytical tools of an engineer. That understanding begins with the following key ideas. Energy is the capability to do work. Energy comes in many kinds. These kinds are capable of conversion from one kind to another. In the course of these conversions, energy is neither created nor destroyed, but rather, conserved. This is a principle that an astute engineer can exploit to advantage in the analysis of complex systems using the concept of a control boundary that separates the universe into that which we are studying and the rest of the universe. Using this concept, we can directly calculate energy flows to and from the control boundary as an aid in analyzing its parts.

EXERCISES

Pay attention to the inferred number of significant figures in your answers! Conversion factors: 1.00 J = 0.738 ft·lbf; 1.00 kg = 2.20 lbm and g_c = 32.2 lbm·ft/lbf·s^2.

1. Determine the translational kinetic energy of the automobile in Example 4.1 if its speed was reduced to 55. miles per hour.

2. Determine the translational kinetic energy in Engineering English units of the automobile in Example 4.1 if its mass was increased to 4.00×10^3 lbm.

3. Determine the translational kinetic energy of the atmosphere in Example 4.2 if the average air velocity increased to 15. m/s.

4. Repeat the calculation of Example 4.2 in Engineering English units. Check that your answers agree with the solution in Example 4.2 using the appropriate conversion factors.

5. What would be the gravitational potential in SI units of the anvil in Example 4.5 if its mass was 100. kg and the cliff was 1000. meters high?

6. Determine the gravitational potential energy (GPE) of an 8.00×10^3 kg truck 30. m above the ground. (**A: 2.4×10^6 J** to two significant figures, since *h* is known only to two significant figures)

7. A spring at ground level—that is, at height = 0.00 m—shoots a 0.80 kg ball upward with an initial kinetic energy of 245 J. Assuming that all the initial TKE is converted to GPE, how high will the ball rise (neglecting air resistance)?

8. Chunks of Earth orbital debris can have speeds of 2.3×10^4 miles per hour. Determine the translational kinetic energy (TKE) of a 2.0×10^3 lbm chunk of this material in SI units. (**A: 4.8×10^{10} J to two significant figures**)

9. An airplane with a mass of 1.50×10^4 kg is flying at a height of 1.35×10^3 m at a speed of 250.0 m/s. Which is larger—its translational kinetic energy or its gravitational potential energy with respect to the Earth's surface? (Support your answer with numerical evidence.) (**A: TKE = 4.69×10^8 J; GPE = 1.99×10^8 J, therefore the TKE is greater than GPE**)

10. Determine the amount of gasoline required in Example 4.7 if the automobile was traveling at 55. miles per hour.

11. Suppose the 1.00 kg book in Example 4.8 fell from a height of 2.50 meters. What would be the final energy of the classroom after the exchange of thermal energy?

12. A vehicle of mass 1.50×10^4 kg is traveling on the ground with a TKE of 4.69×10^8 J. By means of a device that interacts with the surrounding air, it is able to convert 50.% of the TKE into GPE. This energy conversion enables it to ascend vertically. To what height above the ground does it rise?

13. Aeronautical engineers have invented a device that achieves the conversion of kinetic to potential energy as described in Exercise 12. The device achieves this conversion with high efficiency. In other words, a high percentage of the translational kinetic energy of motion is converted into vertical "lift" with little lost to horizontal "drag." What is the device called? (**Hint:** This is not rocket science.)

14. A hypervelocity launcher is an electromagnetic gun capable of shooting a projectile at very high speed. A Sandia National Laboratory hypervelocity launcher shoots a 1.5 gram projectile that attains a speed of 14. km/s. How much electromagnetic energy must the gun convert into TKE to achieve this speed? Solve in SI. (**A: 1.5×10^2 kJ**)

15. Solve Exercise 14 in Engineering English units. (Check your answer by converting the answer to Exercise 14 into Engineering English units.)

16. Micrometeoroids could strike the International Space Station with impact velocities of 19 km/s. What is the translational kinetic energy of a 1.0 gram micrometeoroid traveling at that speed? (**A: 1.8×10^5 J**)

17. Suppose a spaceship is designed to withstand a micrometeoroid impact delivering a TKE of a million joules. Suppose that the most massive micrometeoroid it is likely to encounter in space has a mass of 3 g. What is the maximum velocity relative to the spaceship at which the most massive micrometeorite can be traveling for the spaceship to be able to withstand its impact?

18. A stiff 10.0 g ball is held directly above and in contact with a 600.0 g basketball and both are dropped from a height of 1.00 m. What is the *maximum* theoretical height to which the small ball can bounce?

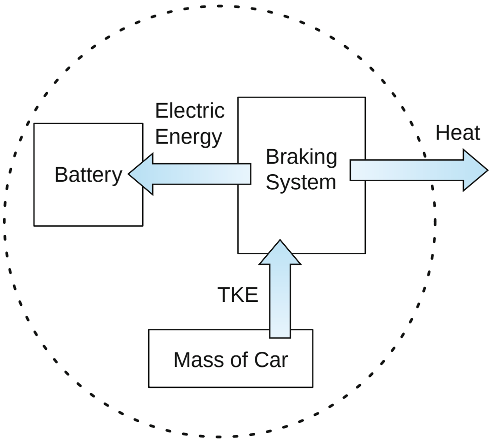
19. What would be the power required by the lightbulb in Example 4.6 if it sustained a voltage drop of 120. V?

20. What would be the current in the lightbulb in Example 4.6 if it sustained a voltage drop of 120. V and required a power of 100. W?

21. An electric oven is heated by a circuit that consists of a heating element connected to a voltage source. The voltage source supplies a voltage of 110. V, which appears as a voltage drop across the heating element. The resulting current through the heating element is 1.0 A. If the heating element is perfectly efficient at converting electric power into thermal power, what is the thermal power produced by the heating element? (**A: 1.1 $\times$ 10^2 W to two significant figures**)

22. A truck starter motor must deliver 15 kW of power for a brief period. If the voltage of the motor is 12 V, what is the current through the starter motor while it is delivering that level of power?

23. A hybrid car is an automobile that achieves high fuel efficiency by using a combination of thermal energy and electrical energy for propulsion. One of the ways it achieves high fuel efficiency is by regenerative braking. That is, every time the car stops, the regenerative braking system converts part of the TKE of the car into electrical energy, which is stored in a battery. That stored energy can later be used to propel the car. The remaining part of the TKE is lost as heat. Draw a control surface diagram showing the energy conversions that take place when the hybrid car stops. (**A: See diagram below**.)



24. Suppose the car in Exercise 23 has a mass of 1000. kg and is traveling at 33.5 miles per hour. As it comes to a stop, the regenerative braking system operates with 75% efficiency. How much energy per stop can the regenerative braking system store in the battery? Illustrate with a control boundary showing the energy flows.

25. Suppose the car in Exercises 23 and 24 has stored 1.00 $\times$ 10^2 megajoules (MJ) of energy in its battery. Suppose the electric propulsion system of the car can convert 90. percent of that energy into mechanical power. Suppose the car requires 30. kW of mechanical power to travel at 33.5 miles per hour. How many miles can the car travel using the energy in its battery? (**A: 28 miles**)

26. Determine the amount of gasoline consumed per second by the automobile in Example 4.9 if it was travelling at 41. m/s. (Assume power required increases as the cube of speed)

27. In order to maintain a speed v on a horizontal road a car must supply enough power to overcome air resistance. That required power goes up with increasing speed according to the formula:

$$P = \text{Power in kW} = K \times v^3$$

where v is the speed measured in miles/hour and K is a constant of proportionality. Suppose it takes a measured 7.7 kW for a car to overcome air resistance alone at 30.0 mph.

a. What is the value of K in its appropriate units?

b. Using a spreadsheet, prepare a graph of power (kW on the y axis) as a function of speed (mph on the x axis) for speeds from 0 mph to 100 mph.

28. Review Exercises 14 through 17 in Chapter 3 concerning the dynamics (and consequent fate) of bungee jumpers. Draw a control surface around the jumper and cord. Show the various forms of energy possessed by the jumper and cord, along with arrows showing the directions of energy conversion inside and across the control surface: (a) when the jumper is standing on the cliff top, (b) when the jumper is halfway down, and (c) when the cord brings the jumper to a safe stop.

29. After working for a company for several years, you feel you have discovered a more efficient energy conversion method that would save your company millions of dollars annually. Since you made this discovery as part of your daily job you take your idea to your supervisor, but he or she claims it is impractical and refuses to consider it further. You still feel it has merit and want to proceed. What do you do?

a. You take your idea to another company to see if they will buy it.

b. You contact a patent lawyer to initiate a patent search on your idea.

c. You go over your boss's head and talk to his or her supervisor about your idea.

d. You complain to your company's human resources office about having poor supervision.

30. Your course instructor claims that energy is not really conserved. He or she uses the example of a spring that is compressed and then tied with a nylon string. When the compressed spring is put into a jar of acid, the spring dissolves and the energy it contained is lost. How do you react?

a. Ignore him or her and follow the established theories in the text.

b. Go to the department chairperson and complain that the instructor is incompetent.

c. Say nothing, but make detailed statements about the quality of the instructor on the course evaluation at the end of the term.

d. Respectfully suggest that the energy in his or her spring example really is conserved by measuring the temperature change of the acid as it dissolves the spring.

This page intentionally left blank

CHAPTER

Chemical Energy and Chemical Engineering

5

Source: © iStockphoto.com/Josh Rodriguez

Chemical engineering is a branch of engineering that applies physical sciences and mathematics in the design, development, and maintenance of large-scale chemical processes that convert raw materials into useful or valuable products. Chemical engineers aid in the manufacture of a wide variety of products such as fuels, fertilizers, insecticides, plastics, explosives, detergents, fragrances, flavors, and pharmaceuticals.

Three primary physical laws underlying chemical engineering design are (1) the conservation of mass, (2) the conservation of momentum, and (3) the conservation of energy. The movement of mass and energy around a chemical process are evaluated using mass balances and energy balances. In more advanced texts, some of these basic ideas are further applied in a powerful methodology called *Unit Operations*, but for the moment we will confine ourselves to simpler techniques: The movement of mass and energy around a chemical process are evaluated using mass balances and energy balances as illustrated in the following section.

5.1 CHEMICAL ENERGY CONVERSION

How is the chemical energy of a fuel such as gasoline or natural gas converted into the kinetic energy of an automobile moving down the road or into the electrical energy to power a television or refrigerator? Each is a multistep process that effects the conversion from fuel to mechanical or to electrical energy. In this chapter, we will consider the first step in that process, conversion of chemical energy into thermal energy. This topic is typically the province of the chemical and mechanical engineer.

Combustion is the *oxidation* of a *fuel* to generate heat, perhaps also accompanied by the emission of light. For engineers **oxidation** is basically the chemical reaction of a substance with oxygen.[1] Combustion may be either slow or rapid, depending on the circumstances.

Other processes can be analyzed in a way similar to combustion. Whereas combustion processes require us to keep tabs on the fate of *atoms* of reactant species, some processes such as those found in oil refineries

[1]Chemists have generalized this definition to mean the loss of an electron from an atom or molecule. For most engineering purposes, the definition given above is sufficient.

can be analyzed by keeping tabs on *molecules*. A prime example is distillation, a major process that is used in oil refineries to separate useful compounds from crude petroleum.

The key points to understanding combustion and the field of chemical engineering are the concepts of:

- **Atoms**, **molecules**, and **chemical reactions**
- **kmol**, a name for a very large number of atoms or molecules
- **Stoichiometry**, a "chemical algebra" method
- **Air-to-fuel ratio**, a measure of the amount of air present when combustion occurs
- **Heating value**, the amount of heat produced by the combustion of a fuel

One useful principle that is very helpful is the discovery of **Amado Avogadro** that equal volumes of gases at the same conditions of temperature and pressure contain the same number of molecules.

5.2 ATOMS, MOLECULES, AND CHEMICAL REACTIONS

An **atom** (from the Greek meaning "cannot be sliced") is the smallest possible piece of a chemical element. Although now known to be sliceable, atoms are still used as the basic building blocks of matter in almost all engineering models. In combustion, as in all chemical reactions, the number and type of participating atoms must remain constant. A **molecule** is the smallest possible piece of a chemical compound. Molecules are made of atoms. In combustion, as in other chemical reactions, the number of molecules present typically does not remain constant, since molecules can be "sliced." This mathematics of molecules, where two plus one can equal two, is presented symbolically in chemical equations, such as the combination of two hydrogen molecules with one oxygen molecule to form two molecules of water:

$$2H_2 + O_2 \rightarrow 2H_2O \tag{5.1}$$

This equation means that two molecules of hydrogen (symbol H_2, meaning a molecule contains two atoms of hydrogen, each symbol H) combine with one molecule of oxygen (symbol O_2, which is a molecule that contains two atoms of oxygen, each symbol O). The result of these molecules reacting is to form two molecules of water, a molecule with symbol H_2O, each molecule of which contains two atoms of hydrogen and one of oxygen. Note that there are four hydrogen and two oxygen atoms to the left of the arrow and also four hydrogen and two oxygen atoms to the right of the arrow.

5.3 THE MOL AND THE KMOL

Since molecules are extremely small entities, it takes enormous numbers of them to provide useful amounts of energy for powering automobiles or performing any macroscopic task. So rather than counting molecules by ones or twos, they are counted in very large units called **mols**,[2] or even larger units called **kmols**[3] (thousands of mols).

> *The mol is defined to be the amount of substance containing as many elementary entities as there are atoms in exactly 0.012 kg of pure carbon-12. (The kmol is a factor 10^3 larger.)*

[2]The "mol" is an abbreviation for "mole," which in turn is an abbreviation for the word "molecule." Without a prefix, mol/mole always means a *gram* mole that contains Avogadro's Number of elementary entities.

[3]Note that we can also define a kilogram mole (or kg mole, further abbreviated as kmol) as 1000 mols. Because of its virtually universal acceptance, the fundamental molar unit is the mol rather than the more logical SI-compatible kmol.

Just as a dozen eggs is a way of referring to exactly 12 eggs, a mol is a way of referring to 6.0221367×10^{23} molecules, which is the number of elementary entities in exactly 0.012 kg of carbon. This number of elementary entities is very large indeed and is referred to as **Avogadro's Number**, symbol N_{Av}. Obviously in a kmol, the number of elementary entities is 6.0221367×10^{26}.

Elementary entities may be such things as atoms, molecules, ions, electrons, or other well-defined particles or groups of such particles. The mole unit is thus nothing but an alternate unit to counting individual elementary particles, and it will be useful in the analysis of chemical reactions. Continuing our dozen-egg analogy, the elementary entities might consist of five individual chicken eggs and seven individual turkey eggs. If so, notice that not every egg will have the same mass. The **atomic masses**[4] of some common elements correct to three significant figures are given in Table 5.1. They are measured *relative* to the mass of carbon–12 (written C^{12} or C-12) being exactly $12.0^{\bullet}$ (the superscript $^{\bullet}$ meaning the zero reoccurs to infinite length). In addition, the number of moles $\boldsymbol{n}$ of a substance with a mass $\boldsymbol{m}$ that has a molecular mass $\boldsymbol{M}$ is given by:

$$\boldsymbol{n = m/M} \tag{5.2}$$

Table 5.1 Atomic Masses of Some Common Elements to Three Significant Figures

Hydrogen, H	1.00	Nitrogen, N	14.0
Oxygen, O	16.0	Helium, He	4.00
Carbon, C	12.0	Argon, Ar	40.0
Sulfur, S	32.1	Chlorine, Cl	35.5

Many gases are divalent (i.e., chemically combined as a paired set) such as hydrogen, oxygen, and nitrogen molecules, written H_2, O_2, and N_2, respectively (and their molecular masses are 2.00, 32.0, and 28.0, respectively). Thus, every kmol of water has a mass of approximately 18.0 kg, since the atomic mass of every hydrogen atom is (approximately) 1.00 kg/kmol, and the atomic mass of every oxygen atom is (approximately) 16.0 kg/kmol.

Example 5.1

(a) How many mols of water are in 10.0 kg of water?
(b) How many kmols of water are in 10.0 kg of water?

Need: Number of mols, kmols in 10.0 kg of H_2O.

Know: Atomic masses of O and H are 16.0 and 1.00, respectively.

How: From Equation (5.2), the number of moles $\boldsymbol{n}$ of a substance with a mass $\boldsymbol{m}$ that has a molecular mass $\boldsymbol{M}$ is given by $\boldsymbol{n = m/M}$.

Solve: The molecular mass of water (H_2O) is $M = 2 \times (1.00) + 1 \times (16.0) = 18.0$ kg/kmol $= 18.0$ g/mol.[5] Then for 10.0 kg of water:

(a) $10.0 \text{ kg} = 100. \times 10^2$ g, then $n = m/M = 100. \times 10^2$ [g]/[18.0 g/mol] = **556 mol.**

(b) $n = m/M = 10.0$ [kg]/[18.0 kg/kmol] = **0.556 kmol.**

[4]The term "g atom" (gram atom) sometimes is used to refer to the mass of atoms in Avogadro's number of the particular atom in question so that 1 (k)g atom of oxygen atoms has a mass of 16.0 (k)g. However, the definition of kmol in terms of elementary entities is inclusive, and we will not have to use the g atom terminology.

[5]And equally = 18.0 lbm/lb mole = 18.0 tons/ton mole, etc.

Example 5.2

Determine the effective molecular mass of air assuming it is composed of 79% nitrogen and 21% oxygen.

Need: Molar mass of air with 21% O_2 and 79% N_2. Thus, air is a mixture. A kmol of a mixture of elementary entities must still have Avogadro's number of elementary particles, be they oxygen or nitrogen molecules. Thus, we need the combined mass of just these two kinds of elementary entities in the correct ratio, each of which has a different mass.

Know: Molar mass of O_2 is 32.0 kg/kmol, and the molar mass of N_2 is 28.0 kg/kmol.

How: Proportion the masses of each constituent according to their concentration.

Solve: $\boldsymbol{M}_{\mathbf{air}}$ = fraction $N_2 \times M_{N_2}$ + fraction $O_2 \times Mo_2 = 0.79 \times 28.0 + 0.21 \times 32.0 =$ **28.8 kg/kmol**.

Note: We have defined a kmol of air (even though air molecules *per se* do not exist), but in so doing we have preserved the notion that every mole should have Avogadro's number of entities.

5.4 STOICHIOMETRY

The goal of many important engineering models is determining the energy that can be provided by the combustion of a particular kind of fuel. For example, a previous chapter posed the question of how an automobile engine achieves by combustion the conversion of some 80 kJ per second of chemical energy into the same quantity of thermal and mechanical energy. The model of combustion presented in this chapter will help answer that question.

A first step in the model is writing the sort of symbolic chemical reaction described earlier. A combustion reaction has the following general form:

$$\text{Fuel} + \text{Oxygen} \rightarrow \text{Reaction Products} \tag{5.3}$$

For example, the equation for the formation of water $2H_2 + O_2 \rightarrow 2H_2O$ is a combustion reaction, with hydrogen as the fuel, and water (H_2O) as the reaction product.

For most of the combustion reactions we will consider, the fuel will be a **hydrocarbon**, a mixture of compounds containing only two chemical elements: hydrogen and carbon. For example, natural gas is mostly methane (CH_4), which has four hydrogen atoms bonded to every carbon atom. Gasoline is a mixture of more than a hundred different hydrocarbons, with chains of carbon and hydrogen atoms containing from 4 to 12 carbon atoms. For many purposes it can be conveniently modeled by considering it to consist only of molecules of isooctane, C_8H_{18}, with a molar hydrogen-to-carbon ratio[6] of 18/8 = 2.25.

The energy history of the modern world might be summed up as an increase in the hydrogen-to-carbon ratio of the predominant fuel. Coal, the dominant fuel in the nineteenth century, has about one atom of hydrogen per atom of carbon. The twentieth century saw increasing use of petroleum-based fuels (with about two hydrogen atoms per carbon atom). At the turn of the twenty-first century, wealthy societies rely increasingly on natural gas (with about four hydrogen atoms per carbon atom). A possible goal for the twenty-first century is the hydrogen economy, with an infinite hydrogen-to-carbon ratio (that is, no carbon in the fuel at all). Advocates of this continuing reduction of the carbon content of fuels point out that that combustion of carbon produces carbon dioxide, a major greenhouse gas implicated in global warming. Nuclear, solar, wind, and hydroelectric energy are possible contenders for primary noncarbon energy sources (indeed these are also the first two entries in the National Academy of Engineering List of Engineering Challenges for the twenty-first century).

[6]By mass, the hydrogen-to-carbon ratio is approximately 0.158.

In general, in an engineering analysis we will know the fuel we want to burn and (most of) the reaction products that result. But we will not initially know the number of kilogram moles of oxygen needed to combine with the fuel, and we will not know the number of kilogram moles of each kind of reaction product that will result.

We can express this situation symbolically by assuming that we have just one kilogram mole of fuel and putting undetermined coefficients in front of the other chemicals present to represent those unknown amounts. For the example we have been considering, this would look as follows:

$$H_2 + aO_2 \rightarrow bH_2O \tag{5.4}$$

Determining the numerical value of the coefficients a and b is done by a chemical algebra called **stoichiometry** (Greek for "component measuring"). It relies on the key fact mentioned earlier: The number of each kind of atom in a chemical reaction remains constant. So we simply write an equation for each type of atom that expresses this equality. Then we solve for the unknown coefficients just as we solve any other set of simple algebraic equations. In the reaction shown in Equation (5.4), for example, we have:

Hydrogen equation: $2 = 2b$ (because there are two kmols of hydrogen atoms in the hydrogen molecules on the left, and 2b kmols of hydrogen atoms in the b kmols of water molecules on the right).

Oxygen equation: $2a = b$ (because there are 2a kmols of oxygen atoms in the oxygen molecule on the left and b kmols of oxygen atoms in the b kmols of water molecules on the right).

Solving those two equations with two unknowns yields $b = 1$ and $a = ½$. Substituting this back into our original reaction results in the **stoichiometric equation**:

$$H_2 + 0.5O_2 \rightarrow H_2O \tag{5.5}$$

You will note that this equation looks slightly different from the one that was presented earlier in Equation (5.1) ($2H_2 + O_2 \rightarrow 2H_2O$), but you can reassure yourself that it is actually the same in meaning as the earlier equation by dividing each of the coefficients of that equation by 2. Note, too, that 0.5 kmol of O_2 makes perfectly good sense, whereas there is ambiguity[7] if we think we can equate ½ molecule of oxygen (i.e., $0.5O_2$) with one atom of O.

The process of finding the stoichiometric coefficients can be more difficult as the number of atoms in the participating molecules increases and yet more complicated when there are repeat molecules on both sides of the chemical reaction. Consider some common fossil fuels: natural gas (methane, CH_4), coal[8] (which we will assume is approximately $CHO_{0.1}S_{0.05}N_{0.01}$ + ash), oil $\sim CH_2$, and many others. Most novices incorrectly believe that they can reliably find stoichiometric coefficients by inspection. Experienced engineers will always use a systematic method. We will use a systematic method using tabular entries.

Example 5.3

Common ether, better identified as diethyl ether, can be written as $H_3CH_2COCH_2CH_3$ but can be more simply written[9] as $C_4H_{10}O$. What are the stoichiometric coefficients to burn it completely?

$$C_4H_{10}O + aO_2 \rightarrow bCO_2 + cH_2O$$

[7]We will learn later that the *energetics* of the reaction is dependent on the *state* of the reactants and products; O_2 is markedly different in this respect from 2O just as CO_2 is different from C and O_2.

[8]Coal is very variable and has significantly different composition and properties, depending on its geologic age and its location.

[9]Chemical composition of organic chemicals can be summarized as Hill formulae, in which we first write the number of carbon atoms, then the number of hydrogen atoms, then the rest of the atoms in alphabetical order. What this loses in chemical structure, it makes up in simplicity for retrieval of data pertaining to a complex molecule.

Need: Stoichiometric coefficients, a, b, and c.
Know: You could certainly find a, b, and c by trial and error.
How: The calculation can be systematized into a simple tabular method of finding three equations in the three unknowns a, b, and c, as per Table 5.2.

Table 5.2 Tabular Method of Solving Stoichiometric Combustion Problems

Atoms	LHS[10]	RHS[10]	Solution
C	4	b	b = 4
H	10	2c	c = 5
O	1 + 2a	2b + c	a = b + c/2 – ½ = 6

Solve: a = 6, b = 4, and c = 5. Therefore, $C_4H_{10}O + 6O_2 \rightarrow 4CO_2 + 5H_2O$

Notice you have three unknowns (a, b, and c) and three equations—one for C, one for H, and one for O. Hence you can solve it uniquely by equating the LHS with the RHS for each element.

In this book we exclusively use this tabular method of determining stoichiometric coefficients. One secret is to make sure that simple balances, such as for C and for H in this illustration, each involving only one unknown, should be solved *before* those that involve more than one variable, such as for O in this instance.

Make a quick scan of your solution when complete to confirm that you have an atomic balance—that is, that you have neither created nor destroyed atoms. In this case you can quickly spot that there are 4 C atoms (or kmols, etc.) 10 H atoms and 13 O atoms on both sides of the equation. If there were other elements present, you can use one additional row in the table to balance each element (one equation per unknown).

5.4.1 The Air-to-Fuel Ratio

Although rockets in space chemically react fuel with an oxygen source that they have to haul on takeoff, most combustion reactions on Earth react fuel with the oxygen in the air. In almost all their combustion models, engineers treat air as a mixture that has 3.76 kmols of nitrogen molecules (N_2) for every kmol of oxygen molecules (O_2), ignoring the other molecules such as carbon dioxide, water vapor, and argon that are present in much smaller quantities. This approximate composition of air works out to be 79% N_2 and 21% O_2 expressed equally as volume fractions or mole fractions. (These are the same if Avogadro's law is obeyed.[11])

Many models also incorporate the fact that in typical combustion systems 100–200 parts per million (ppm) of the nitrogen in the air react with oxygen in the air to form nitrogen oxides that are an important cause of acid rain. However, in this introduction to engineering, we will ignore this quantitatively small but important

[10]LHS = left-hand side of the equation and RHS = right-hand side.
[11]After an extensive period of experimentation, the Italian chemist Count Amado Avogadro (1776–1856) proposed in 1811 that equal volumes of different gases at the same temperature and pressure contained equal numbers of molecules. It was not generally accepted by the scientific community until after Avogadro's death. The law is equivalent to claiming that each component in the gaseous mixture behaves as an *ideal gas*, a constraint that is best assumed at low pressures and high temperatures for most gases.

reaction and treat all the nitrogen in products as in its divalent (and inert) form N_2. Thus, the N_2 in air will be considered totally unreactive in this simple treatment of combustion. Thus, too, the N_2 in the reactant air remains unchanged and leaves as N_2 as a product.

In order to design efficient combustion systems, it is important to know the amount of air needed to burn each gallon or kilogram of fuel. This amount of air is determined by the **air-to-fuel ratio**. To determine it, we begin by including those 3.76 kmols of nitrogen that accompany each kmol of oxygen in the stoichiometric equation. This is done simply by multiplying the stoichiometric coefficient for oxygen by 3.76 and using that as the stoichiometric coefficient for nitrogen. Since, in our model, the nitrogen does not take part in the combustion reaction, it *usually*[12] has the same coefficient on both sides of the reaction. Thus:

$$H_2 + [\tfrac{1}{2} O_2 + \tfrac{1}{2} \times 3.76 N_2] \rightarrow H_2O + \tfrac{1}{2} \times 3.76 N_2$$

or

$$H_2 + \tfrac{1}{2}[O_2 + 3.76 N_2] \rightarrow H_2O + 1.88 N_2 \qquad (5.6)$$

The term $O_2 + 3.76N_2$ (in the square brackets) is *not* 1 mol of air, even though only 1 mol of it, the oxygen, is actually participating in the reaction; in fact, it is 4.76 mols of an oxygen plus nitrogen mixture. It can be thought of as being 4.76 mols of air in the sense that 1.00 mol of air, as is any mol, will always contain Avogadro's number's worth of elementary entities regardless of their identity as an oxygen or as a nitrogen molecule, and thus, 4.76 "mols of air" contain $4.76 \times N_{Av}$ elementary particles.

The preceding stoichiometric equation makes it possible to determine the air-to-fuel ratio in two different ways: as a ratio of numbers of molecules (called the **molecular** or **molar** air-to-fuel ratio) and as a ratio of their masses (called the **mass** air-to-fuel ratio). We will use the nomenclature $(A/F)_{molar}$ and $(A/F)_{mass}$ to distinguish between these two dimensionless ratios. Since these ratios have no units, they are best written explicitly as $(A/F)_{molar}$ = [kmol of air/kmol of fuel] and $(A/F)_{mass}$ = [kg of air/kg of fuel], respectively, to emphasize what is being stated.

Example 5.4

Determine the molar and mass stoichiometric air-to-fuel ratios for the combustion of hydrogen.

Need: $(A/F)_{molar}$ = ___ mols of air per mol of hydrogen and $(A/F)_{mass}$ = ____ mass of air per mass of hydrogen.
Know: Stoichiometric equation: $H_2 + \tfrac{1}{2}[O_2 + 3.76N_2] \rightarrow H_2O + 1.88N_2$.
How: Use the stoichiometric equation to ratio mols; then multiply by relative masses of atoms from Table 5.1.

Solve: $$(A/F)_{molar} = \frac{\tfrac{1}{2}(1 + 3.76)}{1} = \mathbf{2.38\ [kmol\ air/kmol\ H_2]}$$

To express the *mass* ratio, simply assign the molecular masses to each component—that is, $\boldsymbol{(A/F)_{mass}}$ = ½ × (1 × 32.0 + 3.76 × 28.0)/(1 × 2.00) **{**[kmol O_2] [kg O_2/kmol O_2] + [kmol N_2][kg N_2/kmol N_2]**}**/[kmol H_2] [kg H_2/kmol H_2] = **34.3 [kg air/kg H_2]** of mass in which we have identified the air term as that in braces { }.

Since the unit method is clumsy in this example, we can use the concept of a molar air mass of M_{Air} = 28.8 kg air/kmol air from Example 5.2 to simplify the calculation.

Hence, $(A/F)_{mass}$ = 2.38 × 28.8/(1 × 2.00) [kmol air][kg air/kmol air]/[kmol H_2][kg H_2/kmol H_2] = **34.3 [kg air/kg H_2]** (as earlier).

[12] The exception being the fuel-bound nitrogen case—e.g., burning ammonia gas according to $NH_3 + 1.5[O_2 + 3.76N_2] \rightarrow 3H_2O +$ **6.64** N_2 (and not **5.64** N_2).

The fuel-to-air ratios are simply the inverses of the air-to-fuel ratios:

$$(F/A)_{molar} = 1/(A/F)_{molar} = 1/2.38 = 0.42 \text{ [kmol H}_2\text{/kmol air]}$$
$$(F/A)_{mass} = 1/(A/F)_{mass} = 1/34.2 = 0.029 \text{ [kg H}_2\text{/kg air]}$$

The molar and the mass air-to-fuel ratios have different uses in engineering analysis. For example, the molar air-to-fuel ratio is useful to a design engineer determining the volume of air involved[13] in a combustion process and therefore is the basis for the dimensions of air intake and exhaust passages. The mass air-to-fuel ratio is useful to the engineer who wants to calculate how much fuel to provide, perhaps in estimating a vehicle's fuel economy. An environmental engineer may also calculate the masses of different types of pollutants that might result from a combustion process.

The stoichiometric fuel-to-air ratio is invariant for a given fuel composition, and actual fuel-to-air ratios can be varied by an engineer deciding just how much air or fuel to add. The term **equivalence ratio**, symbol φ (Greek phi), is defined as $\varphi = (F/A \text{ actual})/(F/A \text{ stoichiometric})$; it is useful because it quickly tells you the overall conditions of the burn. Stoichiometric combustion means that the equivalence ratio is 1.0. In the jargon of the combustion engineer, excess fuel is described as *rich* and excess air as *lean*—for example, equivalence ratios of $\varphi = 1.05$ and $\varphi = 0.95$, respectively.

5.5 THE HEATING VALUE OF HYDROCARBON FUELS

Chemistry can also answer the question, "How much energy can be obtained by the combustion of one kilogram of fuel under stoichiometric conditions?" That in turn is a key part of answering the question, "How many miles per gallon of fuel is it possible for an automobile or truck to travel?" Let us briefly consider the first of those two questions.

In this section we will discuss only a simplified method and leave the general case for the end of this chapter. To determine the amount of energy that can be obtained by the combustion of one kilogram of fuel under stoichiometric conditions, the type of fuel being used must be specified. Assume that fuel is composed only of a molecule with x atoms of carbon and y atoms of hydrogen, and its combustion is thus described by the stoichiometric equation,

$$C_xH_y + (x + (y/4))O_2 \rightarrow xCO_2 + (y/2)\,H_2O \tag{5.7}$$

We can think of this as taking place in three imaginary steps: The first is the decomposition of the fuel to its constituent atoms, and the second and third are their respective combustion reactions to the stable oxides CO_2 and H_2O. Hence, the conceptual model says that:

$$C_xH_y \rightarrow xC + yH \ \ldots \text{ is followed by:}$$

$$xC + xO_2 \rightarrow xCO_2 \ \ldots \text{ and}$$

$$yH + y/4O_2 = (y/2)H_2O$$

The energy released must be that released by burning x moles of carbon atoms and y moles of hydrogen atoms *less* the energy in step 1 to dissociate the original fuel into its elements xC and yH.

[13] Recall Avogadro's law that moles and volumes are proportional to each other.

Some fuels have different molecular structure than others—for example, aromatic fuels such as benzene (C_6H_6) form hexagonal molecules and are structurally unlike the more linear aliphatic fuels such as isooctane (C_8H_{18}) and as cetane ($C_{16}H_{34}$). We are mostly interested in common transport fuels such as gasoline (as modeled by isooctane) and by diesel fuel (as modeled by cetane).

We will simplify our study of the energy content of hydrocarbon fuels by *asserting* that our typical hydrocarbon fuel has an average heating value of 45,500 kJ/kg irrespective of precise composition (see Table 5.3). This says the energy to dissociate each step of these hydrocarbons is about equal. This is obviously not true of the smallest hydrocarbon member, CH_4, or indeed if the fuels also contain significant oxygenates (such as alcohols), but it is sufficiently accurate for our current purposes and can easily be corrected.

Table 5.3 Heating Values of Hydrocarbon Fuels

Fuel	Heating Value, kJ/kg
Methane, CH_4	55,650
Propane, C_3H_8	46,390
Isobutane, C_4H_{10}	45,660
Gasoline, C_8H_{18}	45,560
Diesel fuel, $C_{16}H_{34}$	43,980
Benzene, C_6H_6	42,350
Toluene, $C_6H_5CH_3$	42,960

The amount of oxygen consumed by our hydrocarbon, C_xH_y, in Equation (5.7) is:

$$\textit{Mass of } O_2 = \frac{32.0(x + {}^y\!/_4)}{(12.0\,x + 1.00\,y)}\ \frac{\text{kg } O_2}{\text{kg Fuel}}$$

The amount of carbon dioxide produced is:

$$\textit{Mass of } CO_2 = \frac{44.0\,x}{(12.0\,x + 1.00\,y)}\ \frac{\text{kg } CO_2}{\text{kg Fuel}}$$

The amount of water produced is:

$$\textit{Mass of } H_2O = \frac{9.00\,y}{(12.0\,x + 1.00\,y)}\ \frac{\text{kg } H_2O}{\text{kg Fuel}}$$

Thus, for typical gasoline-like compounds $x = 8$, $y = 18$, one kilogram of this compound of hydrogen and carbon reacts with 3.5 kilograms of oxygen to release 45,500 kilojoules of energy and to produce 3.1 kg of carbon dioxide and about 1.4 kg of water. This knowledge enables us, to a first approximation, to determine how many miles it is possible for an automobile to travel on one gallon of fuel. That calculation is one of the examples in this chapter.

5.5.1 The Heating Value of Fuels: The General Case

The bottom line of our stoichiometric chemistry is answering the question, “How much energy can be obtained by the combustion of one kilogram of *any* fuel under stoichiometric conditions?” So far in this book, we have asserted an approximation for hydrocarbons that the heating value HV = 45,500 kJ/kg. However, nonhydrocarbon molecules do *not* obey this simple rule, as shown in Table 5.4.

Table 5.4 Some Nonhydrocarbon Fuels	
Fuel	**Heating Value, kJ/kg**
Ethyl alcohol, C_2H_5OH	27,904
Hydrogen, H_2	120,000
Carbon monoxide, CO	10,100
Carbon, C	32,800

Why are these fuel heating values different from the hydrocarbons? What is essentially different is that the energy consumed in making these fuels from their constituent atoms is not equal to the corresponding energy in making hydrocarbons. In the strict language of thermodynamics, the **heat of formation** of a compound is the heat of reaction when that compound is formed from its elements at the same temperature and pressure. In the case of elements such as H_2, O_2, N_2, and plain carbon, their energy of formation is taken as zero.

The numbers in Table 5.5 reflect the exothermicity (heat out) or endothermicity (heat in[14]) of the (often imaginary[15]) act of putting these molecules together starting with their elements (e.g., C and ½ O_2 in the case of CO, and simply H_2 in the case of H_2,[16] etc.). As long as the energies are all calculated on the same basis (such as at the same temperature and pressure), these heats of formation can be comingled.

Table 5.5 Heats of Formation of Various Fuels ((s) means solid, (g) means gas, and (l) means liquid)	
Substance	**kJ/kg**
Carbon, C (s)	0
Nitrogen, N_2 (g)	0
Oxygen, O_2 (g)	0
Hydrogen, H_2 (g)	0
Carbon Monoxide, CO (g)	–3,946
Carbon Dioxide, CO_2 (g)	–8,942
Water, H_2O (g)	–13,423
Methane, CH_4 (g)	–4667
Acetylene, C_2H_2 (g)	+8,720
Hexane, C_6H_{14} (g)	–1,945
Ethanol, C_2H_5OH (l)	–5,771
Benzene, C_6H_6 (l)	+629

[14]For reasons we will ignore, exothermicity is negative and endothermicity is positive.
[15]Usually we have to devise an indirect chemical path to make these molecules from their constituents.
[16]This is why its energy of formation is zero.

The heat of combustion or the heating value of the fuel[17] is defined as:

$$\textbf{Heating Value} = \textbf{Energy in fuel and oxidizer} - \textbf{Energy in products of combustion} \quad (5.8)$$

The energy in the fuel and its combustion products are their respective heats of formation, so we can use the heats of formation in Table 5.5 in conjunction with the stoichiometric method we used earlier. In order to determine a heating value of a fuel, we need to know (1) the amount and type of fuel used and (2) the composition of the combustion products. To do this we need to determine the chemical reaction for the combustion process. This is done by calculating the stoichiometric coefficients of each chemical in the reaction on both sides of the reaction equation using the tabular method. For example, in the combustion of ethanol in pure oxygen, the combustion reaction equation is:

$$C_2H_5OH + 3O_2 \rightarrow 2CO_2 + 3H_2O + \text{``Heat''}$$

Table 5.6 Heating Value Method

Item	LHS	LHS	RHS	RHS
Molar quantities	C_2H_5OH (l)	O_2 (g)	CO_2 (g)	H_2O (g)
Mass, kg	46.0	3 × 32.0	2 × 44.0	3 × 18.0
Mass m, kg/kg of fuel	1.00	2.087	1.913	1.174
ΔH_f, kJ/kg	–5,771	0.000	–8,942	–13,423
m × ΔH_f kJ/kg of fuel	–5,771	0.000	–17,106	–15,759

To systematically carry out this process of determining the heating value of fuels, yet another tabular method is suggested in Table 5.6; indeed, it should also be useful for practicing engineers, since it is a visual and systematic way to avoid algebraic (and particularly sign) errors. Using Equation (5.8), Heating Value = Energy in fuel and oxidizer − Energy in the products of combustion (see Table 5.6), then

$$HV = -5,771 - (-17,106 - 15,759)$$

or

$$HV = +27,100 \text{ kJ/kg of fuel}$$

This method is exact, since we are no longer hiding the heat of formation of different fuels so that non-hydrocarbons are dealt with on the same basis as we can use for simple hydrocarbons.

The original method to determine the heating value of a hydrocarbon fuel was to assume it is fixed at 45,500 kJ/kg; our updated more general method takes into account the energy of formation of the fuel. In effect, our original approximation is valid because the energy to form $H_2C—CH_2$ bonds does not vary much in simple hydrocarbons; thus, the energy of formation for these bonds per mass of CH_2 also does not vary much either. However, for other fuels where the energy of formation is quite different, the more complicated calculations are then required.

[17]There are two common conventions in thermodynamics in which the heat of combustion is positive in the engineering literature and negative in the scientific literature. Both systems are logically consistent.

5.6 HOW DO YOU MAKE CHEMICAL FUELS?

Why don't you pull your car up to an oil well and fill it up? Perhaps you don't have an oil well in your back yard. But suppose that you did. Could the oil that comes directly out of the ground run your car? The short answer to the question is no. Your car would sputter to a stop if filled up from an oil well. In response to that negative answer, chemical engineers design oil refineries that convert crude oil from the well into gasoline at the pump (and many other useful and related compounds). Gasoline is simply one component of crude oil that has been tailored to keep your car running smoothly and efficiently.

We have seen that chemical energy typically is supplied to automobiles mostly in the form of hydrocarbons. These molecules, made of hydrogen and carbon atoms, combine with oxygen to yield carbon dioxide, water, and energy in the form of heat. The most convenient sources of hydrocarbons lie concentrated in deposits under the ground, the remains of animals and plant matter that lived hundreds of millions of years ago. The name for these hydrocarbon deposits, crude oil, encompasses a wide range of substances. At one extreme, hydrocarbon deposits contain just light gases containing only a few carbon atoms per molecule of hydrocarbon. Sometimes just one carbon atom will join to four hydrogen atoms forming a hydrocarbon molecule called methane or "natural gas." At the other extreme, crude oil contains a thick viscous sludge[18] with molecules containing dozens of carbon and hydrogen atoms. Some typical large hydrocarbon molecules are shown schematically in Figure 5.1.

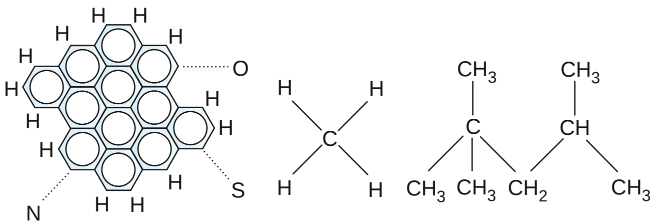


FIGURE 5.1 Some of the Molecules Present in Crude Oil

Crude oil from an oil well would fill your tank with an impractical mixture of molecules. The heavy viscous ones would clog up the fuel system and they would not ignite in the combustion chamber. Crude oil also contains significant amounts of foul-smelling, smog-producing, sulfur and nitrogen compounds. But the lightest hydrocarbons in crude oil evaporate easily and would explode readily. A better choice is the Goldilocks' solution: fill the tank with molecules that are not too big, not too small, but just right. For automotive engines this means molecules that have about 5 to 10 carbon atoms and 10 to 25 hydrogen atoms. This class of molecules includes the compounds collectively called gasoline. The ideal molecule for providing chemical energy to the Otto Cycle, the one used as a standard, is **isooctane**, whose composition is given by C_8H_{18}.

The principal tool introduced earlier in this chapter is **stoichiometry**, the conservation of matter measured in appropriate aggregates. It is used with control boundaries chosen to emphasize input and output streams of various processes. In this section, we will analyze a basic chemical engineering process, a highly simplified process of **distillation**.

[18]Some crude oils are even solids at room temperature.

Example 5.5

An oil deposit contains crude oil that is a mixture of, by **moles** (number of molecules), 7.0% C_5H_{12}, 33.% C_8H_{18}, 52.% $C_{16}H_{34}$, and 8.0% $C_{38}H_{16}$. What is the percentage of each substance by mass?

Need: By mass C_5H_{12} =____%, C_8H_{18} = ___ %, $C_{16}H_{34}$= ____%, and $C_{38}H_{16}$= ____%.
Know–How: Imagine a total of 100 kmols. Divide them among the various species.

$$7.0 \text{ kmols } C_5H_{12},\ 33.\text{ kmols } C_8H_{18},\ 52.\text{ kmols } C_{16}H_{34}, \text{ and } 8.0 \text{ kmols } C_{38}H_{16}.$$

Determine mass of each species by multiplying number of [kmols] by molecular mass in [kg/kmols].

$$7.0 \times (60 + 12) = 500 \text{ kg } C_5H_{12}$$

$$33. \times (96 + 18) = 3{,}800 \text{ kg } C_8H_{18}$$

$$52. \times (192 + 34) = 12{,}000 \text{ kg } C_{16}H_{34}$$

$$8.0 \times (456 + 16) = 3{,}800 \text{ kg } C_{38}H_{16}$$

For a total of 500 + 3,800 + 12,000 + 3,800 = 20,000 kg.

Solve: By mass

$$C_5H_{12} = (500/20{,}000) \times 100 = 2.5\%$$

$$C_8H_{18} = (3{,}800/20{,}000) \times 100 = 19.\%$$

$$C_{16}H_{34} = (12{,}000/20{,}000) \times 100 = 60.\%$$

$$C_{38}H_{16} = (3{,}800/20{,}000) \times 100 = 19.\ \%$$

(Use a tabular format to solve this problem—it's especially helpful to use Excel.)

5.6.1 Process Engineering

In the early decades of the twentieth century, a new idea emerged. Chemical engineers, such as Arthur D. Little, an independent consultant, and W.W. Lewis, a professor at MIT, looked at chemical manufacturing plants in a new way. They viewed the plants, however complicated, as arrangements of simple elements. These elements, which served as a sort of alphabet for chemical plant construction, came to be known as **process steps** and analyzed using the concept of *Unit Operations*. It turned out that many process steps seemed quite different, but were closely related in their basics. Thus, distillation, absorption, stripping, and liquid extraction are all described by essentially the same mathematics.

Distillation

The first major process in most refineries is a set of distillation columns (see Figure 5.2). This is a purely *physical* separation process that relies on the different relative volatilities among the various components of crude oil. Generally they are arranged in sequence with the most volatile components removed first (such a methane, pentane, isooctane, etc.), and in the last distillation columns the least volatile such as asphaltenes are removed. But, because distillation is a physical treatment, the various *molecules* retain their identity. This is *not* true of chemical treatments in which the molecules are broken and reformed into other molecules, such as in a combustion process.

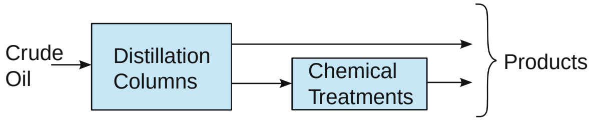


FIGURE 5.2 Process Diagram for a Refinery

Distillation is conceptually a simple process operation. The key unit operation is the fact that condensed liquid (**condensate**) flows down the distillation column while vapor of a different composition flows up (Figure 5.3).[19]

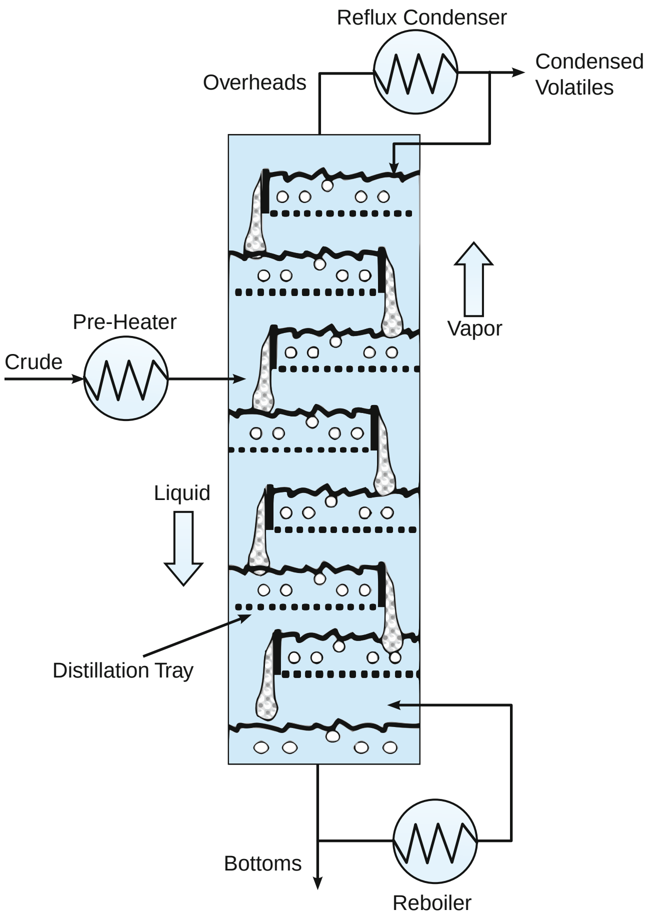


FIGURE 5.3 Schematic of a Typical Distillation Column

[19]The driving force for distillation is the difference in composition between the liquid condensate and the surrounding vapor. A distillation column is made up of pipes, pumps, and heat exchangers, plus a large vertical column containing trays. The trays temporarily hold back the liquid condensate to expose it to up-flowing vapor of a different composition. The vapor contacts the liquid by flowing

Two process diagram symbols for a distillation unit are shown in Figure 5.4. The more detailed one is the more literal including, as it does, the reflux condenser and the reboiler; however often we do not care about the internal flows and we just need the inlet and outlet flow streams. Again, you should imagine the distillation column symbol as surrounded by a control surface dotted line, penetrated by the arrows denoting the flows. That control surface reminds us over enough time, the mass flow into a distillation unit must equal the mass flow out. For these two control surfaces, the internal flows are irrelevant and the two diagrams are completely equivalent.

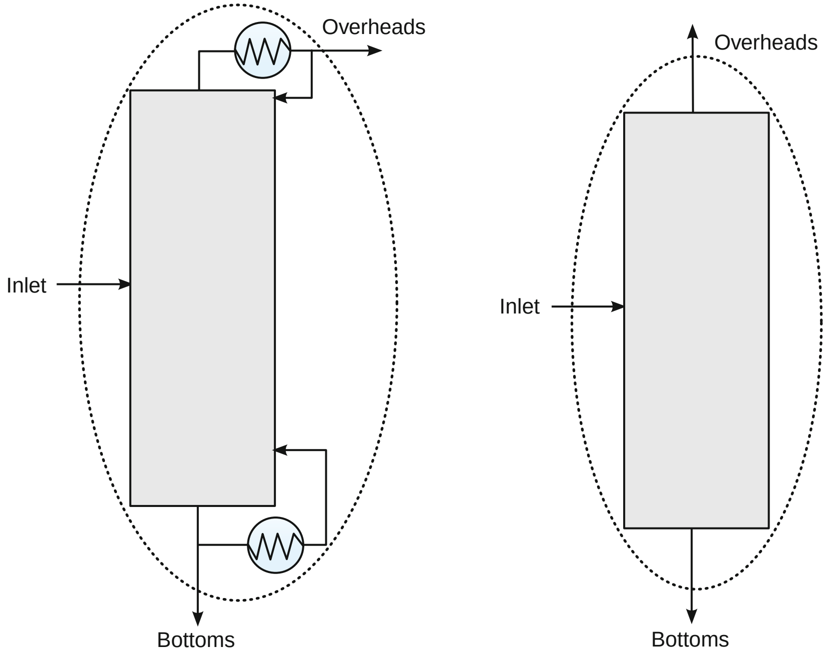


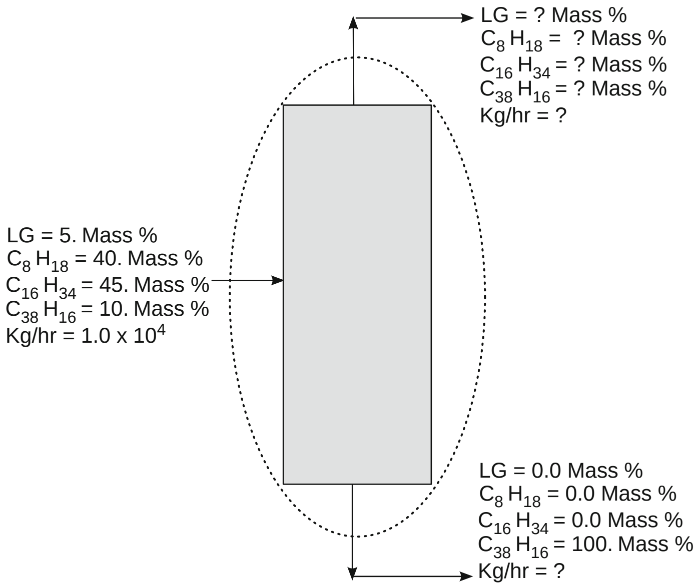
FIGURE 5.4 Process Diagram of a Distillation Column

through an arrangement that does not allow passage of the liquid. For the feed, a heat exchanger adds enough heat to evaporate some of the crude oil into vapor. At the middle of the column, the hot crude oil and some of its vapor are pumped in. The vaporized components go up the column and are washed by condensate flowing down in order to extract any remaining amounts of the less volatile components. The condensate is produced by condensing the vapor at the top of the column by a reflux condenser (a heat exchanger). The more of the vapor that is condensed at the top of the column, the purer the overheads become. The bottom part of the column operates similarly. Condensate runs down the column and contacts up-flowing hot vapor with the result that additional volatile components are extracted from it. This vapor is produced by heating the product bottoms in a reboiler (just another heat exchanger). The more of the bottoms that are heated and vaporized, the more volatiles are stripped from the bottoms. The overhead product at the top of the column is then richer in the volatile components and the bottoms product is richer in the heavier components than the original crude oil. Although 100% separation can never be achieved, we can get close by having many trays and large reflux condensers and reboilers.

Example 5.6

A crude oil contains three components by mass: 10.% asphaltenes ($C_{38}H_{16}$), 5.0% light gases, 40.% isooctane (C_8H_{18}) and 45.% cetane ($C_{16}H_{34}$). An input stream of 1.00×10^4 kg/hr of this crude oil is fed to the distillation column in an oil refinery. Assume the distillation process separates out a top and a bottom stream. If the bottoms consist of 100% asphaltenes, what are the flow rates and composition of all the process lines?

Need: Composition and rate of flow of each component in each line.



Know: Principle of conservation of mass across control volume. Also that the distillation process is a physical one and therefore the type and quantities of molecules are preserved.

How: Use a process diagram showing inputs and outputs of each stream and calculate the required unknowns.

Solve: Since we know that the mass flow rate of the input stream is 1.00×10^4 kg/hr of crude oil, we can easily calculate the amount of each component in the feed:

Inlet mass flow rate of light gases (LG) $= (5.0/100) \times 1.00 \times 10^4 = 5.0 \times 10^2$ kg/hr
Inlet mass flow rate of isooctane $= (40./100) \times 1.00 \times 10^4 = 4.0 \times 10^3$ kg/hr
Inlet mass flow rate of cetane $= (45./100) \times 1.00 \times 10^4 = 4.5 \times 10^3$ kg/hr
Inlet mass flow rate of asphaltenes $= (10./100) \times 1.00 \times 10^4 = 1.0 \times 10^3$ kg/hr

In this case the bottoms are just the asphaltenes, so the total overheads are $1.0 \times 10^4 - 1.0 \times 10^3 = 9.0 \times 10^3$ kg/hr, and the percent composition can be deduced from the inlet mass percentages as:

mass % of light gases (LG) in the overheads is $(5.0 \times 10^2)/(9.0 \times 10^3) \times 100 = 5.6\%$
mass % of isooctane in the overheads is $(4.0 \times 10^3)/(9.0 \times 10^3) \times 100 = 44.\%$,
mass % of cetane in the overheads is $(4.5 \times 10^3)/(9.0 \times 10^3) \times 100 = 50.\%$

Flow Stream	Percent by Mass			
	Light Gases	Isooctane	Cetane	Asphaltenes
Inlet	5.0%	40.%	45.%	10.%
Overheads	5.6%	44.%	50.%	0.0%
Bottoms	0.0%	0.0%	0.0%	100%

Flow Stream	Flow Rate, kg/hr			
	Light Gases (LG)	Isooctane	Cetane	Asphaltenes
Inlet	5.0×10^2	4.0×10^3	4.5×10^3	1.0×10^3
Overheads	5.0×10^2	4.0×10^3	4.5×10^3	0.0
Bottoms	0.0	0.0	0.0	1.0×10^3

Notice the overhead stream is not pure; it consists of three components in the ratios of light gases : isooctane : cetane = 5.6 : 44.4 : 50. by mass. Since these components are potentially each specialized (and valuable) products, they would be further refined in standard refinery practice. Normally a series of several distillation columns would follow.

SUMMARY

Determining useful engineering parameters, such as the miles-per-gallon achievable by an automobile, requires calculating the amount of thermal energy (heat) available from the chemical energy contained in a fuel. This calculation requires the mastery of several concepts from the field of chemistry. Those concepts include **atoms**, **molecules**, and **chemical reactions**; a name for a very large number of atoms or molecules called the **kmol**; a chemical algebra method called **stoichiometry** for keeping track of the numbers of atoms that occur in those reactions; a measure of the amount of air needed for combustion called the **air-to-fuel ratio**; and finally, a way to understand the **heating value** of that fuel.

The term **equivalence ratio**, φ, defined as **(*F*/*A* actual)/(*F*/*A* stoichiometric)**, is used to quickly assess combustion conditions: a fuel-rich process operates at $\varphi > 1$ and a fuel-lean one at $\varphi < 1$.

The principles of stoichiometry and stoichiometric balances has further utility when applied to physical, as opposed to chemical, processes. Many oil refinery processes fall into this category including distillation to separate useful products from crude oil.

EXERCISES

1. How many kmol are contained in 3.0 kg of ammonia, NH_3? (**A: 0.18 kmols**)
2. How many kmol are contained in 1.0 kg of nitroglycerine, $C_3H_5(NO_3)_3$?
3. What is the mass of 5.0 kmol of carbon dioxide, CO_2? (**A:** $\mathbf{2.2 \times 10^2}$ **kg**)
4. What is the mass of 1.00 kmol vitamin B1 disulfide, $C_{24}H_{34}N_8O_4S_2$?

5. The effective molecular mass of air is defined as the mass of a kmol of elementary particles of which 78.09% are nitrogen molecules, 20.95% are oxygen molecules, 0.933% are argon atoms, and 0.027% are carbon dioxide molecules. What is the effective molecular mass of air? (Watch your significant figures!) What other factor could affect the effective molecular mass of air?

6. A gallon of gasoline has a mass of about 3.0 kg. Further, a kg of gasoline has an energy content of about 45,500 kJ/kg. If an experimental automobile requires just 10. kW of power to overcome air resistance at steady speed of 30. miles an hour, and *if* there are no other losses, what would the gas mileage of the car be in miles per gallon? (**A: 110 mpg to only two significant figures as the problem is stated**)

7. Determine the value of the stoichiometric coefficients for the combustion of an oil (assumed molecular formula CH_2) in oxygen: $CH_2 + aO_2 = bCO_2 + cH_2O$. Confirm your answer is correct! (**A: a = 1.5, b = 1, c = 1**)

8. Determine the value of the stoichiometric coefficients for the combustion of coal in oxygen given by the stoichiometric equation:

$$CHN_{0.01}\,O_{0.1}S_{0.05} + aO_2 = bCO_2 + cH_2O + dN_2 + eSO_2$$

9. Determine the value of the stoichiometric coefficients for the combustion of natural gas in air:

$$CH_4 + a(O_2 + 3.76N_2) = bCO_2 + cH_2O + dN_2$$

10. Using the stoichiometric coefficients you found in exercise 9, determine the molar air-to-fuel ratio $(A/F)_{molar}$ for the combustion of natural gas in air. (**A: $(A/F)_{molar}$ = 9.52 kmol of air/kmol of fuel**)

11. Using the results of exercises 9 and 10, determine the *mass* air-to-fuel ratio $(A/F)_{mass}$ for the combustion of natural gas in air. (**A: 17.2 kg of air/kg of fuel**)

12. Determine the mass air-to-fuel ratio $(A/F)_{mass}$ for the combustion of an oil (represented by CH_2) in air.

13. Determine the mass air-to-fuel ratio $(A/F)_{mass}$ for the combustion of coal in air represented by the equation:

$$CHN_{0.01}O_{0.1}S_{0.05} + a(O_2 + 3.76N_2) = bCO_2 + cH_2O + dN2 + eSO_2$$

14. Determine the mass air-to-fuel ratio $(A/F)_{mass}$ for the combustion of isooctane C_8H_{18} in air. Its stoichiometric equation is

$$C_8H_{18} + 12.5(O_2 + 3.76N_2) = 8CO_2 + 9H_2O + 47.0N_2$$

Exercises 15 through 20 use engineering considerations to give insight on the global warming issue.

15. Assume one kg of isooctane produces 45,500 kJ of energy. How much carbon dioxide is released in obtaining a kJ of energy from the combustion of isooctane in air? (**A: 6.79×10^{-5} kg CO_2/kJ**)

16. Assume a kg of hydrogen in combustion produces 1.20×10^5 kJ of thermal energy. How much carbon dioxide is released in obtaining a kJ of energy from the combustion of hydrogen in air?

17. The amount of carbon dioxide levels in the atmosphere has been increasing for many decades. Many scientists believe that the increase in carbon dioxide levels could lead to global warming, and might trigger abrupt climate changes. You are approached by a future US president for suggestions on what to do about this problem. Based on your answers to exercises 15 and 16, what might be the *simplistic* approach to reducing the rate of increase of the amount of carbon dioxide in the atmosphere due to combustion in automobiles and trucks? (**A: Replace isooctane with hydrogen.**)

18. Give three reasons why your suggested solution in exercise 17 to global warming has *not* already been adopted.

19. A monomeric formula for wood cellulose is $C_6H_{12}O_6$ (it repeats many times as a hexagonal structure based on this formula). The energy content by burning wood is approximately 1.0×10^4 kJ/kg. Determine the amount of carbon dioxide released in obtaining a kJ of energy by the combustion of wood in air. (**A: 1.5×10^{-4} kg of CO_2 released/kJ of energy produced**)

20. Repeat exercise 19 for the combustion of coal. For simplicity assume coal is just carbon, C, with a HV of 32,800 kJ/kg.

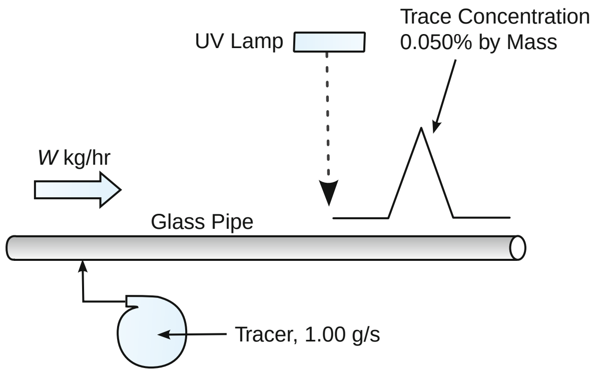
21. Given the heat of formation of liquid methanol, $CH_3OH(l)$, is −238,000 kJ/kmol, what is its heat of combustion in kJ/kg?

22. In 1800 the main fuel used in the United States was wood. (Assume wood is cellulose, whose representative repeating formula is $C_6H_{12}O_6$.) In 1900, the main fuel used was coal (assume coal in this example can be approximated by pure carbon, C). In 2000, the main fuel used is oil (assume it can be represented by isooctane). Assume that in 2100 the main fuel will be hydrogen (H_2). Use a spreadsheet to prepare a graph of carbon dioxide released per kJ of energy produced (y-axis) as a function of year (x-axis) from 1800 to 2100.

23. A pipe carries a mixture of, by kmols (which are proportional to the number of molecules), 15.% C_5H_{12}, 25.% C_8H_{18}, 50.% $C_{16}H_{34}$, and 10.% $C_{32}H_{40}$. What is the percentage of each substance by mass?

24. A crude oil contains just three components by weight: 15.% asphaltenes, 1.0% light gases (containing about 1 – 5 carbon atoms + hydrogen) and the remainder light distillate that is pure isooctane. This crude is fed to the distillation columns in an oil refinery at a rate of 1.0×10^4 kg/hr. Assume the distillation process perfectly separates the stream of gas between a mixture of light gas and isooctane in the top stream. Draw a process diagram of this distillation column and indicate on the drawing the flow rates and composition by component of the output streams.

25. The glass pipe system shown below has a liquid flowing in it at an unknown rate. The liquid is very corrosive and flow meters are correspondingly very expensive. Instead you propose to your boss to add a fluorescent dye via a small pump and to deliver 1.00 g/s of this tracer and activate it by UV light. By monitoring the dye fluorescence downstream, you determine its concentration is 0.050 wt. %. What is the unknown flow rate in kg/hr? (This is a useful method of indirectly measuring flows.)



26. You have worked for a petroleum company producing automotive fuels for several years, and in your spare time at home you have developed a new fuel composition that has a higher heating value than ordinary gasoline. When you began work at this company, you signed a confidentiality agreement that gave them all of your intellectual property. What are your obligations to your employer?

a. It is your work on your time, so you have no obligations to your current employer.

b. Everything you used to develop your new fuel you learned on the job at your company, so your work really belongs to them.

c. Your company's confidentially agreement requires you to provide them with all your work.

d. Give your idea to a friend and let him or her pursue it while keeping you as a silent partner.

Give your solution in Engineering Ethics Matrix form.

27. As a production engineer for a large chemical company, you need to find a new supplier for a specific commodity. Since this contract is substantial, the salespeople with whom you meet are naturally trying to influence your purchasing decision. Which of the following items are ethical in your opinion?[20]

a. Your meeting with a salesperson extends over lunch, and he or she pays for the lunch.

b. In casual conversation at the sales meeting you express an interest in baseball. After the meeting, a salesperson sends you free tickets to your favorite team's game.

c. After your meeting, a salesperson sends a case of wine to your home with a note thanking you for the "useful" meeting.

d. As a result of the sales meeting, you are invited on an all-expense-paid trip to China to visit the salesperson's manufacturing facility.

Give your solution in Engineering Ethics Matrix form.

28. You are attending a national engineering conference as a representative of your company. A supplier to your company has a booth at the conference and is passing out small electronic calculators to everyone who comes to their booth. The calculators are valued at about $25, and you are offered one. What do you do?

a. Accept the gift and, since it has small value, there is no need to report it.

b. Accept the gift and report it to your supervisor.

[20]How do you distinguish between a token gift and a bribe? Where do you draw the line?

c. Decline the gift, explaining that you work for one of their customers.
d. Ask someone else to get one of the calculators for you.
Give your solution in Engineering Ethics Matrix form.

29. It is December 1928 and your name is Thomas Midgley, Jr.[21] You have just invented a new miracle refrigerant composed of chlorinated fluorocarbons (CFCs) that will make your company a lot of money. At midnight on Christmas Eve you are visited by three spirits who show you your past, present, and future. In your future you discover that the chlorine in your miracle refrigerant will eventually destroy the Earth's ozone layer and put the entire human race at risk. What do you do?
a. Claim the entire vision was due to a bit of undigested meat and forget it?
b. Run to the window and ask a passerby to go buy a large goose?
c. Destroy all records of your CFC work before anyone can use it?
d. Start work developing nonchlorinated hydrocarbon refrigerants?
Give your solution using the Engineering Ethics Matrix.

30. It is 2021 and you are a process engineer at a large oil refining company. The world is rapidly moving toward a hydrogen energy economy and your company has been trying to develop an efficient and cost-effective way to extract hydrogen from crude oil. You have found a low-cost process, but it produces a considerable amount of undesirable chemical by-product pollutants. However, during your work on this project you also have discovered an effective and inexpensive way to extract hydrogen directly from seawater. You realize that revealing this process effectively would eliminate the world demand for petroleum and probably would cause serious financial damage to your company. What do you do?
a. Quit your job and start your own hydrogen producing company.
b. Talk to your supervisors and reveal your process to them to see if they wish to pursue implementing it as part of their company.
c. Contact a patent lawyer not associated with your current employer and try to patent this potentially lucrative new process.
d. Without your employer's permission, publish an article in a well-read chemical or energy magazine revealing your process and giving it to the world free of charge.

Give your solution using the Engineering Ethics Matrix.

[21]See an abbreviated story of his life at http://www.uh.edu/engines/epi684.htm

CHAPTER

Mechanical Engineering

6

Source: Courtesy of Daimler-Chrysler

Who are **Mechanical Engineers** and what do they do? Many of the engineering disciplines described in this text are sufficiently specialized that a brief definition of their principal activities can be captured quickly. However, mechanical engineering is one of the oldest and possibly the most diverse engineering disciplines and cannot be described so easily.

Because this is an introductory textbook, the engineering disciplines are described in a single chapter. However, there are several chapters with topics that overlap mechanical and other engineering disciplines. For example, Chapter 13, "Engineering Kinematics"[1] later in this text is common to both the mechanical and civil engineering disciplines.

Mechanical engineering traditionally covers the broad fields of energy conversion, properties and use of materials, and machine design. Consequently, mechanical engineers need a solid understanding of kinematics, fluid mechanics, solid mechanics, energy, and design. Mechanical engineers use these principles in the design and analysis of mechanical systems such as cars, trucks, aircraft, ships, spacecraft, industrial machinery, heating and cooling systems, medical equipment, and much more.

In this introductory chapter we look at the energy conversion side of mechanical engineering by studying the operation of automobile engines (the **Otto cycle**), truck engines (the **Diesel cycle**), and aircraft and ship engines (the **Brayton cycle**). We will also model the **power output** of an Otto cycle and consider the use of that power output to achieve **motion**. We will conclude by considering some proposed improvements to these engines, involving either making them better or replacing them with a different type of propulsion source.

6.1 THE OTTO CYCLE

Central to the drive train of the modern automobile is the type of engine developed in the 1860s by a French engineer named Alphonse Eugene Beau de Rochas and then independently reinvented by a German engineer named Nikolaus Otto in 1876. The Otto cycle, as this invention is now universally called, has been victorious

[1]We dealt with *kinetics* in an earlier chapter; it involved the response of mass to applied forces. *Kinematics* on the other hand is the study of relationships among distance, time, speed, and acceleration without regard to the forces.

so far over rival ideas for powering automobiles ranging from the electric-and-steam powered propulsion systems of the early twentieth century to the gas turbine and rotary Wänkel engine of the 1960s and 1970s and recently reintroduced by Mazda. In the process, the Otto cycle has been improved through inventions ranging from the float-feed carburetor in the 1890s to the computer-controlled fuel-injection systems of today. This chapter will provide engineering models for understanding this ubiquitous and essential element of modern civilization.

6.1.1 Operation of the Otto Cycle

The success of the Otto cycle rests on its combination of power, simplicity, and efficiency. That in turn results from its carrying out all the processes needed to convert chemical energy into mechanical energy within a cylinder. These processes include the four **strokes** of the Otto cycle: intake, compression, expansion, and exhaust. Since the process of combustion, which occurs between the compression and expansion strokes, is also within the cylinder, the Otto cycle belongs to the class of engines called **internal combustion engines**. This contrasts with the previously dominant prime mover, the steam engine, which is an **external** combustion engine. The steam engine requires, in addition to the cylinder, separate structures such as a firebox, a boiler, and, in most cases, a condenser.

The essence of energy conversion in the Otto cycle is a series of controlled combustion processes within the cylinder, about 10 to 50 per second. Each combustion event converts about 1 to 2 kJ of chemical energy into heat, depending on the size of the cylinder. The heat and the extra volume of gases generated during the combustion processes expand the gaseous combustion products to push down a piston, thereby converting a portion of the heat energy into translational kinetic energy. Use of a **crank mechanism** converts that translational kinetic energy of the piston into the rotational kinetic energy of a **crankshaft**, providing the type of energy needed to turn the wheels of an automobile. A conceptual model of a four-cylinder Otto cycle engine is shown in Figure 6.1.

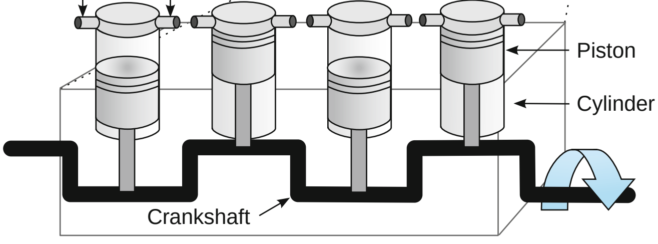


FIGURE 6.1 Conceptual Model of an Automobile Engine

Understanding the operation of this engine requires looking more closely at the cylinder and what happens inside it. The piston is cylindrical in shape and has a very tight fit within the cylinder. For the purposes of an engineering model, the cylinder and the piston moving within it are described by their nearly common **diameter**, by the **stroke** (or distance traveled by the top of the piston from its highest to lowest point, "top and bottom dead center," respectively), by the **displacement**, which is the volume swept out by the piston when going through its stroke, and by the **clearance volume**, which is the volume within the cylinder above the highest point reached by the piston (see Figure 6.2).

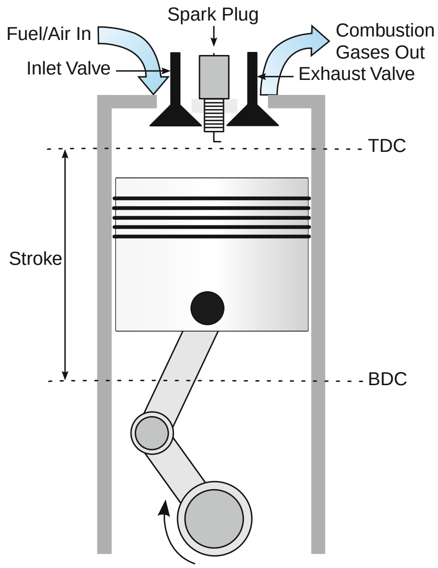


FIGURE 6.2 Stroke of an Engine Cylinder: TDC/BDC = Top/Bottom Dead Center

At the top of the clearance volume there are two ports that can be opened and closed by valves for the purpose of letting fuel and air into the cylinder and letting combustion products out. Within this cylinder the four strokes of the Otto cycle occur sequentially, once per two crankshaft revolutions.[2]

Figure 6.3 illustrates the **intake** stroke. In the first stroke, the intake valve is opened, the exhaust valve is closed, and the downward movement of the piston draws a mixture of air and fuel into the cylinder. Ideally, this is very close to a stoichiometric mixture as defined in the previous chapter.[3] Although one of the virtues of the Otto cycle is its ability to run on a wide range of fuels, for the purposes of this model we will assume the fuel is gasoline, which we can model by the compound isooctane, here denoted by the Hill formula C_8H_{18}. The stoichiometric equation also determines the mass stoichiometric **air-to-fuel ratio** $(A/F)_{Mass}$, also previously defined, and expressed as kilograms of air per kilogram of fuel.

[2]An Otto cycle animation can be found at http://techni.tachemie.uni-leipzig.de/otto/index_e.html

[3]In practice, an Otto engine normally is run a few percent lean. This means that there is up to 5% excess air to ensure that combustion is complete and fewer pollutants are emitted from the tailpipe.

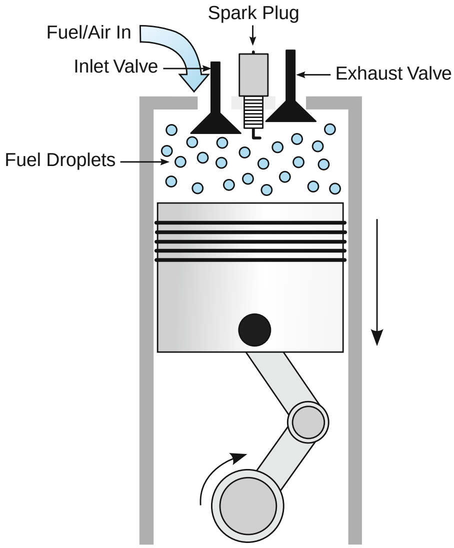


FIGURE 6.3 Intake Stroke

In the second stroke, **compression**, the piston moves upward, compressing the fuel–air mixture from the full cylinder displacement up to the clearance volume. This increases the pressure of the fuel–air mixture from nominally atmospheric pressure to a value greater than the product of (atmospheric pressure) × (compression ratio). The additional pressure is due to the fact that the compression process is rapid and thus heats the gases during the compression stroke.

Figure 6.4 illustrates the **compression** stroke. Following completion of the compression stroke, both valves are closed. A spark is produced by the spark plug at the correct instant. This spark ignites the fuel, and combustion is initiated.[4]

As we have seen, combustion is a chemical reaction that converts the mixture of fuel and oxygen into a mixture of carbon dioxide and water, while leaving almost all the nitrogen in its initial form and converting the chemical energy in the fuel–air mixture into heat.

During combustion the **power** stroke occurs. Both valves remain closed. The piston is driven downward by the expanding gases, delivering its mechanical energy to cause rotation of the crankshaft (see Figure 6.5).

After the piston reaches its lowest point, the intake valve remains closed, and the exhaust valve is opened. The piston is then moved upward by rotation of the crankshaft, pushing the combustion gases out the exhaust port. The **exhaust** stroke is shown in Figure 6.6.

When the piston reaches its highest point of the exhaust stroke, the four strokes of the cycle are complete, and a new cycle begins as the piston proceeds down and more air/fuel enters through the inlet valve

[4]A successful engine and fuel design is such that the combustion process is not explosive *per se* but burns in a controlled manner during the power stroke. Predetonation explosions or “knock” will destroy an engine.

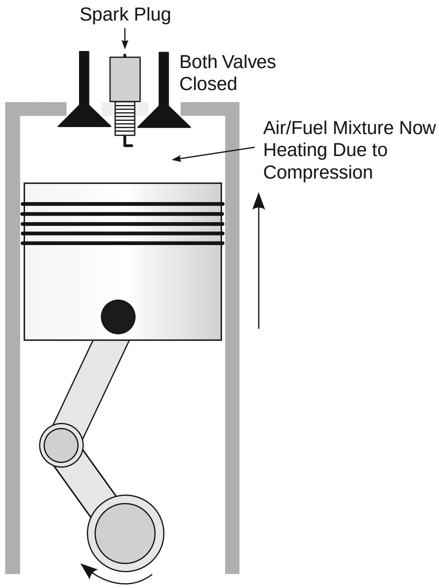


FIGURE 6.4 Compression Stroke

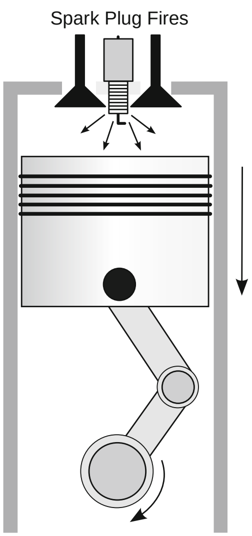


FIGURE 6.5 Power Stroke

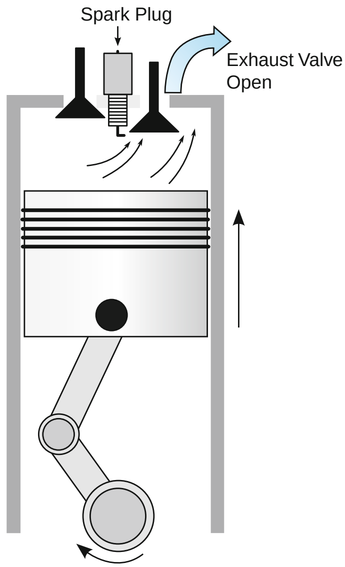


FIGURE 6.6 Exhaust Stroke

(Figure 6.3). Notice the crankshaft has rotated through two complete revolutions and the piston has traveled a total of four sweeps of the cylinder (up and down each being a single sweep or stroke)—hence the name, four-stroke engine. These four sweeps of the Otto cycle are summarized in Figure 6.7.

It is hard to get the dynamic picture of what is occurring from static pictures. There are several online animations that are useful and graphic. With animated graphics, it is easy to visualize that the crankshaft has to rotate twice to complete the entire Otto cycle.

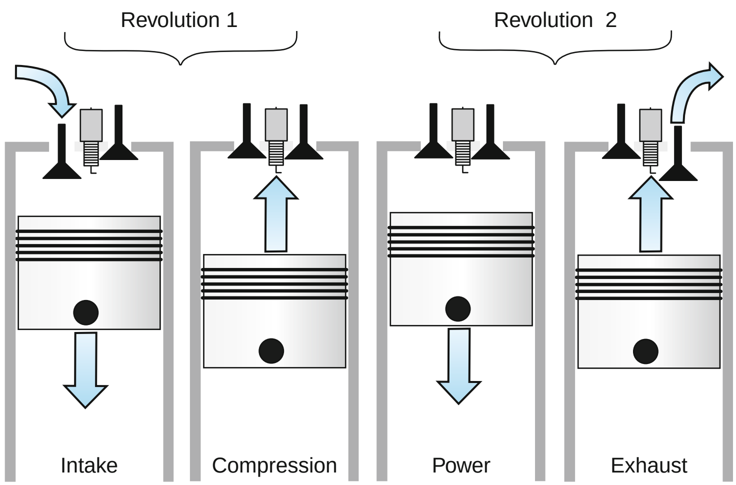


FIGURE 6.7 Summary of the Four-Stroke Engine Cycle

6.2 MODELING THE POWER OUTPUT OF THE OTTO CYCLE

To study the performance of the Otto cycle, the events that occur within it can be expressed in a mathematical model. Engineers summarize this model using the following equation:

$$\text{Ideal Power} = \frac{\eta_{\text{Otto}} \times \rho_{\text{air}} \times D \times N \times HV}{120 \times (A/F)_{\text{Mass}}} \tag{6.1}$$

where:

ρ_{air} = Density of air at cylinder conditions (normally in kg-air/m^3)
η_{Otto} = Otto cycle thermal to mechanical energy conversion efficiency
D = Cylinder displacement (*volume* swept by piston in cylinder)
N = Crankshaft speed in revolutions per minute, RPM
$(A/F)_{\text{Mass}}$ = Air-to-fuel mass ratio (usually slightly lean but close to the stoichiometric value, in kg-air/kg-fuel)
CR = Compression ratio, $(D + CV)/CV$, where CV = Valve clearance volume
HV = Heating value of the fuel in consistent units (normally kJ/kg-fuel)
and the constant '120' has units of strokes/revolution $\times$ seconds/minute

These variables must all be expressed *in a consistent set of units*: In fact, it often helps clarify the terms in such an equation by carrying a set of consistent units in its development.

The preceding equation looks somewhat frightening, but it is simply a matter of bookkeeping and is easily derived starting with the idea of representing the energy release in the Otto cycle as a series of relatively small "deflagrations." (A deflagration is a controlled explosive release of energy occurring at a controlled rate in which a "flame-front" advances in a progressive manner starting at the spark ignition point.) A certain number of such deflagrations occur every second, and each deflagration converts a certain amount of energy from chemical energy to heat energy. That conversion is carried out with a certain efficiency. Breaking down into the elements each of these concepts—frequency of deflagrations, energy release, and efficiency—yields a quantitative model for operation of the Otto cycle. This model enables the engineer to derive that formidable looking equation from simple principles (see Exercise 6).

In SI units the equation will give the answer in kW; in English units, using some additional conversion factors, the result will usually be expressed in horsepower, abbreviated HP. This unit of power goes back to James Watt, an eighteenth-century Scottish engineer. One HP = 0.746 kW, so that it is of comparable magnitude to (what we hope is now) the familiar kW.

The final expression may be termed an "ideal" Otto cycle analysis, since it does not account for losses due to friction that limit the delivered power of an Otto cycle engine.

The Otto cycle theoretical efficiency term in the model, η_{Otto}, can be derived from the material in the field of thermodynamics. The result is

$$\eta_{\text{Otto}} = 1 - CR^{-0.40} \tag{6.2}$$

This result is directly related to the increase in absolute temperature during the compression of the air/fuel charge; it is a subject for more advanced courses. However, the point of the equation is clear: the higher the compression ratio, the higher the efficiency. Even if carried out with no frictional or thermal losses, an ideal Otto cycle will see its efficiency limited by its compression ratio. The ideal efficiency will always be less than one, although it will increase toward 100 percent with an increase in the compression ratio.

There are other mechanical losses associated with friction in the internals of the engine. We can define a mechanical efficiency as:

$$\eta_{\text{Mech}} = \frac{\text{Actual Power Delivered by the Engine}}{\text{Ideal Power Generated by the Engine}} \tag{6.3}$$

Unfortunately, in real engines, this term may only be 50 percent so that one half of the mechanical power developed in the cylinders may not make it through to the final output shaft. (It will ultimately appear as parasitic heat generated somewhere in the engine.) We can define an overall Otto cycle engine efficiency that includes this effect by:

$$\eta = \eta_{\text{Mech}} \times \eta_{\text{Otto}} \tag{6.4}$$

Therefore,

$$\text{Actual Power Delivered by the Engine} = \frac{\eta_{\text{Mech}} \times \eta_{\text{Otto}} \times \rho_{\text{air}} \times D \times N \times HV}{120 \times (A/F)_{\text{Mass}}} \tag{6.5}$$

The value of η_{Mech} may be empirical (meaning determined by experiment), or it may have some theoretical underpinnings. In either case, it suggests that there is still room for considerable improvement in overall gasoline engine efficiency to challenge another generation of engineers.

The simple model of Equation (6.5) enables the engineer to draw conclusions about the effects of changing various engine parameters. For example, the equation predicts that, *other things being equal*, increasing the displacement,[5] compression ratio (as it occurs in the Otto efficiency), and the speed N of the engine will increase its power, whereas increasing the air-to-fuel ratio will decrease its power. However, in this simple model, only by increasing the compression ratio can the efficiency of the engine be increased.

Example 6.1

A new four-cylinder Otto cycle engine has a total displacement of 5000. cm^3 and a compression ratio of 9.00. Its fuel has a heating value of 45,500 kJ/kg and it has a mass air-to-fuel ratio of 15.0. During a test, the actual power delivered by the engine was 65. kW at 4000. RPM. The air density at the engine was 1.10 kg/m^3. Determine (1) the engine's Otto cycle efficiency, (2) the engine's ideal power, and (3) its mechanical efficiency.

Need: (1) η_{Otto}, (2) ideal power produced, and (3) η_{Mech} for this engine.
Know–How: Equations (6.1), (6.2), and (6.3) give the desired results.
Solve:

(1) From Equation (6.2) we have: $\boldsymbol{\eta_{\text{Otto}}} = 1 - CR^{-0.40} = 1 - (9.0)^{-0.40} = 0.58 =$ **58%**

(2) From Equation (6.1) we have: Ideal Power $= \dfrac{\eta_{\text{Otto}} \times \rho_{\text{air}} \times D \times N \times HV}{120 \times (A/F)_{\text{Mass}}}$

where $\eta_{\text{Otto}} = 0.58$, $\rho_{\text{air}} = 1.10$ kg/m^3, $D = 5000.$ cm$^3 = 0.0050$ m^3, $N = 4000.$ RPM, $HV = 45{,}500$ kJ/kg, and $(A/F)_{\text{Mass}} = 15.0.$ kg-air/kg-fuel

[5]For a single cylinder, $D = \pi \times \text{radius}^2 \times \text{stroke} = \pi/4 \times \text{diameter}^2 \times \text{stroke}$. (In engineering, diameter is preferred to radius since that is what a measuring instrument would determine.) In multiple cylinder engines we can multiply this result by the number of cylinders or simply use the total engine displacement, something the manufacturer usually provides (as in a "2.4 liter engine," etc.).

Then, **Ideal Power** = (0.58)(1.10)(0.0050)(4000.)(45,500)/(120×15.0) [kg-air/m^3] × [1/min] × [kJ/kg-fuel]/[s/min] × [kg-air/kg-fuel] = 330. kJ/s = **330. kW**.

(3) From Equation (6.3) we have: $\eta_{\text{Mech}} = \dfrac{\text{Actual Power Delivered by the Engine}}{\text{Ideal Power Generated by the Engine}}$

Therefore, $\boldsymbol{\eta_{\text{Mech}}}$ = 65. kW/330. kW = 0.20 = **20. %**

Like all simple models, this one has limitations. It would suggest, for example, that increasing the rotational speed of the crankshaft (which really means carrying out more deflagrations of fuel per second) would increase the power without limit. This increase indeed occurs at revolution rates up to a few thousand per minute. But it ceases to be true at higher crankshaft revolution rates. At those higher rates, it becomes impossible to move matter (even including low-density matter such as air) or heat around fast enough to allow energy conversion to occur efficiently. Instead, as rotation rate is increased, power output first reaches a maximum, and then decreases. So to mathematically model the performance of an Otto cycle for a wider range of rotational speeds, engineers add other correction terms, sometimes empirical terms guided by experiments but often based on solid theoretical underpinnings that show how power output rises and then falls. This additional term is the subject for one of the exercises at the end of this chapter.

6.3 THE DIESEL CYCLE

Dr. Rudolf Christian Karl Diesel (1858–1913) was a well-educated linguist and social theorist, but most of all he was a remarkable engineer. He was born in Paris, but he received his technical education in Munich under Karl von Linde (1842–1934), a renowned pioneer in mechanical refrigeration.

Though the actual thermal efficiency of Otto's engine was quite good, it was still barely competitive with the ever improving nineteenth-century steam engine. Diesel felt that he could eliminate the electrical ignition system of the Otto cycle engine if he could compress the air to the point where its temperature would be high enough to cause the fuel to ignite spontaneously. This would raise the maximum temperature of the cycle and consequently improve its thermal efficiency. He also felt that a higher combustion temperature would allow cheaper, heavier hydrocarbon fuels (such as kerosene, a common lamp oil in the late nineteenth century) to be used. On August 10, 1893, Diesel's first compression ignition engine ran under its own power for the first time, and by 1898 Diesel had become a millionaire simply by selling franchises for the industrial use of his engine.

Diesel's 1893 test engine compressed air to 80 atmospheres, a pressure never before achieved by a machine. He was nearly killed when one of his engines subsequently exploded.

Diesel wanted to create a constant temperature combustion process to reduce heat losses and improve the engine's thermal efficiency. Unfortunately, he was not able to do this; instead, his cycle has a constant pressure combustion process.

One crucial difference between the Otto and the Diesel cycles is that in the latter, fuel is not added until the end of the compression stroke. This is because the high pressures in the Diesel cycle produce high compression temperatures that would explode the fuel if present at the start of the compression stroke, hence the late addition of fuel.

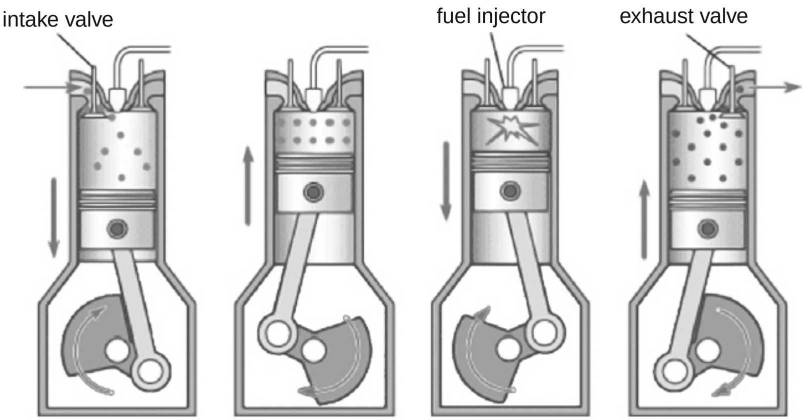


FIGURE 6.8 Operation of a Diesel Cycle

The Diesel cycle efficiency, η_{Diesel}, is also derived in **thermodynamics**. The result is

$$\eta_{Diesel} = 1 - \frac{CR^{-0.4}(CO^{1.4} - 1)}{1.4(CO - 1)} \quad (6.6)$$

where CR is the compression ratio and CO is the cut-off ratio.[6]

Example 6.2

A new four-cylinder, four-stroke Diesel cycle engine with a compression ratio of 19.0 and a cut-off ratio of 2.37 has a total displacement of 4.50×10^{-3} m^3 (4.50 liters). The engine burns kerosene with a heating value of 45,500 kJ/kg and has a mass air-to-fuel ratio of 30.0 kg-air/kg-fuel. During a test, the intake air density was 1.20 kg/m^3 when the engine was running at 2500. RPM and producing 33.1 horsepower. Determine (1) the efficiency of the engine, (2) the ideal power output of the engine, and (3) the mechanical efficiency of the engine.

Need: The engine's efficiency, ideal power output, and actual power output.
Know–How: These results are given by Equations (6.6), (6.1), and (6.3).
Solve:

1. The engine's efficiency is given by Equation (6.6) as:

$$\eta_{Diesel} = 1 - 19.0^{-0.4}(2.37^{1.4} - 1)/(1.4(2.37 - 1)) = 0.623 \text{ or } \mathbf{62.3\%}$$

[6]The phrase "cut off" is an archaic steam engine term that has been absorbed into modern Diesel engine terminology. It was introduced in the 1780s by James Watt to describe the process of cutting (or shutting) off the steam to a steam engine. Today this phrase is used to indicate the point where the combustion process "cuts off" (i.e., stops) in a Diesel engine. It is determined by the geometry of the combustion chamber and the fuel charge. Also, cut off is a popular synonym for "shut off" in the southern United States. It was probably introduced into this region by the British soldiers stranded there after the American Revolutionary War.

2. The ideal power output is the same as given by Equation (6.1) with η_{Diesel} replacing η_{Otto}. This produces

$$\text{Ideal Power} = \frac{\eta_{Diesel} \times \rho_{air} \times D \times N \times HV}{120 \times (A/F)_{Mass}}$$

or

$$\textbf{Ideal Power} = 0.623 \times (1.20) \times (4.50 \times 10^{-3}) \times (2500.) \times (45{,}500)/(120 \times 30.)$$
$$[\text{kg-air/m}^3][\text{m}^3][1/\text{minute}][\text{kJ/kg-fuel}]/[\text{kg-air/kg-fuel}][\text{s/minute}] = \textbf{106.kW}$$

3. The actual power output of the engine is 33.1 hp, or 33.1 × 0.746 [hp][kW/hp] = 24.7 kW, and the mechanical efficiency of the engine is given by Equation (6.3) as

$$\eta_{Mech} = 24.7/106. = 0.233 = \textbf{23.3\%}$$

6.4 THE BRAYTON CYCLE

George Brayton (1830–1892), an American engineer, adopted the dual reciprocating piston technique using one piston as a compressor and the second piston to deliver power. A chamber was inserted between the two pistons to provide a constant pressure combustion process.

The original Brayton cycle was conceived as an external combustion piston engine. However, it was found to run more reliably when it was converted into an internal combustion piston engine. Much later it was discovered to be an adequate model for gas turbine engines, shown in Figure 6.9.

The development of the reciprocating piston Brayton cycle engine was stifled by its complexity, and the cycle quickly might have become obsolete if it had not been for a new technology that was being developed for steam, the turbine. By replacing steam with combustible gas, a new type of engine, the gas turbine, was produced.

One difficult characteristic of a gas turbine engine is that it requires the inlet air to be compressed. Early air compressors were very inefficient, and this single fact proved to be a major stumbling block in the development of gas turbine engine technology.

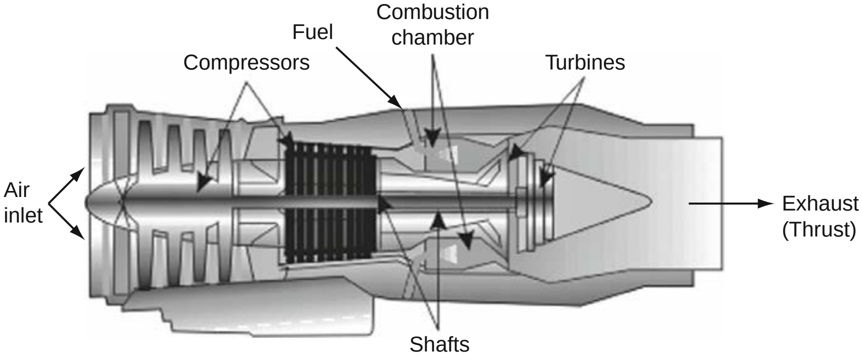


FIGURE 6.9 Schematic of a Gas Turbine Engine

The first Brayton cycle gas turbine unit to produce a net power output (11 hp) was built in 1903. It had a very low efficiency (about 3%) and could not compete economically with the existing engines of its time. Compressor efficiency problems continued to plague gas turbine technology, and many new prototype gas turbine engines were designed and built that still could not produce a net power output as late as the 1930s. Since the exhaust thrust produced by an aircraft engine is not considered to be part of the engine's shaft work output (thrust being kinetic energy, not shaft work), aircraft engines do not necessarily need high thermal efficiencies to be effective. It was in this industry that the gas turbine engine first became successful.

The Brayton cycle efficiency, η_{Brayton}, is also derived in thermodynamics. The result is

$$\eta_{\text{Brayton}} = 1 - PR^{-0.286} \tag{6.7}$$

where PR is the compressor pressure ratio.

Example 6.3

A jet engine needs to produce only enough net output power to drive the aircraft's accessories (fuel pump, hydraulics, generator, etc.) and consequently it need not have a very high efficiency. The first successful Brayton cycle turbojet aircraft was the German Heinkel-178, which flew for the first time on August 27, 1939. The engine weighed 364 kg, produced a thrust of 1,100 lbf (4900 N) at 13,000 rpm, and had a compressor pressure ratio of 3.0. Determine its efficiency.

Need: The Heinkel WWII turbojet engine efficiency.
Know: The Brayton cycle efficiency is given by Equation (6.7).
How: The engine's efficiency is given by Equation (6.7) as: $\eta_{\text{Brayton}} = 1 - PR^{-0.286}$
Solution: Therefore, $\boldsymbol{\eta}_{\textbf{Brayton}} = 1 - 3.0^{-0.286} = 0.27$ or **27%**. However, most of the energy of the fuel went into the engine's exhaust thrust, which is not considered in this calculation.

During the 1960s, a turbojet design called the "turbofan," or "fan-jet" engine was developed wherein some of the inlet air bypassed the combustion chamber and was mixed with the combustion products at the turbine inlet. This cooled the turbine inlet gases slightly and helped to complete the combustion process. Earlier, in 1945, a turbojet engine called the "turboprop" was developed wherein a gas turbine was used to drive a propeller as well as provide exhaust thrust. The first commercial jet airliners used turboprops in the 1950s, but this type of turbojet engine is used only on short range aircraft today.[7]

6.5 MOTION

Another application of an engine model is to help understand why stepping on the accelerator increases a car's engine power output. Stepping on the accelerator turns a plate positioned in the air intake to the engine in a way that enables more air to be sucked into the cylinder with each intake stroke (Figure 6.10).

In our model Equation (6.1), the air density entering the engine maximizes at that of the surrounding atmosphere. Once air is sucked into the engine by the intake stroke, its pressure is partially controlled by the throttle plate. Opening the throttle plate more means an effective increase in ρ_{air}, the air density. Assuming the air-to-fuel ratio remains constant (modern cars have sophisticated controls that can ensure this is true), the fuel will increase. As a result, each deflagration within the cylinder will result in a greater power output.

[7] In the early 1960s Chrysler Corporation attempted to introduce a gas turbine powered automobile to the mass market without success.

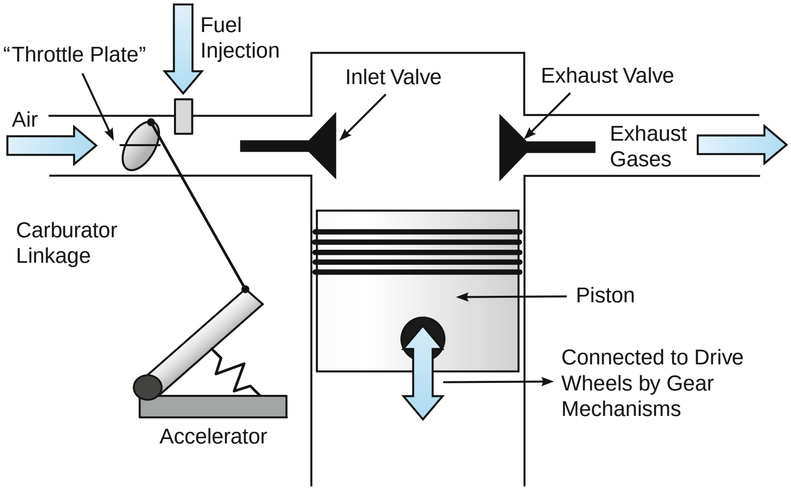


FIGURE 6.10 A/F System Schematic

6.6 IMPROVING THE OTTO, DIESEL, AND BRAYTON CYCLES

The Otto, Diesel, and Brayton cycles' supremacy through the twentieth century has not convinced all engineers that they will remain the power source of the twenty-first century. There are several competitors that stand particularly prominent in offering a challenge due either to higher fuel efficiency, lower pollution emissions, or both.

The first is offered by the **hybrid** propulsion system. It uses a conventional Otto or Diesel cycle as its prime mover, combined with an electrical generator and an energy storage system. The gasoline or Diesel engine is run at its most efficient performance level, or turned off altogether, and the increased power needed to achieve acceleration is drawn from the energy storage system (typically an electric battery, but conceivably from another type of storage system such as a flywheel or from an ultracapacitor[8]).

A more radical challenge to these engine cycles is the **fuel cell**. It is a system for converting chemical energy *directly* to electrical energy, which then can drive an electric motor to turn the wheels. Because this conversion does not require the release of heat, it can in principle offer higher efficiency than the Otto, Diesel, or Brayton cycle, while eliminating the production of such pollutants as carbon dioxide and nitrogen oxides (the latter being produced at "hot spots" in the cylinder). To achieve these advantages to the fullest extent, engineers have proposed running fuel cells on hydrogen, a concept that was developed for the fuel cells used in the 1960's space program. This would indeed result in a clean-running, efficient automobile propulsion system with a vehicle range comparable to that of today's cars. However, hydrogen is not readily available today as a fuel and is difficult to store in a compact form. So use of hydrogen fuel cells would require either a completely new fuel infrastructure of storage facilities and hydrogen "gas stations," or a component called a "fuel reformer" that would enable each vehicle to convert gasoline (or another hydrocarbon fuel) into hydrogen. This in turn introduces significant system complications, as well as major cost and pollution issues.

[8]An ultracapacitor is a hybrid of an electrical capacitor and a battery; it can deliver a limited amount of energy as a burst of power (see Chapter 8, "Electrochemical Engineering and Alternative Energy Sources").

The least radical approach to more efficient vehicles that might extend the life of the Otto, Diesel, and Brayton cycle engine is to simply reduce the vehicle's weight.

Today, compact American cars weigh about 2000 lbf and larger cars weigh about 4000 lbf (and SUV's weigh as much as 7500 lbf). Figure 6.11 shows clearly that a vehicle's weight directly affects its fuel consumption—the heavier the vehicle the larger the fuel penalty.[9] In fact, whether sedans, SUV's, minivans, or crew cab trucks, Figure 6.11 shows that about 90 percent of the fuel economy is determined solely by the vehicle's weight!

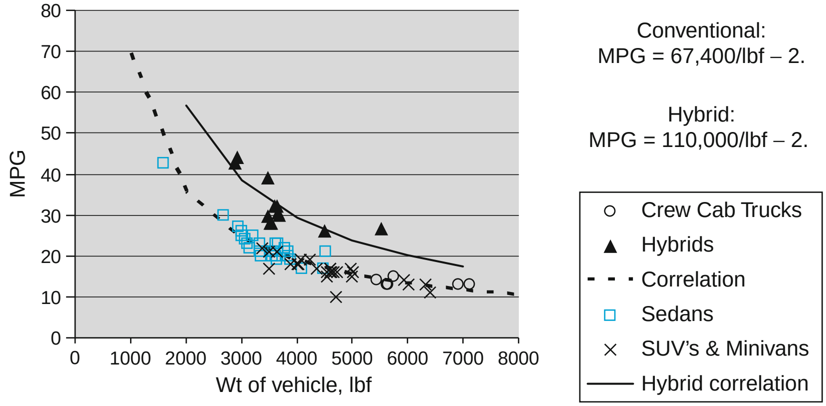


FIGURE 6.11 Weight and Mileage (based on Consumer Reports data, April 2008)

The curve for standard vehicles in Figure 6.11 can be expressed by the following equation:

$$MPG \text{ of a standard vehicle} = \frac{67,400}{\text{Vehicle weight in lbf}} - 2,$$

and the curve in Figure 6.11 for hybrid vehicles can be expressed by the following equation:

$$MPG \text{ of a hybrid vehicle} = \frac{110,000}{\text{Vehicle weight in lbf}} - 2.$$

Thus, a 3000 lbf car, minivan, or truck would be expected to have a fuel economy of about 21 mpg, whereas a hybrid of the same weight would have a fuel economy of 35 mpg. Of course, if lighter materials such as low density metal alloys and more plastics were used in hybrid vehicles, their fuel economy could be even better. Indeed, lighter materials may be a technologically less challenging way to quickly achieve higher fuel economy than a hybrid route. One prototype that could mark this future is an experimental VW vehicle shown in Figure 6.12 that has an 8.5 HP engine and weighs just 290 kg (640 lbm). It is capable of an astounding nearly 300 mpg,[10] and appreciably beats the previous correlation for conventional vehicles.

[9]These data are for combined in-town and highway driving.
[10]http://www.canadiandriver.com/news/020419-1.htm

FIGURE 6.12 Volkswagen Experimental—Near 300 mpg—Car

Example 6.4

The cost of fuel is rising and you want to calculate the cost of one year's fuel if the price is $6.00 per gallon and you drive 12,000 miles per year. If you are thinking of buying a 2,750 lbf weight conventionally powered vehicle, how much is the projected cost of fuel?

Need: Cost of fuel for 12,000 mile, 2,750 lbf vehicle = __?
Know: Relation between weight and mpg: $MPG = 67400/lbf - 2.$
How: \$/year = (gallons/mile) × (miles/year) × (\$/gallon)
Solve: **MPG** = 67400/lbf – 2. = 67,400/2,750 – 2. = **22.5 mpg**
Therefore, **\$/year** = (1/22.5) [gallons/mile] × 12,000 [miles/year] × 6.00 [\$/gallon] = **\$3,200 \$/year**

6.7 ANOTHER VISION OF THE FUTURE

The Honda FCX Clarity (Figure 6.13) is a next-generation, hydrogen-powered fuel cell vehicle. The vehicle's only emission is water, and its fuel efficiency is three times that of a modern gasoline-powered automobile. The vehicle's performance includes a driving range up to 280 miles,[11] and a fuel economy of 72 miles/kg-H_2 (or 72 miles per Gasoline Gallon Equivalent (GGE) a measure of the fuel's energy content), and which may be competitive with conventional fuels. Table 6.1 lists the GGE calculation for several alternative fuels.

[11]Based on official 2008 EPA estimated range and fuel efficiency values.

FIGURE 6.13 The Honda FCX Clarity Fuel Cell Car

Table 6.1 Alternative Fuel Conversion Factors to GGE[12]

Fuel Type	Measurement Unit	Conversion Factor	Gasoline Gallon Equivalent (GGE) Calculation
Biodiesel B100	Gallons	1.015	GGE = B100 in gal × 1.015
Biodiesel B20	Gallons	1.126	GGE = B20 in gal × 1.126
Diesel	Gallons	1.147	GGE = Diesel in gal × 1.147
Compressed Natural Gas (CNG)	Gallons at 3000 psi	0.225	GGE = CNG in gal (at 3000 psi) × 0.225
E-85	Gallons	0.72	GGE = E-85 in gal × 0.72
Electric	kWh	0.03	GGE = Electricity in kWh × 0.03
Gasoline	Gallons	1.0	GGE = Gasoline in gal × 1.0
Hydrogen[13]	kg	1.0	GGE = H_2 in kg × 1.0
LNG	Gallons at 14.7 psi and –234°F	0.66	GGE = LNG in gal × 0.66
LPG	Gallons	0.74	GGE = LPG in gal × 0.74

[12]See http://alternativefuels.about.com/od/resources/a/gge.htm

[13]Did hydrogen cause the Hindenburg to explode? The fire that destroyed the hydrogen-filled German blimp Hindenburg in 1937 gave hydrogen a bad reputation. The hydrogen (used to keep the airship buoyant) initially was blamed for the disaster; however, an investigation in the 1990s showed that the airship's fabric was coated with chemicals similar to solid rocket fuel, and easily ignitable by an electrical discharge. The Zeppelin Company, builder of the Hindenburg, since has confirmed that the flammable, doped, outer cover was the probable source of the fire. *Source*: National Hydrogen Association.

SUMMARY

This chapter does not imply that a study of engines is *all* that mechanical engineers do. This topic has been chosen to illustrate just one area of energy conversion technology within the discipline. It does illustrate the thinking process that goes into solving an engineering problem.

Note that the internal combustion engine energy conversion technology was a seminal invention of the twentieth century that built global economies and created the age of oil. Thus the future of the automobile and the Otto cycle are not merely engineering issues but social, economic, and political issues as well. The topics touched on range from global product competition, to the geopolitics of relying on the Middle East and elsewhere for imported oil, to the environmental impacts of increased atmospheric levels of carbon dioxide. Engineering models of energy conversion, such as the ones described in this chapter, cannot by themselves decide such issues. They remain, however, essential guides for providing realistic assessment of technical possibilities as society considers alternative energy futures. So engineers' abilities to understand the operation of the **Otto, Diesel**, and **Brayton cycles**, model the power output of those cycles, use that power output to achieve motion, and assess proposed improvements to these engines contribute not only to their technical expertise but also to their capability to participate in an important public debate.

EXERCISES

1. What would be the Otto cycle efficiency in Example 6.1 if the engine's compression ratio was 10.0?

2. What would be the ideal power produced in Example 6.1 if the engine's total displacement was 8000. cm^3 and it was running at 3000. RPM?

3. What would be the ideal power produced in Example 6.1 if the engine had a turbocharger that boosted the inlet air density to 3.7 kg/m^3?

4. An Otto cycle engine has a compression ratio of 8.0. What is the engine's thermal efficiency? (**A: 0.56 or 56%**)

5. Suppose you want to design an Otto cycle engine with a thermal efficiency of 45 percent. What compression ratio must you use?

6. An Otto cycle engine has a total displacement of 0.0040 m^3. Its compression ratio is 9.0. It is fueled with gasoline having a specific heating value of 45,500 kJ/kg. Its mass air-to-fuel ratio is 16.0. Assume the density of air is 1.00 kg/m^3. Determine its ideal power output at $N = 4.00 \times 10^3$ RPM, in kW and in HP. (**A: 2.2×10^2 kW, 3.0×10^2 HP (kg air)/(kg fuel)**)

7. For the engine in exercise 6, suppose you wanted to increase the power output at 4.00×10^3 RPM by 10 percent. One way to do this would be by changing only the compression ratio. What would the new compression ratio be?

8. For the engine in exercise 6, suppose you wanted to increase the power output at 4.00×10^3 RPM by 50 percent while retaining a compression ratio of 9.0. One way to do this would be by changing only the displacement. What would the new displacement be (in cm^3)? (**A: 6.0×10^3 cm^3**. This is a large engine for a car; it's easier and cheaper to achieve this performance by increasing ρ_{air} in the equation for engine power, since engine power is proportional to ρ_{air}.)[14]

[14] If you increase air density by compression from its nominal value of $\sim$1.00 kg/m^3 (an approximate value at one atmosphere) to $\sim$1.50 kg/m^3 by adding a turbo charger (compressor) to bring the air pressure up to 1.50 atmospheres (the volume of air falls by Boyles' law proportional to the increase in pressure—ergo, the density of a fixed mass of gas increases in the same proportion). Many cars do have some form of turbo charger to get additional power from a smaller engine.

9. The derivation of Equation (6.1) relies on the method of canceling units in brackets [. . .] as in the following steps. Replace each of the steps in words with its dimensionally correct mathematical equivalent, and so derive the Otto Cycle Power Output Equation.

Step 1: How much air is drawn into the cylinder? The actual volume of air drawn into the cylinder is D as the piston cycles from top dead center to the bottom dead center. The mass of air drawn into the cylinder is thus . . .?

Step 2: How much fuel? (**Hint:** Use $(F/A)_{\text{Mass}}$ ratio)

Step 3: How much thermal energy is released by this amount of fuel for just one power stroke? (**Hint:** Needs the HV of the fuel)

Step 4: Of this heat, only a portion is converted into mechanical work. (**Hint:** The mechanical work is related to the thermal input by η_{Otto})

Step 5: How many power strokes are there? As we have seen, there is just one per every two engine revolutions, and N is the number of revolutions/minute.[15] (**Hint:** Multiply the mechanical work by number of power strokes/s)

Step 6: Hence, the rate of working is $\eta_{\text{Otto}} \times [\rho_{\text{air}} \times D \times (A/F)_{\text{Mass}}] \times HV \times N/(2 \times 60)$ [kJ of work per power stroke] [power strokes per second] = [kJ (work)/s] = [kW].

Step 7: Finally, cast this in the suggested form:

$$\text{Ideal Power} = \frac{\eta_{\text{Otto}} \times \rho_{\text{air}} \times D \times N \times HV_{\text{fuel}}}{120 \times (A/F)_{\text{Mass}}}$$

Exercises 10 through 12 involve the world's fastest truck record holder, a 1940 Ford fire truck. Its top speed is 404 mph! It was powered by two Rolls Royce 601 Viper jet engines with afterburners, developing an equivalent to 1.20×10^4 HP. It consumes 2.0×10^2 lbm of fuel per second (see http://www.jetfiretruck.com).

10. Suppose the fuel used by the fire truck contains 45,500 kJ/kg of chemical energy. What is the overall efficiency (ratio of mechanical power out to chemical power in) of the truck's engines?

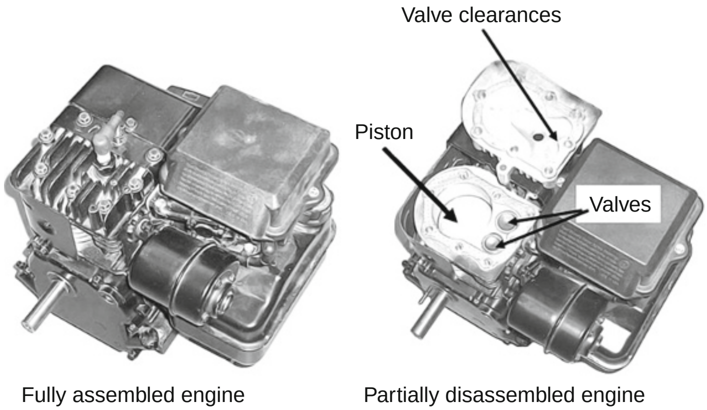
11. Suppose you wanted to design an Otto cycle engine with the same power output as the two jet engines powering the fire truck. Suppose your Otto cycle engine was able to achieve the far higher efficiency of $\eta_{\text{mech}} = ½$ and $\eta_{\text{Otto}} = 2/3$. Suppose you operated the engine at 4.00×10^3 RPM and $(\text{A/F})_{\text{Mass}} = 16.0$ (kg air)/(kg fuel). Suppose your fuel was isooctane, with $HV_{\text{fuel}} = 45{,}500$ kJ/kW. How large would the displacement of the Otto cycle engine have to be in order to achieve the fire truck's power output of 1.20×10^4 hp? (**A: 0.28 m^3**)

12. To get an idea of how realistic it would be to build an Otto cycle engine with the displacement you calculated in exercise 11, assume that your engine had cylinders each the size of a two-liter soft drink bottle (a very large cylinder for an Otto cycle engine designed for automobile use!). How many cylinders would be required to produce the desired power?

Exercises 13 through 17 refer to a 3.5 horsepower lawn mower engine at a nominal speed of 2.50×10^3 RPM. A typical lawn mower engine, and the inside of its cylinder head, is shown below. **Note:** the depression

[15]"Minutes," a non-SI unit, will introduce a numerical factor of 60 somewhere in the equation provided the rest of the variables contain seconds—so look for a numerical constant to appear in the final result. (The other factor of 2 is there because there are *two* rotations of the crank per power stroke.)

in the cylinder head is the clearance for the engine's valves. If such an engine is available to you in class, you can disassemble it and measure the bore, stroke, and clearance volume, CV, of the engine.



13. If available, based on your measurements of bore (cylinder diameter), stroke (TDC to BDC), and clearance volume, CV, what is the displacement, D, of the lawnmower engine? (Use experimental numbers if available; otherwise, use these numbers: stroke = 0.100 m, bore = 0.050 m, and $CV = 2.5 \times 10^{-5}$ m^3.) (**A: 196 cm^3**)

14. Based on your measurements and calculations (exercise 13), what is the compression ratio, CR, of the lawnmower engine?

15. Based on your measurements and calculations so far, what is the Otto efficiency of the lawnmower engine? (**A: 0.58 = 58%**)

16. Based on your measurements so far, if the mechanical efficiency of the lawnmower engine is 0.50, what is the power output of the lawnmower engine, in kW, when it is operating at 2.50×10^3 RPM with a mass air-to-fuel ratio of 15 (kg air)/(kg fuel) with a fuel that gives 45,500 kJ/kg of energy?

17. The power output of the lawnmower engine is not directly proportional to the number of RPMs, as the simple formula Equation (6.1) suggests. It actually reaches a peak at about N = 2500 RPM and declines for higher RPM. This behavior can be approximated by the formula: $\eta_{Mech} = \exp(-1.26 \times 10^{-7} \times N^2)$.[16] Prepare a graph of power in kW versus RPM from 0 to 8000 in increments of 500; comment on the shape of the graph.

18. What would the Diesel cycle engine efficiency in Example 6.2 be if the cut-off ratio was reduced to 1.9?

19. What would the ideal power produced by the Diesel engine in Example 6.2 be if the air-to-fuel ratio was changed to 25?

20. Determine the Diesel cycle efficiency of an engine that has a compression ratio of 18.6 and a cut-off ratio of 2.2.

21. Using highly advanced materials, a research laboratory has produced a new Diesel cycle engine with a compression ratio of 58.9 and a cut-off ratio of 3.9. What is the Diesel cycle efficiency of this engine?

[16]This can be written in spreadsheet form as η_{Mech} = EXP((−1.26E-7) * Nˆ2) where N is in RPM. Note that η_{Mech} varies from 1.0 at $N = 0$ RPM to 0.0 as N becomes very large.

22. What would the Brayton cycle efficiency in Example 6.3 be if the pressure ratio was increased to 3.5?

23. What pressure ratio would be required to produce a Brayton cycle efficiency of 75 percent?

24. A gas turbine engine has been developed with a compressor pressure ratio of 6.7. What is the Brayton cycle efficiency of this engine?

25. Estimate the amount of water to generate H_2 if all the cars in the United States were hydrogen-powered fuel cell vehicles. Assume a vehicle fuel economy of 60. miles per kg of hydrogen, total vehicle miles traveled $= 2.6 \times 10^{12}$ miles/year, and that 1.0 gallon of water contains 0.42 kg of hydrogen.

26. For the conventionally powered vehicle in Example 6.4 weighing 2,750 lbf and ranging 12,000 miles/year, what is the total annual cost of fuel if it costs $10.00 per gallon?

27. Using a spreadsheet analysis, plot the annual cost of fuel for the vehicle in Example 6.4 as the cost of gasoline ranges from $1.00 to $10.00 per gallon.

28. What is the Gasoline Gallon Equivalent of 5.0 gallons of
- **a.** Diesel fuel
- **b.** E-85
- **c.** Compressed natural gas at 3000 psi
- **d.** Liquid natural gas (LNG)

29. As a young engineer in the transmission division of a large automotive company, you are given an engineering assignment that originated in the engine division. Your supervisor is experienced and has a good idea of what the engineer in the engine division wants. Your solution is not what your supervisor expected, and he or she asks you to change your report. What do you do?
- **a.** Fully discuss the issues with your supervisor, and if you sincerely feel that you are right, then do not change your report.
- **b.** As long as your supervisor assumes responsibility for your report, then change it as requested.
- **c.** Since you are not very experienced, change your report as long as the changes are still correct engineering practice.
- **d.** Report your supervisor's behavior to the human resources office as harassment.

30. As an engineering student you are offered a lucrative summer job at a large tire manufacturing company. On your first day of work you are asked to sign an agreement assigning to the company all patent rights for your summer and the succeeding year. What do you do?
- **a.** Sign the agreement, since your summer work will probably not produce anything patentable.
- **b.** Ask that the agreement be altered to omit patentable work in areas outside those included in the summer job (e.g., thesis work).
- **c.** Sign the agreement and hope the company does not hold you to it.
- **d.** Refuse to sign and look for another job.

CHAPTER

7 Electrical Engineering

Source: © iStockphoto.com/Lisa F. Young

Electrical engineering is a field of engineering that deals with the study and application of electricity, electronics, and electromagnetism. One of the most convenient ways to make energy useful is by converting it to electromagnetic energy, usually called electricity. In a previous chapter, electromagnetic power was introduced briefly as the result of multiplying the instantaneous current by the voltage drop that produced it. This chapter expands that introduction by giving a more detailed, but still highly simplified, discussion of electromagnetic energy and some of its uses.

Though electricity is everywhere around us at all times, the easiest way of getting electricity to go where it is needed is by means of an arrangement called an electrical circuit. In this chapter you will learn what **electric circuits** are, and you will apply a very simple model for analyzing their operation. The model has as its basic variables **charge**, **current**, **voltage**, and **resistance**. This simplified treatment relies on a model consisting of **Ohm's Law**, the "**Power Law**," and **Kirchhoff's Voltage** and **Current Laws**. Using these laws, we can analyze the operation of two classes of direct current circuits called **series** circuits and **parallel** circuits.

The field of electrical engineering is devoted to the design and analysis of electrical circuits to meet a wide range of purposes, from melting metals to keeping food frozen, from projecting visual images to powering cell phones, from taking elevators to the top of skyscrapers, to moving submarines through the ocean depths. Rather than attempt to describe the entire range of electrical engineering applications, we will focus on one case study: The development of ever-faster **switches**. Different kinds of switches can be implemented in many ways, using properties of electricity such as the ability of a current to produce **magnetism**, and the use of electric charge to control the flow of current.

7.1 ELECTRICAL CIRCUITS

A circuit is a *closed* loop of wire connecting various electrical components such as batteries, lightbulbs, switches, and motors. The word *circuit* should bring to mind the idea of a circle. We first recognize a natural phenomenon of **electric charge**, measured in a unit called **coulombs**, which flow through the electrically conductive wires like water flows through pipes. This is the essence of the simple model we will use for understanding electrical circuits. And just as water is actually made up of very small particles called water molecules, each of which has the same mass, charge in a wire actually is carried by very small particles called **electrons**, each of which has the same very tiny electrical charge of -1.60×10^{-19} coulombs (notice the minus sign). Because it is inconvenient to measure amounts of water in numbers of molecules, instead we use gallons, liters, or whatever, which contain a huge number (about 10^{27}) of molecules. It is similarly inconvenient to measure charge in number of electrons, so we measure charge in coulombs. The coulomb is a huge number (about 10^{20}) of electrons. A car battery might hold 1 million or more coulombs of charge, and the mid-sized C cell used in a flashlight might hold 10,000 coulombs.

Water would not be useful to us if it just sat in a reservoir, and electric charge is not useful to us if it just sat in a battery. Think of a Roman aqueduct system in which water from a high reservoir is allowed to flow down to the city; it progressively loses its original potential energy as it flows. Electrons similarly flow from a high potential to a lower one. We call this potential **voltage**. Thus electric charges are *pumped* by a battery or generator through the wires for use in electrical components that we value such as lights and air conditioning, and then they return to the battery by means of wires for their reuse.

Just as the flow of water from a hose can be separated into two or more exiting streams totaling the amount of entering water, if trillions of electrons flow every second into an electrically conductive branch junction, you can be sure that the total number of electrons leaving the junction is exactly equal to the number of arriving electrons, a principle called **conservation of charge**. The flow of charge is a variable called **current** and is measured in **amperes**, where one ampere is equal to one coulomb per second (and correspondingly equal to the flow of many trillions of electrons every second).

The hollow interior of a water pipe permits the water flow. Wires are not hollow, but they do allow the flow of electricity in the same way that a hollow pipe allows the flow of water when subjected to upstream pressure. We call materials used in wires **conductors**, which include metals. Some other materials do not allow the significant flow of electricity through them. These materials are called **insulators**. They include such nonmetals as ceramics and plastics. Other materials, such as the **semiconductors**, have properties somewhere between those of conductors and insulators. We distinguish among conductors, semiconductors, and insulators by a property known as **resistance**. Electric current, in the form of electrons, flows because of the voltage that is applied to the electrons. The electrons in metal conductors flow from the negative to the positive direction in potential because the negatively-charged electrons are repelled from the negative end of the conductor. However, traditionally (and unfortunately), the current is considered to flow from the positive pole of the battery to the negative one.[1] It was discovered well after this convention had been adopted that the current carriers in metals are electrons and flow in the opposite direction (see Figure 7.1).

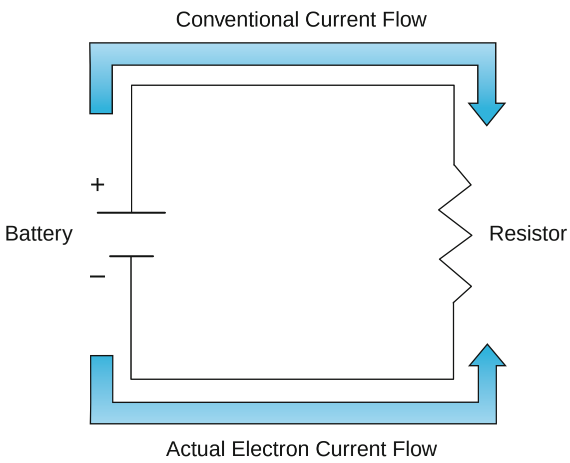


FIGURE 7.1 Blame Ben Franklin!

[1]This convention was suggested by Benjamin Franklin in the mid-eighteenth century and is still used today.

The motion of water through a pipe may be caused by a pump, which exerts pressure on the water, forcing it to move through the pipe. In the same way, a battery, in addition to holding electrical charge, serves as a sort of electrical pump to force electrons to move through a wire. This electrical pressure is the variable we call **voltage**, and it is measured in a unit called **volts**. Just as a hand pump might be used to apply a pressure of say 12 pounds per square inch to move water through a pipe, a car battery might be used to apply 12 volts to move current through a wire.

Resistance is just the ratio of the voltage drop to the current. High resistance means that a large voltage drop is required to achieve a given current. Low resistance means that only a small voltage drop is required to achieve a given current. Resistive devices are called **resistors** (and often colloquially just "resistances"), and may be made from coils of wire alloy, pieces of carbon, or other materials. In fact, all materials except electrical superconductors[2] exhibit resistance to current flow. Some metals, such as pure silver and pure copper, have a very low resistance to electron flow. Other metals, such as aluminum and gold, have somewhat higher resistance but are still used as conductors in some applications. We usually use copper for connections, since we do not want much resistance there. Metal alloys and semiconductors offer appreciable resistance to electron flow, and we use those where we wish to exploit that property.

To summarize: In our model of an electric circuit, the voltage (i.e., the electrical pressure) pushes a moving current of charge though a conductor. This charge (i.e., electrons) never disappears but simply circles around the circuit again and again, doing its various functions each time around the circuit. As the electrons travel around the circuit, they are pumped to high potential energy by the voltage source and then lose this energy as they perform their functions.

Symbolically, the simplest possible electrical circuit consists of a battery, a wire, and an electrical component located somewhere along the wire, as shown in Figure 7.2, which shows the visual symbol for a battery, with the longer line indicating the "plus" side of the battery and the shorter line representing the "minus" side of the battery.

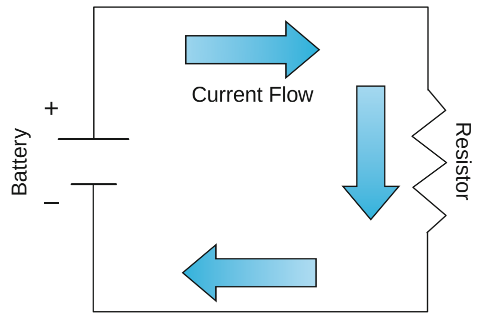


FIGURE 7.2 Current Flow in Wires

The zigzag line represents a resistor, and the straight lines are the wires connecting it to the battery. As previously stated, the current is considered to flow from the positive pole of the battery to the negative one. Since a circuit goes in a circle and ends up where it started, the voltage must also return to its initial value after a trip around the circuit. Since the voltage is raised across the battery, it must be lowered back somewhere else in the circuit. In our simplest model, we will assume that none of this voltage "drop" occurs in the wires, but rather all of it occurs in the component or components in the circuit, although we will soon relax that restriction.

[2]These materials allow the *unrestricted* flow of electrons and therefore have no voltage drop associated with current flow through them.

7.2 RESISTANCE, OHM'S LAW, AND THE "POWER LAW"

As water flows though a pipe, it experiences friction by coming into contact with the walls of the pipe, and would tend to slow down the flow. Pressure is needed to keep water moving though the pipe, and the magnitude of the change of the pressure is determined by the friction between the water and the walls.

In the same way, as charge flows through a wire it experiences interactions that tend to reduce the voltage that is "pushing" the charge. We will initially assume that these interactions do not occur in the connecting wires but are confined to the components that are hooked together by the wires. The interactions that reduce the voltage are described by resistance. It is measured in units called **ohms** and abbreviated with the uppercase Greek letter omega, **Ω**. The voltage drop (that is, an electrical pressure drop) needed to keep the charges moving through the wire is determined by the resistance. This relation between current, voltage, and resistance constitutes the first basic law of our model of electrical circuits, **Ohm's Law**. It states that the resistance, which is defined as the ratio of voltage drop to current, remains *constant* for all applied voltage drops. Mathematically stated,

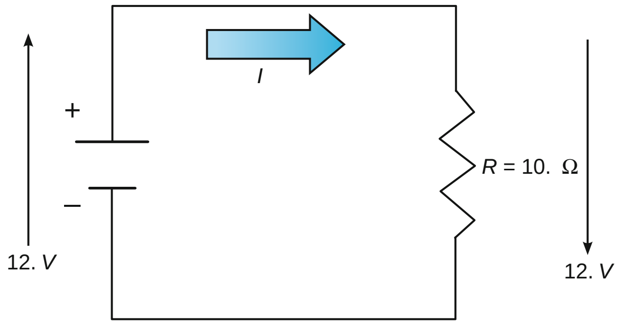
$$\boldsymbol{R = V/I} \tag{7.1}$$

If V is in volts and I is in amperes, R is defined in the units of ohms $[\Omega] = [V]/[A]$. For example, consider a 2.000 ohm resistor that is made of a material that obeys Ohm's Law. If a voltage drop of 1.000 V is applied across the resistor, its current will be exactly 0.500 A. If a voltage drop of 1000. V is developed, the current will be exactly 500. A.

Few if any real materials obey Ohm's Law in this exact manner. The law is, however, a very useful approximation for most practical electric circuits. We now have a way of using numbers in our model of a circuit.

Example 7.1

Find the current flowing in amperes through the wire in a circuit consisting of a 12. volt battery, a wire, and a 10. ohm resistor.



Need: Current flowing through the wire in amperes.

Know: Voltage provided by the battery is 12. volts, and the resistance is 10. ohms.

How: Sketch the circuit and apply Ohm's Law.

Solve: The voltage drop across the resistor thus is 12.0 volts and, as given by Ohm's law:

$V = I \times R$ or $I = V/R = 12./10. =$ **1.2 A** (two significant figures).

Ohm's law implies a second important property about electric circuits. Recall that for the circuit as a whole, the total drop in voltage in the components other than the battery must be equal in magnitude but opposite in sign to the increase in voltage caused by the battery. Ohm's Law then enables us to find this voltage drop including, if necessary, its implied sign.

In continuing our water analogy, hydraulic resistance to water flow may be thought of as flow restrictions in the water pipe—think of a water valve half shut. However, opening the valve, say, from one quarter open to one half open reduces its hydraulic resistance. Since water is forced through pipes under pressure, it provides energy that might be used, for example, to power a water wheel to run your washing machine. To attempt this, it would be useful to know the relation between water pressure, amount of water flow, and power output from the water wheel. In the same way, electric charge can provide energy to run a wide variety of electrical appliances, from lightbulbs to computers. To attempt this, it is useful to know the relation between voltage, current, and power.

That relation is the second leg of our model—the "**Power Law**"— which was stated in a previous chapter but bears repeating. We have seen that electrons with some charge—say, Q coulombs—"fall" through an electric potential—say, ΔV—in a resistor. This is analogous to a mass m falling through a gravitational potential Δh to produce an energy change $mg\Delta h$. The equivalent electric work done is $Q\Delta V$. The power produced is thus $Q/t \times \Delta V$ in which Q/t is the rate of charge flow, which we call current. The "Power Law" thus simply states that electric power is given by **Power = current × voltage**, or

$$\boldsymbol{P = I \times V} \tag{7.2}$$

Note that power is not a new variable, but it is the same old one that we used earlier in talking about energy. We have inserted quotes around "Power Law" to emphasize that it is not an independent law but perfectly derivable from first principles using the Newtonian definition of power as the rate of working. Power is still measured in watts, where a watt is a joule per second. This fact allows us to relate our model of electrical circuits to our earlier energy models. The analogy between water flow and electricity is thus reasonably complete (see Table 7.1).

Table 7.1 Analogy Between Water Flow in a Pipe and Electricity Flow in a Wire

Water Flow in Hose	Electricity Flow in Wire
Pressure	Voltage
Flow rate	Current
Hydraulic resistance	Electrical resistance
Power in water stream = pressure drop × flow rate	Power in electrical circuit = voltage drop × current

Evidence of the basic "Power Law" relation is provided by a second important property of a resistor in addition to its ability to limit the flow of electricity. That property becomes evident if you force enough charge through the resistance at a high enough voltage and observe that the resistor may then "glow" (unless it melts first!). This is the basis of a very useful invention, the lightbulb (technically the incandescent lamp), invented not, as popularly believed, by Thomas Edison in 1879, but by Humphrey Davy in England and August De La Rive in France in about 1810.[3]

7.3 SERIES AND PARALLEL CIRCUITS

There are two basic types of electric circuits. In a **series circuit**, shown in Figure 7.3A, the current takes a single path. In a **parallel circuit**, shown in Figure 7.3B, the wire branches out into two or more paths and these paths subsequently join up again. In a parallel circuit, envision the current as being divided at the point where the wires branch out and then joining together again where the wires come together.

[3]However, neither scientist took his observations to a practical conclusion as did Edison.

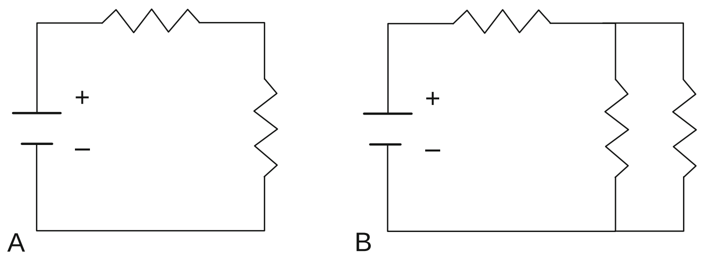


FIGURE 7.3A,B Series Circuit and Parallel Circuit

As previously stated, the current in each of the two branches has to add up to the amount that went in by the conservation of charge principle, and conversely, the total that goes out of the two branches has to equal the total that went in.

The mathematics for determining the voltages and currents in series and parallel circuits rests on two very simple principles: (1) two resistors in a series will have the same current passing through each of them, and (2) two resistors in a parallel will have the same voltage drop across each of them.

Example 7.2

Consider the series circuit in Figure 7.3A, with the battery voltage equal to 12. volts and each resistor with a resistance of 100. ohms. What is the current in the circuit and the voltage drop across each resistor?

Need: Current in circuit and the voltage drop across each resistor.
Know: Battery voltage = 12. volts; resistance of each resistor = 100. ohms.
How: Call the unknown current I and the resistance in each resistor R. By the first principle, the current I must be the same in each series resistor. So by Ohm's Law, the voltage drop across each resistor must be $I \times R$. Since the total voltage drop around the circuit must be zero, $V(\text{battery}) + V(\text{resistor}_1) + V(\text{resistor}_2) = 0$, or $V(\text{battery}) - I \times R_1 - I \times R_2 = 0$. Solve for the unknown I.
Solve: 12. volts − I × 100. ohms − I × 100. ohms = 12. V − I × 200. ohms = 0, or I = **0.060 Amps**.
Then the **voltage drop** across each resistor is $V = I \times R = 0.060 \times 100. =$ **6.0 V**.

Notice that we simply could have added the two 100. ohm resistances together and used their sum of 200. ohm to directly get the current by Ohm's law in one step: $I = 12./200. = 0.060$ A. In a series circuit, the equivalent series resistance R is always the sum of the individual resistances: $R_{eq} = R_1 + R_2 + R_3 + \ldots$.

Example 7.3

Consider the parallel circuit in Figure 7.3B with the battery voltage = 12. volts and each resistor having a resistance of 100. ohms. What is the current drawn from the battery?

Need: Current drawn from battery.
Know: Battery voltage = 12. volts; resistance of each resistor = 100. ohms.
How: Call that current the unknown, I. By the second principle, the voltage drop must be the same across each parallel resistor. Since the total voltage drop around the circuit must be zero, $V(\text{battery}) = V(\text{resistor}_1) = V(\text{resistor}_2)$. Now assume that the current divides between the two resistors with I_1 going through one of

the resistors and I_2 going through the second. By Ohm's law, $V(\text{battery}) = V(\text{resistor}_1) = I_1 \times R_1 = V(\text{resistor}_2) = I_2 \times R_2$. Solve for I_1 and I_2. Since the current coming out of the branches is the sum of the current in each branch, $I = I_1 + I_2$.

Solve: 12. volts $= I_1 \times 100.$ ohms $= I_2 \times 100.$ ohms. Therefore, $I_1 = I_2 = 0.12$ A (since both resistors have the same value here). Then, $\boldsymbol{I} = I_1 + I_2 =$ **0.24 Amps**.

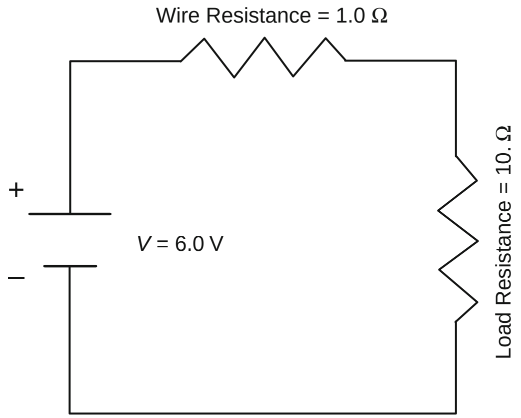
Note that in this case the parallel circuit draws more current than the series circuit. This will generally be true for circuits with similar resistances. It's the same thing as adding a second hose to a garden faucet; provided the supply pressure remains constant, you will spray more water (analogous to an electric current) with two hoses than with one.

The value of a single resistor equivalent to a group of resistors connected in parallel can be easily computed by observing that the total current I is equal to the sum of the currents in each of the parallel resistors, or $I = I_1 + I_2 + I_3 + \ldots$. Since each resistor is exposed to the same voltage, then using Ohm's Law as $I = V/R$ gives $I = V/R_{eq} = V/R_1 + V/R_2 + V/R_3 + \ldots$, or $\mathbf{1/R_{eq} = 1/R_1 + 1/R_2 + 1/R_3 + \ldots}$ where R_{eq} is the equivalent resistance of all the resistors in parallel. Thus the equivalent resistance in Example 7.3 is $1/R_{eq} = 1/100. + 1/100. = 2/100. = 1/50.0$, so $R_{eq} = 50.0$ ohms. Then, $I = V/R_{eq} = 12./50.0 = 0.24$ A as before.

One particular form of series circuit enables us to add more realism to our simple model. Earlier, wires were modeled as conductors with zero resistance. However, a real wire has a small resistance, which is typically proportional to the length of the wire and inversely proportional to its cross-sectional area. In more realistic circuit models, this resistance of the wire is modeled by a resistance element (our familiar wiggly-line symbol) inserted at an arbitrary location into the wire. This resistance will henceforth be called the **wire** resistance, and the resistance of a component such as a lightbulb will be called the **load** resistance.

Example 7.4

For the circuit shown, assume (1) the wire resistance (actually distributed along the whole length of the wire) totals 1.0 ohms, (2) the load resistance is 10. ohms, and (3) the battery voltage is 6.0 volts. Compute the efficiency of the circuit, where the efficiency is defined as the power dissipated in the load divided by the power produced by the battery.



Need: Efficiency of circuit = (power dissipated in load)/(power produced by battery).

Know: Battery voltage = 6.0 volts; wire resistance = 1.0 ohms; load resistance = 10. ohms.

How: Computing efficiency requires two computations of power. Computation of power using our "Power Law" $P = I \times V$ requires knowing the voltage across and the current through each circuit element. So first find the voltage across and current through each element using Ohm's Law. Then compute the power. Then compute the efficiency.

Solve: In a series circuit, the total resistance is the sum of the individual resistors.

So $R = R(\text{wire}) + R(\text{load}) = 10. + 1.0 = 11.$ ohms. The battery voltage is 6.0 volts, so, by Ohm's Law, $I = V/R = 6.0/11. = 0.55$ A.

Now, applying the "Power Law," $P(\text{battery}) = V(\text{battery}) \times I = 6.0 \times 0.55$ [V][A] = 3.3 W. This is the total power drained from the battery.

Also applying Ohm's Law to the wire's resistance, its overall voltage drop is $V(\text{wire}) = 1.0 \times 0.55$ [Ω][A] = 0.55 V.

Applying the "Power Law" to wire losses, $P(\text{wire}) = 0.55 \times 0.55$ [V][A] = 0.30 W.

Therefore, the "Load" power = 3.3 − 0.30 = **3.0 W.**

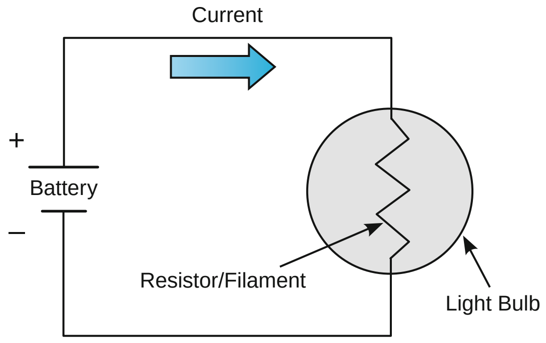
So the efficiency is 3.0/3.3 = **0.91** (note that efficiency, being a ratio of watts to watts, has no units).

One task of an engineer might be to maximize the **efficiency** of application of power, since higher efficiency generally results in lower cost, and lower cost generally is desired by customers. The previous exercise gave one version of that challenge. The next example will give another.

It was Edison, who in 1879, applied the "Power Law" and as a result made the first *useful*[4] lightbulb. He did so by recognizing that a lightbulb is simply a resistor that converts electricity first into heat and then renders part of that heat (typically only about 5%) into visible light. In the example that follows, treat the resistor (indicated by the zigzag line in the figure) inside the bulb (indicated by a circle) as an ordinary, if very high-temperature, resistor that obeys our two electrical laws. The lightbulb's resistor is called its *filament*.

Example 7.5

In the diagram shown, assume the battery delivers a constant 10.0 A of electric current whatever the voltage across the filament may be. If we wish to design the lightbulb to have a power of 100. W, what should the resistance of the lightbulb be? What is the voltage drop across the filament?



Light Bulb Circuit

[4]Edison's major achievement came from considering the entire electrical production, distribution, and lighting problem as a single system design problem. The choice of a filament resistance was only a part of this greater problem. His rivals were unsuccessful because they were concerned only with producing lamps.

Need: $R =$ ____ Ω and $V =$ _____ V.
Know: P and I.
How: $P = I \times V$ and $V = R\,I$.
Solve: $P = I \times V$ or $\boldsymbol{V} = P/I = 100./10.0$ [W]/[A] = **10.0 volts** to three significant figures.
Therefore, $\boldsymbol{R} = V/I = 10.0/10.0$ [V/A] = **1.00 ohms** (to three significant figures).

If the wire in this example also had a 1.00 ohm resistance, as much power would be dissipated in it as in the bulb's filament. It would have needed another 10.0 volts of voltage, so the battery would have had to supply 20.0 volts.

Having resistance in the wire comparable to that of the filament complicated the problem. The key to Edison's problem of maximizing the bulb's efficiency was in maximizing the efficiency of the *entire* circuit, rather than concentrating on the bulb alone. Higher efficiency meant lower costs, eventually low enough for electricity and the lightbulb to become mass consumer products.

In all the circuits we have discussed so far, the current always flows in the same direction (shown, according to convention, flowing from positive to negative around a circuit). These are called **direct current circuits**. However, there is another class of electrical circuits that allows the current to reverse its direction many times a second. These **alternating current circuits** are the basis for the vast number of electrical technologies beyond the lightbulb. They range from the 1000-megawatt generators that supply electricity, to the motors that put electricity to work, to the radio wave generators that enable us to send out radio signals, or to thaw frozen food in minutes in our microwave ovens.

7.4 KIRCHHOFF'S LAWS

Gustav Robert Kirchhoff (1824–1887) was a German physicist who contributed to the fundamental understanding of electrical circuits, spectroscopy, and the emission of thermal radiation by heated objects. Kirchhoff formulated his circuit laws through experimentation in 1845, while still a university student.

7.4.1 Kirchhoff's Voltage Law

Kirchhoff's Voltage Law is a form of the Conservation of Energy Law, and can be stated as:

> ***The algebraic sum of the voltage drops in a closed electrical circuit is equal to the algebraic sum of the voltage sources (i.e., increases) in the circuit.***

Mathematically, Kirchhoff's Voltage Law is written as:

$$\Sigma\, V_{(\text{closed loop})} = \Sigma\, IR_{(\text{closed loop})} = 0 \tag{7.3}$$

where Σ is the summation symbol.

A closed loop can be defined as any path in which the originating point in the loop is also the ending point for the loop. No matter how the loop is defined or drawn, the sum of the voltages in the loop must be zero. Loop 1 and loop 2 in Figure 7.4 are both closed loops within the circuit. The sum of all voltage drops and rises around loop 1 equals zero, and the sum of all voltage drops and rises in loop 2 must also equal zero. There is also a third closed loop in this circuit, the one that goes around the outside from A to B to C to D and back to A again. But this loop is not independent of the other two, so it does not provide any additional information.

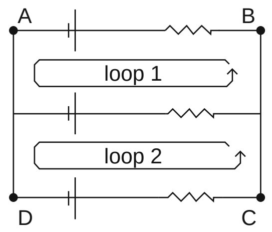


FIGURE 7.4 Closed Loops in a Simple Electrical Circuit

Voltage Dividers

If two resistors are in series, there is a voltage drop across each resistor, but the current through both resistors must be the same. A simple circuit with two resistors in series with a source (voltage supply) is called a **voltage divider**.

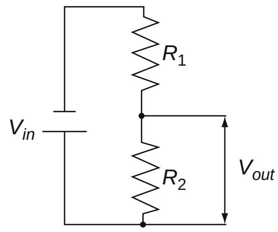


FIGURE 7.5 Voltage Divider Circuit

In Figure 7.5, the voltage V_{in} is dropped across resistors R_1 and R_2. If a current I flows through the two series resistors then by Ohm's Law: $I = V_{in}/(R_1 + R_2)$. Then, $V_{out} = IR_2$ or

$$V_{out} = V_{in} \times [R_2/(R_1 + R_2)] \tag{7.4}$$

Hence, V_{in} is divided, or reduced, by the ratio $R_2/(R_1 + R_2)$.

Example 7.6

Suppose you want to build a voltage divider that will reduce the voltage by a factor of 15. You visit your local Radio-Shack store and find that they have 5.0, 10., 15., and 20. ohm resistors in stock. Which ones do you choose to make your voltage divider in the simplest and most economical way?

Need: An inexpensive voltage divider such that $V_{out}/V_{in} = 1/15. = 0.067$.

Know: The voltage divider Equation (7.4), and a supply of resistors.

How: Since we know that $V_{out}/V_{in} = R_2 / (R_1 + R_2) = 0.067$, we could just try the values of the resistors in stock to see if we can get close to the required ratio. This is a trial-and-error method and in the general case well suited to a spreadsheet analysis.

Solve: By inspection, we see that R_1 must be much bigger than R_2 to make V_{out}/V_{in} a small number. So we choose the smallest available resistor, 5.0 ohms, for R_2 and then solve the voltage divider equation for R_1. This gives $R_1 = 70.$ ohms. Using the available resistors we can come close to this value by connecting seven 10. ohm resistors in series with a 5.0 ohm resistor. This will make $R_1 = 70.$ ohms and then $V_{out}/V_{in} = 0.067$. Of course, if each of the 10. ohm resistors has a *tolerance* of ± 1. ohm, the range on $V_{out}/V_{in} = 0.06$ to 0.08.

7.4.2 Kirchhoff's Current Law

Kirchhoff's Current Law is a form of the Conservation of Charge Law, and can be stated as:

The algebraic sum of all the currents at a node must be zero.

A node is any electrical junction in a circuit. Kirchhoff's Current Law can be written mathematically as:

$$\Sigma\, I_{(\text{node})} = 0 \tag{7.5a}$$

where Σ is the summation symbol. Stated differently as:

The sum of all the currents entering a node is equal to the sum of all the currents leaving the node.

or

$$\Sigma\, I(\text{entering a node}) = \Sigma\, I(\text{leaving the node}) \tag{7.5b}$$

Current flows through wires much like water flows through pipes. The amount of water that enters a branching junction in a pipe system must be the same amount of water that leaves the junction. The number of pipes in the branching junction does not change the net amount of water (or current in our case) flowing (see Figure 7.6).

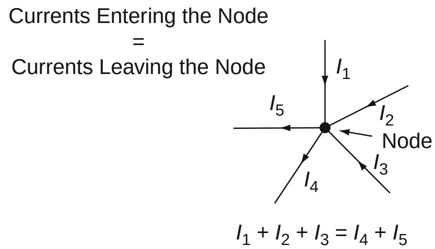


FIGURE 7.6 Currents Entering and Leaving a Circuit Node

Current Dividers

If two resistors are in parallel, the voltage across them must be the same, but the current divides according to the values of the resistances. A simple circuit with two resistors in parallel with a voltage source is also called a current divider.

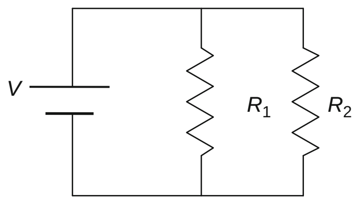


FIGURE 7.7 Parallel Resistor Current Divider

The equivalent resistance of two resistors R_1 and R_2 in parallel in Figure 7.7 is:

$$1/R_{eq} = 1/R_1 + 1/R_2$$

which can be algebraically rearranged as

$$R_{eq} = (R_1 + R_2)/(R_1 \times R_2) \quad (7.6)$$

Now, if the voltage across these resistors is V, then the current I flowing in the circuit before the division is, according to Ohms Law:

$$I = V/R_{eq} = V \times (R_1 + R_2)/(R_1 \times R_2)$$

The current through R_1 according to Ohms Law is:

$$I_1 = V/R_1$$

Dividing this equation by the previous one and solving for I_1 gives:

$$I_1 = I \times R_2/(R_1 + R_2) \quad (7.7)$$

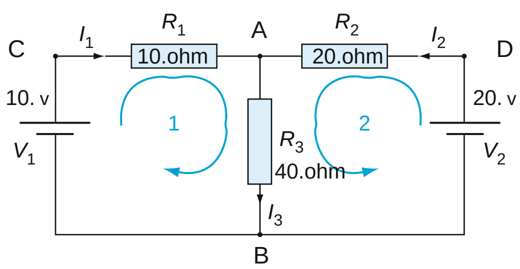
You can also show that

$$I_2 = I \times R_1/(R_1 + R_2) \quad (7.8)$$

Thus, the current I has been divided, or reduced, by the resistance ratio.

Example 7.7

Find the current through and voltage across the 40. ohm resistor, R_3, in the following circuit.



Need: $I_3 =$ ____ A and $V_3 =$ ____ V.

Know: Kirchhoff's current and voltage laws, Equations (7.3) and (7.5).

How: This circuit has four nodes (A, B, C, and D) and two independent loops. Write the current law equations for the two nodes and the voltage law equations for the two loops. Then solve the resulting set of algebraic equations for the currents.

Using Kirchhoff's current law, for node A the equation is $I_1 + I_2 = I_3$

Using Kirchhoff's voltage law around loops 1 and 2, the equations are:

$$\text{Loop 1:} \quad 10. = R_1 \times I_1 + R_3 \times I_3 = 10. \times I_1 + 40. \times I_3$$

$$\text{Loop 2:} \quad 20. = R_2 \times I_2 + R_3 \times I_3 = 20. \times I_2 + 40. \times I_3$$

Since I_3 is the sum of $I_1 + I_2$ we can substitute this into the loop equations and rewrite them as:

$$\text{Loop 1:} \quad 10. = 10. \times I_1 + 40. \times (I_1 + I_2) = 50. \times I_1 + 40. \times I_2$$

$$\text{Loop 2:} \quad 20. = 20. \times I_2 + 40. \times (I_1 + I_2) = 40. \times I_1 + 60. \times I_2$$

Solve: Now we have two simultaneous equations that can be solved algebraically to give the values of I_1 and I_2. Algebraically solving the Loop 1 equation for I_2 and substituting this expression into the Loop 2 equation allows us to solve for I_1 as: $\boldsymbol{I_1} = -\mathbf{0.14\ A.}$

Now substituting this value of I_1 into either Loop equation allows us to solve for I_2 as $\boldsymbol{I_2} = +\,\mathbf{0.43\ A}$. Since $I_3 = I_1 + I_2$, the current flowing in resistor R_3 is: $\boldsymbol{I_3} = -0.14 + 0.43 = \mathbf{0.29\ A}$, and, from Ohm's Law, the voltage drop across the resistor $\boldsymbol{R_3}$ is $0.29 \times 40. = \mathbf{11.\ V}$.

The negative sign we got for I_1 in this example simply means that we initially assumed the wrong direction for the current flow, but the numerical value is still correct. In fact, the 20-volt battery is actually charging the 10-volt battery (i.e., putting electrical energy back into the battery by forcing an electric current through it).

7.5 SWITCHES

Here we will concentrate on a humble circuit element whose importance was appreciated by Edison: the switch. In our water flow model, a switch is equivalent to a faucet. It turns the flow on or off. Surprisingly, it is this simplest of components that is the basis of the most sophisticated of our electrical technologies, the computer. In its essence, a computer is nothing more than a box filled with billions of switches, all turning each other on and off. The faster the switches turn each other on and off, the faster the computer can do computations. The smaller the switches, the smaller the computer can be. The more reliable the switches, the more reliable the computer. The more efficient the switches, the less power the computer uses. The criteria engineers have used in developing better switches are higher speed, smaller size, greater reliability, and increased efficiency.

In the late nineteenth century, a few farsighted individuals, such as the British economist William Stanley Jevons and the American mathematician Allen Marquand, proposed that the humble switch could be used to do something remarkable: carry out calculations in the field of philosophy and mathematics, called logic. (The mathematics of logic will be discussed in a later chapter.) Here we will stick to the underlying electrical technology: the switch.

Unlike the elements we have discussed so far, the switch has two states. As indicated in Figure 7.8A, a switch can be open, in which case it presents an infinite resistance to the flow of electric current (that is, no current flows at all), or a switch can be closed, in which case ideally it presents zero resistance to an electric current.

Jevons and Marquand envisioned using switches to automate logic, but the technology of their time was not up to the task. One of the pioneers who turned their idea into twentieth-century reality was George R. Stibitz. Along with other inventors in the 1930s, he accomplished automation of computation using a type of switch component called a **relay**.

A relay is a switch operated by an electromagnet (Figure 7.8B). The relay simply closes the switch against the spring when the external power source is turned on. When the external power source is turned off, the spring returns the switch to its open position. It got its name from its original use, relaying telegraph messages over longer distances than could be accomplished by a single circuit. The relay consists of two parts. The first

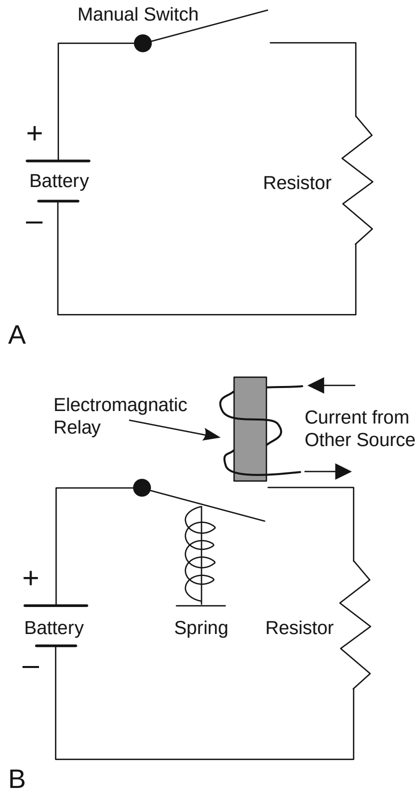


FIGURE 7.8A, B Simple Switches

is an electromagnet that is powered by one circuit, which we will call the driving circuit. The second part is another circuit with a metal switch in it, called the driven circuit. The driven circuit is placed so that when the driving circuit is on, its electromagnet attracts the moving metal part of the driven circuit's switch. This results in closing the gap in the driven circuit and turning it on.

Relays continue to be used in control systems because they are highly reliable, but they are high in power demands, slow, and bulky. In the twentieth century the basis of a faster switch emerged. It was based on an invention used in radios, a circuit element called the vacuum tube. Vacuum tubes were used as switches because they were fast. However, they were inefficient, bulky, and unreliable. By the 1940s, an even more revolutionary idea was already in the wings. It emerged in initial form in 1946, when three physicists at Bell Laboratories—John Bardeen, Walter Brittain, and William Shockley—invented the transistor, a solid state switch developed to replace the vacuum tube in communications. How the transistor works is a topic for an advanced course in electronic engineering or physics. For our purposes, we can carry on the analogy with the relay and the vacuum tube. Figure 7.9 shows one important type of transistor, the metal oxide semiconductor field effect transistor (MOSFET), a type that came along several years after the original one invented by Bardeen, Brittain, and Shockley. In the MOSFET, a driving circuit called a gate plays the role the magnet plays in the relay, and the source plays the role of a heated metal grid in the vacuum tube. Together they open and close (like a switch) a driven circuit that determines the state of the output.

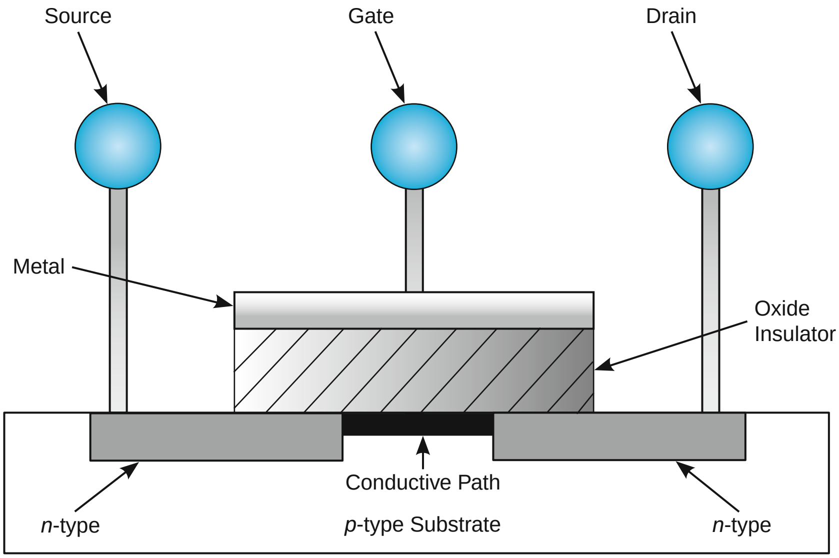


FIGURE 7.9 Principles of MOSFET Construction

A transistor is made of semiconductors—materials that have properties midway between those of a conductor and those of an insulator. Most of today's transistors use semiconductors made out of the element silicon. By "doping" silicon—that is, by adding trace amounts of other elements—its conductivity can be controlled. As a simple model of a particularly useful type of transistor called the field effect transistor, we can imagine two such pieces of doped silicon, a source that serves as a source of electrical charge, and a drain that serves as a sink into which the charge flows. The source and drain are connected by a semiconductor channel, which is separated by an insulator from a metal electrode called a gate. Some of the semiconductor is doped with material that sustains an excess of electrons and is called *n*-type (*n* for negative). Other semiconductors are doped with materials that sustain a deficiency of electrons called *p*-type (*p* for positive). When the gate is given a positive voltage in the *npn* MOSFET shown in Figure 7.9, this promotes the flow of charge from source to drain, effectively making a thin layer of continuous *n*-type conductive channel between the source and the drain. When the gate is given a negative voltage, current cannot flow from source to drain, and the transistor acts as an open switch.

Transistors (switches) are smaller, faster, more efficient, and more reliable than vacuum tubes. They made possible practical hearing aids, pocket-sized portable radios, and computers that were the size of refrigerators instead of filling entire rooms. But to get to the modern computer, one more step was needed. This was the **integrated circuit**, independently co-invented by Nobel Prize–winning engineer Jack Kilby and Intel cofounder Robert Noyce in the early 1960s.

A modern integrated circuit contains millions (or billions) of transistors connected together to form a logic circuit on a piece of silicon the size of a fingernail. This is done by a process called **photolithography**, which literally means "using light to write on a stone." The circuit designer first makes a large drawing of the circuit. Optical lenses then are used to project a very small image of the drawing onto a fingernail-sized chip of silicon, which is coated with special chemicals.

Modern integrated circuits might contain the level of detail that would be seen on a street map of the entire state of California. Packing all this complexity into a small space launched one of the most dramatic technological revolutions in history. Since 1960, the number of electronic computations possible per second and the number of transistors that can be packed onto a single chip have doubled every one or two years (and since 1990, rather than slowing, the pace has actually speeded up).[5]

Integrated circuits are the hardware that made personal computers possible. But integrated circuits by themselves do not constitute a computer. A crucial distinction needed to understand computers is **hardware versus software**. Hardware is the collection of mechanical, electrical, or electronic devices that make up a computer. Among the hardware in a modern computer, two elements are crucial: (1) a system called a **central processing unit** or **CPU**, which carries out binary logic and arithmetic; and (2) memory, often called **RAM** (random access memory) or **ROM** (read-only memory), which store the ones and zeroes corresponding to inputs and outputs of the CPU. Both CPU and memory are integrated circuits composed of many millions of switches.

What enables that collection of switches to carry out computations? That is the job for a sequence of instructions, called **software**. Two important applications of software, control and computation, will be discussed in later chapters in this text where we will model the fuel control system of an automobile using computer logic.

SUMMARY

The flow of electricity through wires can be modeled in the same way we might treat the flow of water through pipes. The model has as its basic variables **charge**, **current**, **voltage**, and **resistance**. The simplest model consists of **Ohm's Law**, the "**Power Law**," and **Kirchhoff's Voltage** and **Current Laws**. Using this model, we can analyze the operation of a particular class of electric circuits called **direct current** circuits and, in particular, two classes of direct current circuits called **series** circuits and **parallel** circuits.

An important type of electric circuit component is the **switch**. On the hardware side, the development of the computer can be viewed as the search for ever-faster switches, based first on **relays**, then on **vacuum tubes**, then on **transistors**, and finally on the form of electronics called **solid state integrated circuits**. This form of electronics provides the **hardware**, which in conjunction with a sequence of instructions called **software**, makes modern computation and control possible.

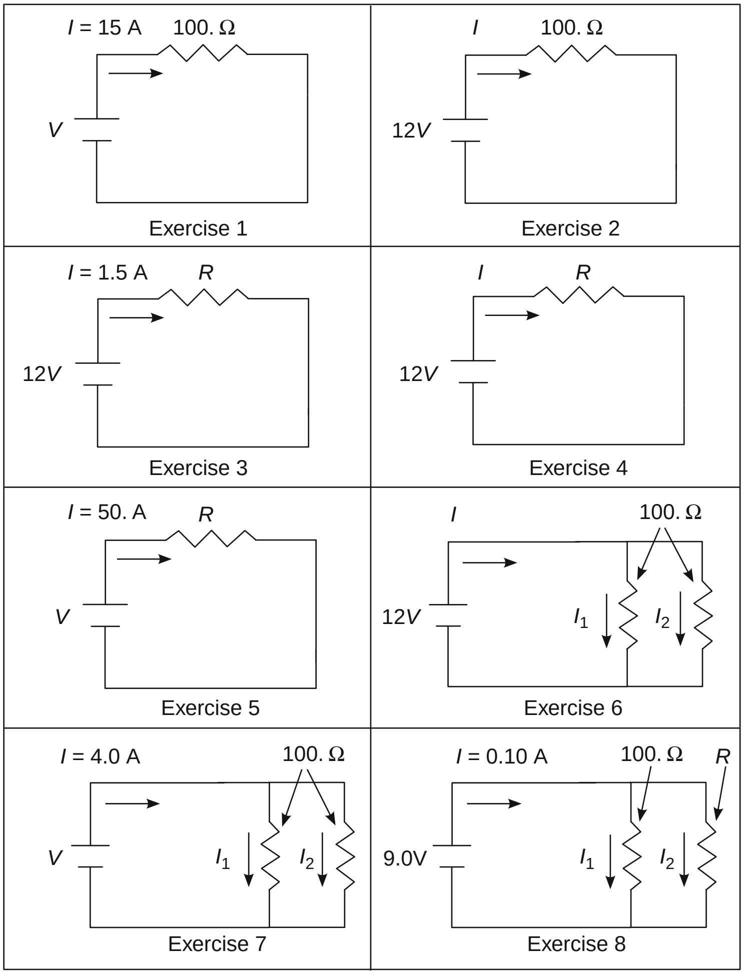
EXERCISES

If no circuit sketch is already given, it is highly recommended you *first draw each circuit* to answer these problems. It is generally useful to notate your diagram with what you think you know, such as voltages and currents at each point.

Note the equivalence among these electrical units: $[\Omega] = [V]/[A]$, $[W] = [V][A] = [\Omega][A]^2 = [V]^2/[\Omega]$ and all other combinations of these.

[5] This is Moore's Law. In 1965 by Gordon Moore, co-founder of Intel, noted that the number of transistors per square inch on integrated circuits had doubled every 12 months since the integrated circuit was invented. Moore predicted that this trend would continue for the foreseeable future. In subsequent years, transistor density has doubled approximately every 18 months, and this is the current definition of Moore's Law. Most experts, including Moore himself, expect Moore's Law to hold until at least 2020.

The circuits for **exercises 1 through 8** are given in the table.



1. For the circuit shown, find the voltage V. (**A: 15** $\times$ **10^2 V**)

2. For the circuit shown, find the current I.

3. For the circuit shown, find the resistance R.

4. For the circuit shown, a power of 100. watts is dissipated in the resistor. Find the current I. (**A:** $=$ **8.3 A**)

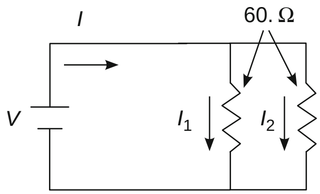
5. For the circuit shown, a power of 100. watts is dissipated in the resistor. Find the resistance, R.

6. For the circuit shown, find the current I. (**A: 0.24 A**)

7. For the circuit shown, find the voltage V.

8. For the circuit shown, find the resistance R.

9. For the circuit shown, a power of 100. watts is dissipated in each resistor. Find the current I and the voltage V.



For **exercises 10 through 19**, draw the circuit and solve for the unknown quantity requested.

10. A circuit consists of a 3.0 V battery and two resistors connected in series with it. The first resistor has a resistance of 10. ohms. The second has a resistance of 15. ohms. Find the current in the circuit. (**A: 0.12 A**)

11. A circuit consists of a 12. V battery and a resistor connected in series with it. The current is 105. A. Find the resistance.

12. A circuit consists of a 9.0 volt battery and two parallel branches, one containing a 1500 ohms resistor and the other containing a 1.0×10^3 ohm resistor. Find the current drawn from the battery. (**A: 0.015 A**)

13. A circuit consists of a 12. volt battery attached to a 1.0×10^2 ohm resistor, which is in turn connected to two parallel branches, each containing a 1.0×10^3 ohm resistor. Find the current drawn from the battery.

14. An automobile's 12. V battery is used to drive a starter motor, which for several seconds draws a power of 3.0 kW from the battery. If the motor can be modeled by a single resistor, what is the current while the motor is operating? (**A: 250 A**)

15. An automobile's 12. V battery is used to light the automobile's two headlights. Each headlight can be modeled as a 1.00 ohm resistor. If the two headlights are hooked up to the battery in series to form a circuit, what is the power produced in each headlight? Why would you *not* wire car lights in series?

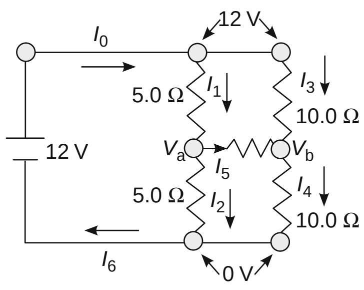
16. An automobile's 12. V battery is used to light the automobile's two headlights. Each headlight can be modeled as a 1.00 ohm resistor. If the two headlights are hooked up to the battery in parallel to form a circuit, what is the power produced in each headlight? Why is a parallel circuit preferred for this application?

17. An automobile's 12. V battery is used to light the automobile's two headlights. Each headlight can be modeled as a single resistor. If the two headlights are hooked up to the battery in parallel to form a circuit, and each headlight is to produce a power of 100. W, what should the resistance of each headlight be? (**A: 1.4 ohms per bulb**)

18. Suppose one of the headlights in exercise 17 suddenly burns out. Will the power produced by the other headlight increase or decrease?

19. Suppose a car has headlights operating in parallel but with different resistances—one of 2.0 ohms and the other of 3.0 ohms. Suppose the headlight parallel circuit is connected in series with a circuit for a car stereo that can be modeled by a 1.0 resistor. What is the current being drawn from a 12. V battery when both the lights and the stereo are on? (**A:** $\boldsymbol{I}$ **= 5.5 A**)

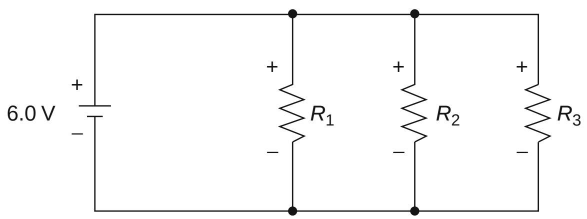
20. Consider the circuit shown with exercise 9. What are the currents I_0 through I_6? (**Hint:** What are the voltages V_a and V_b? Note the symmetry between $R_1/R_2 = 5.0/5.0$ and $R_3/R_4 = 10.0/10.0$.)



21. Determine the output voltage, V_{out}, in Figure 7.5 if $V_{in} = 90.$ V, $R_1 = 10.$ ohms, and $R_2 = 50.$ ohms.

22. Use the same RadioShack stock resistors in Example 7.6 to design the construction of a voltage divider that will reduce the voltage by a factor of 30.

23. Use Kirchhoff's voltage and current laws to determine the voltage drop and current in each of the following three resistors if $R_1 = 10.$ ohms, $R_2 = 20.$ ohms, and $R_3 = 30.$ ohms.



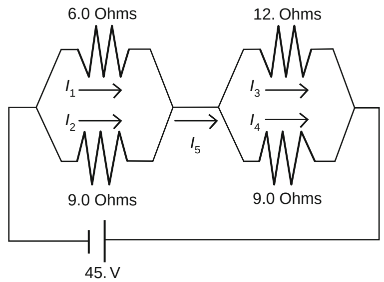
24. If the resistors in exercise 23 are all equal to 10. ohms ($R_1 = R_2 = R_3 = 10.$ ohms), what is the current supplied by the 6.0 volt battery?

25. Determine the currents I_1 and I_2 in the current divider illustrated in Figure 7.7 if $V = 50.3$ volts, $R_1 = 125.$ ohms, and $R_2 = 375.$ ohms.

26. Repeat Example 7.7 with the resistor R_1 changed to 100. ohms and all other values unchanged.

27. Repeat Example 7.7 with V_1 increased to 100. volts and V_2 reduced to 5.0 volts.

28. Determine the currents I_1, I_2, I_3, and I_4 in the following figure.



29. It is the last semester of your senior year and you are anxious to get an exciting electrical engineering position in a major company. You accept a position from company A early in the recruiting process, but you continue to interview, hoping for a better offer. Then your dream job offer comes along from company B. More salary, better company, more options for advancement, it is just what you have been looking for. What do you do?

a. Accept the offer from company B without telling company A (just don't show up for work).
b. Accept the offer from company B and advise company A that you have changed your mind.
c. Write company A and ask them to release you from your agreement.
d. Write company B thanking them for their offer and explain that you have already accepted an offer.

Summarize using an Engineering Ethics Matrix.

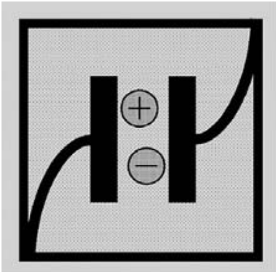
30. A female student in your class mentions to you that she is being sexually harassed by another student. What do you do?

a. Do nothing; it is none of your business.
b. Ask her to report the harassment to the course instructor.
c. Confront the student accused of harassment and get his or her side of the issue.
d. Talk to the course instructor or the college human resource director privately.

Summarize using an Engineering Ethics Matrix.

CHAPTER 8

Electrochemical Engineering and Alternate Energy Sources



Source: http://www.gdch.de/strukturen/fg/aelchem_e.htm

Alternate energy sources are the hopes for a tomorrow when we can eliminate, or at least relieve, the heavy dependence we have today on petrochemical fuel sources. One large vision, still under development, is the use of electrochemical engineering systems to propel electric cars, to improve electronic devices, and to supplement solar energy systems.

But what is electrochemical engineering and why do engineers care about it? And, as a follow-up question: what kinds of engineers are engaged in the study and development of electrochemical engineering? If we answer the first of these questions then the answers to the subsequent ones make logical sense.

8.1 ELECTROCHEMISTRY

Electrochemistry is the science behind batteries and fuel cells, and as such, plays an essential role in the development of novel energy sources. We will get to fuel cells in good time; batteries are common enough, so we will start there.

You are probably aware there is a heavy duty battery in every automobile (to activate the starter at the twist of the ignition key as well as for auxiliary uses in headlights and in car radios, etc.). But there are also small and familiar batteries that power flashlights, electronic devices, computers, and much more. In addition, today there are also special batteries that are an integral part of a hybrid automobile. (Hybrid here has the meaning that an automobile has two sources of motive power, an electric motor and a relatively small assisting gasoline engine.)

There are batteries intimately involved in various alternate energy concepts. For example, battery electrical storage units may be used in solar energy systems, be they small systems for individual houses or large ones for central power stations (Figure 8.1).

A problem posed by such schemes is that they are subject to the vagaries of the weather as well as fluctuations in the demand for the electricity that is produced because electricity is the ultimate "consumable." It has to be used immediately because it can't be stored in the power grid. In other words, you can make electricity when the Sun is shining (or the wind is blowing if you use windmill power). But what happens if you don't have an immediate use for that energy? Do you somehow discard it even though this is a very wasteful thing to do?

And what happens when you do need more power than you are producing at that moment? Suppose you use electric power for cooking the morning and evening meals. Your highest demand probably is in the early

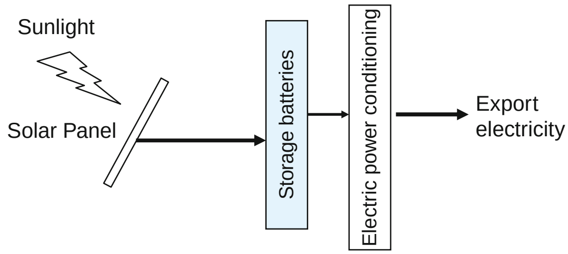


FIGURE 8.1 Solar Power Schematic.

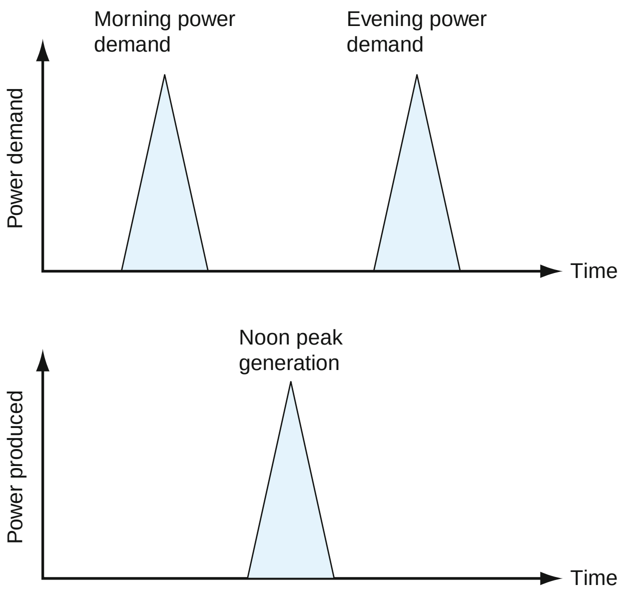


FIGURE 8.2 Mismatch of Daily Power Production and Need.

morning and evening before the Sun has come up and after the Sun has gone down (or when the winds are quiet). Figure 8.2 shows the problem.

Obviously what you want to do is store the excess energy when it is abundant (noon in Figure 8.2) and use it when it is required (early morning and evening in Figure 8.2). Batteries are a relatively simple way to achieve this harmony. Individual house solar installations may rely on this tactic, known as "load leveling," as shown in Figure 8.3.

The economics of solar power will depend on the efficiency of battery storage units that save the electric energy for when it is needed. Consequently, batteries will be a main technological and economic element of alternative energy practices.

The other major topic of this chapter related to batteries is fuel cells. What are they and why do we need them? Fuel cells are simply continuously refueled batteries. Normal batteries such as the battery in a car, or a flashlight battery, or a rechargeable computer battery, have a finite amount of chemical energy stored in them. That's why batteries go dead after extended use—the potential energy stored in their electrodes/electrolyte (the chemical heart of a battery) has been extracted and the equivalent amount of electrical energy has been

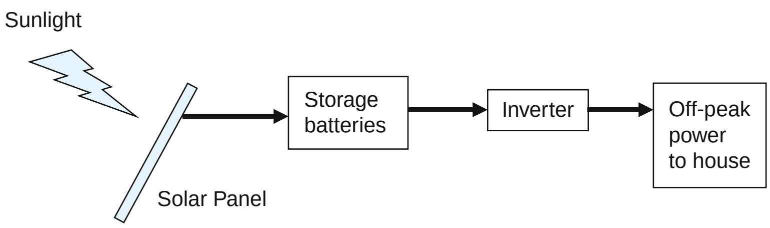


FIGURE 8.3 Load Leveling for Solar Energy. An Inverter Converts Direct Current (DC) into Alternating Current (AC).

used. A fuel cell does not suffer this limitation. It is constantly replenished with fresh chemicals and thus fresh chemical energy as more electricity is demanded. The high energy feed chemicals (e.g., hydrogen gas or methanol) produce effluents that are low energy such as water or carbon dioxide.

We can now answer the initially posed question of what kind of engineers are interested in electrochemical engineering. Since chemistry is heavily involved, **chemical** and **materials engineers** have been at the vanguard along with **mechanical** and **electrical engineers** who integrate such systems into useable energy resources.

Example 8.1

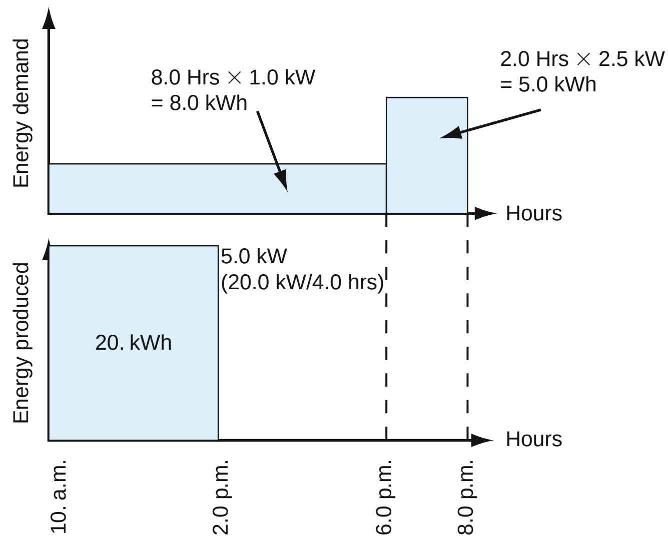
Between 10:00 AM and 2:00 PM on an otherwise very cloudy day the Sun comes out and a solar-powered house produces and stores 20. kWh of electric energy. (**Note**: 1.000 kWh is the energy produced by 1.000 kW for 1.000 hour and is equal to 3,600. kJ.) The rest of the day it produces nothing. The house requires 1.0 kW all the time except between 6:00 and 8:00 PM when it requires 2.5 kW. Will a 10. kWh storage battery be large enough to supply the house needs between 6:00 and 8:00 PM?

Need: Size of storage battery for load leveling.

Know: House uses 1.0 kWh except between 6:00 and 8:00 PM when it requires 2.5 kW. It stores 20. kWh between 10:00 AM and 2:00 PM.

How: We need an energy balance relating the energy flow and the net energy stored. We'll make the assumption that there was no energy stored until 10:00 AM.

Solve: The **easiest** way to solve this kind of problem is to graph it.



As the schematic shows, there is the need to store 1.0 kW for 4.0 hours (4.0 kWh) from 2:00 PM until 6:00 PM and to store another 5.0 kWh between 6:00 PM and 8:00 PM for a total of 9.0 kWh. We produced a net 20. – 4.0 = 16. kWh of electric power, much of which we will have to dispose of, but our 10.0 kWh battery storage is large enough for our needs. The additional 7.0 kWh may be added to the electric power grid if the local utility allows it.

8.2 PRINCIPLES OF ELECTROCHEMICAL ENGINEERING

Why does a battery work at all? Where do the electrons come from that carry the current in the external wires? To partially answer we will briefly digress into elementary chemistry. First, many simple compounds are **ionic**, meaning they are held together by the balance of positive and negative electrical charges. The perfect example is common salt, which can be symbolically written either as NaCl or, more accurately, as Na^+Cl^-. Ionic compounds are formed by the transfer of an electron from an electrically neutral atom or molecule to another electrically neutral atom or molecule. In a *crystal* of common salt, these charges strongly hold the crystal lattice together and are called Columbic forces, which are nothing but the forces of electrostatics governed by Coulomb's law.

$$F = \frac{e^2}{kr^2}$$

Here F is the force between two electron charges of magnitude $e = 1.60 \times 10^{-19}$ coulombs held apart at a distance of r. The constant k is the **dielectric constant** of the medium in which the compound is immersed. When these charges are of opposite signs this force is strongly attractive and is responsible for the crystalline structure of common table salt.

If an ionic crystal is immersed in water instead of in air, the relative value of k increases by a factor of 80 and the corresponding force between atoms in the Na^+Cl^- crystal is loosened by the same factor of 80. The result is that positive ions of Na^+ and negative ions of Cl^- will dissolve in water and the ions will drift apart. Thus a solution of Na^+Cl^- becomes a solution of two kinds of ions, Na^+ and Cl^- in equal proportion (and thus neutral overall). The trick in any electrochemical device is to physically separate these charges and create a voltage potential; this potential can then drive electrical charges through the external circuit.

Before we describe any electrochemical applications, it is necessary to define the terms electrolyte and electrodes (the latter being further subdivided into anodes and cathodes). Figure 8.4 shows these concepts.

Ben Franklin arbitrarily decided that electric current flowed from the positive to the negative terminal. Much later the British physicist J.J. Thompson discovered that it is negatively charged electrons that carry the current in the opposite direction assumed by Franklin, i.e., from negative to positive. It's much easier to think in terms of electrons than conventional current flows. For a discharging battery the key concept is that the anodes churn out electrons and cathodes consume them. Conventional current during discharge flows into the anode (so it will be marked as negative) and away from the cathode (so it will be marked as positive).

8.3 LEAD-ACID BATTERIES

As our first electrochemical example, consider today's lead-acid batteries. The principles of their operation are conceptually similar to those for most electrochemical devices. The lead-acid battery was invented more that 150 years ago and has been much improved with detailed understanding of its internal chemistry.

The electrodes are basically lead (chemical symbol Pb) plus several other metals added to improve its performance. The other electrode is lead coated with a lead oxide, known as litharge. The electrolyte is about one

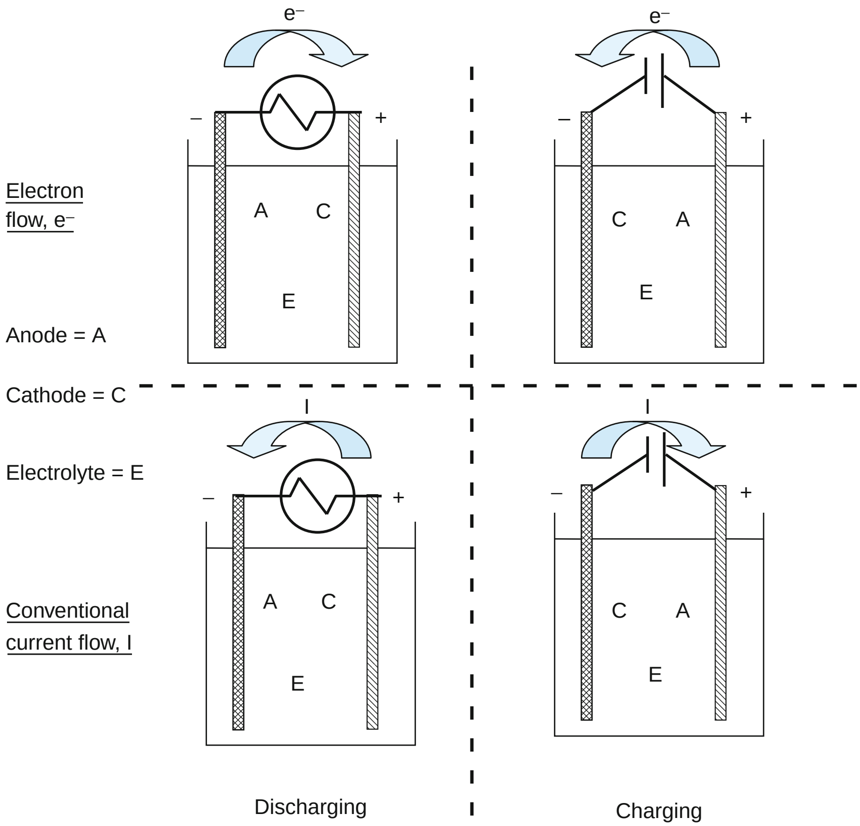


FIGURE 8.4 Some Electrochemical Terms.

third sulfuric acid (H_2SO_4) in water. The sulfuric acid solution in water may be thought of as dissociating into its constituent ions $2H^+$ and $SO_4^=$.

The discharge chemistry at the anode will dissolve some lead and put positive lead ions into solution (since electrons are removed from anodes and flow into the external circuit).

Anode: The anode is made of "pure" lead. The principle anodic reaction is:

$$Pb = Pb^{++} + 2e^-$$

The lead ions Pb^{++} are doubly charged and will combine with the sulfate ion from sulfuric acid and will immediately precipitate as highly insoluble salt lead sulfate.

$$Pb^{++} + SO_4^= = PbSO_4$$

The removal of sulfate ions from solution means they will no longer match the hydrogen ions concentration and the excess H^+ ions will migrate across the electrolyte toward the cathode.

In the case of a lead-acid cell, the reactions near the anode produce a voltage of about 0.36 V and near the cathode another 1.69 V for a total cell voltage of about 2.0 V (see Figure 8.5). There is also a voltage loss

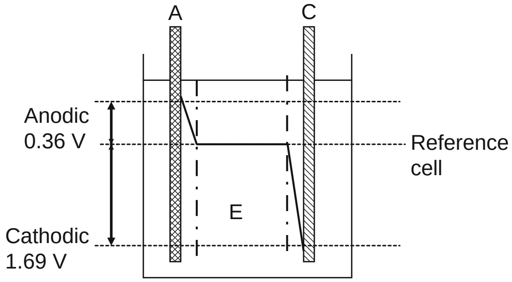


FIGURE 8.5 Cell Potentials for Lead-Acid Half Cells

across the electrolyte, which we have ignored, that is referred to as Ohmic loss because it is explained by Ohm's Law. It occurs only when current is being drawn from the battery.

Cathode: The cathode is made of a sheet of lead coated with a paste of lead oxide (PbO_2). During discharge, the cathode receives electrons from the external circuit, which, in turn, helps to convert the lead oxide to its insoluble sulfate.

$$PbO_2 + 4H^+ + SO_4^= + 2e^- = PbSO_4 + 2H_2O$$

Both cathodic and anodic reactions remove lead ions and sulfate ions from the electrolyte and thus deplete the charge. The battery can be recharged several hundred times by applying a reverse voltage slightly in excess of 2.05 V per cell. Eventually the repeated solution/dissolution of the lead sulfate will produce debris that falls to the bottom of the cell and will eventually fatally short it.

A car battery has six lead-acid cells in series and weighs about 25 kg. It can deliver 200 to 300 A at about 12 V for several minutes. The principal reason for its heavy weight is that the density of lead is more than 11 times that of water. This fact is important for the future use of lead-acid batteries as energy storage devices.

Table 8.1 Measures of a Lead-Acid Battery Efficiency[1]

Battery mass	25 kg
Energy content	3,000 kJ
Mass energy storage density	120 kJ/kg
Volumetric energy storage density	250 kJ/liter
Power	5 kW

Example 8.2

Compare the statistics in Table 8.1 to that in 25. kg of gasoline. Comment on the prospects for all-battery driven cars in the future.

Need: Energy in 25. kg of gasoline and compare to Table 8.1.
Know: Combustion energy of gasoline is the same as its mass energy storage density, or about 46,500 kJ/kg (see Chapter 5).

[1]Derived in part from http://en.wikipedia.org/wiki/Lead-acid_battery and http://www.wdv.com/Hypercars/EAATalk.html

How: We will need the **mass density** of gasoline to compute its volumetric energy storage density. From tables elsewhere we find that the density of gasoline is 740. kg/m^3 or 0.740 kg/liter.

Solve: The energy content of gasoline is 46,500 × 25. [kJ/kg][kg] = **1.2 × 10^6 kJ**. For the volumetric energy storage density, 46,500 × 0.740 [kJ/kg][kg/liter] = **34,400 kJ/liter**.

Property	Lead-Acid Battery	Gasoline
Mass	25. kg	25. kg
Energy	3,000 kJ	1.2 × 10^6 kJ
Mass energy storage density	120 kJ/kg	46,500 kJ/kg
Volumetric energy storage density	250 kJ/liter	34,400 kJ/liter
Power	5 kW	Typically >100 kW

In tabular form, you can easily see the challenge for the all-electric car that uses conventional batteries. Basically, a lead-acid battery is too heavy and has far too little stored energy to compete with gasoline (hence the compromise solution of hybrid gasoline/electric cars).

8.4 THE RAGONE CHART

There is a neat way to compare the energy storage capability of an electrical storage device with its power producing capability known as the **Ragone plot** (Figure 8.6). To show the energy storage and power producing characteristics of electrical storage devices, plot the logarithm[2] of energy storage density (expressed as Wh/kg) against the logarithm of power producing density (expressed as W/kg).

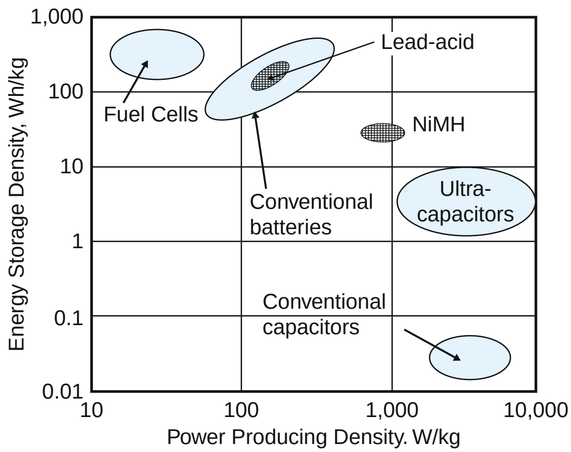


FIGURE 8.6 Ragone Plot Showing Characteristics of Some Batteries and Other Electrochemical Devices (Modified from a graphic of Maxwell Technologies: http://www.maxwell.com)

[2]All logarithms are base 10 in this chapter.

You should want to create an electrochemical device that operates in the upper right-hand corner so you can simultaneously maximize power producing mass density and energy storage mass density and make the smallest possible battery with the least materials. The Ragone plot indicates lead-acid battery is more competitive in its energy storage capacity than it is for its power producing density as compared to the NiMH (nickel metal hydride) battery. There are also fuel cells and ultracapacitors on this plot that we will investigate later, and in the exercises.

8.5 ELECTROCHEMICAL SERIES

To introduce other batteries we will bring in one other useful concept arising from what we have already seen. The **electrochemical series** is defined by the relationship among half cell reactions (such as those developed for the lead-acid battery at each electrode) and put in order of their potentials measured against a standard cell called the hydrogen electrochemical half cell.

Table 8.2 The Electrochemical Series (Note: In principle, these reactions are all reversible)

Half Cell Chemistry	Potential in Volts
$Li^+ + e^- \leftrightarrow Li(s)$	−3.05 V
$Na^+ + e^- \leftrightarrow Na(s)$	−2.71 V
$Mg^{++} + 2e^- \leftrightarrow Mg(s)$	−2.37 V
$Zn^{++} + 2e^- \leftrightarrow Zn(s)$	−0.76 V
$Fe^{++} + 2e^- \leftrightarrow Fe(s)$	−0.44 V
$Ni^{++} + 2e^- \leftrightarrow Ni(s)$	−0.25 V
$2H^+ + 2e^- \leftrightarrow H_2(g)$	0.00 V (Hydrogen ½ cell is defined as zero)
$Cu^{++} + 2e^- \leftrightarrow Cu(s)$	0.34 V
$Cu^+ + e^- \leftrightarrow Cu(s)$	0.52 V
$Ag^+ + e^- \leftrightarrow Ag(s)$	0.80 V
$Pd^{++} + 2e^- \leftrightarrow Pd(s)$	0.95 V

All ions are in aqueous solution; (s) means as a solid and (g) means as a gas.

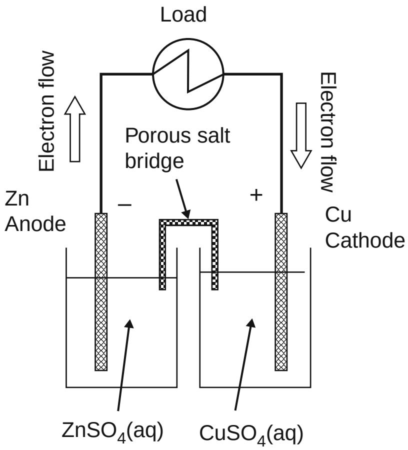
Simple cells operate on principles that use this electrochemical series.

Example 8.3

In a Daniell cell, the electrolytes are $ZnSO_4$(aq) with a Zn anode in its half cell and $CuSO_4$(aq) with a copper cathode in its half cell. The two electrolytes are separated by a porous and inert "salt bridge" that completes the path for the current flow while allowing the separation of ions from each electrolyte. Write down the reactions in each electrolyte and explain what happens in the salt bridge. Finally, what is the voltage produced by the Daniell cell?

Need: Explanation of half cell reactions in Daniell cell and the voltage produced therein.
Know: The electrochemical series in Table 8.2.

How: Draw the cell and use the electrochemical series knowing that the Zn cell will be the anode.



Solve: In the anodic electrolyte, Zn^{++} and $SO_4^{=}$ must be in aqueous solution in balance with each other.

$$ZnSO_4\ (Aq) \leftrightarrow Zn^{++} + SO_4^{=}$$

But the anode is also dissolving and thus yields some locally extra Zn^{++} ions according to:

$$Zn(s) \leftrightarrow Zn^{++} + 2e^- \quad (V = +0.76\ V)$$

Note this is the reverse reaction of that in Table 8.2 so this half cell reaction produces a positive voltage. We will have also produced some additional Zn^{++} ions at a rate of one ion for every two electrons that leave the half cell. A corresponding number of Zn^{++} ions must also move across the salt bridge and be neutralized by a corresponding number of $SO_4^{=}$ ions.

The cathodic electrolyte must be electrically neutral containing equal numbers of Cu^{++} and $SO_4^{=}$ ions; the cathodic reaction must be as written just as in Table 8.2:

$$Cu^{++} + 2e^- \leftrightarrow Cu(s) + 0.34\ V$$

Notice we have removed copper ions from solution; therefore there must be a corresponding reduction in $SO_4^{=}$ ions in the electrolyte. They must move into the salt bridge to exactly counteract the Zn^{++} ions from the anodic side. Thus the cell potential is equal to (+0.76 V) + (+0.34 V) = +**1.10 V**.

There is something else going on that you might have spotted. In the anode, we dissolved one unit of Zn(s) that became Zn^{++} ions and, in the cathode, we precipitated the corresponding number of ions of Cu^{++} ions as Cu(s). This is the type of reaction that is used in the electroplating industry. In that industry, solutions of metal ions are plated out to produce gold, silver, chrome, copper, and nickel-plated objects. In addition, corrosion engineers are interested in the opposite of plating—the dissolution of bulk metals. For example, the rusting of iron occurs substantially because of the half cell reaction $Fe(s) \leftrightarrow Fe^{3+} + 3e^-$.

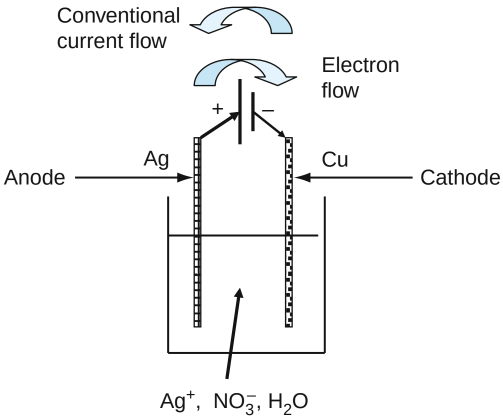
Example 8.4

An electroplater wants to coat a 10.0 cm by 10.0 cm copper plate with 12.5 micrometers of silver. How many electrons must pass in the external circuit? How many coulombs are passed? If the plating takes 1,200. s what's the electrical current in amperes in the external circuit?

Need: Number of electrons and the current flow to deposit 12.5 μm of silver onto 100. cm^2 from a solution containing Ag^+ ions.

Know: Atomic mass of Ag is 108 kg/kmol. Its density is 10,500 kg/m^3. Avogadro's number (N_{Av}) is 6.02×10^{23} atoms/mol or 6.02×10^{26} atoms/kmol. What we call current is nothing but the rate of flow of electrons, so 1.00 A = 1.00 coulomb/s and one electron carries -1.60×10^{-19} C.

How: Sketch a credible electrochemical circuit; use an electrolyte of silver nitrate (which is soluble in water producing equal numbers of Ag^+ and NO_3^- ions). Use a silver anode and the copper plate as a cathode. Finally connect a battery with the polarity chosen as shown.



Solve: The reaction at the anode is $Ag(s) \rightarrow Ag^+ + e^-$ and the reaction at the cathode is $Ag^+ + e^- \rightarrow Ag(s)$; hence one atom of silver dissolves at the anode and one atom of silver is deposited at the cathode. For each atom of silver dissolving at the anode and depositing at the cathode, one electron must circulate in the external circuit.

First we need to know what the mass of silver is in the coating. We will assume the deposited silver atoms fully pack without any internal pores. Then the mass of deposited silver is:

$$\text{Mass} = \frac{12.5 \times 10^{-6} \times 100. \times 10,500}{1.00 \times 10^{4}} [\mu m] [^{m}/_{\mu m}] [cm^2] [kg/m^3]/[^{cm^2}/_{m^2}]$$

$$= 1.31 \times 10^{-3} \text{ kg}$$

$$\text{Next convert to kmols : kmols} = \frac{1.31 \times 10^{-3}}{108} [kg] \left[\frac{kmols}{kg}\right] = 1.22 \times 10^{-5} \text{ kmols}$$

Now convert into the number of silver atoms using Avogadro's number:

$$\text{Number of atoms of Ag(s) deposited} = 1.22 \times 10^{-5} \times 6.02 \times 10^{26} [\text{k mole}] \left[^{atoms}/_{kmole}\right] = 7.32 \times 10^{21} \text{ atoms}$$

This is equal to the number of electrons that have flowed in the external circuit. Hence we have used the services of 7.32×10^{21} electrons.

We know that an ampere is the charge flowing/time; the electrical charge in this case is $7.32 \times 10^{21} \times 1.60 \times 10^{-19}$ [e^-][coulombs/e^-] = 1.17×10^3 coulombs. This charge flows for 1,200. s and hence the **current I** = 1.17×10^3/1,200. [coulombs]/[s] = **0.975 A**.

8.6 ADVANCED BATTERIES

Many batteries such as the common alkaline battery use a solid electrolyte rather than a liquid, but all operate on similar principles. So we must ask, "Are there higher performance batteries available?" And, of course there are. A widely available battery today is the **nickel-metal hydride** (NiMH) battery. They are used both to propel hybrid cars and to start their auxiliary engines. Smaller versions of them are also used for long lasting electronic devices.

In the anode the overall principle reaction that occurs is:[3]

$$Ni(OH)_2 + OH^- \leftrightarrow NiO(OH) + H_2O + e^-$$

and the corresponding overall cathodic reaction is:

$$M + H_2O + e^- = MH + OH^-$$

These reactions are not fundamental, but represent several internal reactions; the M is a metal and MH is a metal hydride. The metal, M, is an intermetallic compound of two metallic elements, one of which is a rare Earth element. A rare Earth element is any element of atomic numbers 58 through 71, characterized by having very similar chemistry. The other metal is one or more of Ni, Co, Mn, or Al. The electrolyte is mostly KOH, a strong alkali.

The basic advantage of this kind of battery is that it has both relatively high energy storage density and power producing density so, when packaged as a common D-size flashlight battery, it can supply more than 10 A for one hour (which is more than 50 kJ). The Ragone plot (Figure 8.6) shows its power performance is considerably better than a lead-acid battery. No wonder these batteries in various sizes are preferred for rechargeable use in higher-end electronic devices. For automotive hybrid use, similar sizes to D cells are packaged in a large array, thus simultaneously providing sufficient voltage and current capacity. But the NiMH battery is being overtaken by the lithium-ion battery whose overall chemistry is a shift of a Li ion, Li^+, from a compound of lithium, cobalt, and oxygen, to lithium carbide. Li-ion batteries have high voltages (approaching four volts compared to most batteries of 1.5–2 volts) and still have a high energy storage density.

Whether NiMH or Li-ion (or the more common NiCd, pronounced Ni-Cad), these batteries are two or three times better than lead-acid batteries in capacity and thus will continue to find use in high-grade electronics and in hybrid cars. None are likely to *completely* replace the gasoline engine because their mass and volumetric energy densities are simply too low.

8.7 FUEL CELLS

If batteries with stationary solid or liquid electrolytes fail to replace gasoline engines completely, then why not use a constant flow of fresh replacement electrochemical fuel and continuously purge the used material? This is what a fuel cell does; it is not a recent invention—a fuel cell based upon phosphoric acid was invented about 175 years ago. Since then there have been many variations. One important class of fuel cells was invented after fundamental research by two GE scientists, Grubb and Niedrach, in the 1950s. Their use in the NASA Moon program in the 1970s led to the near disaster of the Apollo 13 Moon shot when poorly insulated wires electrically shorted, resulting in a hydrogen/oxygen explosion.

The hydrogen fuel cell is the most promoted of all fuel cells for future transportation use since it converts H_2 to H_2O *and nothing else*, and thus would be a welcome relief to the air quality of urban areas currently relying on gasoline vehicles.[4]

[3]http://en.wikipedia.org/wiki/Nickel_metal_hydride

[4]This is a bit misleading since hydrogen must come from somewhere; its manufacture consists of burning hydrocarbons in the presence of air and steam, thus venting significant quantities of CO_2 waste product into the atmosphere at its point of manufacture.

A schematic of the central feature of a fuel cell is shown in Figure 8.7. The reactants, hydrogen and oxygen (actually air), flow to the anode and cathode, respectively, and hot water or steam, the sole product of reaction, is removed.

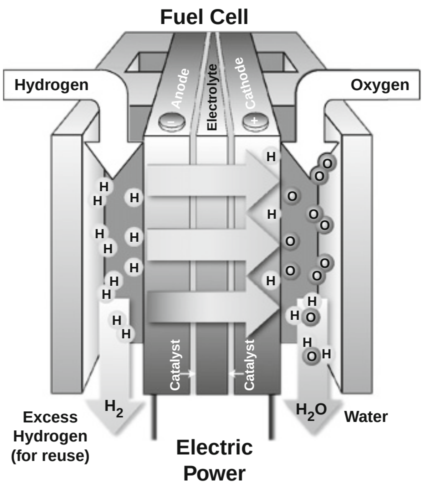


FIGURE 8.7 Principle of a Proton Exchange Membrane

The key step of Grubb and Niedrach was to introduce the Proton Exchange Membrane (PEM) based on a fluorinated polymer membrane. Niedrach showed how to impregnate the membrane with platinum, which is a catalyst for the oxidation of hydrogen by air at ambient temperatures. Hydrogen diffuses into the membrane where it is ionized. The resulting electrons flow from the anode.

$$2H_2 \rightarrow 4H \rightarrow 4H^+ + 4e^-$$

Oxygen gas counter diffuses in the PEM and reacts to form water while absorbing electrons:

$$4H^+ + 4e^- + O_2 \rightarrow 2H^+ + 2OH^- \rightarrow 2H_2O$$

The actual construction of a fuel cell is quite complex (as schematically shown in Figure 8.7) but this is not its principal challenge. The biggest problem for a hydrogen fuel cell is . . . hydrogen.[5]

[5]It's a very flammable and explosive gas. Further as a gas, the amount of hydrogen that can be stored is limited. You can compress hydrogen in pressure vessels, reversibly absorb it in a solid, or liquefy it. At 200 atmospheres pressure (about 3,000 lbf/in^2) its volumetric energy storage density is only 2250 kJ/liter compared to gasoline's 34,000 kJ/liter. The absorption of hydrogen on porous carbon may achieve two or more times the volumetric efficiency compared to compressed gas, but still well short for that of gasoline. The highest *possible* storage energy storage density for hydrogen of 9,700 kJ/liter can be achieved by its liquefaction; unfortunately the liquefaction process is energy intensive (and thus occurs with a CO_2 effluent). Liquid hydrogen boils at −253 °C, which is another engineering obstacle for vehicular use. There is also the practical matter of replacing gas stations with hydrogen stations all across the world. One last problem is that a hydrogen fuel-celled car is not truly "green" because the hydrogen has to be produced somewhere. That process is also net CO_2 generating, although, as previously noted, it will be vented far away from where the hydrogen is used.

8.7.1 Fuel Cells Using Novel Fuels

Although there are lots of electrochemical schemes that function as a fuel cell, one shows some promise because it uses a fuel that can be stored as a liquid—methanol, CH_3OH. The anode feed for such a cell is a mixture of methanol and water vapor onto a PEM membrane. With the appropriate catalyst on the PEM, the methanol is broken down, producing electrons and protons (H^+).

$$CH_3OH + H_2O \rightarrow CO_2 + 6H^+ + 6e^-$$

The electrons are removed via the anode and the protons transport across the PEM as in a hydrogen cell. CO_2 is vented, a serious disadvantage. At the cathode, the reaction is:

$$^3/_2O_2 + 6H^+ + 6e^- \rightarrow 3H_2O$$

Overall the reaction is:

$$CH_3OH + ^3/_2O_2 \rightarrow CO_2 + 2H_2O$$

This cell operates at reasonable temperatures from 60°C to 120°C; its biggest disadvantage is that it produces CO_2; this is offset by the fact that an all-electric drive is roughly twice the efficiency of a gasoline powered car (thus doubling its mpg). As a further use of these fuel cells, Figure 8.8 shows a prototype methanol fuel cell designed to boost the recharging interval of laptop computers to 10 or more hours.

FIGURE 8.8 Fuel Cells for Laptops (http://science.nasa.gov/headlines/y2003/images/fuelcell/notebook_med.jpg)

If we plot the electrical characteristics of fuel cells on a Ragone plot they will lie toward the upper left-hand corner because their energy storage density is high (perhaps 500 Wh/kg) but their power producing density is low (perhaps 20–40 W/kg).

Example 8.5

A PEM cell is fed with 100. standard (meaning measured at atmospheric pressure and 0 °C/273K temperature) ml/min of $H_2(g)$ and its stoichiometric equivalent of air. What's the cell voltage at zero current given that the half cathodic cell produces 1.23 V, and what's the maximum possible electrical output of the cell?

Need: Hydrogen cell voltage and current capability.

Know: Anode half cell: $2H_2 \rightarrow 4H \rightarrow 4H^+ + 4e^-$ for which $V = 0.00$ V (see Table 8.2), and the cathode half cell $4H^+ + 4e^- + O_2 = 2H_2O$ for which $V = +1.23$ V.

$N_{Av} = 6.02 \times 10^{26}$ molecules/kmol and the electronic charge $= 1.60 \times 10^{-19}$C.

How: Voltage from the half cell potentials and current from the assumed 100 percent conversion of the energy in 100. standard ml/min of $H_2(g)$. Convert standard ml to moles using $pV = nR_uT$ where R_u is the **universal** gas constant, 8.31×10^3 J/kmol·K.

Solve: The voltage for the cell is found by algebraically adding the half cell potentials = 0.00 + 1.23 V = **1.23 V**.

So, 100. standard ml/min of hydrogen requires 50.0 standard ml/min of O_2 for the reaction $H_2 + ½O_2 = H_2O$. The amount of air needed is 4.76 × 50.0 standard ml (see Chapter 5) = 238 standard ml/min.

The net reaction uses 100. standard ml/min of H_2. The number of kmols/min of H_2 consumed is: $\dot{N} = p\dot{V}/R_uT = (1.00 \times 10^5) \times (100. \times 10^{-6})/(8.31 \times 10^3 \times 273)$ [N/m²][ml/min][m³/ml]/[J/kmol·K][K] $= 4.41 \times 10^{-6}$ kmols/min.[6]

This contains $4.41 \times 10^{-6} \times 6.02 \times 10^{26} = 2.65 \times 10^{21}$ molecules/min of $H_2(g)$ or 4.42×10^{19} molecules/s flowing in.

The maximum current is: $2.00 \times 4.42 \times 10^{19} \times 1.60 \times 10^{-19}$ [e⁻/molecule] [molecule/s] [C/e⁻] = **14.1 A**. (**Note**: The factor of two appears because there are two electrons needed/molecule of H_2.)

If methanol is an interesting fuel for a fuel cell, can we use a hydrocarbon such as gasoline in a fuel cell? In due course a fuel cell that could use gasoline directly could be advantageous since present infrastructure could be used; an all-electric gasoline-fueled vehicle has the potential to at least double today's gasoline combustion engine mileage. A major scientific and engineering obstacle is to develop a low temperature catalyst that would readily break the sturdy chemical barriers between C–C atoms that form most of the compounds in gasoline. An alternative way is a two-step process that uses a built-in hydrogen generator starting with gasoline. The hydrogen generator is complex with two discrete reaction steps: a *reformer* and a *shift reactor* (Figure 8.9).

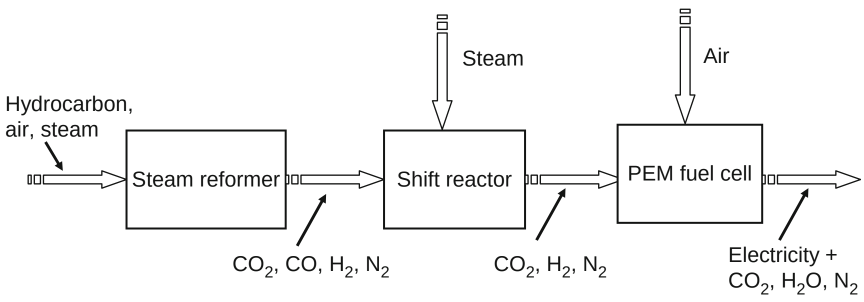


FIGURE 8.9 Reforming Hydrocarbons for a Hydrogen Fuel Cell

[6]The overhead '˙' or over-dot notation as in $\dot{N}$ is used to denote rate.

That this is a complicated affair is shown by a state-of-the-art reformer that produces 250 kW (335 HP) or enough for a large truck (Figure 8.10). The final module is very large compared to a diesel fuel engine and fuel tank for the same purpose.

FIGURE 8.10 A 250 kW Ballard Power Systems Fuel Cell with a Built-In Natural Gas Reformer

The reformer is a high temperature (~1,000°C), high pressure (~ 20–40 bars) reactor containing a nickel metal catalyst on which a portion of a hydrocarbon is burnt. The partially oxidized hydrocarbon is then reacted with hot steam to form a mixture of CO_2, CO, and H_2. If we represent petroleum by the Hill formula, CH_n (with n ~ 2 while noting this is not a compound formula but a molecular ratio of the elements C and H), ignoring the inert nitrogen, the reformer reaction conceptually is:

$$CH_2/O_2/H_2O \rightarrow CO/CO_2/H_2$$

in which the products of the reforming reaction are CO_2, CO, and H_2, the ratios of which are given by the ratios of the feeds: CH_n: O_2: H_2 as set by the reactor's designer. Since we want to minimize the CO_2 as our product stream we need only sufficient oxygen for a substoichiometric reaction.

When the resulting gas mixture of CO_2, CO, and H_2 and N_2 is cooled, it is further reacted by the shift reaction, so called because it shifts CO to H_2:

$$CO + H_2O = CO_2 + H_2$$

The importance of this step is that the precious metal catalyst in the fuel cell is poisoned by CO and it must be removed to about 10 parts per million by volume. The shift reaction is conducted at about 250 °C and in the presence of a large excess of steam (as much as 50 times stoichiometric). The hydrogen thus produced is then fed to a PEM-type fuel cell as previously described.

8.8 ULTRACAPACITORS

The common electrical capacitor uses physical principles (instead of chemical ones) to store electrical energy in the form of electric charges. It works by charging a *dielectric* (which is essentially a nonconductor of electricity such as a polymer or glass) by applying a voltage across it and by sandwiching it between two metal foils or plates. If the dielectric material has the appropriate structure, electrons can be deformed out of their normal orbits to induce net charges at the interface between the metal and the dielectric material. Common electrical capacitors have an energy storage density of less than 0.1 Wh/kg, which is too small for virtually all battery storage operations, but is often useful for timing applications in electric circuits.

Ultracapacitors (sometimes called supercapacitors) are a cross between a common electrical capacitor and a battery. Like common capacitors they store electrical charges on surfaces between charged electrical conductors. Unlike regular capacitors the charges accumulate in porous carbon immersed in oil (see Figure 8.11).

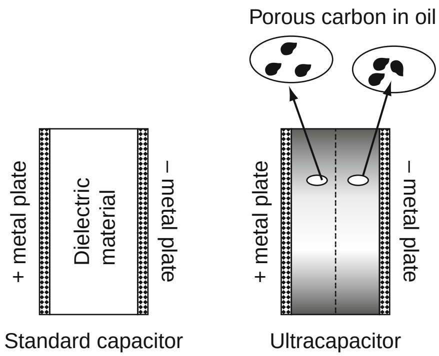


FIGURE 8.11 Comparison of Conventional Capacitor and Ultracapacitors

A detailed description of ultracapacitors reveals that their charges are in double layers at interfaces and thus sometimes they are called electric double-layer capacitors. In ultracapacitors, porous carbon particles are the electrical storage medium because they can have internal surfaces of 100 to 300 m^2/g. If nonporous, carbon's external surfaces would total only a few cm^2/g.

Ultracapacitors are constructed with a separator and two compartments so + and – charges can be separated and collected on their external plates. There is a factor of about 10,000 in total charge if the charges can fill the interstices of the porous carbon instead of just external surface areas. This huge multiplier compared to a standard capacitor means the ultracapacitor has a huge energy storage capacity. Indeed ultracapacitors have energy storage densities of 1 to 10 Wh/kg and their capability to deliver this energy as power can be as high as several thousand W/kg. Unlike batteries, the charge/discharge cycle has no *net* chemical changes and thus ultracapacitors are capable of thousands of charge/discharge cycles.

The Ragone chart in Figure 8.6 shows the areas of applicability of electrochemical batteries, fuel cells, and ultracapacitors. Ultracapacitors can deliver short pulses of high power at reasonable energy storage densities. The region of applicability for ultracapacitors is unique; short bursts of power for high power applications coupled to their energy storage capability is suitable for small power demands such as personal electronics.

SUMMARY

Electrochemical systems are a part of most alternate energy schemes to store electrical energy for load leveling purposes. Batteries work because of the electrical nature of matter. They all have anodes (producing electrons) and cathodes (consuming them). The lead-acid battery is familiar because of its automotive uses but is limited by weight and its energy storage capacity.

Batteries are made up of two half cells, the polarity of which is a result of the position of the half cells in an electrochemical series. Such half cells are also important in electroplating and in metal corrosion.

Modern battery systems, for example such as NiCd, MNiH, and Li-ion, have higher energy storage density than traditional batteries but are inadequate for a purely electrical vehicle. They do find application in electronic equipment that needs longer intervals between recharging. For longer term or continuous use, a fuel cell is required. Practical fuel cells include the hydrogen PEM cell and, to a lesser degree, the PEM methanol fuel cell.

Both ultracapacitors and fuel cells fill power/energy niches not satisfied by standard battery systems. A Ragone plot offers a quick assessment of the region that a given electrochemical engineering technology will best fill.

EXERCISES

1.00 kJ = 0.278 Wh; 1.00 Wh = 3.600 kJ; 1.00 US Gallon = 3.79 liters; 1.000 metric ton (tonne) = 1,000. kg; one electron charge = 1.60×10^{-19} coulombs; Avogadro's number = 6.022×10^{26}/kmol; standard molar volume = 22.4 m^3/kmol; or *equivalently*, R_u is the **universal** (i.e., per kmol) gas constant, 8.31×10^3 J/K·kmol.

1. A vehicle has a 15. US gallon gas tank and can be filled from empty in 60. seconds.
   a. What is the rate that power is transferred to the vehicle?
   b. If the vehicle is converted to an all-battery system using a battery pack to replace the gas tank, what is the rate of power transferred to the batteries if it takes 4.0 hours to charge?

(Assume the density of gasoline is 740 kg/m^3, its heat of combustion is 46,500 kJ/kg, and the battery's energy storage density is 525 kJ/liter). (**A:** (a) **33. MW,** (b) **2.1 kW**)

2. A windmill produces mechanical power according to this formula: $P = \frac{1}{2}\eta\rho V^3 A$ where η is its efficiency (assume η = 60.%), ρ = density of air (1.00 kg/m^3), V = wind speed in m/s (assume 5.0 m/s), and A is the cross-section of the mill that faces the wind (assume it is circular with a radius of 35.0 m).
   a. How many kW does the windmill produce?
   b. To load level, you have a battery storage device that can store 2.00 MWh. How long will you be charging it?
   c. If the energy storage density you can achieve is 125 Wh/kg, how big is this storage battery in tonnes?

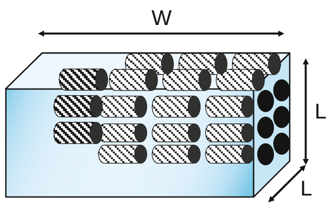
3. You wish to store 2.00 MWh as an emergency power supply for a "big-box" store. If the gross energy storage density of the battery is 425 kJ/liter how big is the storage battery in m^3? If the density of the battery averages 2.5×10^3 kg/m^3 is this reasonable compared to other batteries on the market? (**Hint**: See Figure 8.6.) (**A**: **2.60 meter cube; yes**)

4. In a proposed hybrid car, the battery is reinforced by an ultracapacitor with the following characteristics: 5.0 Wh/kg and 5.0 kW/kg. If the car wants to draw upon its ultracapacitor for 10. kW for 10.0 s, is it power or energy limited?

5. Will a Ragone chart confirm the conclusions of exercise 4? Describe how by referencing Figure 8.6. (**Hint**: How long to discharge the ultracapacitor?)

6. Draw lines of constant discharge time on the Ragone plot. Interpret these lines. (**Hint:** Energy/power = time. On a log-log plot since time is represented only to the first power, its slope is 1, which is represented as one ordinate decade to one abscissa decade.)

The following figure applies to **exercises 7, 8, and 9**.



7. Design a battery pack composed of D-cells (6.35 cm long × 3.18 cm diameter each weighing 0.100 kg). The final package (see diagram) must supply 42.0 V (the coming standard for all cars) at 30.0 A for 2.00 hours. Each cell produces 3.0 V and has a power producing density of 125 W/kg. Estimate *W* and *L* for the package (ignore its energy storage density for now).

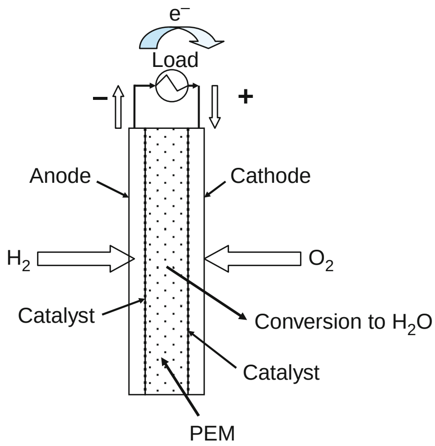
8. Design a battery pack composed of D-cells (6.35 cm long × 3.18 cm diameter each weighing 0.100 kg). The final package (see diagram) must supply 42.0 V (the coming standard for all cars) at 30.0 A for 2.00 hours. Each cell produces 3.0 V and has an energy storage density of 125 Wh/kg. Estimate *W* and *L* for the package (ignore its power producing density for now).

9. Design a battery pack composed of D-cells (6.35 cm long × 3.18 cm diameter each weighing 0.100 kg). The final package (see diagram) must supply 42.0 V (the coming standard for all cars) at 30.0 A for 2.00 hours. Each cell produces 3.0 V and has a power producing density of 125 W/kg *and* an energy storage density of 125 Wh/kg. Estimate *W* and *L* for the package.

10. Compare a battery made from the following pairs of aqueous half cells: $Mg^{++} + 2e^- \leftrightarrow Mg(s)$ and $Cu^{++} + 2\ e^- \leftrightarrow Cu(s)$ with $Fe^{++} + 2\ e^- \leftrightarrow Fe(s)$ and $Ag^+ + e^- \leftrightarrow Ag(s)$. What will be the voltages and can you speculate on the relative weights of the batteries if the densities of Mg(s), Cu(s), Fe(s), and Ag(s) are 1740, 7190, 7780, and 10,500 kg/m^3, respectively? (**Hint:** A significant part of the weight of an electrochemical cell is the weight of its electrodes.)

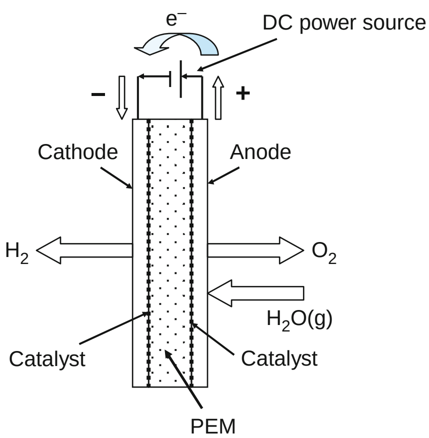
11. Consider a graph of the half cell potentials in Table 8.2 vs. the density of the electrode metal in the half cell. Explain whether or not it is possible to design a cell that is both light and high voltage (say, more than 3 V).

12. An industrial fuel cell has to supply a continuous 20. MW to a perfectly efficient inverter at 123.V. Assuming it uses a hydrogen gas/air/PEM system, what are the needed flows of H_2 and air to the cells in standard m^3/s? See the following simplified schematic. Assume each fuel cell produces 1.23 V.

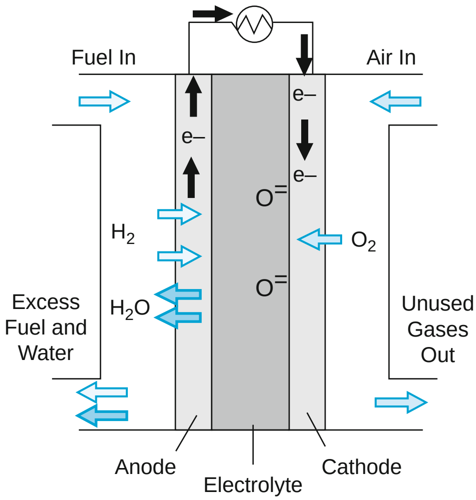


13. In the same fuel cell configuration as in exercise 12, the actual voltage delivered by the cell varies with load; assume the voltage measured at full load is 0.75 V so that the cell electrolytic efficiency is 0.75/1.23 = 61.%. How many cells will we need in series and what is the fate of this inefficiency?

14. If you run a fuel cell in reverse as an electrolysis cell, H_2O is split into its elementary gases, $H_2(g)$ and ½ $O_2(g)$. This way you can make hydrogen and have it stored in case of the loss of outside electric power. The hydrogen can then be fed back to your electrolysis unit acting as a fuel cell. (1) If it takes 1.23 V to electrolyze water, what is the overall efficiency defined as (Theoretical Power Required in MW)/(Actual Power Required in MW) of this scheme? (2) If the electrolysis requires 1.50 V, what then is its overall efficiency?



15. Solid oxide fuel cells use unique electrolytes—a hard ceramic, based on a solid state solution of zirconium oxide stabilized by the element yttrium, and which will conduct oxygen ions when hot enough (which is about 1,000°C). The anode is a Ni/ceramic material and the cathode is an exotic material such as lanthanum strontium manganite. It works similarly to a PEM fuel cell. Air is applied to the cathode and hydrogen to the anode. It is about 60% efficient and, if it has an application, it will be to sophisticated stationary power plant systems. Write down the anode and cathodic reactions; what is the theoretical voltage of such a cell?



16. A methanol fuel cell is used as a battery for a laptop whose average electrical demand is 20. W at 10.0 V. Each cell produces 0.50 V. How much methanol in ml is needed each day? Methanol density is 792 kg/m^3.

17. A steam reformer is fed 1.00 liters/s of methane, 0.952 liters/s of air, and 1.00 liters/s of steam. If it reaches its stoichiometric maximum extent of reaction, what is the composition of its product gases? Assume all quantities given in standard liters/s (1.00 bar, 0°C basis). (**A: CO = 0.600 liters/s, CO_2 = 0.400 liters/s, H_2 = 3.00 liters/s,** and **N_2 = 0.752 liters/s**)

18. The same reformer as in the previous problem has a problem—the 3.00 liters/s of H_2 is contaminated with 0.600 liters/minute of CO. Assume the shift reactor removes this contaminant down to 10.0 ppm of CO. What is the final composition of gas?

19. For the reformer in the two previous questions, what is its maximum electrical power production?

20. You are an engineer on a team developing a radically new battery invented by your immediate supervisor, a successful and famous inventor. Preliminary tests suggest that the new battery, if used in automobiles, will triple range over the best existing battery. Detailed tests have not been made as to the safety or lifetime of the battery. But the inventor urges immediate introduction to the market, using his personal reputation as a guarantee to customers of the value of the product. Only by testing it in actual use, the inventor argues, can any remaining “bugs” be identified and removed. You are concerned that customers will regard the inventor's confident statements as an implied warranty. You urge delay until safety and life tests can be done. The inventor, your immediate supervisor, rejects this approach. What do you do?

Use the Engineering Ethics Matrix.

Comment: This problem, and the next, are based loosely on experiences of Thomas Edison in developing a new battery initially invented for use in automobiles and submarines.
For details of the actual case, see Byron Vanderbilt, *Thomas Edison, Chemist*.

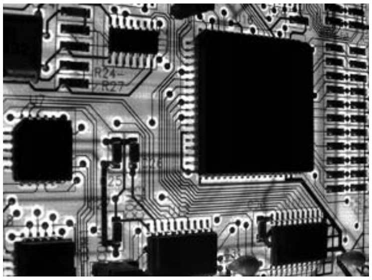
21. You are the CEO of a company that has developed a new fuel cell invented by one of your company's most creative engineers. After thoroughly testing the system for safety and reliability under a wide range of ordinary conditions, you arrange a sale to a customer operating under unusually stressful conditions (undersea exploration to unprecedented depths and pressures) not covered by your tests. As part of the contract, the customer takes full responsibility for any difficulties that arise under these conditions. In use, the fuel cell fails miserably in the application, due to an unanticipated pressure effect. The inventor, embarrassed by the failure, urges that you fully reimburse the customer for the sale price, keep the failure secret, and allow him to correct the problem at company expense. He argues that you are ethically obligated to do this to uphold the company's good name. What is the ethical way to respond to the inventor's request?

Use the Engineering Ethics Matrix.

This page intentionally left blank

CHAPTER

9 Logic and Computers



Source: © iStockphoto.com/Calvinng

Why do cars, trucks, planes, ships, and almost everything else today contain computers? Until the 1960s we got along perfectly well without computers. As late as 1965, the question, “Should a car or truck or plane or ship or anything else contain a computer?” would have seemed ridiculous. In those days, a computer was not only much more expensive than a car, but also bigger.

9.1 MOORE’S LAW

In the late 1960s, the integrated circuit was invented, and computers began to shrink in both size and cost. In the 1990s it also was asserted that “if a Cadillac had shrunk in size and cost as fast as a computer did since 1960, you could now buy one with your lunch money and hold it in the palm of your hand” (although no one has ever explained how you could drive around town in a Cadillac that you could hold in the palm of your hand). The basic message, however, was, and remains, valid. From the 1960s until the present, integrated circuits, the building blocks of computers, have doubled their computing power every year or two. This explosive progress is reflected in a technological trend called *Moore’s Law*, which states that the number of transistors on an integrated circuit “chip” will double every year or two. The name honors electronics pioneer Gordon Moore, who proposed it in 1965. Moore’s Law is not a law of nature, but an empirical rule of thumb, and it has held true for four decades.

Today’s computers are so small that in most applications size is no longer an issue. For example, a computer for controlling the air/fuel mixture in an automobile typically is housed along with its power supply and communication circuitry in an assembly about the size of a small book. As a result, people reliably control not just the hundreds of horsepower of an automobile, but the delivery of energy routinely available to them in modern industrial societies while minimizing pollution emitted in the course of energy conversion.

The subjects of control and binary logic have taken on a life of their own in the form of computers and other information systems. Indeed, such systems have become so pervasive as to lead many people to assert that the industrialized world changed in the late twentieth century from an industrial society to an information society.

The computers that carry out these logical and arithmetical tasks are essentially collections of electrical/electronic circuits. In simpler terms, they are boxes of switches that are turning each other on and off. So the question for this chapter might be rephrased as follows: How can switches turning each other on and off respond to inputs and perform useful calculations? We will give a modest answer to this question by the end of this chapter.

Answering that question introduces the topics of digital logic and computation, which are central parts of computer engineering. The term *digital* can apply to any number system, such as the base ten system used in ordinary arithmetic. However, today's computers are based on a simpler number system: the base two or binary system. Therefore, this chapter focuses on binary logic and computation.

Computation was implemented first by mechanical devices, then by **analog computers**, and today by **digital computers**. A particularly convenient way to use computation to accomplish control is through the use of **binary logic**. A means of summarizing the results of binary logic operations is through **truth tables**. This technique is useful not only in control, but also in other areas of computation, such as **binary arithmetic** and **binary codes**. Binary arithmetic and information are the basis of computer software. Binary logic also enables us to define the engineering variable *information*.

In this chapter you will learn how to (1) carry out **logic operations** on **binary variables**, (2) summarize the results of carrying out logic operations on binary variables in a **truth table**, (3) implement truth tables as **electrical logic circuits**, (4) perform **binary arithmetic**, and (5) communicate the inputs and outputs of binary processes efficiently through **binary codes**. These topics will be illustrated mostly by reference to an automotive application: deciding when a seat belt warning light should be turned on.

As an afterword, the process by which actual computers do these things, the **hardware-software connection**, will be qualitatively summarized.

The topics in this chapter extend far beyond single computers. Millions of *embedded* computers (that is, computers placed within other systems and dedicated to serving those systems) now can be found in hundreds of applications. Electronic and computer engineers have the jobs of reproducing this computer population, which brings forth a new generation every two or three years. The jobs these engineers do range widely. Some develop the electronic circuits that are the basis of electronic computation. Others devise new computer architectures for using those circuits. Yet others develop the "software": the instructions that make computers perform as intended. Many other engineers apply these tools to everything from appliances to space vehicles.

9.2 ANALOG COMPUTERS

For hundreds of years, engineers proved ingenious and resourceful at using mechanical devices to control everything from the rotation of water wheels and windmills to the speed of steam engines and the aiming of guns on battleships. However, such mechanical systems had major disadvantages. They were limited in their speed and responsiveness by the mechanical properties of the components. And they had to be custom designed for every control challenge.

By the twentieth century, this had led engineers to seek more flexible, responsive, and general types of controls. As a first step, engineers put together standardized packages of springs, wheels, gears, and other mechanisms. These systems also proved capable of solving some important classes of mathematical problems defined by differential equations. Because these collections of standard mechanisms solved the problems by creating a mechanical analogy for the equations, they were called analog computers. Bulky and inflexible, they often filled an entire room and were "programmed" by a slow and complex process of manually reconnecting wires and components in order to do a new computation. Despite these drawbacks, in such applications as predicting the tides, determining the performance of electrical transmission lines, calculating the transient behavior of steam-generating nuclear reactors, or designing automobile suspensions, analog computers marked a great advance over previous equation-solving methods.

9.3 FROM ANALOG TO DIGITAL COMPUTING

Meanwhile, a second effort, underway since about 1800, sought to calculate solutions numerically using arithmetic done by people. Until about 1950, the word *computer* referred to a person willing to calculate for wages. Typically these were selected members of the workforce whose economic status forced them to settle for relatively low pay. It was not until about 1900 that these human computers were given access to mechanical calculators and not until the 1920s that practical attempts at fully automated mechanical calculations were begun.

As early as the 1820s, beginning with the ideas of the British scientist Charles Babbage, attempts had been made to do arithmetic accurately using machinery. Because these proposed machines operated on digits as a human would (rather than forming analogies), they were called *digital* computers. Digital techniques were also used to control machinery. For example, the French inventor Joseph Marie Jacquard (1752–1834) used cards with holes punched in them as a digital method to control the intricate manipulations needed to weave large complex silk embroideries.

Humans learn digital computing using their ten fingers.[1] Human arithmetic adapted this ten-digit method into the decimal system. Digital computers can be built on a decimal basis, and some of the pioneers, such as Babbage in the 1840s and the team at the University of Pennsylvania who, in the 1940s built a very early electronic digital computer, the ENIAC,[2] adopted this decimal system.

9.4 BINARY LOGIC

However, computer engineers quickly found that a simpler system less intuitive to humans proved much easier to implement with electronics. This is the binary system, based on only two digits: one and zero. The Jacquard loom was such a binary system, with a hole in a card representing a 1 (one) and the lack of a hole representing a 0 (zero). In other applications, turning on a switch might represent a 1, and turning it off might represent a 0. The logic and mathematics of this system were developed mainly by the British mathematician George Boole (1815–1864), whose contributions were so important that the concept often is referred to as **Boolean algebra**.

Binary logic begins with statements containing variables symbolized by letters such as X or Y. The statement can assert anything whatsoever, whether it is "The switch is open" or "There is intelligent life on a planet circling the star Procyon." The variable representing the statement can be assigned either of two values, 1 or 0, depending on whether the statement is true or false. (We will adopt the convention of using the value 1 for true statements and the value 0 for false ones.)

Binary logic permits three, and only three, operations to be performed:

AND (sometimes called "intersection" and indicated by the symbol $\cdot$ or $*$): Given two statements X and Y, if both are true, then $X \cdot Y = 1$. If either one is false, then $X \cdot Y = 0$. For example, the statement "It is raining (X) and the Sun is out (Y)" is true only if both it is raining *and* the Sun is out.

OR (sometimes called "union" and indicated by the symbol $+$): Given two statements X and Y, if either one or both are true, then $X + Y = 1$; only if *both* are false does $X + Y = 0$. For example, the statement "It is raining (X) *or* the Sun is out (Y)" is true in all cases except when *both* are not true.[3]

[1]The word *digitus* is literally "finger" in Latin, indicating the discrete nature of such counting schemes.

[2]Whether ENIAC was the first electronic computer is open to debate. Tommy Flowers, a British post office engineer, designed and built "Colossus" to break German military codes in WWII; many attribute this to be the first electronic computer.

[3]Technically, this operation is called an *inclusive or* to distinguish it from the operation *exclusive or,* which gives a value 0 when both X and Y are true, as well as when both are false. In our example, this is when it is raining and the Sun is out as well as when it is not raining and the Sun is not out.

NOT (sometimes called "negation" and indicated here by the symbol $'$ as in X$'$, but sometimes indicated in other texts with an overbar as in $\overline{X}$): This operation is performed on a single statement. If X is the variable representing that single statement, then $X' = 0$ if $X = 1$, and $X' = 1$ if $X = 0$. Therefore, if X is the variable representing the statement "It is raining," then X$'$ is the variable representing the statement "It is not raining."

These three operations provide a remarkably compact and powerful tool kit for expressing any logical conditions imaginable. A particularly important type of such a logical condition is an if–then relationship, which tells us that **if** a certain set of statements has some particular set of values, **then** another related statement, often called the **target statement,** has some particular value.

Consider, for example, the statement "If a cold front comes in from the south or the air pressure in the north remains constant, but not if the temperature is above 50°F, then it will rain tomorrow." The target statement and each of the other statements can be represented by a variable. In our example, X can be the target statement "It will rain tomorrow," and the three other statements can be expressed by A = "a cold front comes in from the south," B = "the air pressure in the north is constant," and C = "the temperature is above 50°F." The connecting words can be expressed using their symbols. Thus, this long and complicated sentence can be expressed by the short and simple **assignment** statement:

$$\boldsymbol{X = A + B \cdot C'}$$

This is not a direct *equivalence* in which information flows both ways across the equation. Specifically, in an assignment,[4] the information on the right-hand side is assigned to X but not vice versa. This distinction is required because of the way that computers actually manipulate information.

However, as previously written, the statement still presents a problem. In which order do you evaluate the operations? Does this make a difference? A simple example will show that it *does* make a difference! Consider the case $A = 1$, $B = 0$, $C = 1$. Suppose the symbol $+$ is used first, the symbol $\cdot$ next, and the symbol$'$ last. Then in the preceding statement, $A + B = 1 + 0 = 1$; $(A + B) \cdot C = 1 \cdot C = 1 \cdot 1 = 1$; and $C' = 1' = 0$. Then $X = 0$.

However, if the symbols are applied in the reverse order ($'$ first, $\cdot$ next, and $+$ last), then $C' = 1' = 0$; $B \cdot C' = 0 \cdot 1 = 0$; $A + B \cdot C' = 1 + 0 = 1$, and then $X = 1$.

So to get consistent results when evaluating logic statements, a proper order must be defined. This is similar to standard precedence rules used in arithmetic.

That order is defined as follows:

1. All NOT operators must be evaluated first, then
2. All AND operators (starting from the right if there is more than one), and finally
3. All OR operators (starting from the right if there is more than one) are evaluated.

If a different order of operation is desired, that order must be enforced with parentheses, with the operation within the innermost remaining parenthesis being evaluated first, after which that parenthesis is removed. In our previous example, the proper answer with explicit parentheses would have been written $X = A + (B \cdot C')$ and evaluated as $1 + (0 \cdot 1') = 1$. However, the value of the expression $X = (A + B) \cdot C'$ is $X = (1 + 0) \cdot 1' = 1 \cdot 1' = 1 \cdot 0 = 0$.

[4]In some programming languages such logic statements are written with an *assignment* command, ": =" and not just an "=" command so that our Boolean statement could be written as $X := A + B \cdot C'$ to reinforce the fact that these are not reversible equalities.

Example 9.1

Consider the following statement about a car: “The seat belt warning light is on.” Define the logic variable needed to express that statement in binary logic.

Need: Logic variable (letter) = “. . . .” where the material within the quotes expresses the condition under which the logic variable has the value 1 = true and 0 = false.

Know–How: Choose a letter to go on the left side of an equation. Express the statement in the form it would take if the content it referred to is true. Put the statement on the right side in quotes.

Solve: W = “The seat belt warning light is on.”

This example is trivially simple, but more challenging examples can arise quite naturally. For example, if there are two or more connected constraints on a given action, then the methods of Boolean algebra are surefire ways of fully understanding the system in a compact way.

Example 9.2

Consider the statement involving an automobile cruise control set at a certain speed (called the “set speed”): “Open the throttle if the speed is below the set speed and the set speed is not above the speed limit.” Express this as a logic formula, and evaluate the logic formula to answer the question: *if the speed is below the set speed and the set speed is above the speed limit, then will the throttle be opened*? Answer using these variables: the car’s speed is 50. miles per hour, the set speed is 60. miles per hour, and the speed limit is 45 miles per hour.

Need: A binary logic formula expressing the statement, “Open the throttle if the speed is below the set speed and the set speed is not above the speed limit,” and an evaluation of the formula for the situation when the speed is 50. miles per hour, the set speed is 60. miles per hour, and the speed limit is 45 miles per hour.

Know: Any statement capable of being true or false can be represented by a variable having values 1 = true and 0 = false, and these variables can be connected by AND (·), OR (+), and NOT (′).

How: Define variables corresponding to each of the statements, and use the three connectors to write a logic formula.

Solve: Let X = “throttle is open.”
Let A = “speed is below set speed.”
Let B = “set speed is above speed limit.”

Then the general logic formula expressing the target statement is

$$X = A \cdot B'$$

If the speed is 50. miles per hour, and the set speed is 60. miles per hour, then $A = 1$. If the set speed is 60. miles per hour, and the speed limit is 45 miles per hour, then $B = 1$. Substituting these values, the general logic formula gives

$$X = 1 \cdot 1'$$

Evaluating this in the proper order gives

$$X = 1 \cdot 0 = 0$$

In plain English, if the speed is below the set speed, and the set speed is above the control speed limit, the throttle will not open.

Example 9.3

Suppose we want to find a Boolean expression for the truth of the statement "$W =$ The seat belt warning light should be on in my car," using all of the following Boolean variables:

Case 1:

D is true if the driver seat belt is fastened.
Pb is true if the passenger seat belt is fastened.
Ps is true if there is a passenger in the passenger seat.

Case 2: For actuation of the warning light, include the additional Boolean variable:
M is true if the motor is running.

Need: $W = ?$

Know–How: Put the *W* variable on the left side of an assignment sign $W =$ and then array the variables on the other side of the assignment sign.

Case 1:

a) Put *D*, *Ps*, and *Pb* on the right side of the assignment sign. Thus, the temporary (for now, incorrect) assignment statement is $W = D\ Ps\ Pb$.
b) Connect the variables on the right side with the three logic symbols $\cdot$, $+$, and $'$ so that the relationship among the variables on the right side correctly represents the given statement.

A good way to do this is simply to put the symbols the way they should appear in the if statements. For example, since part of the if statement *Pb* contains the words "the passenger's seat belt is fastened," it should also contain a "not." The corresponding logic variable will appear as Pb'. Also, you only care about the passenger's seat belt if the passenger is sitting in the seat; that is the intersection between these two variables.

Solve: A solution in English is: "If the driver's seat belt is *not* fastened" **or** "if there is a passenger in the passenger seat" **and** "the passenger's seat belt is *not* fastened," **then** "the seat belt warning light should be on in my car," or $\boldsymbol{W = D' + Ps \cdot Pb'}$.

Once written, a logic equation can now be solved for any particular combination of variables. This is done by first plugging in the variable and then carrying out the indicated operation.

Case 2: For activation of the warning light, the light will go on if the motor is running, and if either the driver's seat belt is not fastened and if there is a passenger in the passenger seat and that seatbelt is not fastened. This is written as: $\boldsymbol{W = M \cdot (D' + Ps \cdot Pb')}$

9.5 TRUTH TABLES

It is often convenient to summarize the results of a logic analysis for all possible combinations of the values of the input variables of an "if . . . then" statement. This can be done with a **truth table**. It is simply a table with columns representing variables and rows representing combinations of variable values. The variables for the "if" conditions start from the left, and their rows can be filled in systematically to include *all possible combinations* of inputs. The column at the far right represents "then." Its value can be computed for each possible input combination.

Example 9.4

Consider the condition in Example 9.2: Open the throttle if the speed is below the set speed and the set speed is not above the speed limit. The If conditions are A = "speed is below the speed limit" and B = "set speed is above the speed limit." The Then condition is X = "open the throttle." The truth table is set up as shown here.

A	B	B'	$X = A \cdot B'$

Need: All 16 entries to the truth table.
Know: Negation operator, ′ and the AND operator ·.
How: Fill in all possible binary combinations of statements A and of B.
Solve: One convenient way of making sure you insert all the possible input values is to "count" in binary from all zeroes at the top to all 1's at the bottom. (If you're not already able to count in binary, this is explained on the next page or so.) In this example, the top line on the input side represents the binary number 00 (equal to decimal 0), the second line is 01 (decimal 1), the third is 10 (decimal 2), and the fourth is 11 (decimal 3).[5]

A	B	B'	$X = A \cdot B'$
0	0		
0	1		
1	0		
1	1		

Next the value of the Then ("open the throttle") is computed for each row of inputs (this is done here in two steps, first computing B′, then computing $X = A \cdot B'$).

A	B	B'	$X = A \cdot B'$
0	0	1	0
0	1	0	0
1	0	1	1
1	1	0	0

In English, this truth table is telling us that the only condition under which the control will open the throttle ($X = 1$) is when both the speed is below the set speed and the set speed is below the speed limit. This is the way we *should* want our cruise control to operate!

[5] If the input had three columns, A, B, and C, you would count from 0 to 7 in binary (000 to 111). If four columns, the 16 entries would count from 0 to 15 in binary (000 to 1111), and so on.

Truth tables can also be conveniently expressed as electric circuits. Indeed, this capability is the essence of computing. This capability is further explored in the exercises.

9.6 DECIMAL AND BINARY NUMBERS

In the decimal[6] or base 10 number system, digits are written to the left or right of a dot called the **decimal point** to indicate values greater than one or less than one. Each digit is a **placeholder** for the next power of 10. The digits to the left of the decimal point are whole numbers, and as you move to the left every number placeholder increases by a factor of **10**. On the right of the decimal point the first digit is **tenths** (1/10), and as you move further right every number placeholder is **10 times smaller** (see Figure 9.1).

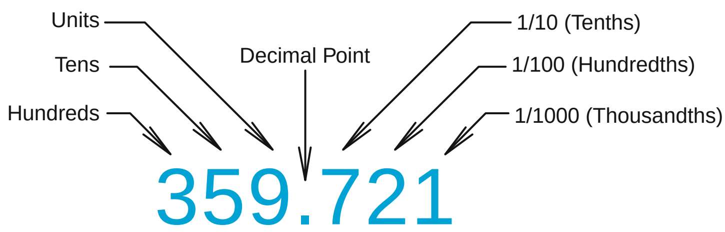


FIGURE 9.1 The Structure of a Decimal Number

The number 6357. has four digits to the left of the decimal point, with 7 filling the units place, 5 filling the tens place, 3 filling the hundreds place, and 6 filling the thousands place. You can also express a decimal number as a whole number plus tenths, hundredths, thousandths, and so forth. For example, in the number 3.76, the 3 to the left of the decimal point is the whole number; the 7 on the right side of the decimal point is in the tenths position, meaning 7 tenths, or 7/10; and the 6 is in the hundedths position. So, 3.76 can be read as "3 and 7 tenths and 6 hundredths."

There is nothing that requires us to have 10 different digits in a number system. The **base-10** number system probably developed because we have 10 fingers, but if we happened to have eight fingers instead, we would probably have a base-8 number system. In fact, you can have a **base-anything** number system. There are often reasons to use different number bases in different situations.

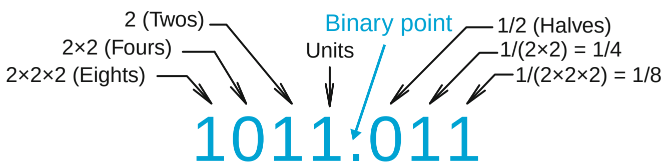


FIGURE 9.2 The Structure of a Binary Number

The binary[7] (base 2) number system is similar to the decimal system in that digits are placed to the left or right of a "point" to indicate values greater than one or less than one (see Figure 9.2). For binary numbers, the first digit to the left of the binary point is called the **units**. As you move further to the left of the binary point,

[6]The word **decimal** means "based on 10" (from the Latin *decima* meaning a tenth part).

[7]The word **binary** comes from "Bi-" meaning two. We use it in words such as "bicycle" (two wheels) and "binocular" (two eyes).

every placeholder increases by a factor of **2**. To the right of the binary point the first digit is **half** (1/2), and as you move further to the right every placeholder number becomes **half again smaller**. Table 9.1 shows a few equivalent decimal and binary whole numbers.

Table 9.1 Decimal and Binary Numbers

Decimal:	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Binary:	0	1	10	11	100	101	110	111	1000	1001	1010	1011	1100	1101	1110	1111

So, how do you convert binary numbers to decimal numbers? Several examples are given next, to show the steps in converting several binary numbers to decimal numbers.

1. What is 1111 in decimal?
- The 1 in the leftmost position is in the 2×2×2 (or 2^3) position, so that means 1×2×2×2 = 8
- The next 1 is in the 2×2 (or 2^2) position, so that means 1×2×2= 4
- The next 1 is in the 2 (or 2^1) position, so that means 1×2 = 2
- The last 1 is in the units (2^0 = 1) position, so that means 1×1 = 1
- Answer: 1111 = 8 + 4 + 2 + 1 = 15 in decimal.

2. What is 1001 in decimal?
- The 1 in the leftmost position is in the 2×2×2 (or 2^3) position, so that means 1×2×2×2 = 8
- The next 0 is in the 2×2 (or 2^2) position, so that means 0×2×2 = 0
- The next 0 is in the 2 (or 2^1) position, so that means 0×2 = 0
- The last 1 is in the units (2^0) position, so that means 1×1 = 1
- Answer: 1001 = 8 + 0 + 0 + 1 = 9 in decimal

3. What is 1.1 in decimal?
- The 1 on the left side of the binary point is in the units (2^0) position, so that means 1×1 = 1
- The 1 on the right side is in the halves (2^{-1}) position, so that means 1 × (1/2) = 0.5
- So, 1.1 is 1-and-1-half = 1.5 in decimal

4. What is 10.11 in decimal?
- The 1 in the leftmost position is in the 2 (or 2^1) position, so that means 1 × 2 = 2
- The 0 is in the units (2^0) position, so that means 0 × 1 = 0
- The first 1 on the right of the point is in the halves (2^{-1}) position, so that means 1 × (1/2) = 0.50
- The last 1 on the right side is in the quarters (2^{-2}) position, so that means 1 × (1/4) = 0.25
- So, 10.11 is 2 + 0 + ½ + ¼ = 2.75 in decimal

A single binary digit (0 or 1) is called a **bit**. For example, the binary number **11010** has five bits. The word ‘bit’ is made from the words **bi**nary dig**it**. Bits are usually combined into 8-bit collections called **bytes**. With an 8-bit byte, you can represent 256 values ranging from 0 to 255, as shown:

0 = 00000000
1 = 00000001
2 = 00000010
....
254 = 11111110
255 = 11111111

Bytes often come with **prefixes** like kilo, mega, and giga, as in kilobyte, megabyte, and gigabyte. Table 9.2 lists the decimal values of these **binary** numbers.

Table 9.2 Binary Number Byte Prefixes

Name	Size
Kilo (K)	$2^{10} = 1{,}024$
Mega (M)	$2^{20} = 1{,}048{,}576$
Giga (G)	$2^{30} = 1{,}073{,}741{,}824$
Tera (T)	$2^{40} = 1{,}099{,}511{,}627{,}776$
Peta (P)	$2^{50} = 1{,}125{,}899{,}906{,}842{,}624$
Exa (E)	$2^{60} = 1{,}152{,}921{,}504{,}606{,}846{,}976$
Zetta (Z)	$2^{70} = 1{,}180{,}591{,}620{,}717{,}411{,}303{,}424$
Yotta (Y)	$2^{80} = 1{,}208{,}925{,}819{,}614{,}629{,}174{,}706{,}176$

A terabyte hard drive actually stores 10^{12} bytes.[8] How could you possibly need a terabyte of disk space? When you consider all the digital media[9] available today (music, games, and video), it is not difficult to fill a terabyte of storage space. Terabyte storage devices are fairly common, and there are probably a few petabyte storage devices floating around somewhere.

9.7 BINARY ARITHMETIC

The value of a bit depends on its position relative to the binary point. For example, the binary number 11010.101 has a decimal value computed from the second and third rows of Table 9.3 as $16 \times 1 + 8 \times 1 + 4 \times 0 + 2 \times 1 + 1 \times 0 + 0.500 \times 1 + 0.250 \times 0 + 0.125 \times 1 = 26.625$.

Table 9.3 Converting the Binary Number 11010.101 to a Decimal Number

Placeholder	2^4	2^3	2^2	2^1	2^0	.	2^{-1}	2^{-2}	2^{-3}
Bit	*1*	*1*	*0*	*1*	*0*	.	*1*	*0*	*1*
Decimal value	16	8	4	2	1		1/2 = 0.500	1/4 = 0.250	1/8 = 0.125

Rules of Binary Addition

$$0 + 0 = 0$$
$$0 + 1 = 1$$
$$1 + 0 = 1$$
$$1 + 1 = 10, \text{ so carry the 1 to the next bit and save the 0}$$

[8]The capacities of computer storage devices typically are advertised using their SI standard values, but the capacities reported by software operating systems use the binary values. The standard SI terabyte (TB) contains 1,000,000,000,000 bytes = 1000^4 or 10^{12} bytes. However, in binary arithmetic, a terabyte contains 1,099,511,627,776 bytes = 1024^4 or 2^{40} bytes.
[9]Just five minutes of digital video requires about 1 gigabyte of storage.

For example adding 010 (digital 2) to 111 (digital 7) gives:

$$\begin{array}{r} 010 \\ +\ 111 \\ \hline 1001 \end{array} \text{ (digital 9)}$$

Binary addition is conceptually identical to decimal subtraction. However, instead of carrying powers of ten, we carry powers of two. Here are the steps to follow:

1. Starting at the right, $0 + 1 = 1$ for the first digit (no carry needed).
2. The second digit is $1 + 1 = 10$ for the second digit, so save the 0 and carry the 1 to the next column.
3. For the third digit, $0 + 1 + 1 = 10$, so save the zero and carry the 1.
4. The last digit is $0 + 0 + 1 = 1$.
5. So the answer is 1001 (digital 9—you can see it is correct since $2 + 7 = 9$).

Binary subtraction is conceptually identical to decimal subtraction. However, instead of borrowing powers of ten, we borrow powers of two.

Rules of Binary Subtraction

$0 - 0 = 0$

$0 - 1 = 1$, and borrow 1 from the next more significant bit

$1 - 0 = 1$

$1 - 1 = 0$

The following examples illustrate "borrowing" in binary subtraction.

$$\begin{array}{r} 10 \\ -1 \\ \hline 1 \end{array} \qquad \begin{array}{r} 100 \\ -10 \\ \hline 10 \end{array} \qquad \begin{array}{r} 1010 \\ -110 \\ \hline 100 \end{array}$$

Can you complete Table 9.4? You can self-check against their decimal equivalents.

Table 9.4 Simple Binary Arithmetic Examples

Example 1		Example 2		Practice 1		Practice 2	
Binary	Decimal	Binary	Decimal	Binary	Decimal	Binary	Decimal
1001	9	1001	9	1011		1011	
+101	+5	−101	−5	+110		−110	
1110	14	100	4				

The process of binary subtraction may be viewed as the addition of a negative number. For example, $3 - 2$ may be viewed as $3 + (-2)$. To do this you must determine the negative representation of a binary number. One way of doing this is with the **one's complement**.

The one's complement of binary number is found by changing all the ones to zeroes and all the zeroes to ones as shown:

Number	One's Complement
10011	01100
101010	010101

To subtract a smaller number from a larger number using the one's complement method you:

1. Determine the one's complement of the smaller number.
2. Add the one's complement to the larger number.
3. Remove the final carry and add it to the result (this step is called the end-around carry).

Example 9.5

Do the following subtraction: 11001 (decimal 25) – 10011 (decimal 19)

Need: 11001 – 10011 = ______? (a binary number)

Know–How:

Step 1: the one's complement of 10011 is 01100

Step 2: Adding the one's complement to the larger number gives 01100 + 11001 = 100101

Step 3: Removing the final carry and adding it to the result gives 00101 + 1 = 00110

Solve: **11001** – **10011** = **110**. To verify that this is correct, convert each base 2 number to decimal and repeat the subtraction, or **25** – **19** = **6.**

To subtract a larger number from a smaller number, the one's complement method is as follows:

1. Determine the one's complement of the larger number.
2. Add the one's complement to the smaller number (the result is the one's complement of the answer).
3. Take the one's complement of the result to get the final answer. Don't forget to add the minus sign since the result is negative.

Example 9.6

Do the following subtraction: 1001 (decimal 9) – 1101 (decimal 13).

Need: 1001 – 1101 = ? (a binary number)

Know–How:

Step 1: The one's complement of the larger number 1101 is **0010**

Step 2: Adding the one's complement to 1001 gives 0010 + 1001 = **1011**

Step 3: Add a minus sign to the one's complement of 1011 to get 1001 – 1101 = –**0100**

Solve: **1001** – **11001** = –**100**. To verify that this is correct, convert each base 2 number to decimal and repeat the subtraction, or **9 – 13** = –**4.**

The rest of the familiar arithmetic functions can also be carried out in binary. **Fractions** can be expressed in binary by means of digits to the right of a binary point. Once again, powers of 2 take the role that powers of 10 play in digital arithmetic. Thus, the decimal fraction 0.5 (i.e., ½) is the binary fraction 0.1 and the decimal fraction 0.25 is the binary fraction 0.01, and so on. A fraction that is not an even power of ½ can be expressed as a sum of binary numbers. Thus, the decimal 0.375 = 0.0011 in binary, which is the sum of decimal 0.25 + 0.125.

Multiplication and **division** can be carried out using the same procedures as in decimal multiplication and long division. The only complication is, again, systematically carrying and borrowing in powers of two, rather than powers of ten.

Rules of Binary Multiplication

$$0 \times 0 = 0$$

$$0 \times 1 = 0$$

$$1 \times 0 = 0$$

$$1 \times 1 = 1, \text{ and no carry or borrow bits}$$

Example 9.7

If a powerful race car has an $(A/F)_{mass}$ of 15 [kg air]/[kg fuel] and the air intake draws in 1.5 kg of air per second, how much fuel must be injected every second? Solve in binary to five significant binary digits.

Need: Fuel rate = ______? kg fuel/s.
Know–How: Fuel flow rate = $(F/A) \times$ 1.5 kg air/s = (1/15) [kg fuel]/[kg air] $\times$ 1.5 [kg air/s] = 0.10 [kg fuel/s].

To illustrate binary arithmetic, let's break down this problem into two separate ones. Although this is unnecessary, it will serve to illustrate binary division and binary multiplication.

The division problem will be 1/15 (decimal) = 1/1111 (binary), and the multiplication problem will take the solution of that problem and then multiply it by 1.5 (decimal) = 1.1 (binary).

Solve: Start with 1/1111, then binary multiply that answer by 1.1.

```
          .00010001        0.00010001
      ___________               ×1.1
1111 | 1.00000000         ----------
        1111               0.00010001
        ----               0.00001000
           10000           ----------
            1111           0.00011001
            ----
               1
```

Therefore, the fuel flow rate = 0.00011001 kg/s (in binary).

Checking in decimal: (1/15) $\times$ 1.5 = **0.10** and 0.00011001 = 0/2 + 0/4 + 0/8 + 1/16 + 1/32 + 0/64 + 0/128 + 1/256 = **0.098** (to get the exact answer of 0.10 we would need to use more significant (binary) figures.

Many times each second, a computer under the hood of an automobile receives a signal from an air flow sensor, carries out a binary computation such as the one shown in this example, and sends a signal to an actuator that causes the right amount of fuel to be injected into the air stream in order to maintain the desired air–fuel ratio. The result is much more precise and reliable control of fuel injection than was possible before computers were applied to automobiles.

9.8 BINARY CODES

We can now see how 0 and 1 can be used to represent false and true as logic values, and of course, 0 and 1 are numeric values. It is also possible to use *groups* of 0s and 1s as codes. The now outdated Morse code is an example of a binary code, whereas the genetic code is based on just the pairings of four, rather than two, chemical entities known as bases.

Suppose we want to develop unique codes for the following nine basic colors: red, blue, yellow, green, black, brown, white, orange, and purple (Table 9.5). Can we do this with a three-bit code (three 0s or 1s [bits])? No, since there are only eight combinations of three bits (note: $2^3 = 8$): namely 000, 001, 010, 011, 100, 101, 110, 111.

Here is a four-bit code that would work, but of course it is only one of several, since there are $2^4 = 16$ possible combinations of a four-bit code. If we have N bits, we can code 2^N different things.

Table 9.5 Assignment of Binary Numbers to Colors

Color	Binary Equivalent	Color	Binary Equivalent
red	0000	brown	0101
blue	0001	white	0110
yellow	0010	orange	0111
green	0011	purple	1000
black	0100		

9.9 HOW DOES A COMPUTER WORK?

How can these abstract ideas of Boolean algebra, binary logic, and binary numbers be used to perform computations using electrical circuits, particularly switches (which we studied in the previous chapter)? The modern computer is a complex device, and any answer we give you here is necessarily oversimplified. But the principles are sufficient to give you some insight. For this discussion we will need to know what a **central processing unit**, or CPU, does and what a computer **memory** is (which can take such forms as read-only memory, or ROM, and random access memory, or RAM).

The "smart" part of the computer is the CPU. It is just a series of **registers**, which are nothing but a string of switches. These switches can change their voltage states from off (nominally no voltage above ground) to on (+5 volts above ground). The early personal computers (or PCs) used only 8-bit registers; modern ones use 64 or 128, but we can think in terms of the 8-bit registers. (The notation x, y in Table 9.6 means either state x or state y.)

This register can contain 2^8, or 256, discrete numbers or addresses. What the CPU addresses is the memory in the computer. You can think of memory as a pigeonhole bookcase with the addresses of each pigeonhole preassigned.[10]

[10]M. Sargent III and R. L. Shoemaker, *The IBM PC from the Inside Out*. Reading, MA: Addison-Wesley Publishing Co. Inc., revised edition, 1986, p. 21.

Table 9.6 Eight-Bit Register

Bit #	**7**	**6**	**5**	**4**	**3**	**2**	**1**	**0**
Voltage	0 or 5 (0,5)	0 or 5 (0,5)	0 or 5 (0,5)	0 or 5 (0,5)	0 or 5 (0,5)	0 or 5 (0,5)	0 or 5 (0,5)	0 or 5 (0,5)
Bits	0 or 1 (0,1)	0 or 1 (0,1)	0 or 1 (0,1)	0 or 1 (0,1)	0 or 1 (0,1)	0 or 1 (0,1)	0 or 1 (0,1)	0 or 1 (0,1)

If we have just 256 of these pigeonholes in our memory, our CPU can address each of them. If we want to calculate something, we write a computer code in a suitable language that basically says something like: Add the number in pigeonhole 37 to that in pigeonhole 64, and then put the contents in pigeonhole 134. The binary bits can represent numbers, logic statements, and so on. All we then need to do is to be able to send out the results of our binary calculations in a form we can read. For example, we can decide that the contents of pigeonhole 134 is an equivalent decimal number.

If, in the previous example, we identify the program input as what is in pigeonholes 37 and 64, and the program output is pigeonhole 134, the computer is constructed essentially as in Figure 9.3 to accomplish its mission.

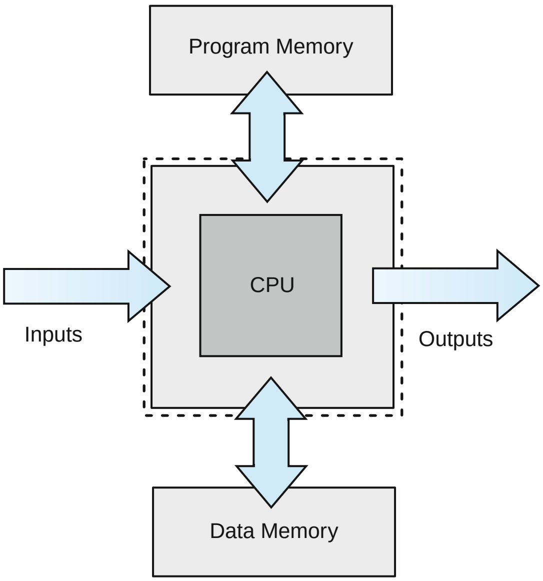


FIGURE 9.3 The CPU

The previous chapter introduced the concept that a computer is made up of **hardware** (electrical circuits containing such components as transistors). How the computer is instructed to execute its functions is managed by **software** (the instructions that tell the components what to do).

Part of the hardware of the computer that we have so far ignored is an internal clock. We might view the software as a list of instructions telling the computer what to do every time the clock ticks. This list of instructions includes the normal housekeeping functions that the computer carries out regularly (such as checking whether a user has entered in any keystrokes on the keyboard in the very short time interval since the last time this was checked). But it can also include downloading into the portion of the computer's memory called **program memory**, a special list of instructions called a **stored program**, and then executing that stored program. Examples of stored programs are a word processor and a spreadsheet program.

Once this stored program is downloaded into the memory, the CPU then carries out this stored program step by step. Some of those steps involve carrying out computations, which are executed in binary arithmetic by the CPU using the methods described in this chapter. In some cases, the result of the computation determines which of the stored program's instructions is the next one that should be carried out. This flexibility regarding the order in which instructions are carried out is the basis of the computer's versatility as an information processing system.

Software in this binary form is called **machine language**. It is the only language that the computer understands. It is, however, a difficult language for a human to write or read. So computer engineers have developed programs that can be stored in the computer that translate software from a language that humans understand into machine language.

The type of language that humans understand is called a **higher-level language**. This language consists of a list of statements somewhat resembling ordinary English. For example, a higher-level language might contain a statement such as "if $x < 0$, then $y = 36$." Examples of higher-level languages are C++, BASIC, and Java. Most computer programs are written initially in a higher-level language.

The computer then translates this higher-level language into machine language. This translation typically is carried out in two steps. First, a computer program called a **compiler** translates the statements of the higher-level language into statements in a language called **assembly language**, which is closer to the language that the computer understands. Then another computer program called an **assembler** translates the assembly language program into a **machine language program**. It is the machine language program that actually is executed by the computer.

The personal computers that we all use, typically by entering information through such devices as a keyboard or mouse, and producing outputs on a screen or via a printer, are called **general-purpose computers**. As the name suggests, they can be used for a wide range of purposes, from game playing to accounting to word processing to monitoring scientific apparatus.

Not all computers need this wide range of versatility. So there is another important class of computers directed at narrower ranges of tasks. These computers are called **embedded computers** (Figure 9.4).

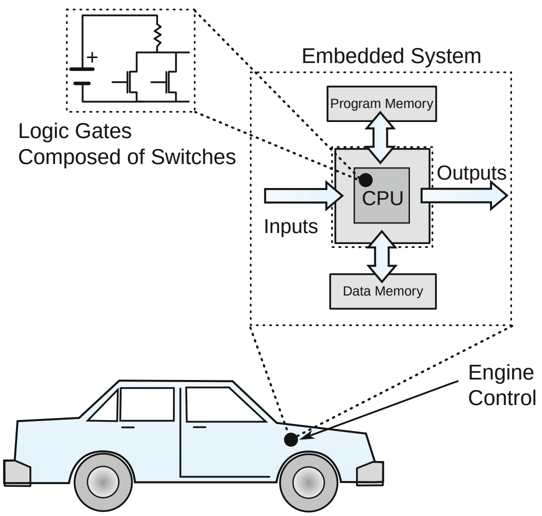


FIGURE 9.4 Schematic for an Embedded Automotive Computer

As their name suggests, these computers are embedded within a larger system. They are not accessible by keyboard or mouse, but rather receive their inputs from sensors within that larger system. In an automobile, there are about a dozen or more embedded computers. There might be, for example, an embedded computer for controlling a car's stereo system, another recording data for automated service diagnostics, and yet another for fuel control. It is such a system for automobile fuel control based on an embedded computer that will be a subject in our next chapter.

SUMMARY

Although analog computers have been of some historical importance, digital computers do almost all important control jobs in today's advanced technologies. So an engineer must understand the principles of digital computation that rest on the immensely powerful concepts of **Boolean algebra, binary logic, truth tables, binary arithmetic**, and **binary codes**. These concepts enable an engineer to make a first effort at defining the concept of information. Binary arithmetic and information are the basis of computer software. These concepts make it possible to move on to the challenge of implementing digital controls using computation.

EXERCISES

A number of these exercises use electrical circuits to effect logical statements. If you are uncertain about electrical circuits, you should review Chapter 7, *Electrical Engineering*.

1. A popular ditty of the late nineteenth- and early twentieth-century railroad era was the following (sung to the tune of *Humoresque* by the nineteenth-century Czech composer Antonin Dvorak[11]):

Passengers will please refrain – From flushing toilets – While the train – Is standing in the station – I love you.

For the preceding ditty, define a variable S expressing whether or not the train is in the station, a variable M expressing whether or not the train is moving ("standing" means "not moving"), and a variable F expressing the fact that the toilet may be flushed. (**A: S = "the train is in the station," M = "the train is moving" (or you could use its negation, M' meaning the train is stationary), and F = "the toilet may be flushed"**)

2. For the ditty in Exercise 1, (a) express as a logic formula the conditions under which one may flush the toilet, (b) evaluate the formula you wrote in (a) for $M = 1$ and $S = 1$, and (c) express in words the meaning of your answer to (b). (Assume that any behavior not explicitly forbidden is allowed.)

3. Rework Example 9.2 to express and evaluate the logic formula to answer the question, "If the speed is below the set speed and the set speed is above the speed limit, then will the throttle be opened?"

4. In Example 9.3, include the additional Boolean variable that for actuation of the warning light, "the driver's door must be closed" (D_{door} is true if the driver's door is closed).

5. Consider the following logic variables for a car:
Db = "the driver's seat belt is fastened"
Pb = "the passenger seat belt is fastened"
W = "the seatbelt warning should be on in my car"

[11] You can find the music at http://www.youtube.com/watch?v=ScSCILXXLnM

Write a sentence in English that expresses the logic equation $W = Db' + Pb'$.
(**A:** W = **"If either the driver's seat belt is not fastened or the passenger's seat belt is not fastened, the seatbelt warning light should be on."**)

6. Consider the following logic variables for a car:
W = "the seatbelt warning light is on"
D = "a door of the car is open"
Ps = "there is a passenger in the passenger seat"
K = "the key is in the ignition"
M = "the motor is running"
Db = "the driver's seat belt is fastened"
Pb = "the passenger seat belt is fastened"

Write a logic equation for W that expresses the following sentence: "If all the doors of the car are closed, and the key is in the ignition, and either the driver's seat belt is not fastened or there is a passenger in the passenger seat and the passenger's seat belt is not fastened, then the seatbelt warning light should be on."

7. Consider the following logic variables for a car:
M = "the motor is running"
Db = "the driver's seat belt is fastened"

In the early 1970s the government ordered all seat belt warnings to be tied to the motor in a manner expressed by the following sentence: "If the driver's seat belt is not fastened, then the motor cannot be running."

a. Write a logic equation for M in terms of Db that expresses this sentence.
b. Write a truth table for that logic equation.

(In practice, this seat belt light logic caused problems. Think of trying to open a manual garage door or pick up the mail from a driveway mailbox. The government soon retreated from an aroused public.)
(**A: (a)** $M = Db$); **(b)**

Db	M
0	0
1	1

Consider the following logic variables:
A = "a customer at a restaurant orders an alcoholic beverage"
I = "the customer shows proper identification"
M = "a customer at a restaurant is over 21"

8. Consider the sentence, "If a customer at a restaurant is over 21 and shows proper identification, then she can order an alcoholic beverage." (a) Express this sentence as a logic equation, and (b) write a truth table for the logic equation.

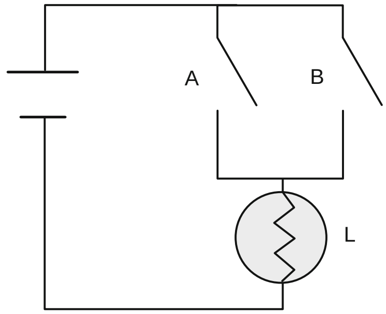
9. Consider the following variables expressing a football team's strategy.
T = "It is third down"
L = "We must gain more than 8 yards to get a first down"
P = "We will throw a pass"

The team's strategy is expressed by this truth table.

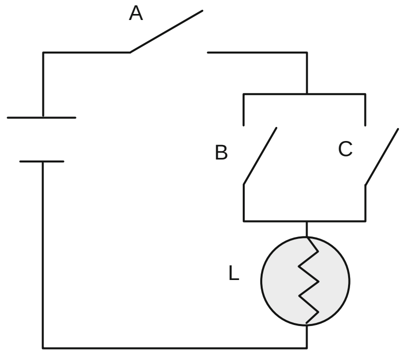
T	*L*	*P*
0	0	0
0	1	1
1	0	1
1	1	1

Write a logic equation for P in terms of T and L. (**A:** $\boldsymbol{P = T + L}$)

10. Consider the following electric circuit: Consider the variables L = "the light is on," A = "switch A is closed," and B = "switch B is closed." Express the relationship depicted by the electric circuit as a logic equation for L in terms of A and B. (**A:** $\boldsymbol{L = A + B}$)



11. Consider the following circuit diagram and the variables L = "the light is on," A = "switch A is closed," B = "switch B is closed," and C = "switch C is closed." Write a logic equation for L in terms of A, B, and C.



12. Consider the following logic variables for a car:
W = "the seatbelt warning light is on"
Ps = "there is a passenger in the passenger seat"
Db = "the driver's seat belt is fastened"
Pb = "the passenger seat belt is fastened"

Draw a circuit diagram for the logic equation $W = Db' + (Ps \cdot Pb')$.

13. Explain the following sentences.

a. There are 10 kinds of people in the world, those who understand binary numbers, and those who don't.

b. Binary is as easy as 1, 10, 11.

14. Convert the following numbers from binary to decimal: (a) 110, (b) 1110, and (c) 101011. **Partial A:**

—	2^5	2^4	2^3	2^2	2^1	2^0	Decimal Equivalent
—	32	16	8	4	2	1	—
a) 110	**0 × 32**	**0 × 16**	**0 × 8**	**1 × 4**	**1 × 2**	**0 ×1**	**6**
b) 1110							
c) 101011							

15. Convert the following numbers from decimal to binary: (a) 53, (b) 446, and (c) 1492. **Partial A:**

Decimal	1024	512	256	128	64	32	16	8	4	2	1
Binary place	2^{10}	2^9	2^8	2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
a) 53	0	0	0	0	0	**1**	**1**	**0**	**1**	**0**	**1**
b) 446											
c) 1492											

16. Do the following binary additions. Check your answer by converting each binary number into decimal. **Partial A:**

Binary	Decimal	Binary	Decimal	Binary	Decimal
1010	**10**	11101		10111	
+110	**6**	+10011		+10	
10000	**16**				

17. Do the following binary subtractions. Check your answer by converting each binary number into decimal. **Partial A:**

Binary	Decimal	Binary	Decimal	Binary	Decimal
1010	**10**	11101		10000	
−110	**−6**	−10011		−1	
0100	**4**				

18. If a powerful race car has an $(A/F)_{Mass}$ of 12.0 (kg air)/(kg fuel), and the air intake draws in 1.000 kg of air per second, how much fuel must be injected every second? Solve in binary to three significant binary digits. (**A: 0.000101 kg**)

19. Suppose we want to devise a binary code to represent the fuel levels in a car:

a. If we need to describe only the possible levels (empty, 1/4 full, 1/2 full, 3/4 full, and full), how many bits are needed?

b. Give one possible binary code that describes the levels in (a).

c. If we need to describe the levels (empty, 1/8 full, 1/4 full, 3/8 full, 1/2 full, 5/8 full, 3/4 full, 7/8 full, and full), how many bits would be needed?

d. If we used an 8-bit code, how many levels could we represent?

20. Construct a spreadsheet[12] that converts binary numbers from 0 to 111 to decimal numbers, print as formulae using the Control-tilde command. Check your spreadsheet against exercise 14. (**A:** e.g., **binary 110** $\equiv \mathbf{1} \times \mathbf{2^2} + \mathbf{1} \times \mathbf{2^1} + \mathbf{0} \times \mathbf{2^0} =$ **decimal 6**).

21. Construct a spreadsheet that converts decimal number 53 to binary. Print as formulae using the Control-tilde command. Check your spreadsheet against exercise 15. (**Hint**: 5 (decimal) can be divided by 2^2 to yield an integer 1 and remainder 1; 1 can't be divided by 2^1 (therefore, integer 0) and 1 can be divided by 2^0 for last integer 1. Check for a spreadsheet function that will divide two numbers and display their result with no remainder.)

22. In the game "Rock, Scissors, Paper" we have the following rules:

Rock breaks Scissors Scissors cut Paper Paper covers Rock

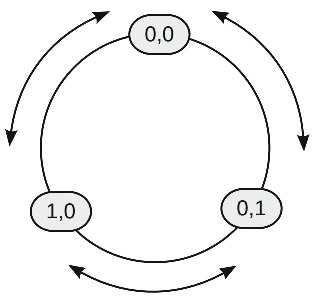
If we were to create a code to represent the three entities, Rock, Scissors, and Paper, we would need two bits. Suppose we have the following code, where we call the first bit X and the second bit Y:

	X	Y
Rock:	0	0
Scissors:	0	1
Paper:	1	0

Now if we have two players who can each choose one of these codes, we can play the game. **Examples**:

Player 1	Rock (0,0)	Player 2	Scissors (0,1)	Player 1 wins
Player 1	Scissors (0,1)	Player 2	Scissors (0,1)	Tie
Player 1	Rock (0,0)	Player 2	Paper (1,0)	Player 2 wins

We see that there are three possible outcomes of the game: Player 1 wins, Tie, Player 2 wins. Complete the following table that describes all the possible outcomes of the game.



[12]Spreadsheets have some built-in functions for decimal conversions to and from binary. It is recommended that you try first to use the actual mathematical functions described in this chapter and then *check* your answers using these functions to confirm your answers.

Player 1 X, Y	Player 2 X, Y	Player 1 wins	Tie	Player 2 wins
0,0	0,0	**0**	**1**	**0**
0,0	0,1	**1**	**0**	**0**
0,0	1,0			
0,1	0,0			
0,1	0,1			
0,1	1,0			
1,0	0,0			
1,0	0,1			
1,0	1,0			
	Totals			

23. A company purchased a computer program for your part-time job with them. The license agreement states that you can make a backup copy, but you can use the program on only one computer at a time. Since you have permission to make a backup copy, why not make copies for friends? What do you do?
a. Go ahead, since your friends use only one computer at a time and these are backup copies.
b. Make the backup copy, but sharing it with anyone clearly violates the license agreement.
c. Ask your supervisor if you can use the backup copy at home, and then make as many copies as you wish.
d. Use the program discretely, since software license agreements can't be enforced anyway.

(**Suggested answer:** Use the Engineering Ethics Matrix. The canons rule out all but option (b).)

24. You are a software engineer at a small company. You have written a software program that will be used by a major manufacturer in a popular product line. Your supervisor asks you to install a "back door" into the program that no one will know about so that he can monitor its use by the public. What do you do?
a. Install the back door, since it sounds like a fun experiment.
b. Tell your supervisor that you can't do it without authorization from the end user.
c. Install the back door but then deactivate it before the software is implemented.
d. Stall your supervisor while you look for another job.

Use the Engineering Ethics Matrix.

CHAPTER

Control System Design and Mechatronics

10

Source: © iStockphoto.com/Konstantin Inozemtsev

How does an airplane's auto pilot, or a farm tractor's GPS steering system, or an automobile's cruise control system work? That question introduces the engineering topic of **control system design**, a central part of the engineering discipline of mechatronics.

10.1 WHAT IS MECHATRONICS?

Mechatronics is the discipline of combining electronics, computer engineering, electrical engineering, and mechanical engineering into an *integrated* program of study. Mechatronics has made it possible to design complex electromechanical systems such as the Segway™ personal transporter (Figure 10.1). This remarkable device automatically keeps perfect balance despite speeding up, slowing down, and maneuvering around obstacles.

Mechatronics enables a robot to tirelessly and accurately carry out an assembly task that would tax the skill, strength, and stamina of a human. Thanks to mechatronic controls, a solid-state camera can automatically sense light levels and adjust its settings to ensure perfectly exposed digital photographs.

As different as these applications are, they share a common topic, control system design; a common discipline, mechatronics; and a common tool, the **block diagram**. A block diagram, the analogue of an energy control volume or an electric circuit diagram, is a way of turning a conceptual sketch of a control system into an engineering model that can be analyzed in terms of variables, numbers, and units. The block diagrams that are the focus of this chapter represent the key components that work together to make up an automatic control system.

A general control system design consists of (1) modeling the system to be controlled as a block diagram; (2) translating that block diagram into a **mathematical model** and obtaining numerical results from the mathematical model, either in tabular, spreadsheet, or graphical form; (3) adding a mathematical model of the effects of external influences; (4) selecting a **control strategy** for the appropriate control of the whole system; and (5) **implementing** the control strategy in hardware.

The principal control strategy used is **feedback**. A feedback device senses the value of a variable (such as vehicle speed) and "feeds" it back to compare it to the desired value of the variable (e.g., 65 mph). If the value of the variable is too low, a correction is made to increase it; if too high, then the correction is made to reduce it (e.g., by easing up on the accelerator if the current speed is greater than the set speed). Engineers often need to cause a variable to stay automatically at a set point—the set point being the desired value. This is a principal concern in the wide-ranging subject of control.

FIGURE 10.1 Personal Transporter. Used by Permission of Segway Inc

Control system design has wide-ranging applications, from controlling the darkness of toast in a toaster to controlling the flow rate of water to a hydroelectric plant. The thermostat in your home heating furnace is a familiar case in point. To be comfortable you want to keep the temperature of a room at 65 to 75°F, depending on your particular comfort level, even if no one is in the room to turn the heat on or off.

There are two subcases of control that are important in control theory: *steady state* and *transient* situations. A steady state situation arises when the inputs and outputs do not change (i.e., they are "steady"). A transient situation involves time-dependent inputs or outputs. Both are important subcases, but the steady state case is the easier one to start a study of the principles of control. Therefore most of this chapter will concentrate on steady state control with just a short digression into transient control theory.

10.2 MODELING THE CONTROL SYSTEM AS A BLOCK DIAGRAM

Developing a control system begins by converting a conceptual sketch of the system being controlled into a block diagram. However, it may be easier to start with a physical sketch and develop the block diagram from what it tells you. Figure 10.2 is a conceptual sketch of a cruise-controlled car. In the sketch, the engine is controlled by the accelerator.

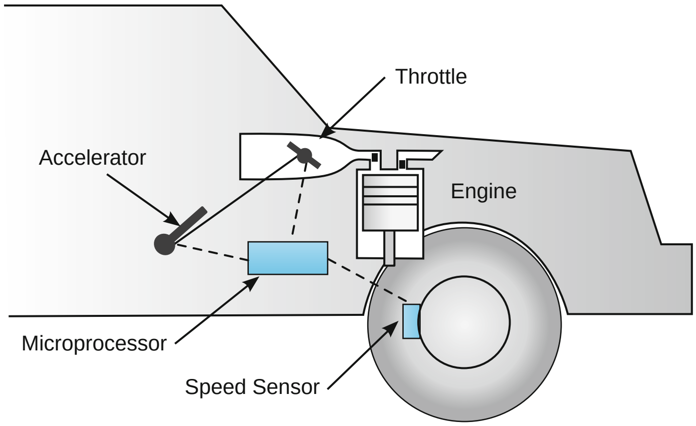


FIGURE 10.2 Speed Control in an Automobile

The accelerator in turn is connected to the throttle and to a microprocessor or microcontroller. The microprocessor is in turn connected to a speed sensor that measures how fast the car is going.

This conceptual sketch provides the basis for creating a block diagram. A block diagram is the next level of abstraction. It is both a visual and a mathematical tool for control analysis. It does for control analysis what control boundaries do for energy conservation analysis and what circuit diagrams do for the analysis of electric circuits—in other words, it focuses your attention on the essential core that you wish to analyze. A block diagram is connected together in the same way as the system it represents. To illustrate this, Figure 10.3 shows a hardware block diagram that represents a cruise control.[1]

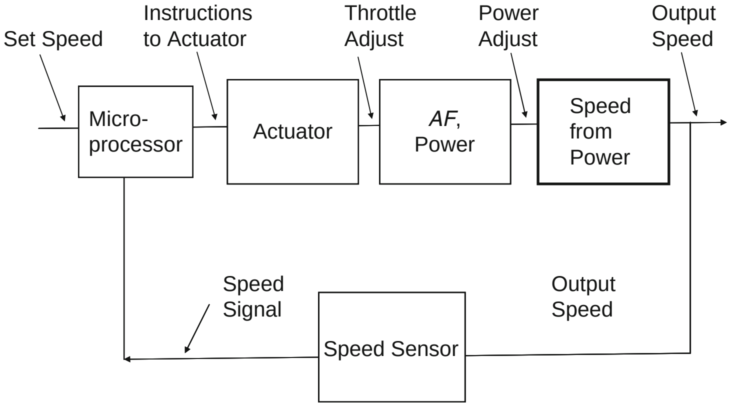


FIGURE 10.3 First Approximation to a Cruise Control

[1]This is called a “hardware” block diagram because it represents a physical rather than functional configuration.

How engineers think about problems such as this starts with a block diagram. The next step is to turn that initial block diagram into a mathematical model of the system. Each block will now express, through a mathematical operation, what the corresponding component of the physical system actually does. For example, if a component did the job of converting one variable to another, it would be represented by the block shown in Figure 10.4. That block consists of an **input variable**—say, 2, entering from the left, the block itself—and an **output variable**—say, 6, exiting on the right. The interior of the block describes the mathematical operation performed on the input to produce the output. We will call the internal mathematical operation a **response function** or a **transfer function**.

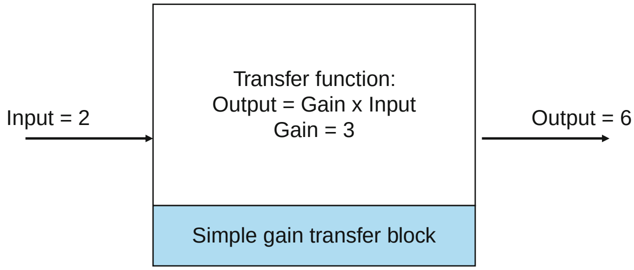


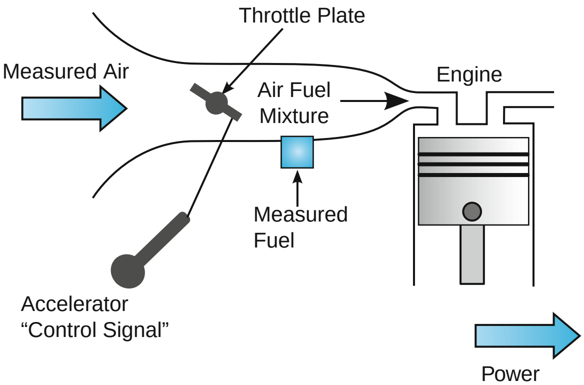
FIGURE 10.4 A Simple Block Diagram Operation

If the transfer function is a simple multiplication, it is called a **gain**. In Figure 10.4 the transfer function operates on the input 2 producing the output 6. The shaded label below the block can be used to give the block a descriptive name.

The current task is to translate each of the elements of the engine and its control system into blocks of this type. As we do, the blocks will be connected in an arrangement similar to the conceptual sketch of the physical system.

Example 10.1

Model the throttle mechanism of a car with a control block.

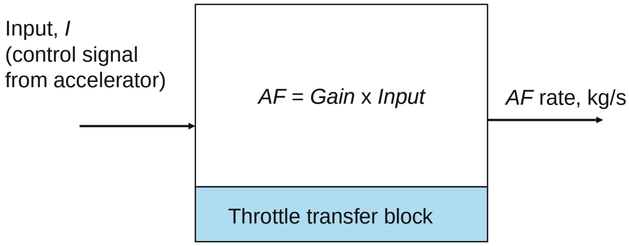


Schematic of Fuel/Engine System

Need: A control block corresponding to the throttle mechanism shown.

Know: The input to the throttle is a control signal from the accelerator (call it ***I***). The output from the throttle is an air-fuel mixture that goes to the engine. Call it ***AF***; it will be in units of flow—for example, in SI units *AF* would be expressed[2] in kg/s. The relation between *I* and *AF* is determined by the amount a given control signal "opens" the throttle. The physical linkage could be a mechanical or electrical signal and a corresponding electrical actuator.

How: Assume that the transfer function consists of the multiplication of *I* by *G* (***G* for gain**). This relation can then be expressed by a block diagram.



Controller Block for Fuel Injection Rate

For example, if the linkage were mechanical, perhaps you move the accelerator just 2.5 mm causing the throttle plate to open by 10° of rotation. The gain is then 4°/mm. The units of gain are therefore quite specific to the particular operation we are modeling.

Solve: The block diagram we are seeking is shown above. Our block for the air-fuel mixture flow rate consists of the accelerator control signal input (basically the effect of the various linkages) entering from the left, the block itself, and an output air-fuel flow rate variable *AF* exiting on the right. The interior of the block describes the mathematical operation performed on the input to produce the output. When the control signal calls for more or less fuel, the amount is $AF = Gain \times Input$, which, in SI units, is in kg/s.

The transfer function described in this block is a simple gain. In addition, since the value of the output *AF* is proportional to the input, it is called **proportional gain**. More complex transfer functions are also common. Since it is *AF* that we are interested in, you can theoretically achieve it by manipulating either the gain or the input, or both (although the gain normally is fixed by the mechanical or electrical design).

For the special case of **on/off control**, the gain simply switches between $gain = 0$ and $gain = 1$. Many on/off control systems exist (some older home thermostats operate this way), but there are much better control systems for most applications. In this book, we will always assume that the gain is fixed at some value chosen to produce the required output for a given input signal.

The next required block is the engine's response to additional fuel and air. The power output of an automobile engine is approximately proportional to its intake rate of air and fuel (see Chapter 6). So engine power can be modeled by another simple block. In this model we will have calculated power output for our engine that says $P(engine) = G_{Engine} \times AF$, in which G_{Engine} is the engine gain (i.e., engine output for a given air–fuel mixture flow rate input).

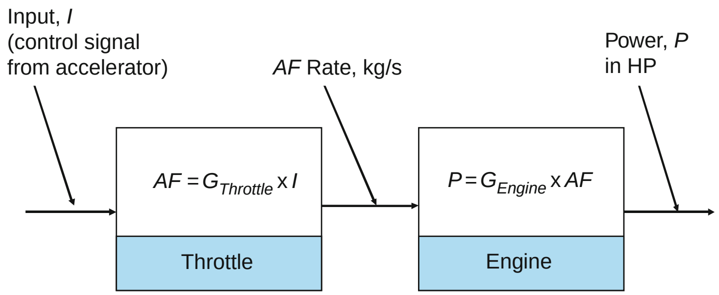
We will now depart from the convenient SI units of power (normally in kW). This is because "horsepower" has become an internationally recognized term for engine power. This is easy to preprogram into our design. Suppose, for example, that our engine produces 40. HP from an input of $AF = 0.05$ kg/s of air and fuel. The engine's gain is thus 800. [HP/kg/s] since 0.05 [kg/s] × 800. [HP/(kg/s)] = 40. HP.

[2]Note that *AF* is *not A*/*F*, mass or molar ratio that we previously used, but is the mass flow rate of air *and* fuel entering the engine.

Example 10.2

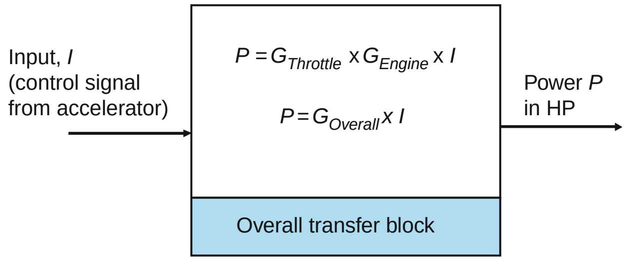
Develop a block diagram that combines the accelerator control signal, throttle setting, and power output.

Need: Overall block diagram for fuel system and for engine power.
Know: Block diagrams for each subsystem.
How: Join the systems together at common points of intersection.
Solve:



Block Diagram for Overall Engine Power

Recognizing that our operations in the blocks are linear, meaning they can be simply multiplied together, the block diagram can be simplified by combining its two blocks into one.



Simplified Block for Overall Engine System

We have written $P = G_{Overall} \times I$ by defining $G_{Overall} = G_{Throttle} \times G_{Engine}$. Let $G_{Overall} = 800.$ [HP/kg/s]. Then if $I = 0.05$ the engine's output is still 40. HP, and if $I = 0.10$ the engine's output is 80. HP.

Note one important feature of the model so far. Assuming we have the necessary physical setup to achieve these values, all we have controlled is the engine's *power*. Remember, however, that our objective is to control the *speed* of the automobile. That variable "speed" does not yet appear anywhere in the diagram! This is a typical situation in control system design. The variable the system directly modifies is often *not* the one that the system ultimately is intended to control. In this case, the system modifies power, but the goal is to control speed.

10.3 SELECTING A CONTROL STRATEGY

A **control strategy** is a method of automatically enabling the system to control itself. Suppose you want to control the speed of your car; you could place a brick on the accelerator pedal and trust it gives what you want. A successful outcome is obviously very unlikely. For example, the car might speed up or slow down as it moves from a flat road to a hilly one. The instantaneous speed S is not necessarily equal to the set speed S_0.

This is an example of a class of strategies called **open loop** control. A conceptual hardware sketch and a block diagram for an open loop controller implementing this strategy are shown as Figures 10.5 and 10.6, respectively.

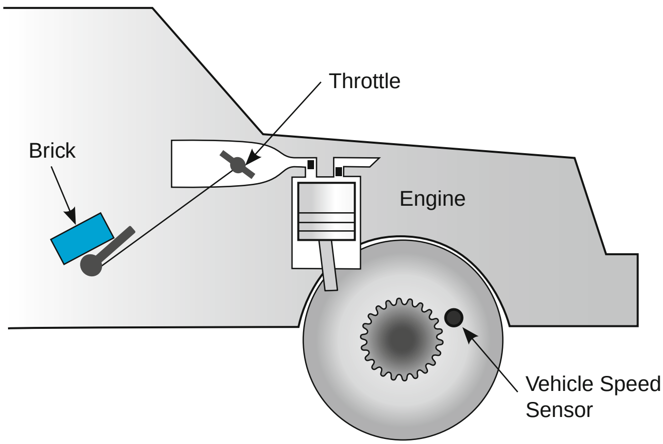


FIGURE 10.5 Open Loop Control

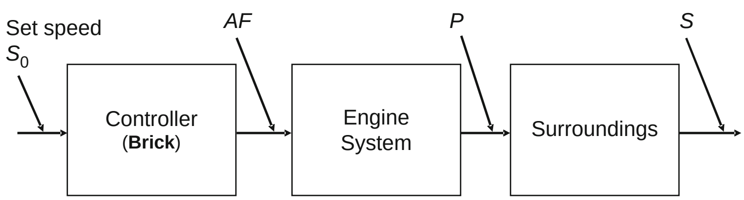


FIGURE 10.6 Open Loop Block Diagram

We sense the vehicle's speed by a device attached to the wheels. Even though we are measuring the quantity we want to control, in open loop strategies we ignore any new information relating to it after we have sought to control it. Another simple example of a system employing an open loop strategy is a light switch. Once it is turned on, the lights go on and stay on, whether it is day or night.

Often, however, an engineer wants the system to respond to changes in the surroundings. For example, a lighting engineer might want to combine a light switch with a light sensor that automatically turns the lights off in the daytime and then turns them back on at night. Systems that sense and respond to the surroundings in this way are called **closed loop** control systems.

In the automobile example, closed loop strategies involve the response of a human driver. A driver attempting to maintain a constant speed regularly compares the desired speed to the actual speed displayed on the speedometer. If that actual speed is lower than desired, the driver steps on the accelerator. If the actual speed is higher than desired, the driver lets up on the accelerator.

How can a closed loop strategy be implemented automatically? This requires putting a new symbol into the control diagram to perform the observation and action done by a human driver. The new symbol is called a **comparator**. The comparator operation, Figure 10.7, is the essence of both how we use a speedometer on a car and how an automatic control system operates. The comparator compensates for the error in the desired value by regularly comparing the measured speed and the desired speed.

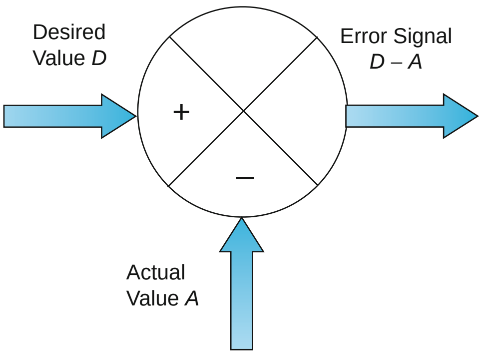


FIGURE 10.7 A Comparator

Obviously the comparator, unlike our previous blocks, has two inputs, the measured speed from the speed sensor and the desired speed called the **set point**. If the output of the comparator is the *difference* of the two input values, as in Figure 10.7, it is called the **error signal** ($D - A$).

Figure 10.8 shows the application of a comparator to a cruise control. As we can see, the loop has been closed. Here we will use the error signal, the difference between the desired and actual speeds, to set the air–fuel mixture, *AF*, and ignore the throttle setting altogether.

For obvious reasons, any control strategy that "feeds back" a signal from the output to alter the input signal is called **feedback control**; it is the principal way that the great majority of modern controllers work.

The + and – signs in the comparator symbol indicate the type of feedback desired. A minus sign as in Figures 10.7 and 10.8 indicates **negative feedback**, in which the feedback signal tends to reduce output, and a plus sign indicates **positive feedback** in which the feedback tends to increase output.[3] Systems that have high-gain feedback can produce oscillations in the output resulting from first positive (overshoot) then

[3]There is also a third type of feedback called **bipolar**, which can either increase or decrease output. Bipolar feedback is present in many natural and human systems. Feedback is usually bipolar—that is, positive and negative—in natural environments, which, in their diversity, furnish synergic and antagonistic responses to the output of any system.

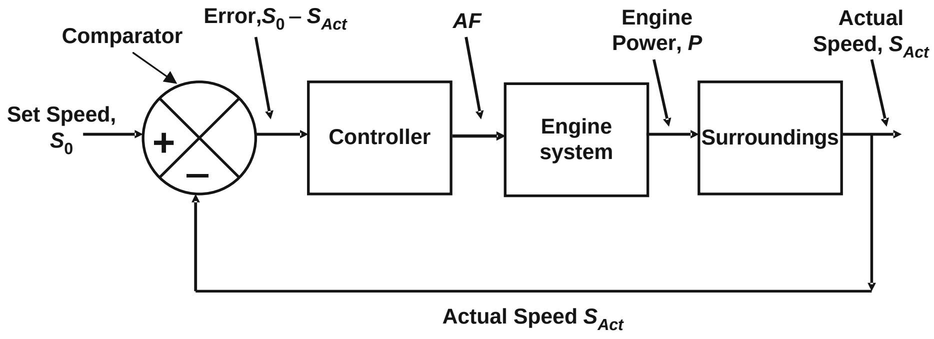


FIGURE 10.8 Closed Loop Control System

negative (undershoot) feedback. This behavior is called *hunting*. Audio feedback sometimes produces this type of oscillation.

To use the comparator, all we need is the speed signal converted into an appropriate format. The controller reads the error between the set speed and measured speed provided by the comparator and then sets the air–fuel supply needed for the engine, which in turn is described by another block whose output is the engine power. The other block, labeled "Surroundings," conceals a number of variables such as road conditions (e.g., flat or hilly), the wind resistance, whether the driver is braking, etc. The resultant is the actual car speed in appropriate units—for we humans, that's mph (or km/h). For the car's control system it's usually an electrical signal proportional to speed. We now need to define what the controller in Figure 10.8 will do.

We will deal here with just one type of controller, a *proportional* controller. As we will see in Example 10.2, a proportional controller simply multiplies its input (the error signal from the comparator) by a constant **gain** (G_p), to produce its output signal. This is written mathematically as:

$$S_{Act} = G_p(S_0 - S_{Act}) \tag{10.1}$$

The reduced closed loop block diagram is shown as Figure 10.9.

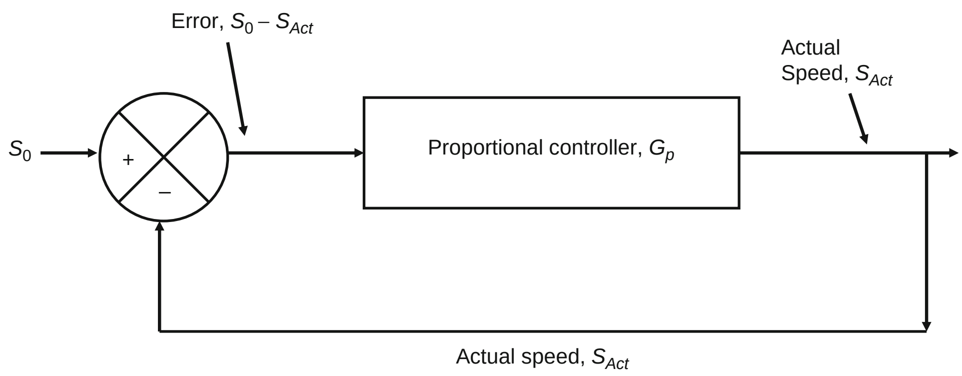


FIGURE 10.9 Closed Loop Proportional Control

Now we can see a problem with purely proportional control; the output from the proportional controller block that simulates the whole system is not what we had hoped for since by rearranging the Equation (10.1) we can produce

$$S_{Act} = \frac{G_p S_0}{(1 + G_p)} \tag{10.2}$$

Only if we turn up the gain will we progressively reduce the steady state error $S_0 - S_{Act}$, to zero since it's only in the limit of $G_p \to \infty$ does $S_{Act} \to S_0$. So for small gains, $G_p \to 0$, therefore, $S_{Act} \to G_p S_0$.

We can now determine the *steady state error* from Equation (10.2) as:

$$S_0 - S_{Act} = S_0 - \frac{G_p S_0}{(1 + G_p)} = S_0 \left[1 - \frac{G_p}{(1 + G_p)}\right] = \frac{S_0}{(1 + G_p)}$$
$$\text{Or} \quad S_0 - S_{Act} = \frac{S_0}{1 + G_p} \tag{10.3}$$

Note that this treatment of control theory *does not measure the transient changes* between S_0 and S_{Act}. Rather it is a *snapshot frozen in time* of the effect of a set point and a proportional controller, which is called "steady state." If you want to understand the transients between any two variables such as S_0 and S_{Act} or, for example, the transient temperature of an oven and its electric heat, you will need a transient model. In the case of an accelerating car, you will at a minimum need Newton's Second Law to describe the interaction of the car's power and its inertia.

Example 10.3

You want to maintain 65. mph in a car using a proportional cruise control. If the proportional gains are (a) 0.10, (b) 1.0, (c) 10., and (d) 100., what are the corresponding actual speeds?

Need: For $S_0 = 65.$ mph, $S_{Act} =$ ______ for $G_p = 0.10$, 1.0, 10. and 100., respectively.

Know: $S_{Act} = \dfrac{G_p S_0}{(1 + G_p)}$

How: Direct substitution.

Solve:

(a) $G_p = 0.10$, so $S_{Act} = (0.10)(65.)/(1 + 0.10) = 5.9$ mph
(b) $G_p = 1.0$, so $S_{Act} = (1.0)(65.)/(1 + 1.0) = 36$ mph
(c) $G_p = 10.$, so $S_{Act} = (10.)(65.)/(1 + 10.) = 59$ mph
(d) $G_p = 100.$, so $S_{Act} = (100.)(65.)/(1 + 100.) = 64$ mph

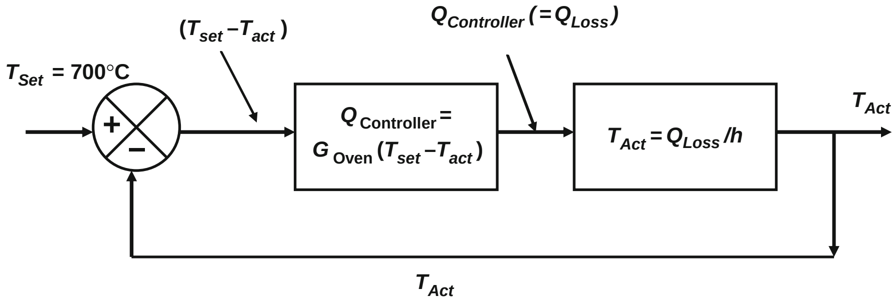
If you have a cruise-controlled car, you know that the controller tracks the driver's setting quite accurately, so the controller probably has a relatively high gain.

Example 10.4

A small oven is to be controlled at a desired temperature of $T_{Set} = 700.°C$. The proportional controller tells the oven to deliver heat at a rate of Q watts according to $Q_{Controller} = G_{Oven}(T_{set} - T_{act})$ where $G_{Oven} = 1250$ W/°C and T_{act} is the resulting oven temperature. Heat losses from the oven are described by $Q_{Loss} = hT_{act}$ where h is called the heat transfer coefficient and has a value of $h = 50.$ W/°C. Draw the steady state block diagram for the oven and controller, then calculate the overall gain $G_{Overall}$ and T_{Act}.

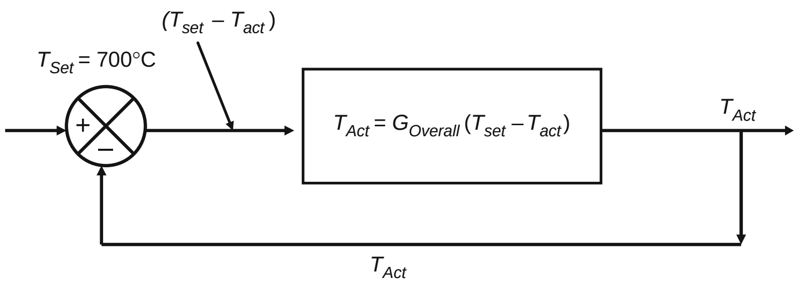
Need: The steady state block diagram for oven, $G_{Overall}$, and T_{Act}.

Know: Oven gain is $G_{Oven} = 1250$ W/°C and heat loss is $Q_{Loss} = hT_{act} = 50.T_{Act}$.
How: The *steady state* heat into the furnace and the heat loss must exactly compensate, or $Q_{Controller} = Q_{Loss}$.
Solve:



The key step is $Q_{Controller} = Q_{Loss}$, so we can write $T_{act} = Q_{Controller}/h = (G_{Oven}/h)(T_{set} - T_{act})$ thus defining the proportional gain as $G_{Overall} = G_{Oven}/h$.

We can redraw the combined block diagram that illustrates this point:



In this form, $T_{Act} = G_{Overall}(T_{set} - T_{act})$ in which $G_{Overall} = G_{Oven}/h = 1250/50.$ [W/°C][°C/W] $= 25.$ Therefore,

$$T_{Act} = \frac{G_{Overall}T_{set}}{(1 + G_{Overall})} = \frac{25. \times 700.}{26.}\frac{[0][^\circ C]}{[0]} = 673^\circ C = \mathbf{670^\circ C}$$

(to two significant figures).

10.4 TRANSIENT CONTROL THEORY

The error predicted in steady state control depends on the size of the proportional gain. It *appears* that we could eliminate the steady state error by simply increasing the proportional gain. Unfortunately this has profound implications when we also try to control transients. The fundamental reason is that every time we update information, the correction is large if the gain is large and may overcorrect causing overshoot and oscillation in the controlled variable. We will attempt to illustrate this in just one example that will have several factors to consider.

If we keep to our theme of a car cruise control, we can write down what we need to include in a transient model: first we know that the car's engine produces power that will maintain the car's speed against various mechanical losses, the principal one being air resistance.

Example 10.5

A wind resistance power loss model for a vehicle is $P_{Losses} = 1.85 \times 10^{-4} \times S^3$ (S in mph, P_{Losses} in HP). How much HP will be lost to the wind at 30.0 to 100. mph in 10.0 mph increments?

Need: Wind resistance as function of speed.
Know: $P(losses) = 1.85 \times 10^{-4} \times S^3$ (S in mph, P_{Losses} in HP)
How: Wind speed in; power losses out.
Solve: The equation, $P_{Losses} = 1.85 \times 10^{-4} \times S^3$ (S in mph, P_{Losses} in HP) is the required response function, since it converts the wind speed to power losses. It predicts the following losses to the wind:

S, mph	30.0	40.0	50.0	60.0	70.0	80.0	90.0	100.
P_{Losses} HP	5.0	11.8	23.1	40.0	63.5	94.7	135.	185.

Notice how quickly the wind losses mount as the speed increases. At 100. mph the losses are nearly 185 HP compared to just 40.0 HP at 60.0 mph and a mere 5.0 HP at 30.0 mph! If not safety, then the additional fuel (with its financial and environmental costs) should dissuade you from speeding!

Even if the controller calls for the extra fuel and air there's another large factor working against us—Newtonian physics. The car has inertial mass so it will not accelerate instantly. Suppose the speed of the car at time t is S_t. The overall block diagram will look like this:

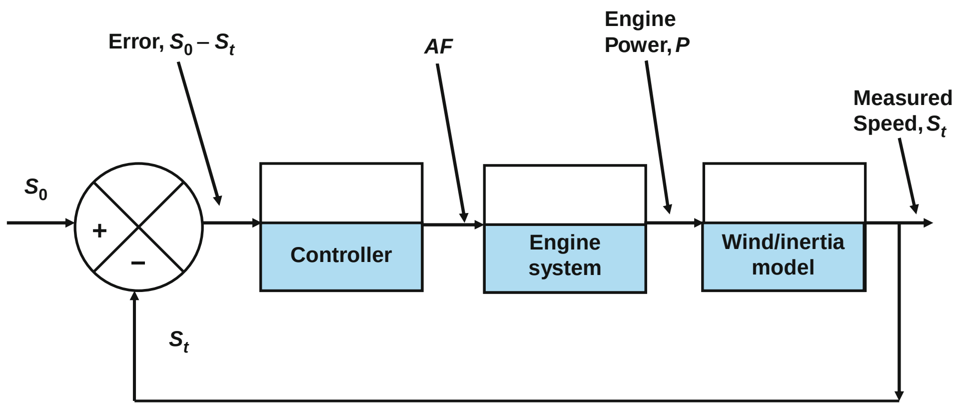


FIGURE 10.10 Transient Model for Cruise Control

In Figure 10.10 we have cheated just a little; the wind/inertia model will have to be "linearized" because the wind power losses are proportional to the *cube* of speed. If we can accept this slight of hand, we can analyze the essential features of transient systems even though transient response can be very complex. For example, Figure 10.11 illustrates the speed of a vehicle as it accelerates under proportional control from an initial speed of 50 mph to a desired speed of $S_0 = 60.$ mph.

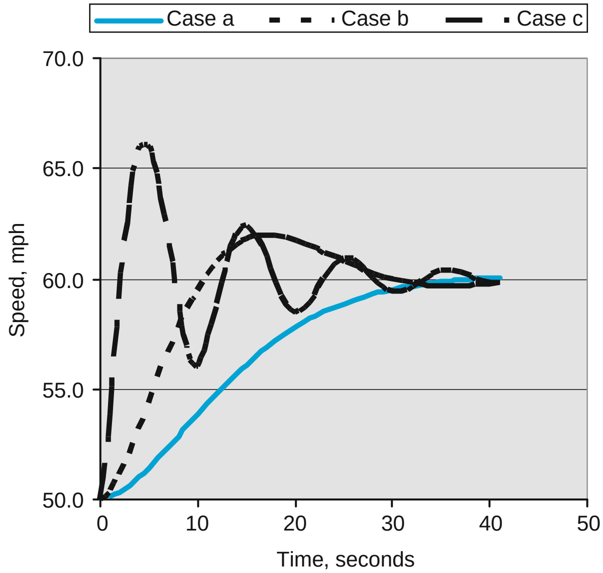


FIGURE 10.11 Transient Proportional Control Can Become Unstable with Increasing Gain. Case (a) $G_{Overall} = 1.5 \times 10^{-4}$ kg/s/mph; Case (b) $G_{Overall} = 5.0 \times 10^{-4}$ kg/s/mph; Case (c) $G_{Overall} = 1.5 \times 10^{-3}$ kg/s/mph

In Case (a) the graph shows how long it will take to achieve the speed we asked for—almost 40 seconds, possibly a dangerously long time. The reason is the incremental fuel addition is very small since $G_{Overall}$ is small. The resulting increment in engine power is accordingly very small. We can call this case "overdamped."[4]

In Case (b) we have approximately tripled the proportional gain. In this case we achieved our 60 mph goal in about 10 s, but the controller visibly overshot the set point of 60. mph before settling. Still, this controller is more or less doing what we asked of it.

So what happens in Case (c) at still higher gain? Case (c) is behaving like an oscillating spring; this is referred to as "underdamped," since it "rings" for a long time before settling down. Such behavior would be very undesirable on the highway. In general, if proportional gains are too large, unsteady behavior will be produced.

In this example, the tradeoff between rapidity of response and stability is accomplished by changing the proportional gain, $G_{Overall}$. In real-world examples, engineers apply control laws that go beyond the on–off and proportional controls so far introduced. With them, plus considerable mathematical analysis, most control systems will behave in acceptable ways.

The key to all of these control systems, however, remains the choice between open loop and closed loop systems. In practice, almost all control systems are closed loop. Most are as shown here—feedback in which a comparator uses an error signal for subsequent control.

[4]**Damping** is any effect, either deliberately engendered or inherent to a system, that tends to reduce the amplitude of oscillations of an oscillatory system. Overdamped systems take longer to react and underdamped systems will have slowly decaying oscillations around the set point.

10.5 GLOBAL WARMING AND POSITIVE FEEDBACK

Many scientists think the atmosphere is warming and it is driven by carbon dioxide (CO_2) emissions from burning fossil fuels. The addition of carbon dioxide to our atmosphere is trapping heat because incoming relatively short wavelength sunlight is converted to long wavelength thermal radiation. Since the atmosphere is relatively opaque to long wavelength radiation, it traps it in the atmosphere. In 1910 the CO_2 content of our atmosphere was 300 parts per million (ppm) by volume; by 2010 it will be 390 ppm and it is now climbing exponentially.

Example 10.6

The atmosphere has warmed about 0.6°C between 1910 and 2000, and in 2000 the average global temperature was 15°C. Develop a block diagram for the feedback system and graph the average atmospheric temperature in each of the years from 2000 to 2050.

Note that this is a *positive* feedback closed loop system, and consequently the output (the atmospheric temperature) will continue to increase.

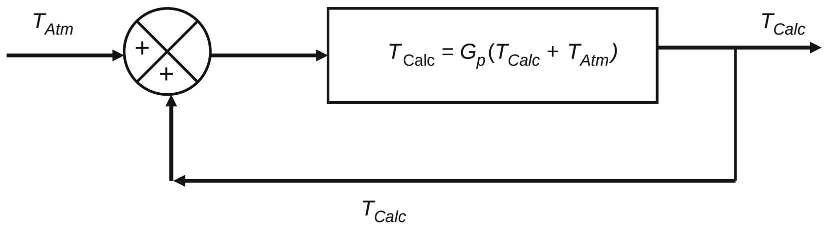
Need: Atmospheric temperature block diagram and the yearly increase in atmospheric temperature between 2000 and 2050.

Know: The rate of increase of atmospheric temperature averages 0.6/(2000 – 1910) = 0.007°C/year with a current atmospheric temperature of 15°C. We can now estimate[5] the gain in the last century to be $G_p = 0.50012$, which is just enough to produce the observed 0.6°C between 1910 and 2000). Because we are adding additional CO_2 to the atmosphere all the time, the current gain should now be larger than the historical record.

How: Draw a block diagram showing the *positive* feedback. Based on the data from 1910 to 2000, assume the mean gain is $G_p = 0.50012$. Assume for the years 2000 to 2050 the mean gain G_p will be: Case (1) 0.50012, and Case (2) 0.5012.

Solve: A highly simplified block diagram is shown below.

Assume that each calculated atmospheric temperature will be next year's average global temperature. Let T_{Calc} be the calculated atmospheric temperature and T_{Atm} be the average global atmospheric temperature. Now repeat this for 50 years using the updated value of T_{Atm}.



From the gain equation in the block diagram,

$$T_{Calc} = G_p(T_{Atm} + T_{Calc})$$

[5]For an initial estimate, the average gain must be larger than the average temperature over that period divided by the sum of the starting and ending temperatures, or 14.8/(14.6 +15) = 0.50.

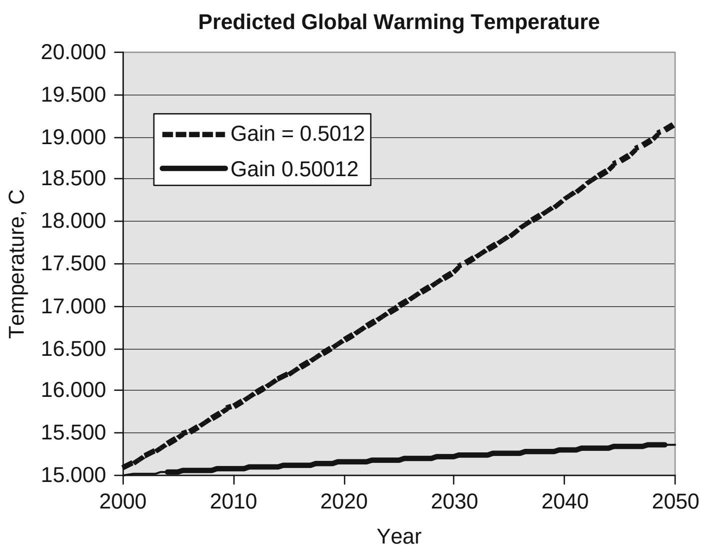
Spreadsheet calculations using this equation:

	D	E	F	G	H
1		G_p	0.5012	0.50012	
2		**Start T, C**	15	15	
3			$T_{Calc} = \frac{G_p T_{Atm}}{(1 - G_p)}$		
4					
5					
6					

7	Year	T atm C	T calc C	T atm C	T calc C
8	2000	15	15.072	15	15.007
9	2001	15.072	15.145	15.007	15.014
10	2002	15.145	15.218	15.014	15.022

	D	E	F	G	H
7	Year	Tamb C	Tcalc C	Tamb C	Tcalc C
8	2000	=F2	=$F$1*E8/(1−$F$1)	=F2	=$G$1*G8/(1−$G$1)
9	=1+D8	=F8	=$F$1*E9/(1−$F$1)	=H8	=$G$1*G9/(1−$G$1)
10	=1+D9	=F9	=$F$1*E10/(1−$F$1)	=H9	=$G$1*G10/(1−$G$1)
11	=1+D10	=F10	=$F$1*E11/(1−$F$1)	=H10	=$G$1*G11/(1−$G$1)
12	=1+D11	=F11	=$F$1*E12/(1−$F$1)	=H11	=$G$1*G12/(1−$G$1)

What do these answers mean? For Case (1) the rise in atmospheric temperature is very close to 50 years $\times$ 0.007°C/year $=$ 0.35°C. The reason is that the annual compounding of the temperature is very small when the assumed positive feedback gain is 0.50012. If we were to set the gain to exactly 0.5000, the predicted atmospheric temperature would remain unchanged at 15°C in perfect balance over the period 2000 to 2050.



For Case (2) with G_p arbitrarily set to 0.5012, the graph shows a predicted atmospheric temperature of more than 19°C by 2050; if you carefully examine the atmospheric temperature line, it has a slight upward curvature. Higher gains would show even more upward curvature; in fact the curve is an exponential whose slope would increase ever more quickly. The apparent flatness of Case (1) is simply a consequence of an exponential curve being linear near its origin.

Case (1) predicts that the Earth's atmosphere will warm by an additional 0.35°C by the year 2050, an estimate less than predicted by many climatologists. The 19°C prediction in Case (2) is much larger than expected. There are a number of other important positive feedback features such as the fact that a warmer atmosphere will retain more moisture. This additional water vapor alone will trap even more thermal radiation from the Sun.

Notice how sensitive the results of Example 10.6 are to minor perturbations that can quickly compound into major events. With positive feedback, the input signal keeps on increasing rather than being damped as with negative feedback. However, ultimately some other parameter will grow large enough and eventually limit it.

10.6 DRIVE-BY-WIRE

Modern technology is being propelled by electronics, which are closely integrated into the operation of a system. Such close integration between mechanics and electronics is called **mechatronics**. "Fly-by wire" and "drive-by-wire" are technologies that rely heavily on mechatronics. Eventually the present mechanical controls in a car (steering, gears, brakes, and accelerator) will all be replaced by integrated mechatronic devices.

In this text so far, we have used mechanical linkages to adjust the A/F system because it is easy to understand. However, some advanced vehicles now do this electronically. One reason is that the rotation of a throttle plate by a mechanical linkage is inherently nonlinear (Figure 10.12).

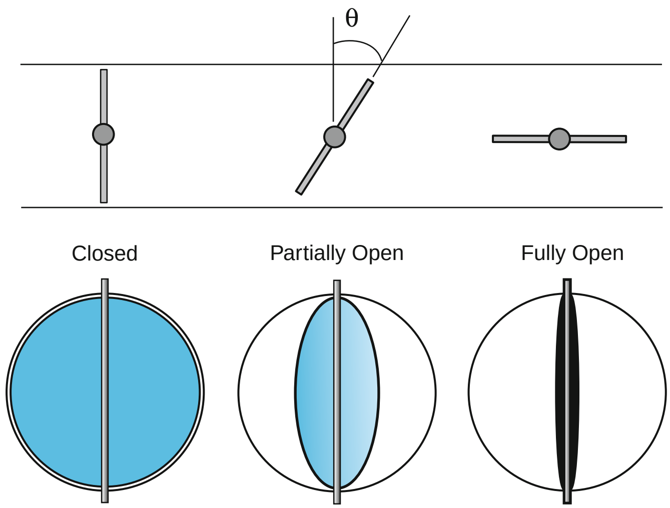


FIGURE 10.12 Throttle Position as a Function of the Angle θ

If you follow the geometry of opening the throttle plate (see exercise 17 at the end of this chapter), the percent open area is $\Delta A/A$ (in %) $= [1 - \cos(\theta)] \times 100$. Table 10.1 shows the corresponding open area versus rotation of the throttle plate. What this table means is that pushing your foot the same amount on the accelerator will produce different results depending on its current position. One solution to this lies in a specialized motor called a stepper motor. Its operation is animated in some web references, and one simplified configuration is shown in Figure 10.13.

Table 10.1 Incremental Opening Angle for 10% Change in Open Throttle Area

Throttle angle position (θ)	0°	25.8°	36.9°	45.6°	53.1°	60.0°	66.4°	72.5°	78.5°	84.3°	90.0°
Percent throttle open area $= [1 - \cos(\theta)] \times 100$	0%	10.%	20.%	30.%	40.%	50.%	60.%	70.%	80.%	90.%	100.%

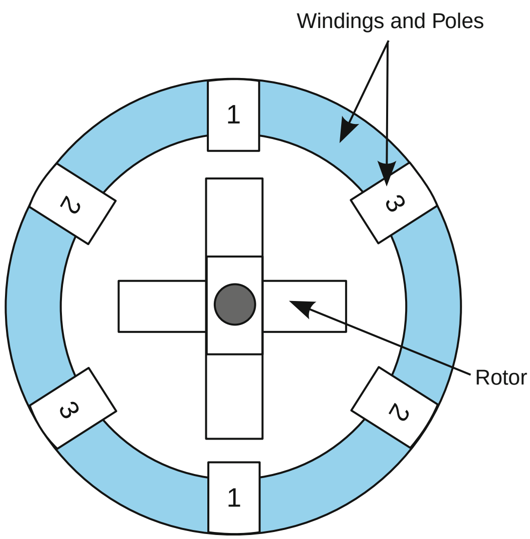


FIGURE 10.13 Principle of a Stepper Motor

The operation consists of pulses of electric power that are applied to the coils that actuate pole pairs 1, 2, and 3 (in that order), causing the rotor to increment counter-clockwise. These "poles" are just pieces of iron that concentrate the local magnetic field. Typical commercial application stepper motors increment by much smaller amounts, typically about 1° to 2° per pulse applied to the rotor. For this reason stepper motors are precise positioning devices and typically are used as actuators in mechatronic devices. As shown in Figure 10.14, they would replace the mechanical linkage.

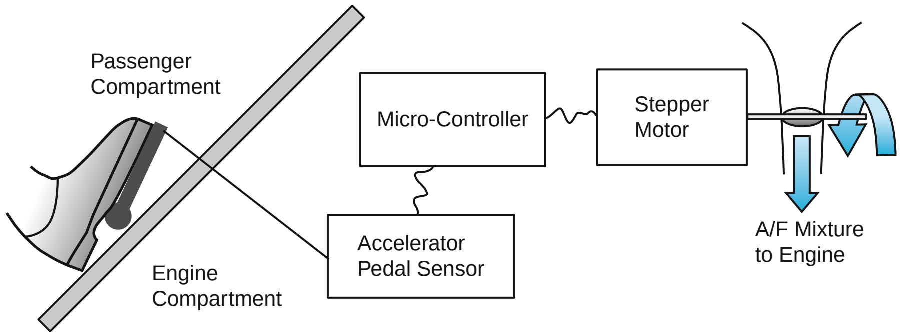


FIGURE 10.14 Drive-by-Wire

Example 10.7

A stepper motor increments by 1.8° per pulse. You wish to use it to ensure that equal movement of the accelerator pedal results in equal rotation of the throttle plate. How many pulses must you supply to open a throttle an extra 10% from an initial position of (a) 10%, (b) 50%, and (c) 80%?

Need: Number of pulses to open throttle by 10% at initial positions of 10%, 50%, and 80%, respectively.
Know–How: Table 10.1 in the text has the opening response as a function of rotation.
Solve:

Case (a): To go from 10% to 20% we need to open the throttle 36.9° – 25.8° = 11.1° or 11.1/1.8 = 6.16 = **6 pulses** (pulses being integers)
Case (b): To go from 50% to 60% we need to open the throttle 6.4° or 6.4/1.8 = **4 pulses.**
Case (c): To go from 80% to 90% we need to open the throttle 5.8° or 5.8/1.8 = **3 pulses.**

As long as the microcontroller will deliver the appropriate number of pulses in response to the signals from the pedal sensor, this will ensure the linear response that the driver expects.

The mechatronic approach would take additional advantage of the integration with electronics; presuming the vehicle also has cruise control, it will use the same stepper motor/throttle plate as shown in Figure 10.14, the difference will be the instructions (in the form of electrical pulses) transmitted to the stepper motor.

10.7 IMPLEMENTING THE CHOSEN STRATEGY IN HARDWARE

The engineer's next step translates the blocks on the control diagram into hardware. Some of that hardware is already installed—for example, actuators such as the throttle mechanism. This section will focus on two other hardware implementations that illustrate the mechatronic approach. Those implementations are the choice of a speed sensor and of a computer system to drive the controller.

For a long time automobiles have possessed a simple and reliable speed sensor. That sensor, the speedometer, traditionally used mechanical linkages to indicate speed. The turning of one of the car's axles was mechanically connected through a series of gears, cables, and shafts to an indicator on the dashboard (the speedometer). That indicator displayed the vehicle speed to the driver.

We could simply adapt this mechanical sensor to the cruise control. But the mechatronic approach suggests a different solution. The control system employs a digital controller, the microprocessor. So why not design a sensor that is also digital? Such a digital speed sensor is shown in Figure 10.15. It is based on the fact that a magnet moving past a coil of wire induces a pulse of electric current in the coil.

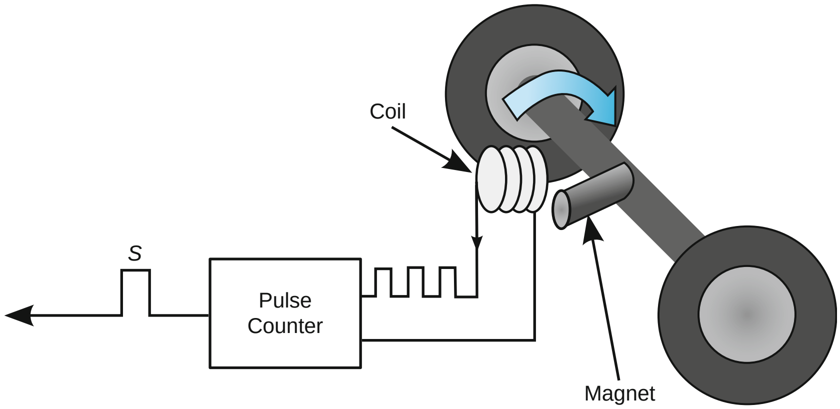


FIGURE 10.15 Electrical Digital Speed Sensor

In the speed sensor in Figure 10.15, a magnet is attached to, and turns with, one of the car's axles. A stationary coil is placed so that the magnet passes the coil each time the axle rotates. When the magnet passes the coil, it induces a current pulse in the coil. So the number of pulses per second in the coil is proportional to the number of revolutions per second of the axle. This provides a direct digital speed measurement for use by the cruise control system. A simple mathematical model of this speed sensor is the subject of one of the exercises at the end of this chapter.

In all modern cars the hardware implementation of the comparator is just a small microcontroller, which also doubles as the proportional control block. This is a design choice, since we could have used a more traditional method of linkages and mechanical devices to achieve the control. But we know we will also need a microprocessor elsewhere in our cars, so why not design dual functionality into the microprocessor controller?

SUMMARY

The control design strategy outlined in this chapter can be applied to any situation where a desired value of a variable is to be maintained. The following table[6] gives a range of applications. Parts of these applications, as well as others, are discussed in the exercises at the end of this chapter.

Application	Controller	Process	Command Signal	Controlled Value
Home heat	Thermostat	Furnace	Heat	Temperature
Drill	Speed dial	Electric motor	Torque	Bit speed
Toilet	Float and lever	Water valve	Water flow	Water level
Economy	Government policy	Treasury and federal reserve	Interest rate, money supply	Gross domestic product (GDP)

[6]Adapted from Raymond T. Stefani et al., *Design of Feedback Control Systems*. New York: Oxford U, 2002, p. 4.

Diverse as these applications are, they all implement similar principles of control system design and can be represented through **block diagrams**. The process of control design consists of the following steps:

1. Modeling the control system as a **hardware block diagram**.
2. Translating that block diagram into a **mathematical block diagram** model and obtaining numerical results from the mathematical model, either in tabular, spreadsheet, or graphical form.
3. Adding a mathematical model of external effects or surroundings.
4. Selecting a **control strategy**—either **open loop** or **closed loop**—for the appropriate control of system-plus-surroundings.
5. Implementing the chosen strategy in hardware.

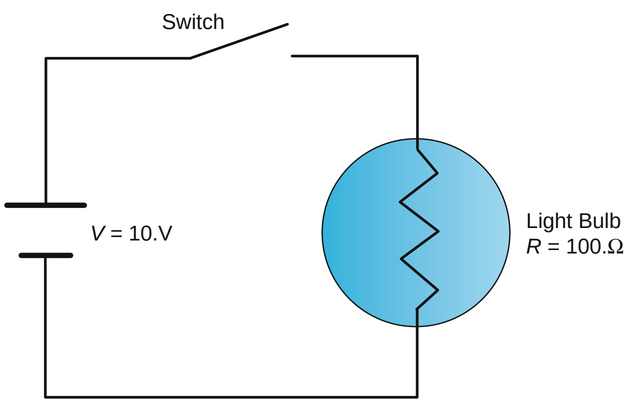
Most closed loop control strategies use negative **feedback**, which corrects the error in the monitored variable. **Proportional gain** is the multiplier of the correction signal in the feedback loop. It's an important variable in control strategies. Too little gain and the desired result may be very slow in coming and the proportional error may be excessive. Too much gain may reduce the error but the system may become unstable. **Underdamping** and **overdamping** are inherent factors in transient control strategies.

Mechatronics is the integration of mechanical and electronic systems into a design. If several objectives are considered simultaneously, the resulting system usually is less complicated than the sum of the individual systems. Control theory underlies the implementation of many mechatronic designs.

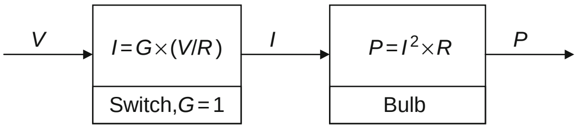
EXERCISES

The answers to these exercises are not necessarily unique since there is some latitude in equivalent ways they can be formulated. In these problems, as well as G for gain, the letter C is used meaning "Command." In some instances, such as on–off control, C and G may have a value of either 0 or 1.

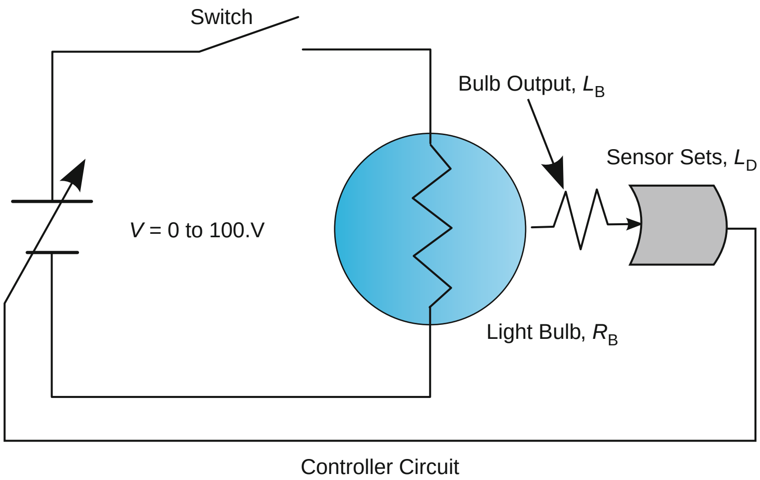
1. The following circuit diagram is an **open loop** control for tuning on or off a light. Draw a block diagram of the control. Why is it considered open loop?



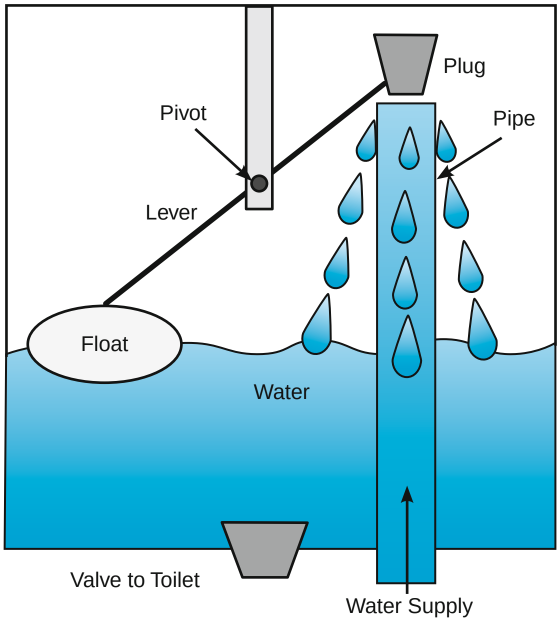
(**A**: The bulb is on and stays on provided the switch is closed, day or night, needed or not, until the switch is turned off manually. Then the same reasoning applies in reverse. That is why this is considered an open loop control.)



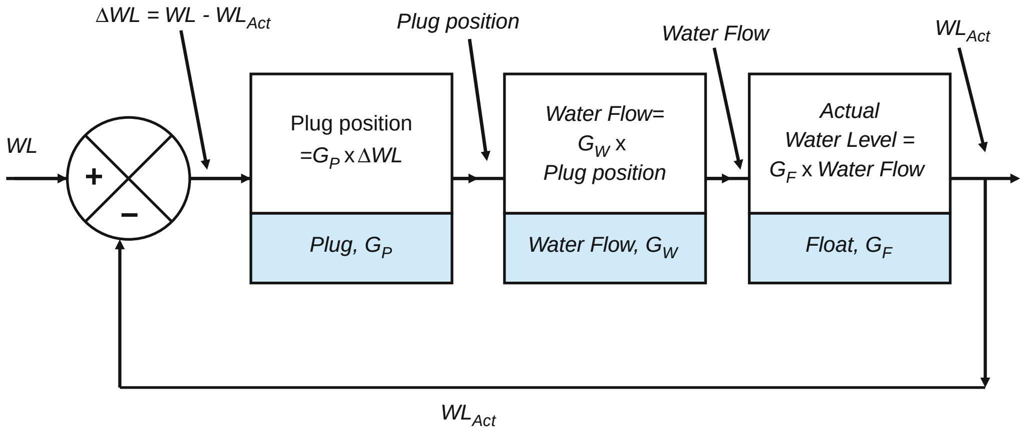
2. The following circuit diagram represents a **closed loop** system for lighting control in an otherwise dark room. Assume you want a lightbulb to provide constant illumination of strength L_D even if the strength of the lightbulb appreciably changes with age. Light control is achieved by a suitable light sensitive sensor that compares the bulb's measured output L_B to the desired light level, L_D. The input signal is the difference between L_D and L_B and is communicated to a variable circuit voltage from 0 – 100.V, which in turn sets the voltage, the current, and finally the bulb's intensity. Draw a block diagram of the control, indicating on the diagram which parts of the conceptual sketch correspond to the blocks. (**Hint:** If the sensor compares two inputs, does it remind you of a comparator?)



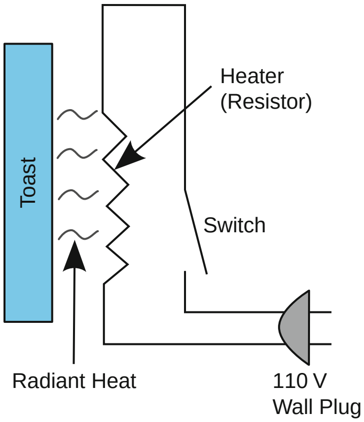
3. Part of the mechanism of a toilet is the system that refills the toilet tank after a flush. A conceptual sketch is shown below. Water enters the tank through an annular opening around the plug; eventually the water level, *WL*, in the tank rises and the float also rises, pushing down the plug and turning off the supply of fresh water. Draw a block diagram of the system, indicating on the diagram which parts of the conceptual sketch correspond to the blocks.


Plug
Pivot
Pipe
Lever
Float
Water
Valve to Toilet
Water Supply

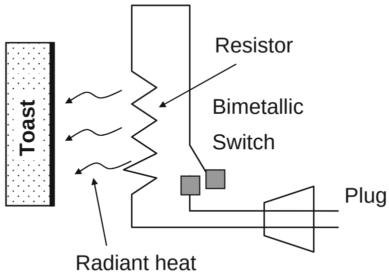
Answer:


ΔWL = WL - WL_Act
Plug position
Water Flow
WL_Act
WL
+
−
Plug position =G_P x ΔWL
Plug, G_P
Water Flow= G_W x Plug position
Water Flow, G_W
Actual Water Level = G_F x Water Flow
Float, G_F
WL_Act

4. The conceptual sketch for an *open* loop control for a toaster is shown. Draw a block diagram of the system, indicating on the diagram which parts of the conceptual sketch correspond to the blocks.

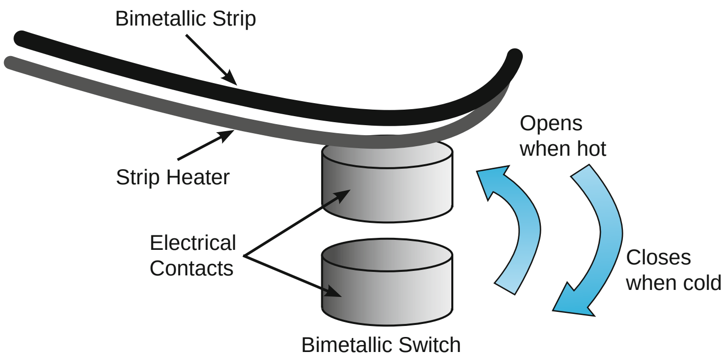


5. The conceptual sketch for a *closed* loop control for a toaster is shown below. The controller is a bimetallic switch. It works by combining two metals of different thermal expansion so that, when heated, it bends away from the more expansive metal and can then open a pair of contacts. The heat is interrupted when the contacts are opened and is reestablished when the bimetallic strip cools.

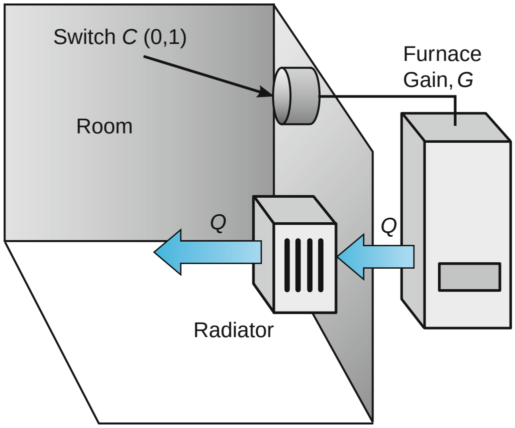


The heat source is "waste" heat generated in the resistor that also heats the toast. The stationary contact is adjustable so that the clearance between the two contacts determines the set point as to when the heater is on or off.

Draw a block diagram of the system, indicating on the diagram which parts of the conceptual sketch correspond to the blocks.

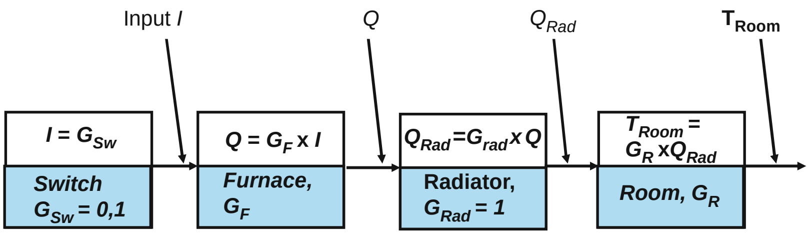


6. The conceptual diagram of an *open* loop control for the heating of a room is shown below.

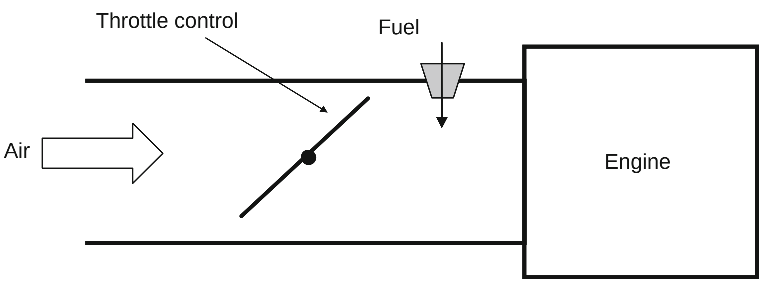


The room achieves a given temperature as a function of the added heat from the radiator and the heat loss Q_L from the room. But, for simplicity, ignore Q_L as being too slow to influence the immediate response. Draw a block diagram of the system that determines T_{Room}, indicating on the diagram which parts of the conceptual sketch correspond to the blocks. Assume the controller switch is a simple off/on (0,1) and the furnace responds with a gain G to produce heat at a rate Q.

Answer:[7]



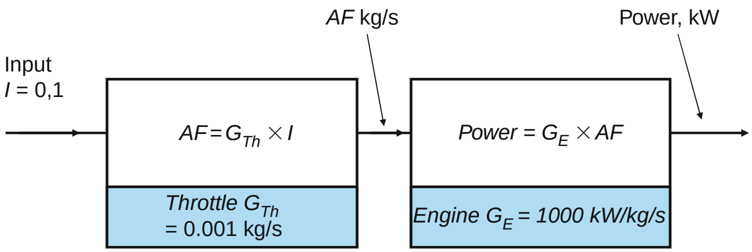
7. The conceptual diagram and block diagram of an open loop control for a lawnmower engine are shown below.



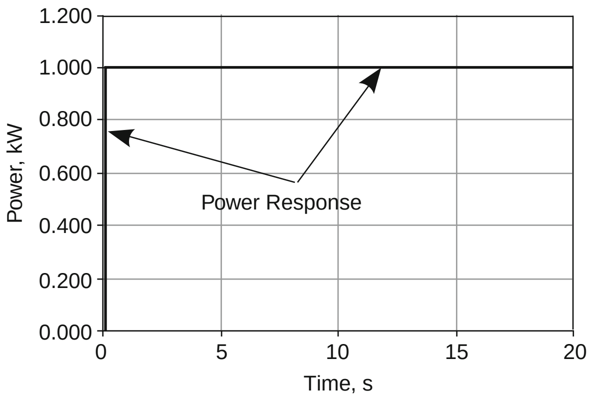
[7]Controllers may actuate their controlled output using either current or voltage. The advantage of current as a control signal is that it is undiminished over long distances (conservation of electric charge), something useful for geometrically large installations.

Assume the control is turned on at time $t = 0$ and left on thereafter. Use the proportional gains for the throttle = 0.001 kg/s and for the engine = 1000. kW/kg/s as shown in the block diagram. Prepare a spreadsheet and graph showing the power of the engine each second for a total of 20 seconds.

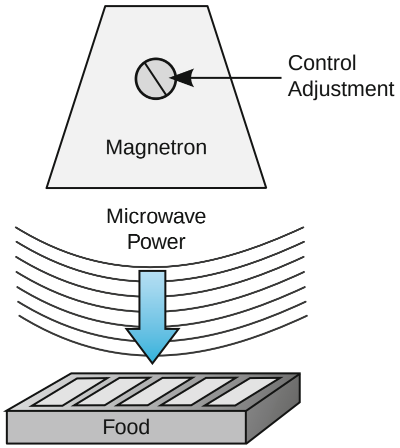
Partial Answer:

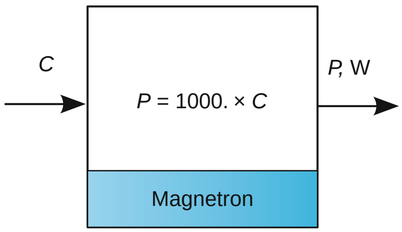


	B	C	D	E
17		Throttle gain	=0.001	
18		Engine gain	=1000	
19				
20	***t* s**	***C* (0 or 1)**	***AF* kg/s**	***P* kW**
21	0	0	=$D$17*C21	=$D$18*D21
22	0	1	=$D$17*C22	=$D$18*D22
23	=1+B22	1	=$D$17*C23	=$D$18*D23
24	=1+B23	1	=$D$17*C24	=$D$18*D24
25	=1+B24	1	=$D$17*C25	=$D$18*D25
26	=1+B25	1	=$D$17*C26	=$D$18*D26



8. The conceptual diagram and block diagram for open loop control of a 1000. W microwave oven are shown below. (The "magnetron" is the device that delivers the microwaves into the oven's cavity.) Assume the control is turned on to value $C = 1$ at time $t = 0$ and left on thereafter. Prepare a spreadsheet and graph showing the power delivered to the food each second for a total of 6 seconds.

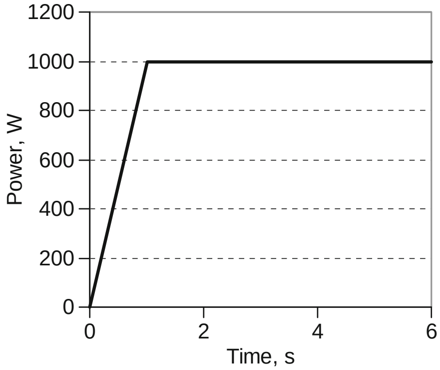




Answer:

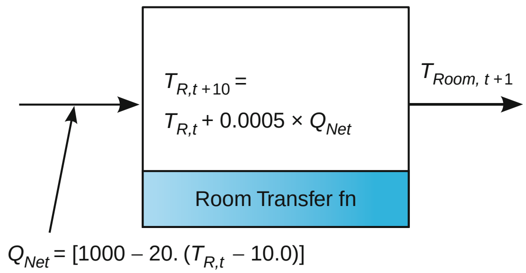
	A	B	C	D	E
16	Magnetron gain		*t*,*s*	*C* (0,1)	*P*, W
17	1000		0	0	0
18			1	1	1000
19			2	1	1000
20			3	1	1000
21			4	1	1000
22			5	1	1000
23			6	1	1000

	A	B	C	D	E
16	Magnetron gain		*t*,*s*	*C* (0,1)	*P*, W
17	1000		0	0	=D17*$A$17
18			=C17+1	1	=D18*$A$17
19			=C18+1	1	=D19*$A$17
20			=C19+1	1	=D20*$A$17
21			=C20+1	1	=D21*$A$17
22			=C21+1	1	=D22*$A$17
23			=C22+1	1	=D23*$A$17



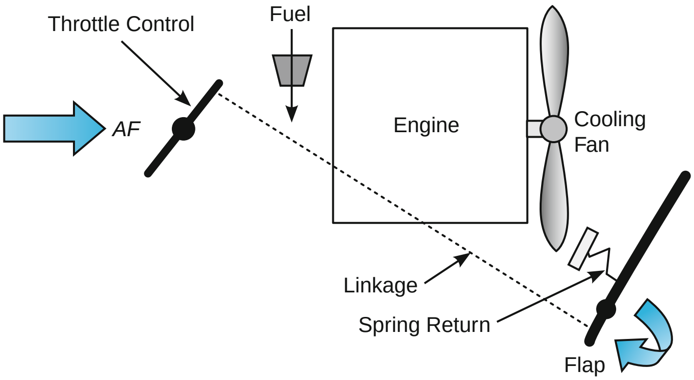
Note: The rate of change of power in the first second is not realistic—its rise time is not defined in this simple model.

9. Suppose a simplified thermostatic room controller operates a furnace that can deliver heat to a room at the rate of 1000. watts. The initial temperature of the room is $T_R = 20.0°C$ and the outside temperature is $T_0 = 10.0°C$.



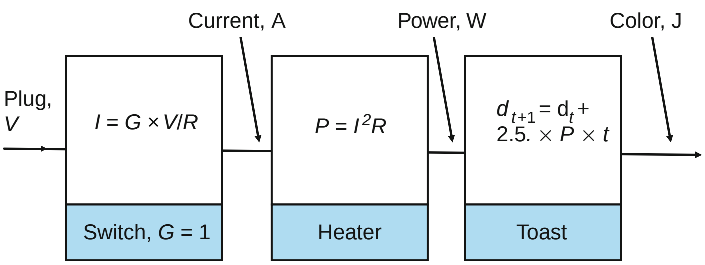
For this example assume the room will lose some heat over a relatively long period when you are tracking the temperature. To simplify, assume the furnace delivers the net heat, which is the difference between what the furnace produces and the heat lost from the room, $Q_{Net} = Q - Q_L$. Draw a single block of a block diagram modeling the surroundings of the heating system under the assumptions that the room gains heat at the rate of $Q_{Net} = Q - Q_L = (1000. - Q_L)$ where $Q_L = 20. \times (T_R - T_0)$ watts, and that the temperature of the room at time $t + 10$ minutes is the temperature at time t minutes plus $0.0005 \times Q_{Net}$. Give the room temperature at intervals of 10 minutes up to 100 minutes. (**Partial Answer: At 100. minutes, $T_R = 23.8°C$**)

10. The conceptual sketch below shows a model for a closed loop control of a lawnmower engine. Unlike the lawn mower motor of exercise 7, there is a feedback device. It takes advantage of a cooling fan that is connected directly to the output shaft of the lawn mower. This fan blows on a flap indicated in the diagram, which is preset to a desired position and kept in that position against a spring tension if the fan is rotating at a desired speed. The desired speed is set by initializing the spring with a tension S_{SP} corresponding to a desired operating condition such as 3,000 RPM.

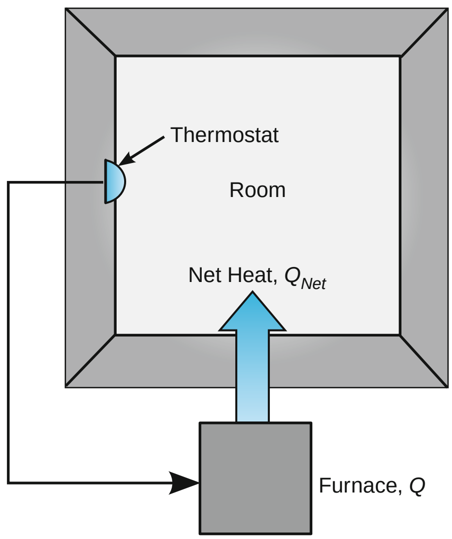


If the fan rotates too fast (i.e., the engine speed is too high), the flap is blown clockwise, causing a mechanical linkage to close the throttle and reduce engine speed. If the speed is too low, the flap is pulled counterclockwise by the spring. Draw a block diagram for this control system, assuming: (1) that the control unit sends an error signal in terms of the spring tension (S_{SP} (set point) – S) to throttle a proportional amount of air/fuel $G_{Th}(S_{SP} - S)$ kg/s; (2) resulting in engine power $P = G_E \times AF$; and (3) the speed response function for the motor speed is $N = G_{SR} \times P$. Finally the controller converts engine speed to spring tension by the relationship $S = G_{ST}N$.

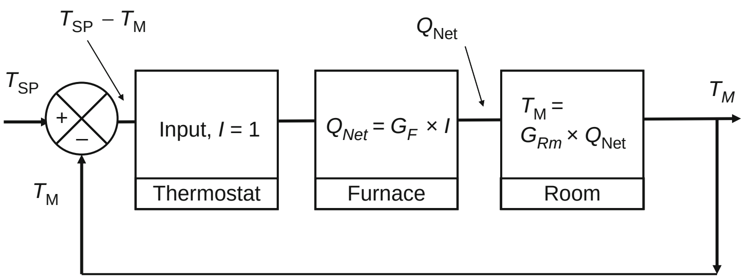
11. Suppose that for the open loop toaster control you sketched in exercise 5, the variable that you really want to control is d (for darkness), where d is the energy absorbed by the toast, in units of joules. Assume the initial energy of the toast (in units of joules) is $d = 0$, and $d_{t+1} = d_t + K \times P \times t$, where P is the power in watts delivered by the heating coil at time t and K is the proportional gain relating d to P. (a) Fill in the toast block to correspond to this model. (b) If the line voltage is 115 V and the heater resistor is 13.2 ohms, and the darkness gain is 2.5, plot d in joules for 10. s. (**Partial Answer: At 10. seconds, darkness function for toast = 2.5 × 10^4 J**)



12. The conceptual sketch below shows a model of a *closed* loop control for heat in a room. Prepare a block diagram for this closed loop control. Label each block on the block diagram with the name of the corresponding component or components on the sketch. To simplify, assume the furnace delivers the net heat being the difference between what the furnace produces and the heat lost from the room, $Q_{Net} = Q - Q_L$. The thermostat operates by comparing T_{SP} (the thermostat's set point) to T_M (the measured room temperature) and sending that signal to the furnace. The furnace is "on" if that difference is positive, and vice versa.



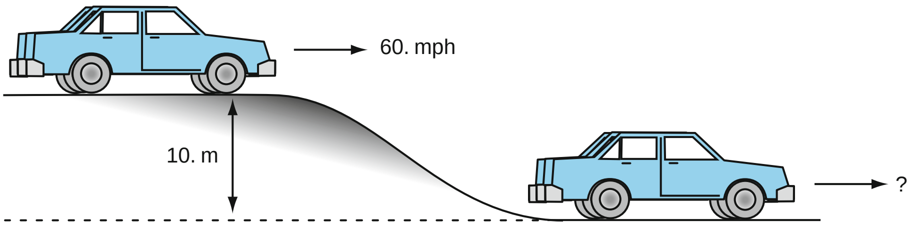
Answer:



13. Use the block diagram developed in the previous exercise. Assume the room is initially at $T_0 = 15°C$ and the set point for the furnace is 22°C. The furnace heat rate is $Q = 1000.$ W and the room losses are $Q_L = 20. \times (T_M - T_0)$. The response function for the room is $0.003 \times Q_{Net}$°C per minute when the furnace is on and $0.003 \times Q_L$°C per minute when it is off. Plot the temperature response of the room for 10 minutes. (**Hint:** Use IF statement, IF (test, value if true, value if false) to indicate when the furnace is on and when it is off.)

14. Suppose a car employs a digital speed sensor as shown in Figure 10.15, and the car's wheels are each 0.60 meters in diameter. If the speed sensor consists of the magnet and coil and the magnet induces one pulse in the coil for every revolution of the axle, how many pulses per second will the coil send to the pulse counter if the car is traveling at 65. miles per hour? (**A: 15. pulses/s if there is one pulse per revolution**)

15. Do you really need a cruise control? Suppose you are driving in the Midwest on a flat road. You are bored with the monotony of the terrain and you set a brick against the accelerator pedal that maintains the vehicle's speed to 60. mph. Unexpectedly, you drive onto a construction road that is 10. m deep. Ignoring friction and wind losses, will your vehicle still be at 60. mph at the bottom of the construction road, and, if not, what will its speed *S* be? (**A: 68 mph**)



16. In this chapter, we have discussed but not derived, an equation for the transient inertial limited response to engine power and to windage losses. The equation is

$$S_{t+1} \cong \sqrt{S_t^2 + \frac{2(\Delta t)(P(engine) - P(losses))}{m}}$$

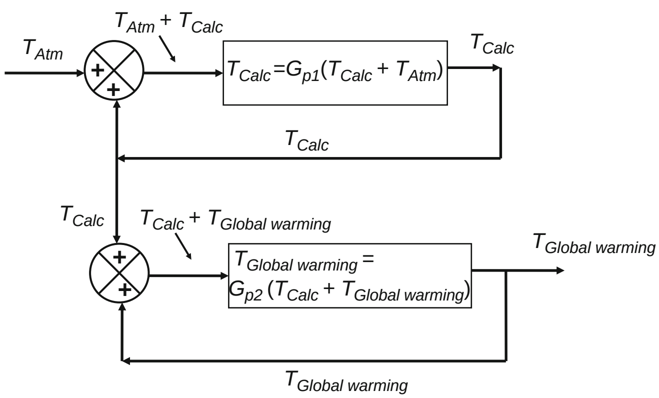
in which $P(engine)$ and $P(losses)$, respectively, are the engine's power output and the air resistance losses. The vehicle has mass m. Starting with Newton's Law of motion, derive this equation in which the vehicle has sped up to S_{t+1} from S_t in a time interval of Δt.

17. Referring to Figure 10.12, when a throttle plate rotates in a circular plenum, its effective blocking area is πab where a, b are the principal axes of an ellipse. Show that

$$\frac{\Delta A}{A} = 1 - \cos\theta$$

where A is the fully open area of the pipe and ΔA is the open area when the plate has been turned through an angle θ.

18. Some skeptics have suggested that we are not entering a man-made period of global warming but instead are entering a mini ice age (as Earth has done many times before in its 4+ billion year history). Consider this simple block model of the conflicting possibilities:



In this model the calculated warm atmospheric temperatures are subject to a cooling feedback, the result of which is the predicted atmospheric temperature of Earth. Show that

$$T_{Global\ warming} = \frac{G_{p1}G_{p2}T_{Atm}}{(1 - G_{p1})(1 - G_{p2})}$$

in which G_{p1} is the gain for global warming and G_{p2} is the gain for global cooling. If K is defined as G_{p2}/G_{p1}, plot the Earth's predicted atmospheric temperature from the years 2000 to 2050 for (a) $K = 1.001$, (b) $K = 1.000$, and (c) $K = 0.999$. Assume $G_{p1} = 0.50012$ as in Example 10.6.

19. You work for the control and guidance systems division of a major aircraft manufacturer with a major government contract. You observe employees who regularly leave work early while being paid for time not worked. What do you do?

a. Ignore it, since there is nothing you can do anyway.
b. Talk to your supervisor about it.
c. Report it to the government representative on the contract.
d. Blow the whistle and talk to a newspaper reporter.

Use the Engineering Matrix in your solution.

20. You are a new engineer working for a motorcycle manufacturer that produces a bike with a known control system instability. This instability can cause the rider to lose control at high speed and crash. Your supervisor says that you should ignore it because everyone knows about it and it would be too expensive to fix. Besides, a new control algorithm would cause more harm, since the drivers expect the bikes to behave in a certain way. What do you do?

a. Attempt to convince your supervisor that it will be cheaper to fix the flaw than pay the subsequent law suits.
b. Suggest that a warning label be put on the bikes about riding at high speeds.
c. Talk to your corporation's legal office about your professional obligations.
d. Blow the whistle and talk to a newspaper reporter.

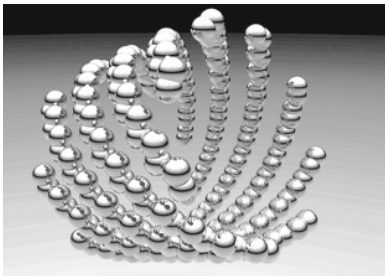
Use the Engineering Matrix in your solution.

This page intentionally left blank

CHAPTER

Materials Engineering

11



Source: © iStockphoto.com/Alwyn Cooper

Engineers select materials. Should a refrigerator case be made of steel or plastic? Should an armor plate be a single sheet of steel or a lighter layered composite alloy? Should a transistor be made out of germanium or silicon? Should a telephone cable be made of copper or fiber optic glass? Should an artificial hip joint be made out of metal and, if metal, should it be titanium or stainless steel, or would a polymer composite be better? These materials selection problems are further examples of constrained optimization. An engineer finds the solution that best meets given criteria while also satisfying a set of constraints.

11.1 CHOOSING THE RIGHT MATERIAL

The criterion to be optimized when choosing the right material for an engineering application might be cost, weight, or performance, or some index such as minimizing weight × cost with a requirement of a particular strength. The constraints (also called design requirements) typically involve such words as **elastic modulus** (also called **stiffness**), **elastic limit**, **yield strength** (sometimes abbreviated as plain "strength"), and **toughness**. Although we all have loose ideas of what is meant by these terms, it is necessary to precisely express them as engineering variables. Those variables, in turn, must contain the appropriate numbers and correct units.

To help develop appropriate variables, numbers, and units, this chapter will introduce a new tool: the **stress–strain diagram**. It is a tool for defining elasticity, strength, and toughness as engineering variables, as well as a tool to extract the numerical values of those variables in materials selection.

Working through the examples and problems in this chapter will enable you to:

1. Define **material requirements**.
2. Relate to two important classes of materials, **metals** and **polymers**.
3. Understand the internal microstructure of materials that can be **crystalline** and/or **amorphous**.
4. Use a stress–strain diagram to express materials properties in terms of the five engineering variables: stress, strain, elastic limit, yield strength, and toughness.
5. Use the results determined for those properties to carry out materials selection.

Humankind's reliance on the properties of materials for various applications, from weapons to shelter, has been around since the dawn of the ascendancy of our species.[1]

[1]Ashby, M. F. Technology of the 1990s: Advanced Materials and Predictive Design. Phil. Trans. R. Soc. London (1987), A322, 393 (1987).

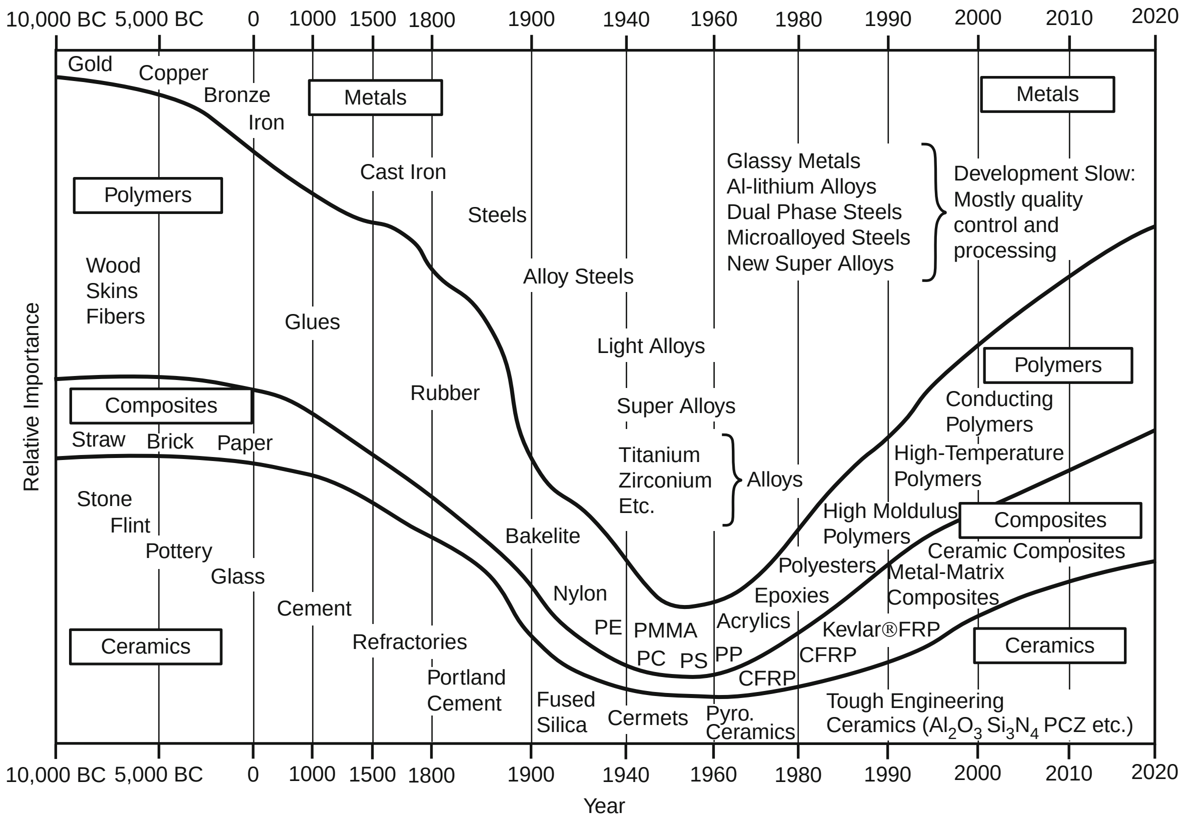


FIGURE 11.1 Materials Since the Ascension of Homo Sapiens (Reprinted by Permission of the Royal Society—London and Professor Ashby)

Figure 11.1 illustrates how various materials have evolved over time. The earliest were naturally occurring elements[2] and compounds, followed by materials discovered by what today is called the "Edisonian"[3] method (i.e., the method of trial and error), and then to today's modern materials designed by systematic investigations.

Our interest here is to discover the principles behind new materials. Notice how the modern materials are classified by type: metals, polymers, composites, and ceramics. For the sake of brevity, we will confine ourselves to just the first two of these classes: metals and polymers.

Presumably you already have a picture of what is meant by a *metal*. Generally, metals are strong and dense. Metals also reflect light, and conduct heat and electricity. On the other hand, polymers, popularly known as plastics, are generally weak, sometimes opaque and sometimes transparent, and generally do not conduct heat or electricity. We will shun the term *plastic* in favor of *polymer* in describing these materials, since, as we shall see, the word *plastic* is reserved to describe a particular behavior in metals.

[2]Most metals were discovered in ascending order of their melting points.
[3]Another homage to that inventive giant of the nineteenth century, Thomas Edison.

11.2 STRENGTH

What are the reasons for a material's strength? Material properties are directly related to their molecular properties and hence to the properties of their constituent atoms. The only fundamental illustration that we will calculate is the breaking strength of a material, such as a piece of pure iron, as determined by its molecular structure. This calculation thus represents the upper limit to its strength. All failures at lower strength values are due to some defect in the material or its structure.

We will make a quick review of material structure such as whether a material is **amorphous** (meaning no structure observable at the micro level—10^{-6} meters and smaller) or whether it is **crystalline** (which means it consists of definite microstructures that can be seen under a low power microscope or even by the naked eye).

Suppose you clamped the ends of a piece of pure iron and pulled it as shown in Figure 11.2 until it eventually breaks.

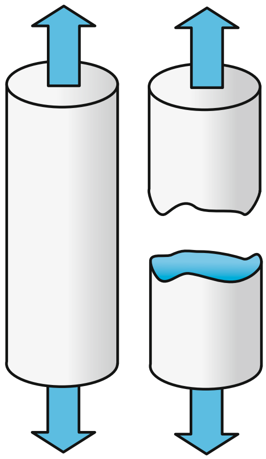


FIGURE 11.2 Failure Under Tension

A straight pull is called **tension**. If the nominal cross-sectional area of the break is known, and the number of atoms in that plane is also known, we can, in principle, calculate the force to separate the atoms at the break zone and also the work required (i.e., force × distance) to break it.

In a crystal of iron the geometry is particularly simple. A **crystal** is a representative repeat pattern of the atoms that make up the structure. Figure 11.3 shows the arrangement of the atoms in a perfect crystal of iron. If the crystals are large enough (i.e., there are trillions and trillions of atoms all arranged in the same pattern), these crystals can often be seen with low-power optical microscopes, and their atomic structure can be deduced using X-rays.

The crystal structure will arbitrarily extend in every direction in this idealized model (as in a single crystal). At the nominal fracture plane shown in Figure 11.3, this works out to contain 1.8×10^{19} atoms per m^2. All we need to know now is the strength of each bond and we will have a model of the theoretical strength of this material. We can estimate this quite easily by doing a simple thought experiment: Imagine you started with a lump of iron and then you heated it until it evaporated (just like boiling water to steam). The total energy absorbed represents the energy to break all the bonds in the solid iron and make individual atoms

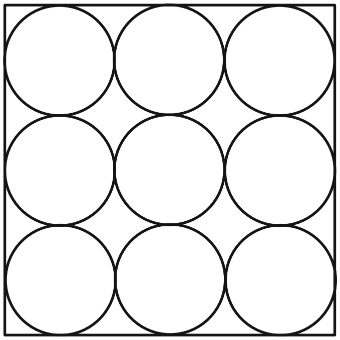


FIGURE 11.3 Structure of an Iron Crystal (the Centers of the Atoms are 0.234×10^{-9} Meters Apart)

detach from the solid, and which we will use as a crude measure of the bond strength of all those atoms in the original piece of iron. This works out to be about 6.6×10^{-19} J/atom of iron. Therefore, the *work* per unit area (or we could equally talk in terms of energy per unit area) required to fracture a piece of iron is calculated by multiplying by the atom density (atoms of iron/m^2) to obtain about 12 J/m^2. This sounds quite modest.

Calculating the force required to break the material is another matter. Recall that work is force × distance. So, how far do the atomic planes in iron need to be pulled apart to consider them fully separated? Atomic forces fall off very rapidly with distance and a reasonable estimate of the minimum distance required to produce fracture is 0.1 – 0.2 nm *beyond* the equilibrium center to center separation of the individual iron atoms (0.234 nm), as shown in Figure 11.4 A, B. At that additional separation, the individual iron atoms will no longer interact, and fracture has occurred.

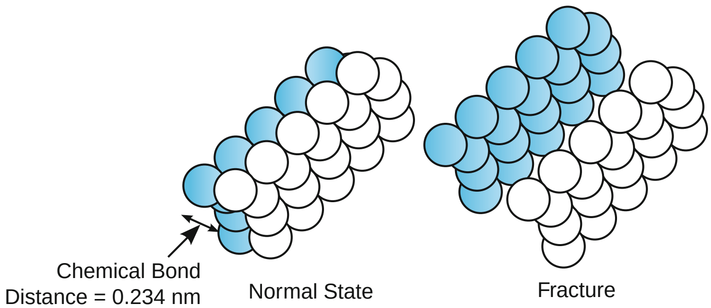


FIGURE 11.4 Before and After View of Fracture Along an Atomic Plane in Iron

The origin of the attractive force in metals is found in their atomic structure; this structure promotes their ability to conduct heat and electricity. Metals have free (negatively charged) electrons inhabiting the spaces between (positively charged) metal ions (see Figure 11.5). This electronic structure provides strong forces between the iron atoms and also explains why it is easy to get electrons to flow in metal wires.

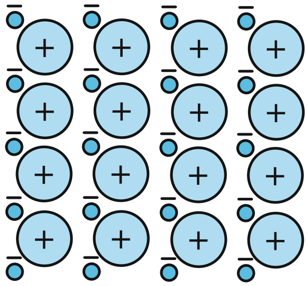


FIGURE 11.5 Metallic Bonds

We will equate the work done in separating the atoms beyond their equilibrium positions to a force × an *assumed* distance of 0.15 nm of separation. Using the 12 J/m^2 fracture work per unit area just calculated,

$$\text{Work/Area}\ (12\ \text{J/m}^2) = \text{Force/Area}\ (\text{N/m}^2) \times \text{Distance}\ (0.15 \times 10^{-9}\text{m})$$

Therefore,

$$\text{Force/Area} = 8.0 \times 10^{10} \approx 100 \times 10^9\ \text{N/m}^2$$

The units N/m^2 have the name **Pascal** (abbreviated as **Pa**), but generally we are interested in large numbers and therefore use the *giga* prefix of 10^9. Our answer is thus 100 GPa (G being the symbol for *giga*). This force per unit area is equivalent to piling 100 billion apples on top of a 1 m × 1 m tray[4] here on Earth, and, it is a much larger force/area than observed in practice.

Why? Real crystals are finite in extent and are usually randomly oriented. Figure 11.6 shows a polished piece of copper as viewed under a microscope. Individual random crystals of copper are quite obvious.

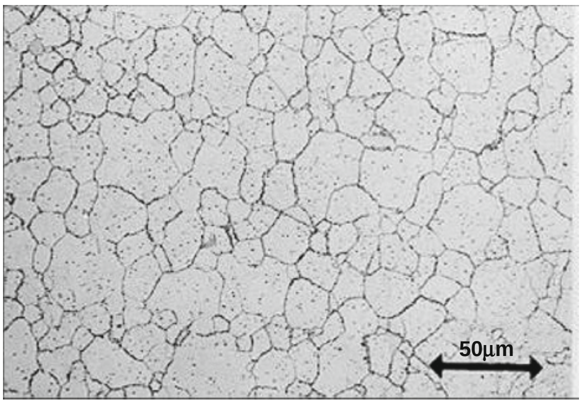


FIGURE 11.6 Polished Copper Showing Individual Crystals (Courtesy of the Copper Development Association, Inc.)

[4]The pressure of one atmosphere is 100,000 Pa, so we have to apply a million atmospheres of tension to fracture the metallic bond.

Fracture may occur between crystals at their relatively weak **grain boundaries** rather than within the crystals, resulting in lower fracture forces than the previous mathematical model.

In addition, not all materials are crystalline. Some have a completely random structure, called **amorphous**. Many materials may also consist of mixed phases such as crystals in a matrix of amorphous material.

The other class of materials we will deal with is **polymers**. Polymers are repeating chains of small molecular assemblage, and the word *polymer* basically means "many molecular pieces." Generally, polymers are made of organic chemical links.[5] For example, the common polymer polyethylene is made from long chains of ethylene[6] molecules written as $-[CH_2-CH_2]-$ strung together, each dash representing the net attraction of electrons between two adjacent carbon atomic nuclei (see Figure 11.7). Each carbon atom also has two off-axis hydrogen atoms and one on-axis C—C bond. Roughly speaking, the strength of a C—C bond is about 10 percent that of an iron-to-iron bond, so you might think that our model apparently implies a force/area of 10 GPa as the upper limit to break a typical polymer bond. However, this is not true.[7]

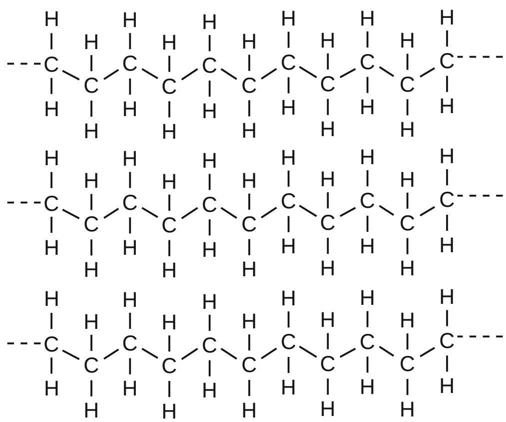


FIGURE 11.7 Regular Array of Polyethylene Molecules

What you need is not the force between individual carbon atoms but that between adjacent polymer chains. Although there are many kinds of possible molecular interactions, for the case of polyethylene, the kinds of forces are lumped under the name van der Waals forces. They are quite weak, about 0.1 percent of the strength of the metal bonds, ~0.1 GPa. The upper limit for the strength of a polymeric material therefore might be of this magnitude.

[5]Meaning based on carbon, hydrogen, and a few other atoms such as nitrogen, oxygen, and sulfur.

[6]The chemical formula of ethylene gas is $CH_2 = CH_2$ with the "=" standing for a "double bond" between the carbon atoms. If this double bond is broken and rejoined to an adjacent carbon, it makes the repeating polyethylene structure found in that solid material. (Milk bottles are commonly made of this material.)

[7]Regular arrays of a polymer also are called crystals. They can coexist with a matrix of an amorphous polymer, in good analogy to the structure of metals. Crystals in polymers are usually the right size to scatter light efficiently, so you know that a white, opaque polymer such as Teflon™ is crystalline, and a clear one such as polyethylene terephthalate or PET (used in soft drink bottles) is amorphous.

11.3 DEFINING MATERIALS REQUIREMENTS

Consider the choice of material for a car's bumper. Defining material requirements for a bumper begins by understanding what a bumper is and what it does. A bumper is a structure often integrated into the main chassis; typically it is 1 to 2 meters wide, 0.1 to 0.2 m high, and 0.02 to 0.04 m thick and is attached to the front or rear of the car about half a meter above the ground.

Figure 11.8 shows a "conceptual bumper" for which a material will be selected. A bumper does two main jobs:

1. It survives undamaged in a very low-speed (1–2 m/s) collision.[8]
2. It affords modest protection for the car, though sustaining major damage to itself, in a moderate-speed (3–5 m/s) collision.

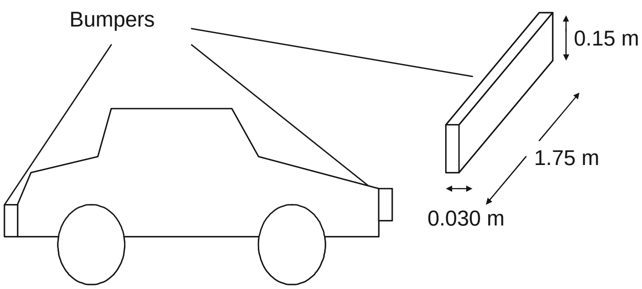


FIGURE 11.8 Conceptual Bumpers

What does an engineer mean by terms such as *collision*, *survive undamaged*, and *protection for the car*? Translation of such language into engineering variables, units, and numbers begins with the concept of energy conversion.

A collision is a rapid process of energy conversion. A moving car is carrying TKE. A collision rapidly converts that TKE into **elastic** and **plastic** energy[9] of materials. Elastic energy is a form of stored potential energy, and plastic energy is converted, via internal collisions of the atoms in the material, into heat. In a properly designed bumper, these two processes of energy conversion will occur in the bumper—not in the car itself and certainly not in the passengers.

In a really low-speed collision, the energy conversion should leave the bumper looking as it did before the collision. In a moderate-speed collision, the bumper may be damaged or destroyed, but it should absorb and release the energy in a way that leaves the car and the passengers undamaged. In a high-speed collision,[10] other methods of protecting the passengers are needed, as discussed in a later chapter.

A wide range of bumper designs can achieve these energy conversion objectives. To select a design from these possibilities, an engineer first picks dimensions for a bumper. For simplicity, consider a bumper to be a

[8]Federal regulations call for protection against damage to the vehicle in a 2.5 mph (1.12 m/s) collision; even though this might seem to be a modest requirement, remember that modern cars are heavy!

[9]We'll properly define these terms later, but for the moment realize that *plastic* in this context is not a word meaning polymers, and that *elastic* means to spring back like an elastic band.

[10]The Insurance Institute for Highway Safety tests passenger cars up to 35 mph (15.6 m/s); see http://www.iihs.org/default.htm

rectangular solid object 1.75 × 0.15 × 0.030 meters (width, height, and thickness, respectively). For a given level of protection, the engineer will then seek the lightest material that will enable a bumper of these dimensions to meet the design requirements.

Why does the engineer seek the lightest material? In the days of metal bumpers, a bumper provided a significant part (about 5%) of the mass of the car. The higher the mass of the car, the lower the gas mileage. Conversely, anything that reduces mass will increase gas mileage. So all other things equal, a car with lighter bumpers will have the higher gas mileage than a car with heavier bumpers. Our optimization problem will be choosing the material that will give the lightest bumper of given dimensions that will do a bumper's job.

Solving that optimization problem begins by using the principle of energy conservation to translate the design requirements from words into variables, units, and numbers.

Example 11.1

Estimate the total energy that a 1.00×10^3 kg car has to absorb in a 2.5 mph (1.12 m/s) collision with a perfectly rigid wall. If all this is absorbed in the bumper, what is the specific energy absorbed (i.e., energy per m^3 of bumper volume) given its dimensions of 1.75 m × 0.15 m × 0.030 m?

Need: Energy absorbed ____ in J and the specific energy absorbed ____ in J/m^3.
Know: Speed of car, dimensions of bumper.
How: The conservation of energy provides the energy absorbed, and the energy absorbed divided by the volume of the bumper gives the specific energy absorbed.
Solve: The bumper must absorb the total TKE of the vehicle from a 1.12 m/s collision. The vehicle's TKE is ½ (mv^2) = ½ $\times 1.00 \times 10^3 \times 1.12^2$ [kg][m/s]^2 = $\mathbf{6.3 \times 10^2}$ **J**.

The bumper volume = 1.75 × 0.15 × 0.030 = 0.0079 m^3. Therefore, the **specific energy** absorbed by bumper = energy absorbed/volume of bumper = $6.3 \times 10^2 \times 10^{-6}/0.0079$ [J][1/m^3][MJ/J] = **0.080 MJ/m^3**.

Notice if we wanted 5.0 mph protection, the total energy absorbed goes to 2,500 J and the specific energy absorbed goes to 0.32 MJ/m^3 and at 35 mph (15.6 m/s) the corresponding numbers are 0.12 MJ and 1.6×10^2 MJ/m^3.

Is there a candidate bumper material that can absorb these amounts of energy elastically and rebound to its original dimensions? Or will there be sustained damage to the bumper, or worse? For us, the choice of a bumper material boils down to a choice between a polymer and a metal. A polymer offers light weight at some sacrifice of strength. A metal offers high strength but with a penalty of higher weight. Which should an engineer choose? Answering these questions begins by considering the types of materials that Nature and human ingenuity offer. Making that choice requires applying a new tool: the **stress–strain diagram**.

The stress–strain diagram is a tool for using measurements of two new variables, **stress** and **strain**, to quantify the terms **modulus of elasticity, elastic limit, plasticity, yield strength**, and **toughness**. Two basic experiments[11] are shown in Figure 11.9. Materials are either stretched under **tension**, as in Figure 11.9A, or put into **compression**, as in Figure 11.9B.

Add more weight, Mg, in either of the situations and the wire stretches or the plate compresses accordingly. In this way we can plot a diagram of the stretch in L or compression in T versus the applied load. The response is considered **elastic** if the material returns to its initial dimensions when the load is removed.

[11]A third mechanical test is to twist the material and measure its torsional material properties.

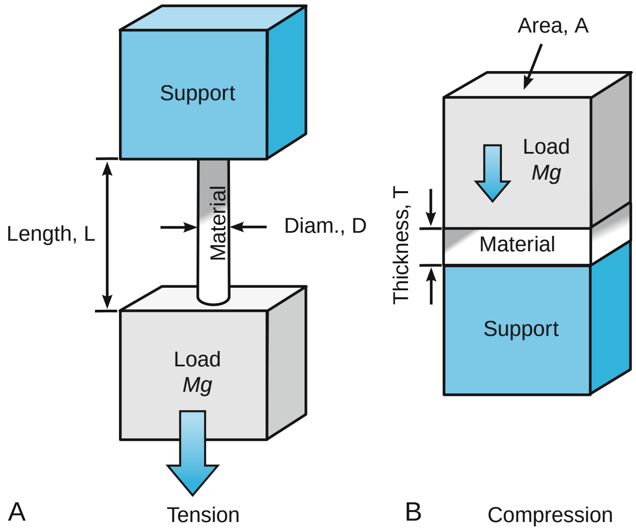


FIGURE 11.9A, B Basic Material Tests

Figure 11.10 is a composite plot of four elastic experiments. In tension, the long rod (or wire, etc.) shown in curve **b** stretched in proportion to its load twice as much as did the shorter one shown in curve **a**. The compression shown in curve **d** is half that shown in curve **c**, and is also proportional to its load. This is an expression of **Hooke's Law**.[12]

$$(L - L_0) \propto L_0 \frac{M_g}{A} \tag{11.1}$$

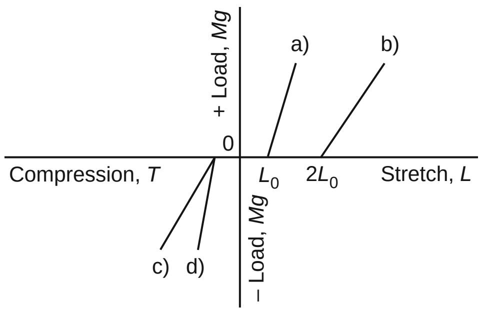


FIGURE 11.10 Elastic Response of a Material to Tension and Compression: a) Tension, Initial Length L_0, Diameter D; b) Tension, Initial Length $2L_0$, Diameter D; c) Compression, Thickness T, Load Area A; d) Compression, Thickness T, Load Area 2A

[12]Robert Hooke was a talented rival of Isaac Newton.

Notice that we have used a convention that tension is a positive load, stretching is a positive response to it, and the opposite is true for compression and the contraction response to it.

But why do we have to plot these data as clumsily as in Figure 11.10? Equation (11.1) is our clue that we can be more inclusive as well as more general. We write it as:

$$\frac{Mg}{A} \propto \frac{(L - L_0)}{L_0} \quad \text{or} \quad \frac{Mg}{A} = E\frac{(L - L_0)}{L_0} \text{ (thus defining } E) \tag{11.2}$$

The fractional (or for the percentage, × 100) stretching or compression—that is, the change in length divided by the original length, $(L - L_0)/L_0$, is called the **strain**. If the material stretches in tension, the strain is positive, and if the material is compressed, the strain is negative. We use the Greek letter epsilon ($\boldsymbol{\varepsilon}$) to denote strain. Note that strain is a dimensionless number, since it is the quotient of two quantities with the same dimension, in SI units of meters/meters but just as well in feet/feet.

When the force causing the strain is divided by the applicable cross-sectional area we get the **stress** in the material, denoted by the Greek letter sigma ($\boldsymbol{\sigma}$). Again, a tensile stress is positive, and a compressing stress is negative. Stress has SI units of N/m^2 (or Pa) and English units of lbf/in^2 (or psi). The constant $\boldsymbol{E}$ in equation (11.2) is called **Young's modulus**[13] or the **Elastic modulus**. Hooke's Law can now be written as:

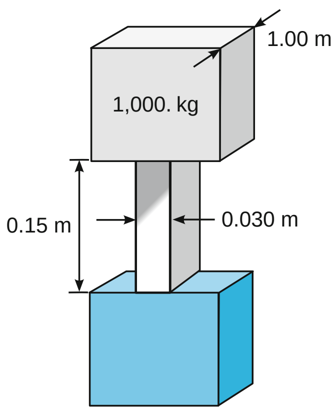
$$\boldsymbol{\sigma} = \boldsymbol{E\varepsilon} \tag{11.3}$$

Equation (11.3) is the way that engineers use Hooke's Law, and it should be committed to memory in this form once it is understood.

Example 11.2

When a mass of 1,000. kg of steel is carefully balanced on another piece of steel of depth 1.00 m, height 0.15 m, and thickness 0.030 m, and sits on a rigid base, the length of the middle piece contracts by 0.000010 m.

a. Is the middle piece of steel in tension or compression?
b. What is the stress in the middle piece of steel?
c. What is the strain in the vertical direction of the middle piece?



[13]After the early nineteenth century scientist and scholar Thomas Young, who first suggested this approach to understanding materials.

Need: Stress ____ N/m^2, strain ____ fraction.
Know: Mass of the object compressing the piece = 1,000. kg. Initial thickness of sample is 0.15 m and it suffers a contraction of 1.0×10^{-5} m when the load is applied.
How: By definition, stress = force/area, and force = weight = *Mg*; the supporting cross-sectional area is $1.00 \times 0.030\ m^2 = 0.030\ m^2$. By definition, strain = (change in length)/(initial/length).
Solve:

a. The middle piece is in compression.
b. **Stress**, $\boldsymbol{\sigma}$ = weight/area = $-(1000. \times 9.81)/0.030$ [kg][m/s^2][1/m^2] = -3.27×10^5 N/m^2 = $-\mathbf{3.27 \times 10^5}$ **Pa**. The negative sign indicates compression.
c. By definition, **strain** is the change in length divided by the initial/length or $\varepsilon = -1.0 \times 10^{-5}/0.15$ [m]/[m] = $-\mathbf{6.7 \times 10^{-5}}$ (a pure number; the − sign indicates compression).

Suppose we made a large number of measurements like those described in Example 11.2 but with differing loads. Suppose we then plotted the results on a graph, with strain as the horizontal axis and stress as the vertical axis. For each axis, the negative direction would be compression, and the positive direction would be tension. The resulting graph is called a **stress–strain diagram**. It captures in one diagram three of the most important properties of a material: elasticity, strength, and toughness.

The stress–strain diagram of an idealized steel[14] looks like Figure 11.11. It also defines some new terms: **yield stress** (or **yield strength**), **yield strain**, and **plastic deformation**.

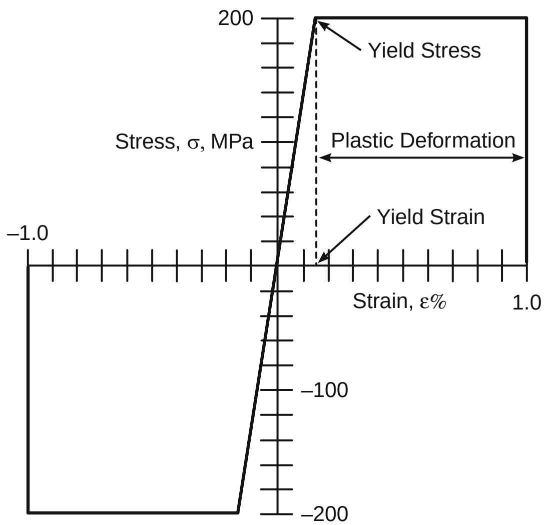


FIGURE 11.11 Simplified Stress–Strain Curves for an Idealized Steel

[14] Basically iron + some additives + some processing.

In the first instance, we have reduced the number of curves we need when we change the geometry of the material, or the applied load, to just two: one curve for tension and one for compression. In doing so, we have taken advantage of Young's equation and collapsed the data into a more general and convenient form. It will no longer matter in this format if we change the sample's size or if we change the load on the sample by increasing the load mass or by reducing its cross-sectional area, and we don't need to worry about the absolute strains changing with the sample length. All the separate curves will collapse to a single curve.

The **yield stress** is the maximum stress at the limit of elastic behavior in which the specimen returns to its initial length when the load is removed. In the case shown in Figure 11.11, it is 200. MPa. The **yield strain** is the strain corresponding to the yield stress. Plastic deformation causes a *permanent* change in the properties of the steel when it passes beyond its elastic limits. Stretching here is called **plastic deformation**. For this idealized steel compressive properties exactly mirror the tensile ones.

Compare the yield stress in Figure 11.11 of 200 MPa to the theoretical failure value for iron calculated earlier at 100 GPa. In other words, our sample of steel is beginning to fail at a stress about 200 MPa or a factor of about 500 lower than its theoretical maximum. Presumably the failure is initiated at inclusions or the crystalline boundaries rather than by pulling apart the rows of atoms of iron in the steel.

The elastic modulus E defined in Equation (11.3) is just the slope of the stress–strain curve. In Figure 11.11 it is $E = 200$ MPa/0.0015 $= 130$ GPa.

Another variable of interest is **toughness**. In ordinary speech, people sometimes confuse it with (yield) stress or strength. To an engineer, however, strength and toughness have very different meanings. A material can be strong without being especially tough. One example is a diamond. It can be subjected to great force and yet return to its original shape, so it is therefore very strong. Yet, a diamond can be relatively easily shattered and therefore is not very tough. A material can also be tough without being strong. An example is polycarbonate, a kind of plastic used in football helmets. It is relatively easily dented and therefore not very strong. Yet, it absorbs a great deal of energy per unit mass without shattering and is therefore very tough.

In this introductory textbook, the stress–strain curve for strains greater than the yield strain will be a horizontal line parallel to the strain axis. In actual stress–strain diagrams, the shape of that portion of the curve is more complicated. This region is called the **plastic** region, and when it is described with a horizontal line it is called a **perfectly plastic** region. This overall model is called **perfectly elastic, perfectly plastic**. It's a useful simplification of how real materials behave.

The horizontal line in the perfectly plastic region extends out to a **maximum strain**. At that strain the material fails. The total strain at failure is as the sum of the elastic strain and the plastic strain. The important point is that a wire, stretched beyond its maximum strain, breaks. A plate compressed to its maximum strain shatters or spreads. The **toughness** of a material is defined as *the area under the portion of the stress–strain curve that extends from the origin to the point of maximum strain.*[15]

Because this area (shaded in Figure 11.12) is the result of stress, measured in N/m^2, and of strain, measured in m/m, toughness has the units of $N \cdot m/m^3$, or J/m^3 (that is, energy per unit volume). It usually is measured in a dynamic experiment involving a strike by a heavy hammer. Toughness is therefore the preferential criterion to be used to predict material failure if a sudden load is applied.

[15]Actually there are several different definitions of *toughness* used by different authors, in particular the ability of a material to withstand sudden impacts. Under certain circumstances this may, in fact, measure the same thing as our definition.

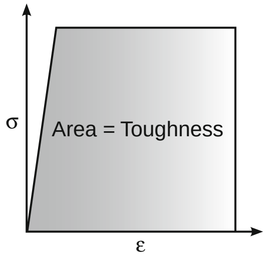


FIGURE 11.12 Toughness

Toughness, defined as the work done until failure, represents the ability of the material to absorb energy. It is a prime variable for a large subset of materials research.

Example 11.3

Given the stress-strain diagram for steel in Figure 11.11, determine (1) its modulus under compression, (2) its yield strength under tension, (3) its toughness under tension to yield, and (4) its toughness under tension to 1.0% strain. Use appropriate SI units.

Need: E = ____GPa under compression, σ = ____ MPa at yield, and toughness at yield and 1.0% strain = ____ MJ/m^3.

Know: Stress–strain curve in Figure 11.11; toughness is area beneath the stress–strain curve.

How: Hooke's law $\sigma = E\varepsilon$, and the stress–strain curves.

Solve:

1. Need slope in compression region for σ/ε curve.
   E = -2.0×10^2 [MPa]/–0.0015 [m/m] = **130 GPa** (to graph-reading accuracy)
2. Under tension, the **yield stress** is $\boldsymbol{\sigma} = \mathbf{2.0 \times 10^2}$ **MPa** (to graph-reading accuracy).
3. Toughness under tension to the yield stress is equal to the triangular area, ½ × 0.0015 [m/m] × 2.0×10^2 [MN/m^2] = **0.15 MJ/m^3**.
4. **At 1% strain**, add rectangular area: (0.01 – 0.0015) × 200. = 1.7 MJ/m^3.
   Therefore, the **toughness** = 0.15 + 1.7 = **1.9 MJ/m^3**.

The no-damage elastic portion of the curve is only a small fraction of the total protection afforded the car's structure. You can also see that plastic deformation, which will manifest itself in permanent deformation or crushing,[16] affords a much better sink to dissipate energy than does the purely elastic behavior. Of course, whether a particular bumper design can distribute this energy deposition uniformly and avoid exaggerated *local* deformation is up to the skill of the engineer.

We're now in a position to see if a steel bumper satisfies our first constraint. What does the stress–strain diagram tell us about its ability to absorb the energy of a collision yet return to its original shape?

[16] No doubt with an accompanying repair bill that will disappoint the vehicle's owner!

Example 11.4

Can the car's bumper from Example 11.1 absorb the TKE of the 2.5 mph collision assumed there? Or will it plastically deform? What if it is traveling at 5.0 mph? Also, what is the weight of the bumper if its density is 7,850 kg/m^3? Assume the mechanical behavior is described by the stress–strain diagram of Figure 11.11.

Need: Energy to be absorbed in bumper during collision = ____ MJ/m^3 compared to energy of collision. Also, weight of bumper = ____ N.

Know: From Example 11.1: TKE = 6.3×10^2 J at 2.5 mph; specific energy absorbed = 0.080 MJ/m^3; steel bumper volume = 0.0079 m^3; density of steel = 7,850 kg/m^3.

At 5.0 mph, TKE = 2,500 J and the specific energy absorbed is 0.32 MJ/m^3. Also from Example 11.3, the elastic toughness = 0.15 MJ/m^3 and plastic toughness is 1.8 MJ/m^3 at 1.0% strain.

How: Compare specific energy to be absorbed to the capability of the bumper.

Solve: Assume the entire surface of the bumper contacts the wall simultaneously (i.e., the force on the bumper is uniform over its entire area).

At 2.5 mph, the specific energy to be absorbed is 0.080 MJ/m^3 of steel. Steel can absorb up to 0.15 MJ/m^3, so it will elastically absorb this amount of energy—again if uniformly applied. If so, the bumper will bounce back to its original shape.

At 5.0 mph, the specific energy to be absorbed goes to 0.32 MJ/m^3 of steel. The bumper will plastically deform since 0.32 MJ/m^3 > 0.15 MJ/m^3 but will not fail since 1.9 MJ/m^3 > 0.32 MJ/m^3.

The weight of this bumper is Mg = density × volume × g = $7{,}850 \times 0.0079 \times 9.81$ [kg/m^3] [m^3][m/s^2] = 610 N, or the weight of a physically fit student!

11.4 MATERIALS SELECTION

The previous section showed how the stress–strain diagram provides the information needed to determine if a material satisfies the design requirements. The question now becomes: Is there a material of less weight than steel that also satisfies the design requirements? To be specific, let us consider a polymer, a material that is much less dense than steel. Can it compete with steel in the bumper application? Again, our tool is the stress–strain diagram. Figure 11.13 and Table 11.1 are for a polycarbonate.

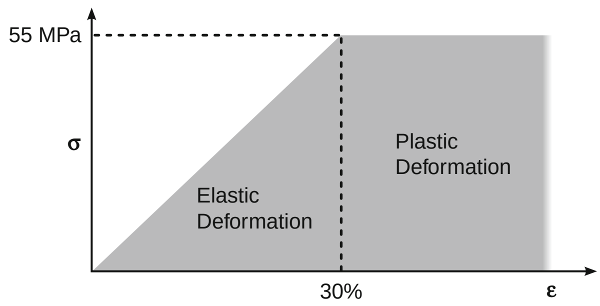


FIGURE 11.13 Stress and Strain for Polycarbonate

Table 11.1 Properties of a Polycarbonate[17]		
Elastic Yield, MPa	**Strain Yield %**	**Density, kg/m3**
55.	30.	1,300.

Notice that the polycarbonate is not as strong as steel because its yield strength is considerably less, but it can be stretched much further while elastically returning to its original strength. Let's repeat Example 11.4 using this material for the bumper.

Example 11.5

If the car's bumper in Example 11.1 is made of the polycarbonate described by Figure 11.13, can it absorb the TKE of the low-speed 2.5 mph collision, or will it plastically deform? What if it is traveling at 5.0 mph? What is the weight of this bumper? The density of polycarbonate is 1,300 kg/m^3, and again assumes the compressive stress–strain diagram is the same as the tensile.

Need: Energy to be absorbed in bumper after collision = ____ MJ/m^3 compared to energy of collision and the weight of a polycarbonate bumper.

Know: From Example 11.4: TKE = 6.3 × 10^2 J; specific energy absorbed = 0.080 MJ/m^3; bumper volume = 0.0079 m^3. At 5.0 mph, TKE = 2,500. J and the specific energy absorbed is 0.32 MJ/m^3.

How: Compare specific energy to be absorbed to capability of material to absorb it. Need to calculate the latter from the $\boldsymbol{\sigma}$, $\boldsymbol{\varepsilon}$ diagram.

Solve: Calculate the elastic toughness by looking at Figure 11.13. It is given at yield by the triangular area, ½ × 0.30 [m/m] × 55. [MN/m^2] = 8.3 MJ/m^3.

Even at 5.0 mph, the specific energy to be absorbed is 0.32 MJ/m^3, which is less than 8.3 MJ/m^3. Thus, this bumper can survive elastically.

The weight of this bumper is Mg = density × volume × g = 1,300 × 0.0079 × 9.81 [kg/m^3] [m^3][m/s^2] = 1.0 × 10^2. N, or much less than the weight of any student!

So, although polymer is not as strong as steel, its ability to compress further without breaking gives it adequate toughness to do the job. In addition, polymers are substantially less dense than steel. So the *polymer is the winner*. In the last two decades, polymers have almost entirely replaced steel as the material from which automobile bumpers are made.

Of course, there are still plenty of options for the materials engineer to design the preferred configuration of the bumper. For example, the metal bumper may be manufactured with springs to absorb the impact. There is surely enough material in our steel bumper that we could use half of it in the form of coil springs, as in Figure 11.14.

A steel spring can be stretched or compressed much further than a flat piece of the same material. It's as if we traded the material for one with a lower modulus (less steep curve) and a much larger strain to yield. Furthermore, there is now plenty of room to design for the randomness of crashes—that is, how the vehicles will interact with an immoveable object such as a pole as indicated in Figure 11.15.

[17]Actually there are many kinds of polycarbonates; the properties of the one shown in Table 11.1 is that of a very common one made from a monomer abbreviated BPA. It is ubiquitous, being used in crash helmets, water bottles, CDs/DVDs, as well as car bumpers. Another version of polycarbonate is used as a lens material for spectacles.

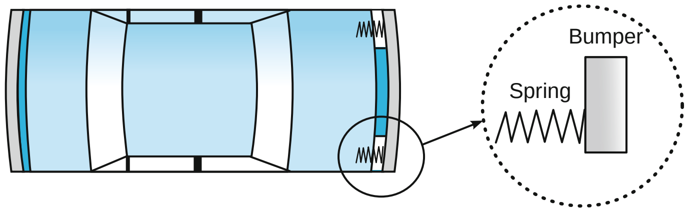


FIGURE 11.14 Mechanical Design Also Influences the Choice of Material

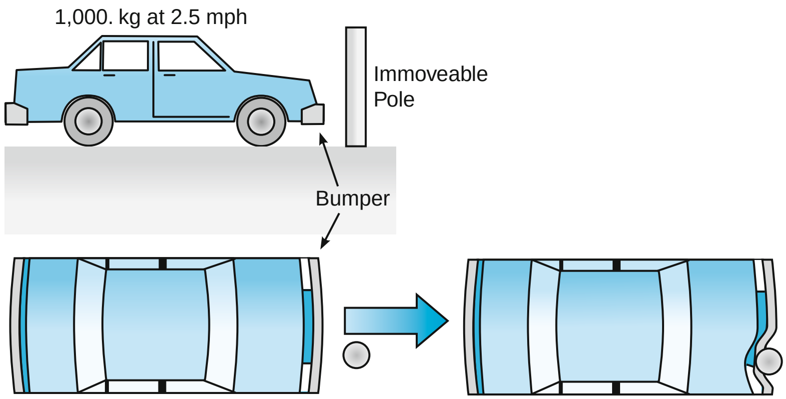


FIGURE 11.15 Not All Crashes are Head-on—The Material May Yield Locally in Low-Speed Impacts

This chapter is just as a glimpse into the most basic considerations that would go into a material design. Of course, materials engineers have become very clever at rearranging the macroscopic properties of materials as they need them. In the car bumper case, the polymer may be backed with a second one that is in the form of foam—the advantage being its strain at yield is mediated by the thousands of gas-filled bubbles within its structure.

11.5 PROPERTIES OF MODERN MATERIALS

This discussion of selecting a material for a bumper gives a feel for the engineer's problem of material selection. However, elasticity, strength, and toughness, though very important, are not the whole story. In the decision whether to use silicon or germanium to make transistors, the electronic and thermal properties of materials play a key role. In the selection of copper versus fiberoptic glass, information-carrying capacity becomes crucial. In the selection of steel versus polymers for refrigerators, considerations of appearance, manufacturability, and corrosion resistance become important. In the selection of a material for a hip joint transplant, compatibility with the human body becomes an essential materials requirement.

In addressing these varied requirements, twenty-first-century materials go far beyond the traditional categories of metals, polymers, and the other classes of materials shown in Figure 11.1. A wide range of composite materials now combine the properties of those original categories. Some are created by embedding fibers of one material within a matrix of a second material. This makes it possible to combine, for example, the high strength of a thin carbon fiber with the high toughness of a polymer matrix.

One of the newest areas of materials engineering are materials whose structure has been engineered at the nanometer scale (1 nm = 10^{-9} m)[18]. These materials are called **nanomaterials**, and are important because the unique properties of materials at the nanoscale enable engineers to create materials and devices with enhanced or completely new characteristics and properties.

Many engineers consider nanotechnology to be the next industrial revolution because it will have enormous social and economic impact. Companies already have introduced nanotechnology in several consumer products. Examples of nanomaterials already on the market include nanoscale titanium dioxide used in some cosmetics and sunscreens, nanoscale silica being used as dental fillers, and nanowhiskers used in stain-resistant fabrics like Eddie Bauer's nanopants. Nanoclays and coatings are being used in a range of products from tennis balls to bikes to cars to improve bounce, strengthen high-impact parts, or render material scratchproof.

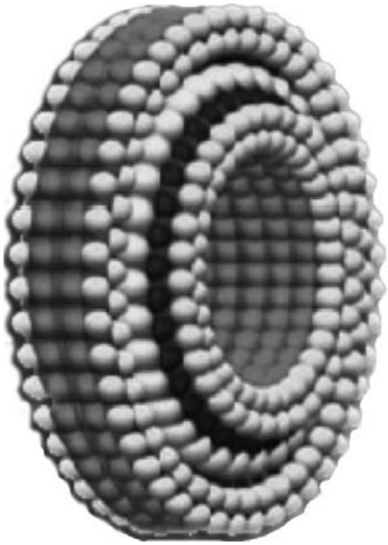


FIGURE 11.16 An Artist's Rendering of a Nanobearing

Nanomaterials are literally tailored atom-by-atom, as in the nanobearing shown in Figure 11.16. This makes it possible, for example, to provide materials with precise combinations of electronic and optical properties. In one sense, however, the emergence of these new composites is actually a matter of going "back to the future." The original materials used by humanity some 10,000 years ago, such as wood, stone, and animal skins, are complicated natural composites. Figure 11.1 sums up this long-term history of humanity's materials use. Even with the new challenges of the twenty-first century, however, the traditional properties of elasticity, strength, and toughness will continue to play a central role in materials selection. As they do, the stress–strain diagram, the best representation of these properties, remains a crucial tool for the twenty-first century engineer.

[18]Atoms, which are the basic building blocks of materials, are on the order of 0.1 to 0.5 nm in size. As small as a nanometer is, it's still large compared to the atomic scale. An atom's nucleus is much smaller—about 0.00001 nm.

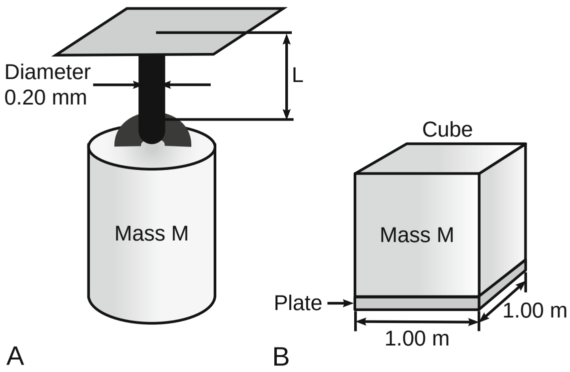
SUMMARY

Materials engineers develop and specify materials. They do so by constrained optimization. One property, such as performance or weight or cost, is to be optimized (maximized or minimized), subject to meeting a set of constraints (design requirements).

Materials selection begins by quantifying design requirements. It proceeds by searching the major classes of candidate materials, metals, polymers, composites, and so on for candidates capable of meeting design requirements. The candidates are then subjected to further screening, based on such characteristics as the **strain** they exhibit in response to **stress**. This particular characteristic is plotted on a graph called the **stress–strain diagram**. This stress–strain, or **σ–ε** diagram can be then used to determine key properties of the material: **modulus of elasticity, yield strength, plastic deformation**, and **toughness**. These properties, expressed in the appropriate units, provide the basis (the "know") for solving particular constrained optimization problems of materials selection.

EXERCISES

The following figures depict the situations described in **exercises 1 through 6**.



1. A mass of $M = 1.0$ kg is hung from a circular wire of diameter 0.20 mm as shown in Figure A. What is the stress in the wire? (**A: 0.31 GPa**)

2. If the wire in Figure A is stretched from 1.00 m to 1.01 m in length, what is the strain of the wire?

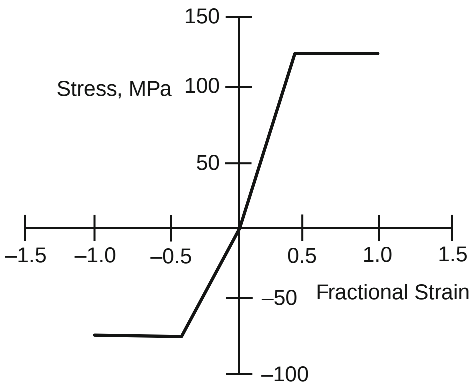
3. In Figure B, when a block of metal 1.00 m on a side is placed on a metal plate 1.00 m on a side, the stress on the plate is 1.00×10^3 N/m^2. What is the mass of the metal cube? (**A: 102 kg**)

4. Suppose the plate described in exercise 3 and Figure B was 0.011 m thick before the cube was placed on it. Suppose that placing the cube on it causes a compressive strain of -0.015. How thick will the plate be after the cube is placed on it? (**A: 0.011, or unchanged to three significant figures**)

5. Suppose the wire in Figure A is perfectly elastic. When subjected to a stress of 1.00×10^4 Pa, it shows a strain of 1.00×10^{-5}. What is the elastic modulus (i.e., Young's modulus) of the wire?

6. A plate of elastic modulus 1.00 GPa is subjected to a compressive stress of -1.00×10^3 Pa as in Figure B. What is the strain on the plate? (**A:** $-\mathbf{1.00} \times \mathbf{10^6}$)

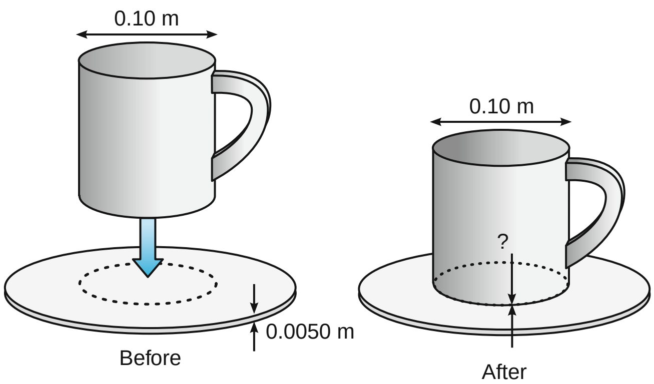
For **Exercises 7 through 9**, assume that a silicone rubber[19] has this stress–strain diagram.



Stress–Strain Diagram for Silicone Rubber

7. What is the yield strength under compression of the silicone?

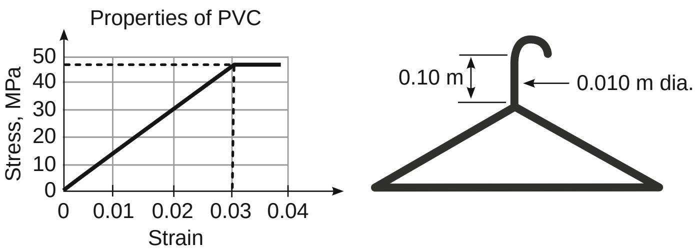
8. A flat saucer made of silicon rubber has an initial thickness of 0.0050 m. A ceramic coffee cup of diameter 0.10 m and mass 0.15 kg is placed on the saucer. What is the final thickness of the saucer beneath the cup, assuming that the force of the cup acts directly downward and is not spread horizontally by the saucer?



9. What is the maximum number of coffee cups in Exercise 8 that can be stacked vertically on the saucer and not cause a permanent dent in the plate? (**A: 4.3** $\times$ $\mathbf{10^5}$ **cups—a tough balancing act**)

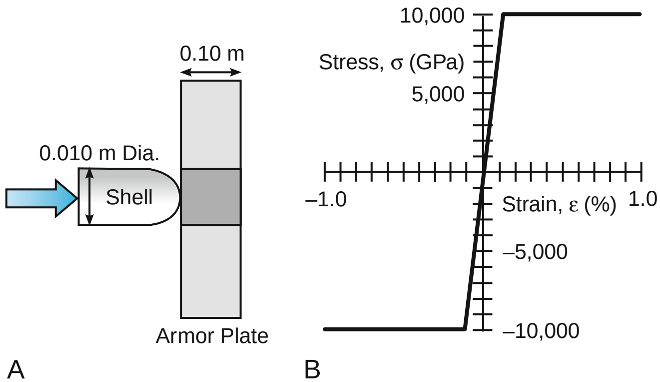
[19]Silicone rubbers are very flexible; their structure consists of polysiloxanes, >Si—O—, in which the Si atom also has two CH_3 chemical groups per atom (not shown). They are off the main chain. Incidentally, you must distinguish *silicon* (a brittle element used in electronics) from *silicone*, the soft rubber described here.

10. A coat hanger is made from polyvinyl chloride.[20] The neck of the coat hanger is 0.01 m in diameter and initially is 0.10 m long. A coat hung on the coat hanger causes the length of the neck to increase by 1.0×10^{-5} m. What is the mass of the coat? The stress–strain diagram is provided. (**A: 1.3 kg**)



11. How many coats having the same mass as those in exercise 10 must be hung on the hanger to cause the neck to remain permanently stretched (i.e., plastically deformed) after the coats are removed?

For **exercises 12 through 15**, imagine the idealized situation as shown in Figure A for an artillery shell striking an armor plate made of an (imaginary) metal called "armory." Figure B is the stress–strain diagram for armory. Assume the material diagram shows armory strained to failure.



12. Suppose the shell has a mass of 1.0 kg and is traveling at 3.0×10^2 m/s. How much TKE does it carry?

13. Assume that the energy transferred in the collision between the shell and the armor plate in exercise 12 affects only the shaded area of the armor plate beneath contact with the flat tip of the shell (a circle of diameter 0.010 m) and does not spread out to affect the rest of the plate. What is the energy density delivered within that shaded volume by the shell? (**A: 5.7×10^9 J/m^3**)

14. Which of the following will happen as a result of the collision in the exercise 13? Support your answer with numbers.

a. The shell will bounce off without denting the armor plate.

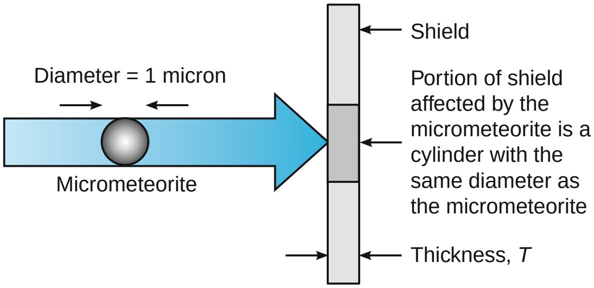
b. The armor plate will be damaged but will protect the region beyond it from the collision.

c. The shell will destroy the armor plate and retain sufficient kinetic energy with which to harm the region beyond the plate. (**A: a—but it is uncomfortably close to yielding and to a permanent set to the armor; an armored tank crew will at least get quite a headache.**)

[20] PVC is a very inexpensive polymer of the monomer $CH_2 = CHCl$ with repeating unit $-CH_2(CHCl)-$.

15. In exercise 14, what is the highest speed the artillery shell can have and still bounce off the armor plate without penetrating it?

Exercises 16 through 19 involve the situation depicted below. Consider a "micrometeorite" to be a piece of mineral that is approximately a sphere of diameter 1.0×10^{-6} m and of density 2.0×10^3 kg/m^3. It travels through outer space at a speed of 5.0×10^3 m/s relative to a spacecraft. Your job as an engineer is to provide a micrometeorite shield for the spacecraft. Assume that if the micrometeorite strikes the shield, it affects only a volume of the shield 1.0×10^{-6} m in diameter and extending through the entire thickness of the shield.



16. Using the stress–strain diagram for steel in Figure 11.11, determine the minimum thickness a steel micrometeorite shield would have to be to protect the spacecraft from destruction (even though the shield itself might be dented, cracked, or even destroyed in the process).

17. Using the stress–strain properties for a polymer (Figure 11.13 and Table 11.1), determine whether a sheet of this polymer 0.10 m thick could serve as a micrometeorite shield if this time the shield must survive a micrometeorite strike without being permanently dented or damaged. Assume the properties of the polymer are symmetric in tension and in compression. (**A: Yes, it will survive unscathed.**)

18. Suppose you were required to use a micrometeorite shield no more than 0.01 meters thick. What would be the required toughness of the material from which that shield was made if the shield must survive a micrometeorite strike without being permanently dented or damaged?

19. Suppose a shield exactly 0.01 m thick of the material in exercise 18 exactly met the requirement of surviving without denting or damage at a strain of −0.10, yet any thinner layer would not survive. What are the yield stress and elastic modulus of the material? (**A:** $\boldsymbol{\sigma_{Yield} = -33}$ **MPa,** $\boldsymbol{E = 0.33}$ **GPa)**

20. Your company wants to enter a new market by reverse-engineering a popular folding kitchen step stool whose patent recently has expired. Your analysis shows that by simplifying the design and making all the components from injection-molded PVC plastic you can produce a similar product at a substantially reduced cost. However, when you make a plastic prototype and test its performance you find that it is not as strong and does not work as smoothly as your competitor's original stool. Your boss is anxious to get your design into production because she has promised the company president a new high-profit item by the end of this quarter. What do you do?

a. Release the design. Nearly everything is made from plastic today, and people don't expect plastic items to work well. You get what you pay for in the commercial market.

b. Release the design but put a warning label on it, limiting its use to people weighing less than 150 pounds.

c. Quickly try to find a different, stronger plastic that can still be injection molded and adjust the production cost estimate upward.

d. Tell your boss that your tests show the final product to be substandard and ask if she wants to put the company's reputation at risk. If she presses you to release your design, make an appointment to meet with the company president.

Use the Engineering Ethics Matrix.

21. As quality control engineer for your company, you must approve all material shipments from your suppliers. Part of this job involves testing random samples from each delivery and making sure they meet your company's specifications. Your tests of a new shipment of carbon steel rods produced yield strengths 10 percent below specification. When you contact the supplier, they claim their tests show the yield strength for this shipment is within specifications. What do you do?

a. Reject the shipment and get on with your other work.

b. Retest samples of this shipment to see if new data will meet the specifications.

c. Accept the shipment, since the supplier probably has better test equipment and has been reliable in the past.

d. Ask your boss for advice.

Use the Engineering Ethics Matrix.

CHAPTER

12 Civil Engineering: The Art and Engineering of Bridge Design

Source: © iStockphoto.com/Jane Norton

When crossing a bridge on a day when traffic is bumper to bumper, or when looking out the window of your airliner and seeing the wing deflect up and down, have you ever wondered if anything was going to break? Does the person who designed the bridge have trouble sleeping at night worrying about the same thing? If not, where does that confidence come from?

12.1 THE BEAUTY OF BRIDGES

Like all **structures**, a bridge is designed to transmit forces while "approximately" maintaining its original shape. The qualification in quotes is needed for the sake of accuracy, as any part or material subjected to forces will undergo small deformations; it's just that most of the time these deformations are imperceptible to the human eye and can be ignored. In structural analysis this is known as the **small displacement assumption**. This is quite different from machines whose parts can be observed to move relative to each other.

The sheer beauty and diversity of bridges around the world suggests an artist's eye at work. So then, who designed these structures? Were they artists or were they engineers? The "art" in existing bridges derived from two sources. In many cases, it was a result of the creative impulses of an architect or designer, with some knowledge of engineering, searching for original ways to express a theme or blend a bridge into its environment. But in more cases than not, the art was a reflection of each bridge's unique political, economic, and functional requirements as rendered by a civil engineer. Sometimes the new designs were inspired by technological advancements, such as emerging new materials and construction methods, and increasingly accurate analytical tools for validating the designs.

In this chapter we will address both the art and engineering of bridge design. The art will be the natural by-products of our attempts to produce an **efficient structure**, one based on minimizing cost or materials. As we will observe, there is a unique and undeniable beauty to structures that have been engineered with functionality and efficiency as the main priorities. The engineering part of bridge design will be represented by the **computational methods** used today to rapidly evaluate and design structures. Many of the fundamentals underlying these mathematical models, such as **free-body diagrams** and the **static equilibrium equations**, have been around for a long time. The American civil engineer Squire Whipple was the first to analyze a truss bridge way back in the 1830s. But the limited computing power of those times placed severe restrictions on the range of bridge designs that could be safely built.[1] With the development of the computer and its ability to solve huge systems of equations quickly, the variety of designs that can be accurately modeled now seems unlimited.

12.2 FREE-BODY DIAGRAMS AND STATIC EQUILIBRIUM

We begin the discussion of bridges with perhaps the most ancient of them all—the rope bridge. A schematic representation of one such bridge is shown in Figure 12.1. The rope is taut and its shape has changed in response to a 200. lbf pedestrian pausing at the center to take in the view. Actually, there are probably two such ropes spanning the ravine with planks laid across them. But the single rope will suffice for evaluating the forces in the rope; the final result can always be divided by two. Our goal is to determine the forces acting on the rope.

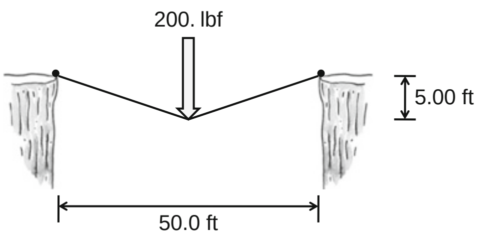


FIGURE 12.1 Rope Bridge

The first step to finding those forces is to draw a **free-body diagram** (**FBD**) of the system. The system can be anything we define it to be. A good choice for this problem is to define the system to be the rope. With that decided, you draw a picture of the system, showing it *isolated* from its environment, while replacing adjoining parts, supports, and weights by the forces they exert on the system. The FBD of the rope is shown in Figure 12.2. The forces pulling on the ends of the rope have unknown magnitudes, represented by **A** and **B**, but known directions, since a rope always aligns itself with the force pulling on it. The weight of the rope is assumed to be very small, and does not appear in the FBD.

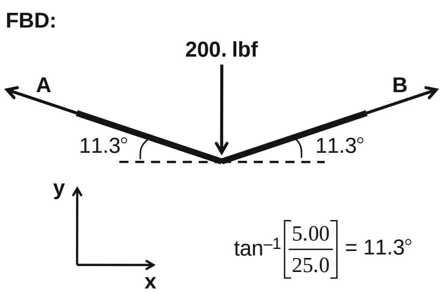


FIGURE 12.2 Free-Body Diagram of the Rope

[1]For a perspective on early American bridge failures see *To Engineer Is Human*, Henry Petroski, Vintage Books, 1992.

The next step is to check that the number of unknowns in the FBD is equal to the number of available equilibrium equations. Only when they match will it be possible to solve the equations. For this system, there are two equations governing **static equilibrium**[2]:

$$\sum_{i=1}^{N} F_{xi} = 0 \qquad \sum_{i=1}^{N} F_{yi} = 0$$

where N is the total number of forces acting on the system; F_{xi} is the x-component of the ith force; F_{yi} is the y-component of the ith force; $\sum$ indicates a *summatio*n over the N force *components*.

What do we mean by components? These are simply a way to project geometric elements of a force, which may act in any direction, onto convenient and useful axes. Convenient axes are simply those with special meaning to us such as the horizontal and the vertical directions, although other convenient directions are possible depending on the particular problem. Usually these axes are denoted by attaching subscripts such as x- and y- to the forces in which x and y are the Cartesian abscissa and ordinate axes, respectively.

The x- and y-components of a force are found by projecting it onto perpendicular axes located at the tail of the force arrow, as illustrated in Figure 12.3. With two equilibrium equations and two unknown forces (**A** and **B**), these equations will be solvable. Such a problem is referred to as being **statically determinate**.

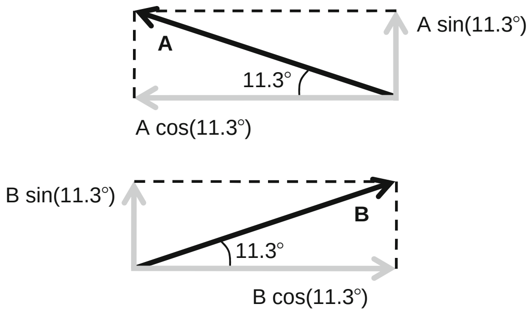


FIGURE 12.3 Components of the Forces A and B

The determinacy check clears the way to write the equilibrium equations. The FBD of Figure 12.2 yields the following two equations in two unknowns describing equilibrium.

$$\sum F_x = 0: \quad -A\cos(11.3^\circ) + B\cos(11.3^\circ) = 0$$

$$\sum F_y = 0: \quad A\sin(11.3^\circ) + B\sin(11.3^\circ) - 200. = 0$$

The signs are determined by the directions of the force components relative to the coordinate axes. Thus, x-components that point to the right are positive and those that point to the left are negative, and y-components that point up are positive and those that point down are negative.

[2]These equations are special cases of Newton's Law of Motion for a system at rest (or at least not accelerating); thus each force component, if not zero, is balanced by at least one other force.

The last step is to solve the equations simultaneously by substitution. The final results are:

$$A = 510.\ \text{lbf} \qquad B = 510.\ \text{lbf}$$

which makes sense given the symmetry of the geometry and loading.

These steps just carried out for the rope bridge—(1) draw the FBD, (2) check determinacy, (3) write the equilibrium equations, and (4) solve the equations—define a general methodology for equilibrium analysis that is applicable to any statically determinate problem. One of the challenges in applying this methodology to bridges is that some designs are composed of hundreds, sometimes thousands, of structural members, each of which has its own FBD and equilibrium equations to solve.

12.3 STRUCTURAL ELEMENTS

Bridges can be idealized as an assembly of different types of structural elements. Each element typically is characterized by its geometry, the forces acting upon it, or both. There are many different types of such elements, ranging from plates and shells to springs, just as there are many different types of Lego bricks in a child's construction set. In this chapter, we will consider three different types of elements—**beams**, **compression members**, and **tension members**.

We begin the discussion of the structural elements with one of the simplest forms that a bridge can take, that of a **beam**, as shown in Figure 12.4. This structural element usually is identified by its long, straight geometry and the transverse loads acting along its length, causing it to bend. Unfortunately, beam bending is one of easiest ways to break a long, straight member.

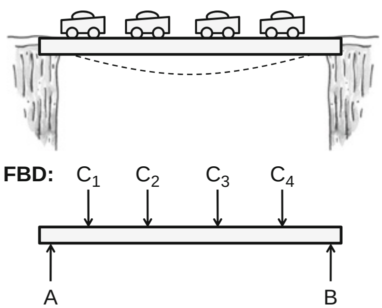


FIGURE 12.4 A Beam (the dashed line represents the deflected shape)

Knowing this, the designer will attempt to limit the bending of a beam by adding supports. For example, we might add a pier to hold up the bridge at mid-span as in Figure 12.5. The beam will still bend under the transverse loads produced by moving vehicles or pedestrians, but with the span effectively cut in half, it will bend about 1/8 as much.

The pier is one example of a **compression member**. Like the beam, a compression member is straight but the loading is different—just end loads acting to compress, or shorten, the member. Short, squat compression members can sustain very high loads, but watch out for long thin compression members. They can fail

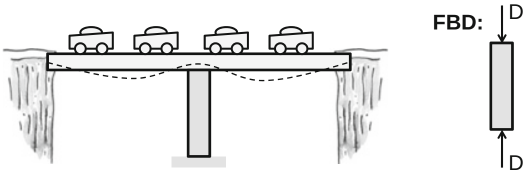


FIGURE 12.5 A Compression Member

spontaneously due to **buckling**. Buckling is a particularly dangerous failure mode, because there are no early warning signs. As the compressive load ramps up, the member will appear to remain perfectly straight with no visible signs of damage, when quite suddenly it will bend and snap as a result of slight, unavoidable imperfections in geometry. To see an example of buckling, place one foot on an upright aluminum soda can and gradually shift your weight onto the can, adding more weight if necessary, until it collapses.

The bridge designer also has the option of supporting the beam from above using **tension members**, shown as cables in Figure 12.6. In this figure, the pier still supports the beam at mid-span, but the height of the pier has been increased to provide anchoring points for the two taut cables that run from the top of the pier to the beam below. The cables are under tension; that is, they are being stretched by opposing colinear forces that pull on the ends of the cable.

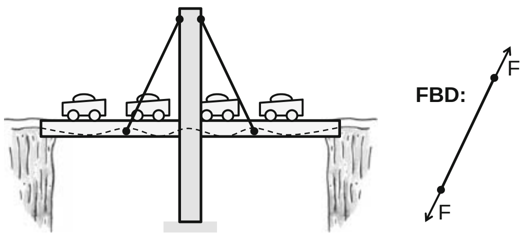


FIGURE 12.6 A Tension Member

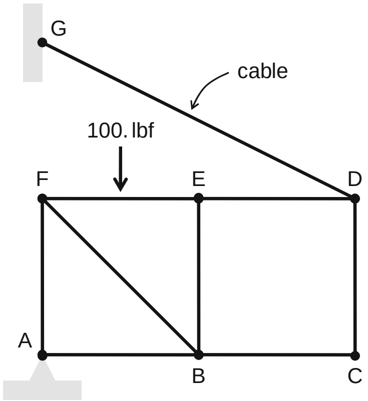
So how do you know from looking at an existing structure, if a structural member behaves more like a beam, a compression member, or a tension member? First, the easy one—if it's a cable, it has to be a tension member. A cable has just too much flexibility to support either a compressive load or bending.[3] If there are transverse loads causing the member to bend, then it has to be a beam. When the bridge is composed of many members joined together at their ends by rivets (see Figure 12.7) or welds to form a lattice, you can distinguish between beams and tension/compression members by examining the local geometry of the structure. Members joined end-to-end to form a triangle tend to act like tension or compression members because bending will be minimal. Members forming other polygonal shapes are much more likely to bend and so should be modeled as beams.

[3] In common parlance, "You can't push a rope."

FIGURE 12.7 Joint Formed by Riveting a Gusset Plate to Converging Members (Used with Permission of Landov Media, New York)

Example 12.1

For the given structure, identify all of its members as beams, compression members, or tension members.



Need: To identify the member types.
Know: Cable; force; geometry of the structure.
How: Look for cables, transverse forces between joints, and triangles.
Solve:

DG is a tension member (because DG is a cable).
FE is a beam (because it will bend under the transverse force of 100. lb).
BC, CD, DE are beams (because they define a square, not a triangle).
AB, BF, AF are tension or compression members (because they define a triangle).
EB is hard to classify because one of the members helping to define EB's triangle is a beam. It is probably safest to assume EB behaves as a beam.

12.4 EFFICIENT STRUCTURES

The goal in designing an efficient structure is to satisfy the requirements of the design at minimum cost. Once construction materials have been selected, this translates to building the bridge using the least amount of material. A bridge designer will employ the following strategies to achieve this goal.

12.4.1 Beams

The shape of a beam's cross-section has as much to do with its ability to resist bending as the choice of materials. To illustrate this, our simple beam bridge will be modeled as a stack of boards glued together lengthwise, as shown in Figure 12.8. When forced to bend, the boards above the horizontal mid-plane (called the **neutral surface**) will shorten (going into compression) and those below will elongate (going into tension) so that the beam can assume its curved shape. The boards furthest from the neutral surface experience the greatest change in length and thus offer the greatest resistance to bending. Or put another way, the most efficient cross-sections for beams are those that concentrate the material furthest from the neutral surface. Figure 12.9 shows how a square cross-section of given area can be split up to form beam cross-sections with higher resistance to bending. The first result is a common cross-section known as an I-beam. The second result, with a much higher resistance to bending, requires connectors between the four isolated pieces of the cross-section. This is a typical cross-section for a truss bridge.

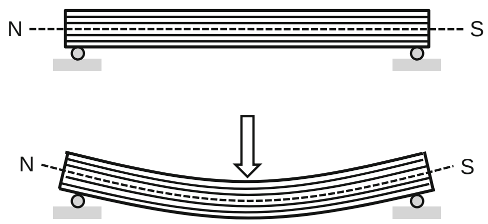


FIGURE 12.8 Deformation of a Beam—The Neutral Surface (NS) Does Not Change in Length

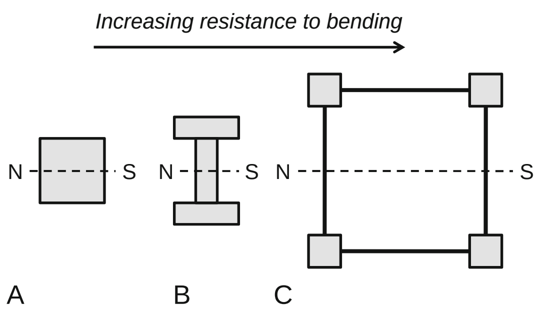


FIGURE 12.9 Equal Area Beam Cross-Sections with Very Different Efficiencies [From least efficient to most efficient: (A) square cross-section, (B) I-beam, (C) truss bridge (end view).]

12.4.2 Trusses

A structure composed of many connected members has to meet the following two conditions to qualify as a truss: (1) it must be composed entirely of tension and compression members and (2) it must be fully triangulated, meaning that every open space within the structure is triangular in shape. In order to meet the first condition, both the **live load**[4] (due to traffic) and the **dead load** (due to the weight of the members) should be applied at the joints. Often, the weights of the members are assumed to be negligible compared to the live load. The bridge in Figure 12.10 is an example of a truss.

FIGURE 12.10 Truss Bridge in Interlaken, Switzerland

A truss can be viewed as simply a highly efficient beam with the cross-section depicted in Figure 12.9(C). Like a beam, when the truss bridge deforms under load, its upper members will shorten as its lower members elongate. The other members tie things together and provide additional resistance to bending. In some cases, these other members experience no load at all and are known as **zero-force members**. These are useful for shortening the effective length of compression members to avoid buckling.

12.4.3 Arches

We can think of an arch as a curved beam that is highest at mid-span and lowest at the ends where it touches ground. This curved beam can be solid or in the form of a truss as in Figure 12.11. Because of the curvature, the transmission of force from the top of the arch to the ground supports is much more direct than in a straight beam, where the live load effectively has to make a 90 degree turn to reach the ends of the beam. As a result, arches are stronger in bending than straight beams with the same cross-section.

[4]In 1987 on the fiftieth anniversary of the opening of the Golden Gate Bridge, the bridge was closed to traffic and opened to pedestrians. So many people (hundreds of thousands) celebrated on the bridge that its live load was the weight of these people. This justifiably scared the bridge operators into rapid stress calculations as the bridge's arch flattened under the load!

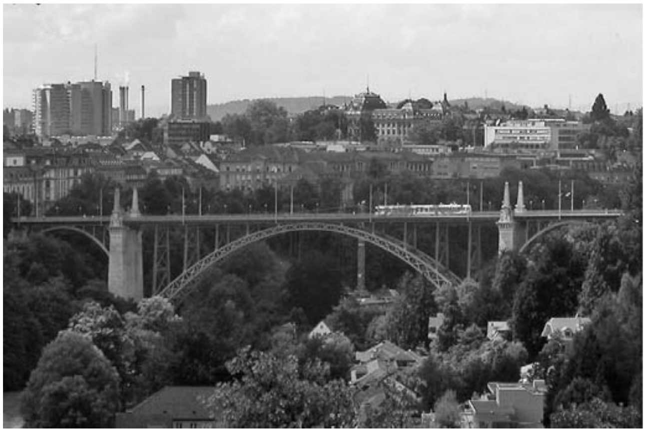


FIGURE 12.11 Truss-Arch Bridge in Bern, Switzerland

12.5 THE METHOD OF JOINTS

Before moving forward with construction of a bridge, the engineer must first prove that the bridge will not fail. Relying on testing the bridge after it has been built is not an option because of the costs involved. Instead, engineers will use mathematical models to analyze the structure. Simple mathematical models, using just a few elements (e.g., beams, tension members, compression members) to approximate the structure have been around since the days of Squire Whipple. But as computing resources have rapidly evolved since the 1960s, so have the mathematical models. The confidence level in these **finite element models** is approaching that of physical testing.

We now introduce one method for developing the mathematical model of a structure. It commonly is referred to as the **Method of Joints**, and is applicable to structures classified as **statically determinate trusses**.

Our goal in applying the method of joints will be to determine the forces on the two members comprising the three-hinged arch of Figure 12.12. The hinges are steel pins that allow the members to undergo small rotations with respect to each other and with respect to the ground supports as the structure deforms under load. This effectively eliminates any tendency for the members to bend, causing them to behave as either tension or compression members. Another practical advantage of the three hinges is that when the members expand in length due to a temperature rise, the central hinge will rise up, relieving the compressive forces that would otherwise build up in the members. The end spans of the Tower Bridge in London are examples of three-hinged arches.

To begin the analysis, we need to know the locations of the joints and the applied forces. Both are represented in Figure 12.12. The force magnitude is based on a conservative estimate of the live load due to pedestrian traffic. It is applied at the joint to avoid inducing bending in the members.

The solution method is the standard one for problems in static equilibrium—draw the FBDs, check for determinacy, write the equilibrium equations, and finally solve the equations. What distinguishes the Method of Joints is that the entire analysis is based on the FBDs of the pins. Thus, implicit in any truss analysis is the assumption that the members can accurately be modeled as being connected by pins. For this reason the joints in a truss commonly are represented as being pin-connected.

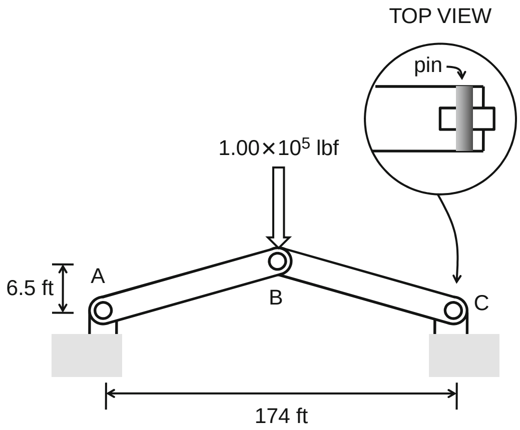


FIGURE 12.12 Three-Hinged Arch

The first step is to draw the FBDs. The FBDs of the members and the pins are shown in Figure 12.13. Both members are assumed to be in tension, in accordance with the recommended sign convention. Subscripts uniquely associate the forces F_{AB} and F_{BC} with the members. Forces exerted by the ground supports on the pins are labeled differently as A_x, A_y, C_x, and C_y, with two force components at each pin since each support constrains the movement of a pin in two directions. The assumed directions of these support reactions are guesses; later, negative signs in the final calculated results will tell us where those guesses were incorrect. Finally and most importantly, the forces exerted by adjoining parts on each other must be shown as equal in magnitude and opposite in direction to insure determinacy and satisfy Newton's Third Law. In this regard, we may note that the tension members are assumed to touch only the pins, and not each other or the ground supports.

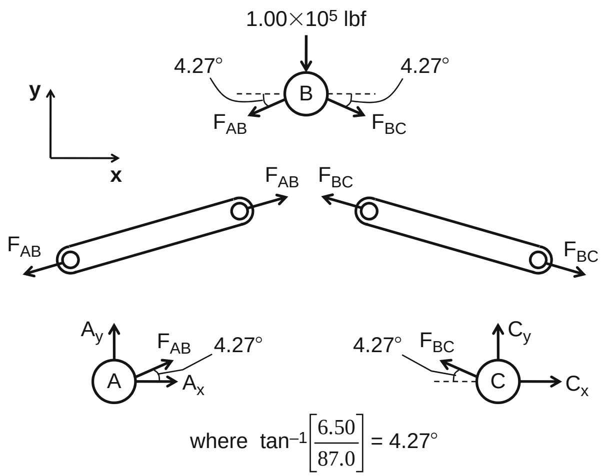


FIGURE 12.13 Free Body Diagrams (FBDs) for the Three-Hinged Arch

The next step is to write the equilibrium equations for each relevant FBD. Equilibrium of the tension members yields the trivial result that the end forces are equal, and so are dropped from the analysis. Checking determinacy of the three remaining FBDs, we find that there are six unknown forces (F_{AB}, F_{BC}, A_x, A_y, C_x, C_y) and a total of six equations (two equilibrium equations per FBD). Since the unknowns and equations balance, we know that we can solve this system of equations. Closer examination reveals that the forces of interest, F_{AB} and F_{BC}, can be found from analyzing just pin B as there will be two unknown forces (F_{AB}, F_{BC}) and two equilibrium equations. This is the solution strategy we will adopt here. The equilibrium equations for the FBD of pin B are:

$$\sum F_x = 0: \quad -F_{AB}\cos(4.27°) + F_{BC}\cos(4.27°) = 0$$

$$\sum F_y = 0: \quad -F_{AB}\sin(4.27°) - F_{BC}\sin(4.27°) - 1.00 \times 10^5 = 0$$

The final step is to solve this system of equations. Solving them simultaneously by substitution leads to the following results:

$$F_{AB} = \mathbf{-6.72 \times 10^5\ lbf} \qquad F_{BC} = \mathbf{-6.72 \times 10^5\ lbf}$$

where the minus sign indicates that our initial guess for the directions of these forces was incorrect; the members are in fact compression members. The symmetry of these results should have been expected given that the geometry and live load share a common plane of symmetry.

12.6 SOLUTION OF LARGE PROBLEMS

For structures with more than two members, solution of the equations by hand is often impractical. The following example introduces a more general procedure for solving large systems of equations.

Example 12.2

The truss bridge of Figure 12.14 is proposed as an alternative to the three-hinged arch. It is required to carry the same live load (1.00×10^5 lbf) and span the same distance (174 ft). We may think of this truss as part arch (top half) and part rope bridge (bottom half). Determine the forces in all the members. The pin at C is mounted on a roller so it can move with the application of the live load.

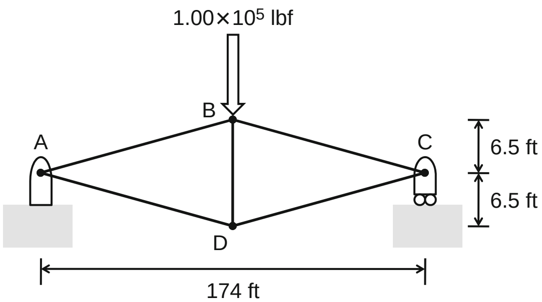


FIGURE 12.14 Truss Bridge

Need: The forces acting on all five members.

Know: The location of the joints and the location and magnitude of the live load. The basic truss assumptions apply; that is, the weights of the members are negligible and the connections can be accurately modeled as pins.

How: By the method joints, with one important difference from the previous example—we will solve the equations using Excel.

Solve: First step is to draw the FBDs of the pins as shown in Figure 12.15. Here again, we assume all members are in tension and guess at the directions of the support reactions. The rollers under the support at C mean this support does not limit the horizontal movement in the x-direction at C (i.e., $C_x = 0$).

Counting unknowns and equations, we find that there are eight unknowns (F_{AB}, F_{BC}, F_{CD}, F_{AD}, F_{BD}, A_x, A_y, C_y) and eight equations (four pin FBDs with two equilibrium equations per FBD). This tells us in advance that the equations will be solvable.

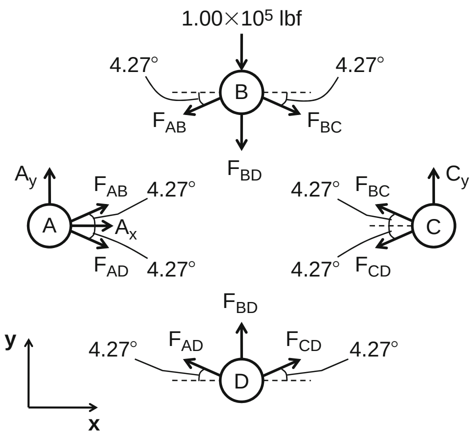


FIGURE 12.15 Pin Free Body Diagrams (FBDs)

The next step is to write the equilibrium equations for each pin FBD. They are presented as:

Pin A: $\sum F_x = 0$, or $F_{AB}\cos(4.27°) + F_{AD}\cos(4.27°) + A_x = 0$

$\sum F_y = 0$, or $F_{AB}\sin(4.27°) - F_{AD}\sin(4.27°) + A_y = 0$

Pin B: $\sum F_x = 0$, or $-F_{AB}\cos(4.27°) + F_{BC}\cos(4.27°) = 0$

$\sum F_y = 0$, or $-F_{AB}\sin(4.27°) - F_{BC}\sin(4.27°) - F_{BD} - 1.00 \times 10^5 = 0$

Pin C: $\sum F_x = 0$, or $-F_{BC}\cos(4.27^\circ) - F_{CD}\cos(4.27^\circ) = 0$

$\sum F_y = 0$, or $F_{BC}\sin(4.27^\circ) - F_{CD}\sin(4.27^\circ) + C_y = 0$

Pin D: $\sum F_x = 0$, or $-F_{AD}\cos(4.27^\circ) + F_{CD}\cos(4.27^\circ) = 0$

$\sum F_y = 0$, or $F_{AD}\sin(4.27^\circ) + F_{CD}\sin(4.27^\circ) + F_{BD} = 0$

The solution of these equations by hand is both cumbersome and subject to error. We therefore turn to a computational approach based on first translating the equations into a **matrix** form that can be read by a computer, and then solving the equations with the matrix operations available in Excel.

Two intermediate steps are required to set up the matrix equations. First, any constants in the equations need to be shifted to the right-hand side of the equal (=) signs. Of course this will change the sign on those constants. Second, the coefficients multiplying the unknown forces need to be identified. The resulting right-hand side constants (*RHS*) and the coefficients of the forces are listed as follows for each equation (*EQ*). The order of the forces across the top of the table is totally arbitrary.

EQ	F_{AB}	F_{BC}	F_{CD}	F_{AD}	F_{BD}	A_x	A_y	C_y	***RHS***
1	cos(4.27)	0	0	cos(4.27)	0	1	0	0	0
2	sin(4.27)	0	0	–sin(4.27)	0	0	1	0	0
3	–cos(4.27)	cos(4.27)	0	0	0	0	0	0	0
4	–sin(4.27)	–sin(4.27)	0	0	–1	0	0	0	1.00×10^5
5	0	–cos(4.27)	–cos(4.27)	0	0	0	0	0	0
6	0	sin(4.27)	–sin(4.27)	0	0	0	0	1	0
7	0	0	cos(4.27)	–cos(4.27)	0	0	0	0	0
8	0	0	sin(4.27)	sin(4.27)	1	0	0	0	0

The matrix form of the equilibrium equations, given here, was extracted directly from this table:

$$\begin{bmatrix} \cos(4.27) & 0 & 0 & \cos(4.27) & 0 & 1 & 0 & 0 \\ \sin(4.27) & 0 & 0 & -\sin(4.27) & 0 & 1 & 1 & 0 \\ -\cos(4.27) & \cos(4.27) & 0 & 0 & 0 & 0 & 0 & 0 \\ -\sin(4.27) & -\sin(4.27) & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & -\cos(4.27) & -\cos(4.27) & 0 & 0 & 0 & 0 & 0 \\ 0 & \sin(4.27) & -\sin(4.27) & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & \cos(4.27) & -\cos(4.27) & 0 & 0 & 0 & 0 \\ 0 & 0 & \sin(4.27) & \sin(4.27) & 1 & 0 & 0 & 0 \end{bmatrix} \begin{Bmatrix} F_{AB} \\ F_{BC} \\ F_{CD} \\ F_{AD} \\ F_{BD} \\ A_x \\ A_y \\ C_y \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \\ 1.00 \times 10^5 \\ 0 \\ 0 \\ 0 \\ 0 \end{Bmatrix}$$

where the brackets [ ] denote a matrix with multiple rows and columns; the braces { } denote a single column matrix commonly referred to as a vector. For ease of reference, we can express this equation using shorthand matrix notation as:

$$[A]\{x\} = \{b\} \tag{12.1}$$

where the coefficient matrix $[A]$ and the right-hand side vector $\{b\}$ come straight from the table. The order of the unknown force magnitudes in the vector $\{x\}$ must correspond to the force labels across the top of the table so that the same multiplications are implied. Because the coefficients of $[A]$ premultiply the variables in $\{x\}$, this operation is known as a **matrix multiplication**.

The unknowns can be isolated by premultiplying both sides of Equation (12.1) by the **matrix inverse** of $[A]$:

$$[A]^{-1}[A]\{x\} = [A]^{-1}\{b\} \tag{12.2}$$

Denoted by $[A]^{-1}$, and having the same dimension as $[A]$, the matrix inverse can be found using one of the special functions available in Excel. Premultiplying a matrix by its inverse is equivalent to premultiplying a number by its reciprocal. Therefore Equation (12.2) simplifies to:

$$\{x\} = [A]^{-1}\{b\} \tag{12.3}$$

Thus, having found $[A]^{-1}$ using Excel, we can then apply another special function in Excel to perform this matrix multiplication and, in so doing, determine the values of the unknown force magnitudes.

The details of how to apply Excel to this example are summarized in the following procedure. Results are shown in the worksheet of Figure 12.16.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	**Solution of the Equations for the Truss Bridge**												
2													
3													
4		0.997	0	0	0.997	0	1	0	0			0	
5		0.0745	0	0	–0.0745	0	0	1	0			0	
6		–0.997	0.997	0	0	0	0	0	0			0	
7	**[A] =**	–0.0745	–0.0745	0	0	–1	0	0	0		**{b} =**	1.00E+05	
8		0	–0.997	–0.997	0	0	0	0	0			0	
9		0	0.0745	–0.0745	0	0	0	0	1			0	
10		0	0	0.997	–0.997	0	0	0	0			0	
11		0	0	0.0745	0.0745	1	0	0	0			0	
12													
13													
14		0	0	–0.7523	–3.3557	–0.5015	0	–0.2508	–3.356			–3.36E+05	$=F_{AB}$
15		0	0	0.2508	–3.3557	–0.5015	0	–0.2508	–3.356			–3.36E+05	$=F_{BC}$
16		0	0	–0.2508	3.3557	–0.5015	0	0.25075	3.3557			3.36E+05	$=F_{CD}$
17	**inv[A] =**	0	0	–0.2508	3.3557	–0.5015	0	–0.7523	3.3557		**{x} =**	3.36E+05	$=F_{AD}$
18		0	0	0.0374	–0.5	0.0747	0	0.03736	0.5			–5.00E+04	$=F_{BD}$
19		1	0	1	0	1	0	1	0			0.00E+00	$=A_x$
20		0	1	0.0374	0.5	0	0	–0.0374	0.5			5.00E+04	$=A_y$
21		0	0	–0.0374	0.5	0	1	0.03736	0.5			5.00E+04	$=C_y$
22													

FIGURE 12.16 Excel Worksheet for Solving the Equations

1. Open an Excel worksheet.
2. Enter the values of $[A]$ into a block of cells with 8 rows and 8 columns.
3. Enter the values of $\{b\}$ into another nearby block of cells with 8 rows and 1 column.
4. Highlight an empty block of cells with 8 rows and 8 columns. Then type:

$$= \text{MINVERSE(B4:I11)}$$

   and, instead of pressing Enter, press the following keys simultaneously:

$$\text{Ctrl} - \text{Shift} - \text{Enter}$$

   to compute the inverse of $[A]$ and have the results appear in the highlighted block of cells. Note that B4:I11 defines the range of cells belonging to $[A]$ and MINVERSE is a function that computes the inverse of the matrix defined by the range. This range could also have been entered by highlighting the cells of $[A]$.
5. Highlight another empty block of cells; this one with 8 rows and 1 column. Then type:

$$= \text{MMULT(B14:I21, L4:L11)}$$

   and press the following keys simultaneously:

$$\text{Ctrl} - \text{Shift} - \text{Enter}$$

   to multiply $[A]^{-1}$ and $\{b\}$ in accordance with Equation (12.3) and have the results appear in the highlighted column of cells. Note that the function MMULT requires us to define two ranges of cells, one for each of the matrices being multiplied.

The results from Excel are displayed along with the structure in Figure 12.17. It does appear that these results make sense. The results are symmetrical and, as should be expected of a beam, the upper members are in compression (as indicated by the minus signs) while the lower members are in tension.

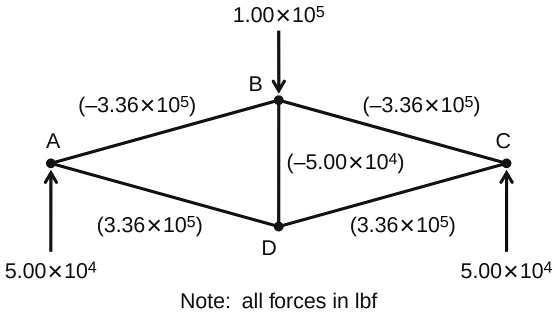


FIGURE 12.17 Results for the Truss Bridge (Member forces are indicated in parentheses)

12.7 DESIGNING WITH FACTORS OF SAFETY

Once the forces acting on the members are known, the members can be checked to determine if they will fail. However, it is not enough just to know the forces. Whether or not a force is large enough to cause the material to fail is also dependent on the cross-sectional area of the member. Therefore the **stress** in each member is calculated as in the previous chapter:

$$\sigma = \frac{F}{A} \tag{12.4}$$

where σ is the stress, F is the force on the member, and A is the cross-sectional area. The sign convention for stress is the same as for the forces on the members; positive for tension and negative for compression.

For ductile materials, like steel, failure is defined to occur when the stress reaches a critical level known from material tests to be associated with the onset of yielding, as expressed by:

$$\textbf{\textit{Failure occurs when}}\ |\sigma| = S_Y \tag{12.5}$$

where S_Y is the **yield strength**, and the absolute value symbol has been introduced to accommodate the possibility of yielding in compression. For other materials, like wood, that do not undergo yielding, S_Y can be replaced by the breaking strength. A more continuous measure of nearness to failure is obtained by rearranging Equation (12.5) and introducing the concept of design **factor of safety** (N):

$$N = \frac{S_Y}{|\sigma|} \tag{12.6}$$

where if $N \leq 1$, the member will probably fail; if $N \geq 1$, the member should be safe from failure, with the level of confidence proportional to the amount by which N exceeds 1. Equation (12.6) can be applied to determine the factor of safety of every member in the truss, provided that the cross-sectional areas are known. The **overall factor of safety** of the truss will be the smallest of these values. In other words, a statically determinate truss is only as strong as its weakest link.

Example 12.3

Determine the overall factor of safety of the truss in Example 12.2, assuming all members have a cross-sectional area of 25.0 in^2 and are made of steel with a yield strength of 36,000 lbf/in^2.

Need: Overall factor of safety.

Know: Forces on the members; $A = 25.0$ in^2 for all members; $S_Y = 36{,}000$ lbf/in^2.

How: First calculate the stress in each member with Equation (12.4). Then substitute stresses into Equation (12.6) to find the factor of safety for each member. Smallest factor of safety in the members defines the overall factor of safety.

Solve: Perform the calculations for member AB.

$$\sigma = \frac{F}{A} = \frac{-3.36 \times 10^5}{25.0}\,[\text{lbf}]\ [1/\text{in}^2] = -13{,}400\ \text{lbf/in}^2$$

$$N = \frac{S_Y}{|\sigma|} = \frac{36{,}000}{|-13400|}\,[\text{in}^2/\text{lbf}] = 2.7$$

Repeat these calculations for the other members and tabulate the results.

Member	Force (lbf)	σ (lbf/in^2)	N
AB	-3.36×10^5	−13400	2.7
BC	-3.36×10^5	−13400	2.7
CD	3.36×10^5	13400	2.7
AD	3.36×10^5	13400	2.7
BD	-5.00×10^4	2000	18.

The overall factor of safety is 2.7, indicating that the truss will safely hold the live load.

In the previous example, analysis was used to evaluate a truss. It can also be used to design a truss. Substituting Equation (12.4) into (12.6), and rearranging, leads to an expression for predicting the cross-sectional area of the members:

$$A = \frac{N|F|}{S_Y} \tag{12.7}$$

where, in this context, N is a target value for factor of safety that the bridge must meet in order to be considered safe. If the cross-sectional areas of all the members are determined in this way, the members will all fail simultaneously, at least theoretically. This is the ideal but unrealizable goal of any structural analysis. As a practical matter, it will prove much more cost efficient to choose from a finite number of readily available standard sizes, while regarding the result of Equation (12.7) as the lower bound on an acceptable size.

Still, there remains the crucial question—how big does N have to be, to be considered safe enough? In general, the answer depends on how sure you are of (1) the material properties, (2) the environmental conditions, including loads, and (3) the accuracy of your analysis. The greater your uncertainty, the larger N should be. For example the aircraft industry employs extensive testing to reduce uncertainty, often enabling the use of factors of safety of less than 2. As a rough guide for use with design projects—start with a value of 3 and add 2 for each technical category (properties, loads, analysis) having a significant amount of uncertainty, or risk, associated with it.

Example 12.4

Given the three-hinged arch of Figure 12.12, determine the cross-sectional areas of the members if the required factor of safety is 5.0 and the members are made of steel with a yield strength of 36,000 lbf/in^2.

Need: The cross-sectional areas of each member.
Know: Force on both members is -6.72×10^5 lbf; $N = 5.0$; $S_Y = 36{,}000$ lbf/in^2.
How: Substitute directly into Equation (12.7) to calculate cross-sectional area.
Solve: Since the force on both members is the same, apply equation (12.7) just once:

$$A = \frac{N|F|}{S_Y} = \frac{5.0\left|-6.72 \times 10^5\right|}{36{,}000}\text{[lbf] [in}^2\text{/lbf]} = \mathbf{93.\ in^2}$$

This is the minimum required cross-sectional area for both members.

Equations (12.6) and (12.7) can be modified to account for the possibility of buckling in compression members. To do this, we introduce an equation that defines the **buckling strength**, S_B, for the special case of a compression member with square cross-section and pinned ends:

$$S_B = \frac{\pi^2 E h^2}{12 L^2} \tag{12.8}$$

in which E is the elastic modulus of the material; h is the width of the square cross-section; L is the length of the compression member. Unlike the yield strength, which is exclusively a material parameter, S_B also depends on the geometry through the ratio h/L. Figure 12.18 illustrates the buckling failure of a long slender piece of dry uncooked pasta. With two possible failure modes (yielding and buckling), and two corresponding strengths to check stresses against, Equation (12.6) becomes:

$$N = \left(\frac{\min(S_Y, S_B)}{|\sigma|}\right) \tag{12.9}$$

where the numerator is taken to be the minimum of the two strengths, as the member will fail by this mode before the other has a chance to occur. The prediction of cross-sectional area is a little trickier. The prediction based on yielding is still given by Equation (12.7). The prediction based on buckling is found from Equation (12.8) by setting $S_B = N|\sigma| = N|F|/A$ and solving for A, where $A = h^2$. The maximum of these two predictions defines the required cross-sectional area of the compression member, as expressed:

$$A = \max\left(\frac{N|F|}{S_Y}, \left(\frac{12\, N|F|L^2}{\pi^2\, E}\right)^{0.5}\right) \tag{12.10}$$

where here again, A should be interpreted as being a lower bound on acceptable values of cross-sectional area. Summing up, when buckling is a possibility and cross-sections are square, tension members should be modeled using Equation (12.6) or (12.7), and compression members should be modeled using Equation (12.9) or (12.10), with the choice of equations depending on whether the goal of the analysis is evaluation or design.

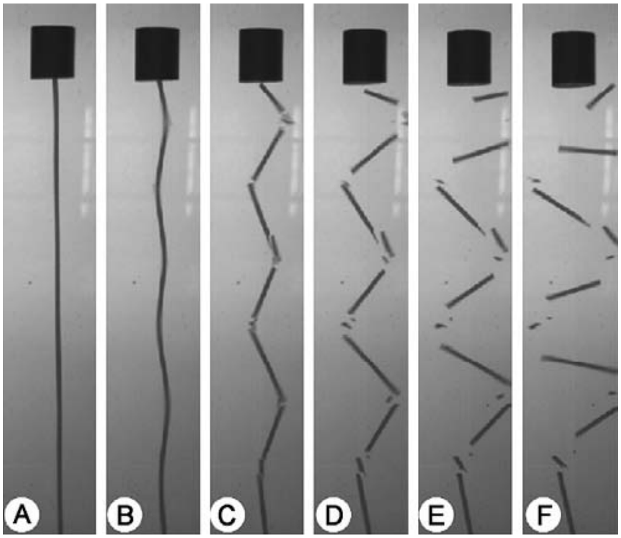


FIGURE 12.18 Sequence of Photos Showing the Buckling of a Piece of Pasta (By Permission of Professor Andrew Belmonte, Department of Mathematics, Pennsylvania State University)

Example 12.5

For the three-hinged arch of the previous example in which the factor of safety is $N = 5.0$ and $S_Y = 36{,}000$ lbf/in^2, determine the required cross-sectional areas once again, but this time consider the possibility of buckling. Assume cross-sections are square and that the elastic modulus for steel is 29. $\times 10^6$ lbf/in^2.

Need: The cross-sectional areas of each member, but this time consider the possibility of buckling.
Know: Force on both members is -6.72×10^5 lbf; $N = 5.0$; $S_Y = 36{,}000$ lbf/in^2; $E = 29. \times 10^6$ lbf/in^2.
How: Substitute directly into Equation (12.10) to calculate the cross-sectional areas of these compression members.
Solve: The estimate of cross-sectional area based on yield is the same as in Example 12.4:

$$A_Y = \frac{N\,|F|}{S_Y} = 93.\ \text{in}^2$$

The estimate based on buckling is:

$$A_B = \left(\frac{12N|F|L^2}{\pi^2 E}\right)^{1/2} = \left(\frac{12\left(5.0\left|-6.72\times10^5\right|(174.)^2\left(12^2\right)\right)}{\pi^2\left(29.\times10^6\right)}\right)^{1/2} = 780\ \text{in}^2$$

where the factor of $(12)^2$ converts [ft^2] to [in^2]. The required cross-sectional area of both members is then given by the larger of these two estimates:

$$A = \max\,(93{,}780) = \mathbf{780\ in^2}$$

This result indicates that buckling, not yielding, is the critical failure mode when designing the cross-sections of these members.

Finally, having analyzed both the three-hinged arch and the truss bridge, which design is better? It's too early to tell. The cross-sectional areas of the members in the truss bridge must first be determined, after which the total volume of the members in each design can be calculated. The design with the smallest total volume wins. But these tasks will be left to the end-of-chapter exercises. For now suffice it to say that a competition was held in Austria to determine the best design for a foot-bridge that had to span 174 ft. The truss, with the dimensions shown in Figure 12.14, won.

SUMMARY

In this chapter, a procedure for designing a **statically determinate truss** was outlined. The steps in this procedure may be restated as follows:

1. Estimate the live load.
2. Establish a target value for the structure's overall factor of safety.
3. Choose a material, and look up its yield strength (S_Y) and elastic modulus (E) in a property table.
4. Propose an efficient, fully triangulated member geometry and assign enough dimensions to uniquely locate all the joints.
5. Apply the **Method of Joints** to determine the forces on the members.
6. Calculate the required cross-sectional areas of the members.

If instead of design, the goal is to **evaluate** an existing statically determinate truss with a known live load, the following alternative procedure should be adopted:

1. Apply the **Methods of Joints** to determine the forces on the members.
2. Calculate the factor of safety of each member.
3. Take the smallest of the factors of safety computed in step 2 to be the overall factor of safety of the structure.

Practicing engineers must never apply computational tools without being fully aware of the assumptions involved. In this chapter, there are just a few assumptions to keep in mind. There are the usual truss assumptions—the weights of the members are negligible and the joints can be accurately modeled as pins. There is also the constraint on consideration of bucking that the members must be square in cross-section. The final assumption always has been implied but never mentioned—that members fail before the joints break.

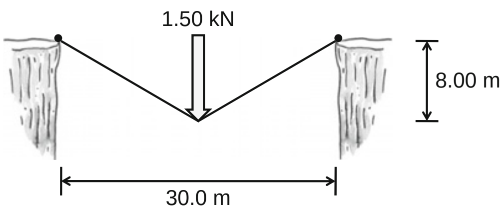
Structural analysis tools similar to these are used in a wide range of applications. Mechanical engineers will use them to analyze the triangulated frame of a formula race car among other things. Material scientists apply truss models, known as statistical lattice models, to simulate progressive damage in materials at very small scales. Bridge design, however, has been and will continue to be a responsibility that belongs to the civil engineers.

EXERCISES

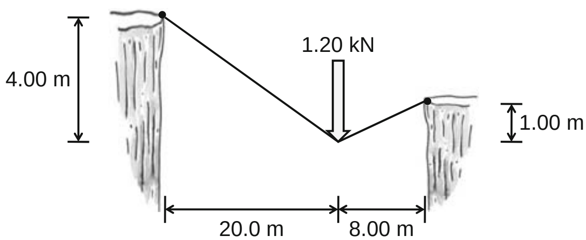
1. When estimating the live load for a new bridge design, you want that estimate to be **conservative**. In other words, you want to error on the safe side by basing the estimate on the worst possible scenario—for example, bumper-to-bumper traffic, the heaviest vehicles, the worst environmental conditions, etc. With this in mind, estimate the live load for the following proposed bridge designs:
   a. A foot bridge with a 174. ft span and two separate 4. ft wide lanes to allow for pedestrian traffic in both directions.
   b. A four-lane highway with a 300. ft span.

   For each of the rope bridges defined in **Exercises 2 through 4**, determine the forces acting on the ends of the rope by (1) drawing the free-body diagram, (2) counting unknowns and equations to check determinacy, (3) writing the equilibrium equations, and (4) solving the equations.

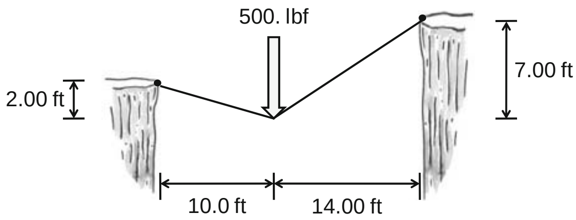
2.



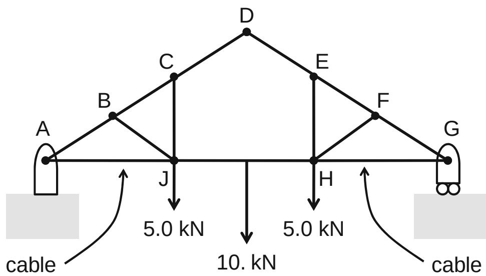
3.



4.



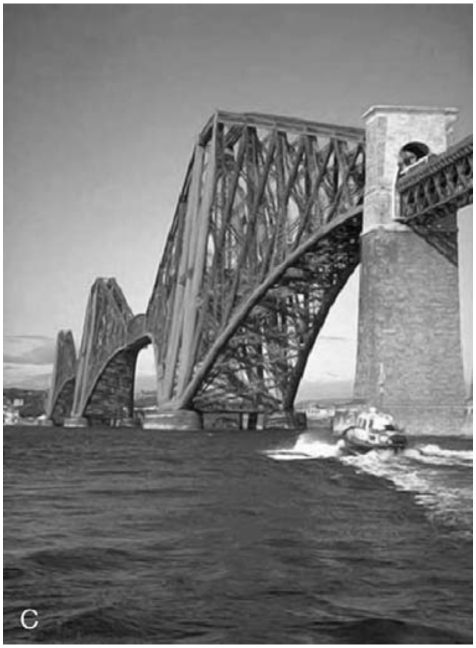
5. For the structure below, do your best to identify all of its members as beams, compression members, or tension members.



6. Photos of four existing bridges are shown. For each bridge, write a paragraph that describes in specific terms the strategies used by the designer to make it an efficient structure. General strategies to draw from include: efficient beam cross-sections, trusses, arches, use of tension members (or cable) to avoid problems with buckling.

A. Akashi Kaikyo Bridge, Japan

B. Bridge of the Americas, Panama

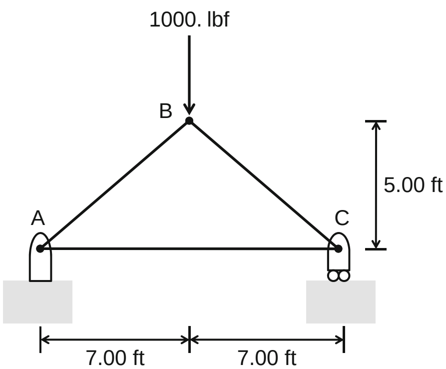


C. Forth Bridge, United Kingdom

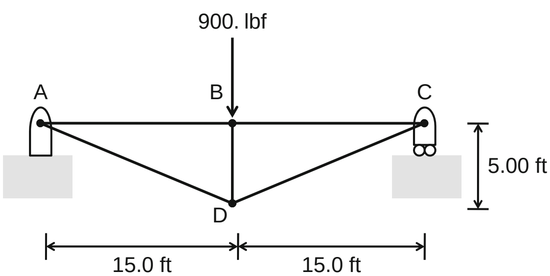
D. Whipple Bridge, USA (By Permission of J. B. Riddel and Associates, www.jbriddel.com.)

For each of the trusses shown in **Exercises 7 through 12**, determine the forces on all members by the method of joints. Use Excel to solve the equations.

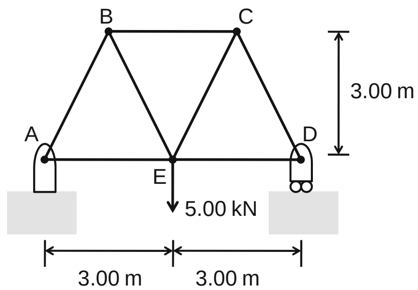
7.



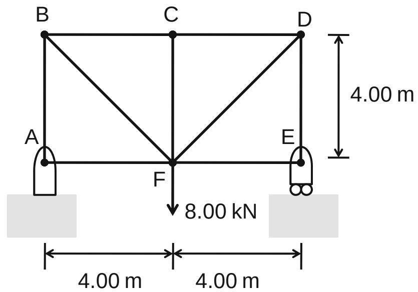
8.



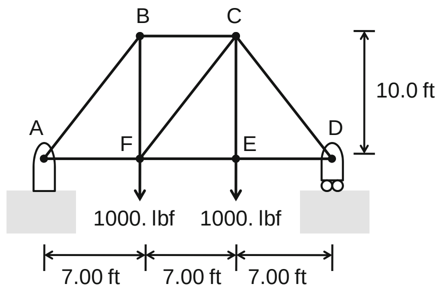
9.


B
C
3.00 m
A
D
E
5.00 kN
3.00 m
3.00 m

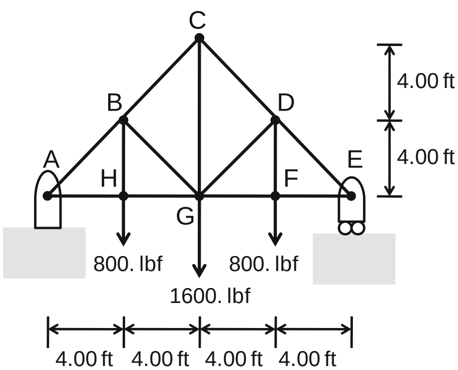
10.


B
C
D
4.00 m
A
E
F
8.00 kN
4.00 m
4.00 m

11.


B
C
10.0 ft
A
F
E
D
1000. lbf
1000. lbf
7.00 ft
7.00 ft
7.00 ft

12.


C
4.00 ft
B
D
A
H
F
E
4.00 ft
G
800. lbf
800. lbf
1600. lbf
4.00 ft
4.00 ft
4.00 ft
4.00 ft

13. A truss that is pinned at one end and on rollers at the other will be statically determinate if it satisfies the following equation:

$$0 = 2j - m - 3$$

where j is the number of joints and m is the number of members. This is a quick way to check for static determinacy without drawing the FBDs. Apply this equation to:

a. Verify that the trusses in Exercises 7 through 12 are statically determinate.
b. Sketch three statically determinate truss bridges with the following numbers of members: (1) 11 members, (2) 15 members, and (3) 19 members.

14. Calculate the overall factor of safety of a truss that has the member forces, lengths, and cross-sectional areas given in the table. Assume steel members ($S_Y = 36. \times 10^6$ lbf/in^2, $E = 29. \times 10^6$ lbf/in^2) with square cross-sections, and consider buckling.

Member	**1**	**2**	**3**	**4**	**5**
Force (lb)	–9060	–9060	12300	12300	5410
Area (in^2)	0.400	0.400	0.500	0.500	0.250
Length (ft)	3.60	3.60	4.20	4.20	3.00

15. Calculate the required cross-sectional areas of the members comprising the truss bridge of Figure 12.14. Refer to Figure 12.17 for the forces and design for a factor of safety of $N = 5.0$. Assume steel members ($S_Y = 36. \times 10^6$ lbf/in^2, $E = 29. \times 10^6$ lbf/in^2) with square cross-sections, and consider buckling.

16. Depending on which of Exercises 7 through 12 you have completed, determine the overall factor of safety of one or more of the following structures. Assume steel members with a yield strength of $S_Y = 36. \times 10^6$ lbf/in^2 (or 250 MPa) and an elastic modulus of $E = 29. \times 10^6$ lbf/in^2 (or 200 GPa). Also assume square cross-sections, and consider buckling.
a. The truss in Exercise 7. Assume $A = 0.50$ in^2 for all members.
b. The truss in Exercise 8. Assume $A = 1.00$ in^2 for all members.
c. The truss in Exercise 9. Assume $A = 4.00$ cm^2 for all members
d. The truss in Exercise 10. Assume $A = 2.25$ cm^2 for all members.
e. The truss in Exercise 11. Assume $A = 1.50$ in^2 for all members.
f. The truss in Exercise 12. Assume $A = 2.00$ in^2 for all members.

17. Depending on which of Exercises 7 through 12 you have completed, determine the required cross-sectional areas of the members comprising one or more of the following structures. Assume steel members with a factor of safety of $N = 3.0$, a yield strength of $S_Y = 36. \times 10^6$ lbf/in^2 (or 250 MPa), and an elastic modulus of $E = 29. \times 10^6$ lbf/in^2 (or 200 GPa). Also assume square cross-sections, and consider buckling.
a. The truss in Exercise 7.
b. The truss in Exercise 8.
c. The truss in Exercise 9.
d. The truss in Exercise 10.
e. The truss in Exercise 11.
f. The truss in Exercise 12.

18. Download the West Point Bridge Designer software from http://bridgecontest.usma.edu and use it to design a truss.

19. In 1985 a judge found the structural engineers for the Hyatt Regency Hotel guilty of gross negligence in the July 17, 1981 collapse of two suspended walkways in the hotel lobby that killed 114 and injured 200 people. Many of those killed were dancing on the 32-ton walkways when an arrangement of rods and box beams suspending them from the ceiling failed.

The judge found the project manager guilty of “a conscious indifference to his professional duties as the Hyatt project engineer who was primarily responsible for the preparation of design drawings and review of shop drawings for that project.” He also concluded that the chief engineer’s failure to closely monitor the project manager’s work betrayed “a conscious indifference to his professional duties as an engineer of record.” Responsibility for the collapse, it was decided, lay in the engineering design for the suspended walkways. Expert testimony claimed that even the original beam design fell short of minimum safety standards. Substantially less safe, however, was the design that actually was used.[5]

Use the Engineering Ethics Matrix to analyze the ethical issues that occurred in this case.

20. Sara, a recent graduate, accepts a position at a small engineering design firm. Her new colleagues form a tightly knit, congenial group, and she often joins them for get-togethers after work.

Several months after she joins the firm, the firm’s president advises her that his wife has objected to her presence on the staff, feeling that it is inappropriate for a young, single female to work and socialize with a group of male engineers, many of whom are married. The president’s wife encouraged him to terminate Sara’s employment, and although the president himself has no issues with Sara or her work, he suggests to her that she should look for another employer.

Everyone in the firm become aware of the wife’s objections, and Sara begins to notice a difference in her work environment. Although her colleagues are openly supportive of her, she nevertheless feels that the wife’s comments have altered their perception of her. She stops receiving invitations to her company’s parties and is excluded from after-hours gatherings. Even worse, although previously she had found her work both interesting and challenging, she no longer receives assignments from the firm’s president and she begins to sense that her colleagues are treating her as someone who will not be a long-term member of the staff. Believing that she is no longer taken seriously as an engineer and that she will have little opportunity to advance within the firm, she begins searching for a new position. However, before she can do so her supervisor announces a downsizing of the firm, and she is the first engineer to be laid off.[6]

Use the Engineering Ethics Matrix to examine potential ethics violations for an engineering employer to exclude and ultimately discharge an employee on the basis of sex, age, or marital status.

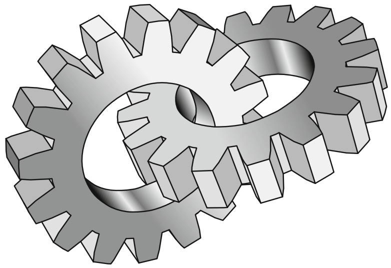
---

[5]Modified from http://wadsworth.com/philosophy_d/templates/student_resources/0534605796_harris/cases/Cases/case68.htm

[6]Modified from http://pubs.asce.org/magazines/ascenews/2008/Issue_01-08/article6.htm

CHAPTER

13 Engineering Kinematics



Source: © iStockphoto.com/Aksenov Vasilii

This subject is part of several engineering disciplines—mechanical, civil, biomedical, to name just three. We mentioned in previous chapters how mechanical engineers use equations of motion. We have included a civil engineering topic in this chapter, highway traffic flow. We also have included the topic of gears and gearboxes, which is of interest to mechanical engineers. Had we the space we also could have included areas of kinematics for bioengineers who are interested in how the human body responds to being subject to abnormal situations (such as in a fall). Electrical, computer, and mechatronic engineers are interested in the motion of robots and use multidimensional kinematic analysis to design them.

13.1 WHAT IS KINEMATICS?

Kinematics is the study of how things move. In this chapter we develop visual and mathematical models that describe the motion of real objects. Much of this chapter is also about how engineers understand the relationships among the important variables of distance, speed, acceleration, and time. These relationships are collectively called kinematics when they do not also involve the forces on objects and the inertia of the objects. Kinematics simply serves as a way to relate the effects of forces on a body to its subsequent motion. In the simplest cases distance, time, speed and acceleration are related to each other by geometric methods but in more complicated situations, by the methods of the calculus.

In this chapter we use kinematic methods to understand the design of on- and off-ramps for highways and the kinematic analysis of why and how roads become crowded. In addition, we use a kinematic-based analysis to describe the first steps in the design of power transmission systems that use gears.

13.2 DISTANCE, SPEED, TIME, AND ACCELERATION

Determining how long an on-ramp should be on a highway requires understanding what an on-ramp is required to do. That requirement might be expressed as follows: An on-ramp should be long enough to enable a driver to speed up gradually and safely from the off-highway speed (probably about 15–30 mph) to the highway speed (probably above 60 mph). As always, the engineers translate this qualitative statement into variables, numbers, and units.

"Long enough" becomes the variable distance. It is measured in units of miles or meters, or whatever scales are appropriate to the problem at hand. As a mental exercise, at this point you might want to estimate in meters, based on your memory, how long you think a typical highway on-ramp is.

In this chapter, only motion in a single direction (often called in techno-ese **one-dimensional**) will be considered. For convenience, think of positive distance as from left to right, and negative distance as from right to left, just as in Cartesian geometry.

Speed is a variable commonly measured in miles per hour (mph) in the United States and in kilometers per hour (kph) in most of the rest of the world. But for engineering design, the SI units of m/s are often more useful (65. mph ~ 105. kph ~ 29. m/s). Speed is calculated by measuring the distance traveled and the time spent and dividing distance by time. In mathematical terms, speed = $\Delta x/\Delta t$, with the uppercase Greek delta Δ symbol meaning "difference." In Cartesian geometry, this is the average slope of the line $x(t)$. In this case speed means the final position minus the initial position divided by the final time minus the initial time (Figure 13.1).

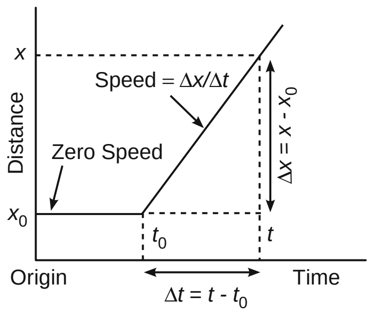


FIGURE 13.1 Constant Speed

Again, only one direction will be considered. We will use the symbol v (as in velocity) for speed (so as not to confuse it with the symbol 's' for seconds). There is an important differentiation between the concepts of *speed* and *velocity*. Speed is the *magnitude* of velocity. Velocity may have more than one component, which is so much in the y direction (as in Cartesian geometry) and so much in the x direction, and so on. Of course, in our one-dimensional analyses, speed and velocity are functionally synonymous. For acceleration we use a, except for that due to gravity, when we use g.

Adding a word to "speed" gets us to the phrase **speed up**. We already have met the speed-up variable, *acceleration*. It is defined as change of speed per unit time or, more mathematically precisely, as the rate of change of speed, and is measured in m/s^2 or ft/s^2. Resist the temptation to express acceleration in such ugly and ultimately less useful hybrid forms as "miles per hour per second."

Finally, there is a fourth variable. It is mentioned only implicitly in the preceding description, but it is so important that, without formally introducing it, we already have used it. This is the variable *time*, measured in seconds in computation but, in the context of highway analysis, often in hours as in miles per hour, contracted as mph.

Strictly speaking, two of the preceding variables should be defined in two ways: in terms of *instantaneous* speed and acceleration and in terms of *average* speed and acceleration. These are important distinctions, as shown in Figure 13.2. An instantaneous variable, here the acceleration, is changing with the change in slope. Your physics courses will enable you to understand how to deal with nonconstant accelerations, but in this introductory treatment we can, and will, get along without nonconstant accelerations.

In an earlier chapter you learned the simplest principles of **kinetics**, the study of motion caused by applied forces. This chapter is concerned with the closely related subject of kinematics, the relationship between

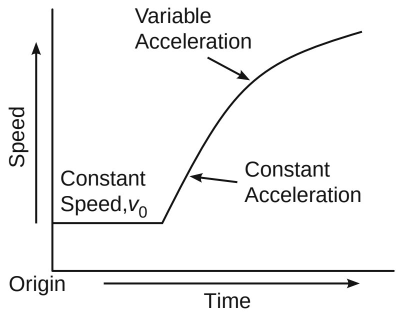


FIGURE 13.2 Acceleration is the *Local* Slope

distance, speed, acceleration, and time, without asking the question of how we can achieve particular values of these variables. (Although, of course, we can calculate the associated momentum changes, forces, and energy given information on the mass or inertia of the bodies being accelerated and so forth.)

The only one of the four concepts of distance, speed, acceleration, and time that gives the typical student any difficulty is acceleration. So before proceeding further, please follow this example.

Example 13.1

A car enters an on-ramp traveling 15. miles per hour. It accelerates for 15. seconds. At the end of that time interval it is traveling at 60. miles per hour. What is its average acceleration? How does that compare to *g*, the acceleration due to gravity?

Need: Acceleration = ______ m/s^2.

Know–How:

Acceleration = (change in speed)/time.

Initial speed = 15. miles per hour = 6.7 m/s.

Final speed = 60. miles per hour = 27. m/s.

Solve: Average acceleration = (27. − 6.7)/15. = 1.3 m/s^2 (in terms of a fraction of *g*, the acceleration due to gravity, this is 1.3/9.8 = 0.13, so you know this is a mild acceleration).

13.3 THE SPEED VERSUS TIME DIAGRAM

We could deal with problems of distance, speed, acceleration, and time using words and equations, but there is a much more intuitive tool. It makes possible insightful descriptions of motion problems. It frequently enables an engineer to solve the problem merely by inspecting the diagram. Formal calculation may be needed only to achieve the correct number of significant figures.

This versatile and essential tool is the speed-versus-time diagram (Figure 13.2). For brevity, we will call our diagram the $v-t$ diagram. The $v-t$ diagram is just what it says: The horizontal axis represents time, with zero being the instant the situation being considered began. For each instant of time, the speed (or one-dimensional velocity) is plotted in the vertical direction.

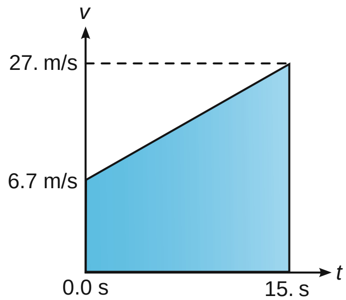
Example 13.2

Plot the $v-t$ diagram for Example 13.1.

Need: $v-t$ diagram describing the situation "a car enters an on-ramp traveling 15 miles per hour. It accelerates constantly for 15 seconds. At the end of that time interval it is traveling at 60. miles per hour."

Know–How: First, convert all speeds to SI. Then plot speed for $t = 0$ (when $v = 15.$ mph $= 6.7$ m/s) and $t = 15.$ s (when $v = 60.$ mph $= 27.$ m/s). Join the two points with a *straight line*. (What tells us the line is straight is the phrase "constant acceleration.")

Solve:



Why is this particular graph so important? There are two reasons.

1. The slope of the $v-t$ graph measures acceleration.
2. The shaded area under the $v-t$ graph measures distance traveled.

You should recall *slope* being defined as the quotient of a vertical rise divided by a horizontal run. For the standard Cartesian x, y diagram, the slope is $\Delta y/\Delta x$. That the slope of the $v-t$ graph gives you acceleration is evident from considering dimensions. The rise of the graph has units of speed, m/s. The run of the graph has units of time, s. So slope = rise/run = $\Delta v/\Delta t$ in units of [m/s]/[s] or m/s^2, valid dimensions of acceleration.

The second statement relating area under the curve to distance traveled is less obvious. In Example 13.2, the *average* speed over the 15-second period of acceleration is $0.5 \times (27. + 6.7) = 17.$ m/s, and so the vehicle will have covered $17. \times 15. = 250$ m. Notice the shaded area is a trapezoid whose area is 0.5 × (sum of parallel sides) × perpendicular distance between them. It is identical to the calculation of distance covered. It also can be approached dimensionally. The "height" or ordinate has units of [m/s]. The horizontal axis has units of seconds. So length × height has units of [s × (m/s)] = [m], a unit of distance.

Those of you familiar with elementary calculus will recognize that saying "the area under the $v-t$ graph measures distance" is the same as saying "the integral of speed over time is distance." Saying that "the slope of the $v-t$ graph measures acceleration" is the same as saying "the derivative of speed is acceleration." These statements are true even when the lines are curved—that is, for the case of nonconstant acceleration. In the general case, calculus is required to evaluate the slopes or areas under curves.

We will, however, get along without calculus. Solving Examples 13.1 and 13.2 using the $v-t$ diagram has done more than enable us to visualize that problem. It has provided a bonus. It has already answered a second question.

Example 13.3

What is the distance traveled in steadily accelerating from 15. mph to 60. mph in 15. seconds?

Need: Distance = ______ m.

Know–How: The distance is the area (shaded in the diagram for Example 13.2) beneath the $v-t$ graph. Geometry tells us we can break the area of a trapezoid into a rectangle with length 15. s and height 6.7 m/s and a triangle with base 15. s and altitude (27. − 6.7) s.

Solve: Distance = Area = area of rectangle + area of triangle

= 6.7 [m/s] × 15. [s] + (½) × 15. [s] × (27. − 6.7) [m/s] = 250 m

Therefore, **Distance = 250 m**

You may already know some formulae you could have used to solve this problem ($x = v_0t + ½\ at^2$, for example), but it is recommended that you use the $v-t$ diagram on the problems and exercises of this chapter. Use of this tool develops a visual and intuitive appreciation of motion problems that cannot be obtained merely by manipulating equations. Formulae are useful once you fully understand the process, but the $v-t$ diagram is the best way to achieve that understanding.

13.4 APPLYING KINEMATICS TO THE HIGHWAY ON-RAMP PROBLEM

Let's now apply the tool to our original problem. Recall its verbal statement: "An on-ramp should be long enough to enable the driver to speed up gradually and safely from the off-highway speed (probably about 15–30 mph) to the highway speed (probably above 60 mph)."

A few words remain untranslated into numbers. In particular, consider the words "gradually" and "safely." What constitutes a gradual and safe acceleration? This is a matter for judgment, for experience, for experiment, but not for definition. Judgment and experience can begin from a well-known acceleration that we have earlier considered. This is the acceleration due to gravity. This acceleration often is called "1 g" or colloquially "1 gee" of acceleration (where $g = 9.8\ m/s^2$ and is not to be confused with the symbol g for "gram.")

Is 1 g too high, too low, or just right for an automobile accelerating along an on-ramp? Experience reveals that a 1 g acceleration is appropriate for amusement park rides, but not for on-ramps. A driver accelerating onto an on-ramp should accelerate with only a fraction of a g.

What fraction? This presents another problem of engineering optimization. If consideration of the value of the driver's time, the cost of real estate, and the cost of concrete did not matter, the fraction could be very small. The driver could accelerate at a leisurely and safe rate along an on-ramp that stretched for many kilometers!

Such considerations do matter, however. These considerations suggest a short on-ramp and high acceleration.[1] Safety suggests a long on-ramp and correspondingly low acceleration. The engineer's job is to compromise between the two. In order to do so, the engineer needs a tool that relates length of the on-ramp to acceleration, and this is exactly what the $v-t$ diagram provides.

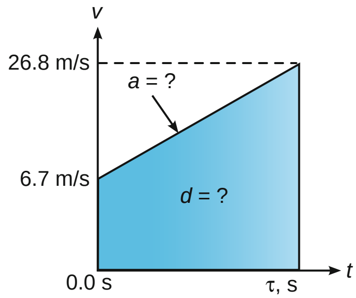
[1]Of course, high acceleration wastes additional fuel, and, anyway, not all vehicles are capable of high acceleration.

Example 13.4

Provide a table and graph of the length of an on-ramp as a function of acceleration for accelerations ranging from 1.0 m/s^2 (about 0.1 g) to 10. m/s^2 (about 1 *g*) in increments of 1.0 m/s^2. Assume that the vehicle enters the on-ramp with a speed of 15. mph (6.7 m/s) and leaves the on-ramp at a speed of 60.0 mph (26.8 m/s).

Need: A table and a graph with entries length of on-ramp *d* in m vs. *a* m/s^2.

Know–How: The *v*–*t* diagram gives the relationships among *a*, *d*, and *t*. Let τ be the (unknown) time you spend on the on-ramp accelerating to highway speed.



From the definition of acceleration as the slope of the *v*–*t* line,

$$a = \frac{(26.8 - 6.7)}{\tau} \quad \text{or}$$

$$\tau = \frac{(26.8 - 6.7)}{a}[m/s][s^2/m] = \frac{20.1}{a}[s]$$

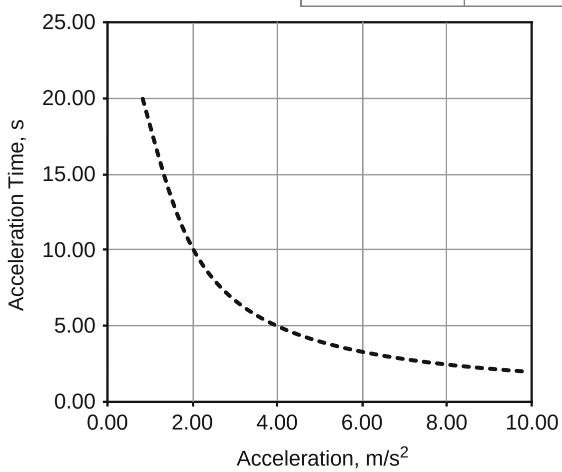
d is the area under the *v*–*t* curve

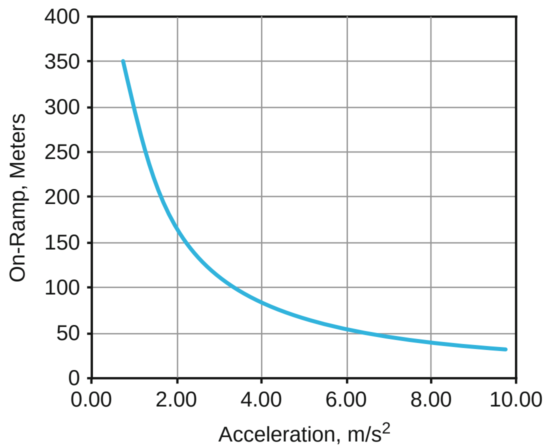
$$d = 6.7\tau + 1/2 \times (26.8 - 6.7)\tau\ [m/s][s] = 6.7\tau + 10.05\tau\ [m]$$

Solve: Now substitute for $\tau = 20.1/a$. Therefore $d = (135/a) + 202/a = 337/a$ in meters.

Take these relationships to a spreadsheet and prepare the table, and from the table, prepare graphs of *a* versus τ and *d* versus τ:

***a*, m/s^2**	***d*, m**	**τ, s**
1.00	337	20.10
2.00	169	10.05
3.00	112	6.70
4.00	84	5.03
5.00	67	4.02
6.00	56	3.35
7.00	48	2.87
8.00	42	2.51
9.00	37	2.23
10.00	34	2.01





This table quantifies the trade-off between acceleration and on-ramp length. That trade-off is then combined with the considerations mentioned earlier. Those considerations, to repeat, include the value of the driver's time, the cost of real estate, and the cost of concrete. Such considerations give which point on the d versus a curve the engineer will actually choose. Experience shows that the trade-off results in a point somewhere around $a = 3.$ m/s^2 (about one-third of a g) for cars, and the on-ramp is then $d = 110$ m. In somewhat more automotive terms, with $a = 3$ m/s^2, the time for a standing quarter (of a mile) is a relatively leisurely 16 s and time to reach 60 mph is 9 s. Obviously, many cars can do much faster than that—which is another trade-off, this one between gasoline consumption and fuel economy.

If the appropriate acceleration for a heavy truck is one-half that for most cars at 1.5 m/s^2, the on-ramp has to be 250 m. This value of about 250 m is probably pretty near the one you guessed from personal experience when you began this chapter. But now you know the right answer for a better reason.

The answers to a number of other engineering problems are also contained in the v–t diagram. These range from the design of intersections, stoplights, and traffic lights, and even to the design of a cannon that might be used to shoot astronauts to the Moon. These subjects are explored in the exercises.

13.5 GENERAL EQUATIONS OF KINEMATICS

The key variables are:

t, the time, typically measured in seconds
x, the distance traveled, typically measured in meters
v, the speed (or velocity in one dimension), typically measured in m/s
a, the acceleration, typically measured in m/s^2

If the speed is constant at $v = v_0$ and the acceleration is zero ($a = 0$), then the distance–time kinematic equation is:

$$x = x_0 + v_0\,(t - t_0) \tag{13.1}$$

If the acceleration is constant, then the distance-time kinematic equation is:

$$x = x_0 + v_0\,(t - t_0) + 0.5a\,(t - t_0)^2 \tag{13.2}$$

and the speed-distance kinematic equation is:

$$v^2 = v_0^2 + 2a\,(x - x_0) \tag{13.3}$$

13.5.1 Simplified Forms

Notice that if $x_0 = 0$ at $t_0 = 0$ in Equation (13.1), then $x = v_0 t$. Also, if $t_0 = 0$ and $x_0 = v_0 = 0$, then Equation (13.2) gives $x = 0.5\ at^2$. And finally, if $x_0 = 0$ and $v_0 = 0$, then Equation (13.3) gives $v^2 = 2ax$.

More general relationships for nonconstant a can be derived using calculus. In any case, we will not be deriving them here—all cases, constant a or not, will be covered in your first-year physics classes.

13.6 THE HIGHWAY CAPACITY DIAGRAM

How many cars can a superhighway deliver to a city at rush hour? Consider a single lane of such a highway. Imagine, at rush hour, you stood beside the highway and counted the number of cars in any one lane that passed you in the space of one hour (assuming you have sufficient patience). What sort of answer do you think you might get?

It's unlikely that your answer was less than 100 cars per hour, or more than 10,000 cars per hour. But how can we narrow down further and justify an answer? As in the previous parts of this chapter, the goal is not a single answer but the basis for a trade-off between the key variables. The key (and closely interrelated) variables in the case of highway capacity are:

Capacity (cars per hour): The number of cars that pass a certain point during an hour.
Car speed (miles per hour, mph): In our simple model we will assume all cars are traveling at the same speed.
Density (cars per mile): Suppose you took a snapshot of the highway from a helicopter and had previously marked two lines on the highway a mile apart. The number of cars you would count between the lines is the number of cars per mile.

You can easily write down the interrelationship among these variables by using dimensionally consistent units:

$$\textbf{Capacity} = \textbf{Speed} \times \textbf{Density} \tag{13.4}$$

$$\text{with units of [cars/hour]} = \text{[miles/hour]} \times \text{[cars/mile]}$$

For simplicity, we will consider these variables for the case of only a single lane of highway. Once again, our initial goal is a tool that provides insight into the problem, and helps us solve it intuitively and visually. As a first step toward that tool, study Example 13.5.

Example 13.5

Three of you decide to measure the capacity of a certain highway. Suppose you are flying in a helicopter, take the snapshot mentioned earlier, and count that there are 160 cars per mile. Suppose further that your first partner determines that the cars are crawling along at only 2.5 miles per hour. Suppose your second partner is standing by the road with a watch counting cars as described earlier. For one lane of traffic, how many cars per hour will your second partner count?

Need: Capacity = ______ cars/hour.
Know–How: You already know a relationship between mph and cars per mile from Equation (13.4).
Capacity = Speed × Density
[cars/hour] = [miles/hr] × [cars/mile]
Solve: **Capacity** = [2.5 miles/hr] × [160 cars/mile] = **400. cars/hour**.

Now let's use the result of Example 13.5 to plot a new graph. It's one of two related graphs, each of which can be called a **highway capacity diagram** (Figure 13.3). It is made by plotting car density on the vertical axis and speed, in mph, on the horizontal axis. In this case, it gives us the highway capacity.

How might Figure 13.3 be made even more useful? If we knew a **rule**[2] relating car density and mph, we could plot a whole curve instead of just one point. This would make the highway capacity diagram a lot more useful.

You possibly already know such a rule either explicitly or implicitly. Let's consider two such rules. One is traditional but no longer recommended. Its implications are, however, worth exploring as a first pass at highway design. That rule is called the **follow rule**. You can see it at work visually in Figure 13.4.

[2]A *rule* is something guided by experience as being useful for a given objective but not based on one of the hard-and-fast laws of Nature.

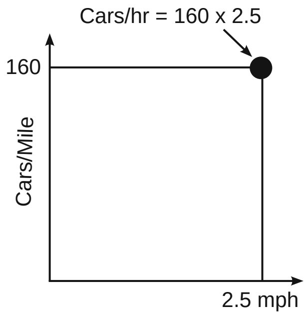


FIGURE 13.3 Highway Capacity Diagram

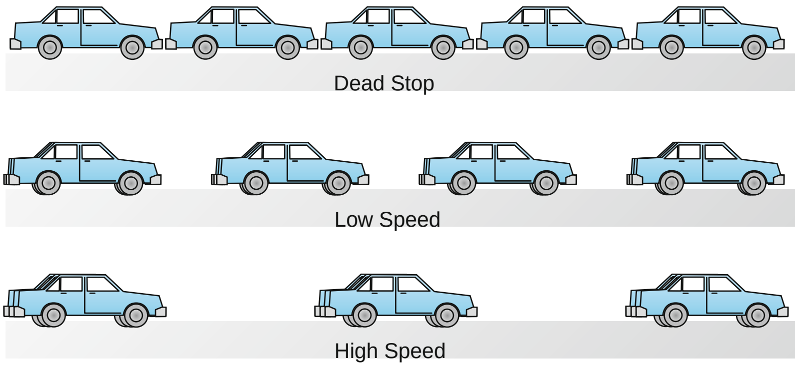


FIGURE 13.4 The **Follow Rule** in Terms of Car Length Separation

In words, one version of this rule states that when following another car, leave one car length between vehicles for every 10. miles per hour. Write the follow rule as Equation (13.5):

$$\textbf{Number of car lengths between cars} = [\textbf{speed in mph}]/[\textbf{10.mph}] \tag{13.5}$$

The effect of this rule is illustrated graphically in Figure 13.4. The separation distance per car is more as the car speed increases. At a dead stop the cars are bumper to bumper. Thus, the car density falls with increasing speed.

In order to use this rule to calculate a highway capacity in cars/hour, it needs to be put into terms compatible with Equation (13.1). The independent variable (subject to the constraints of common sense and the law) is the vehicle's speed. Thus, we need to use the follow rule to relate to the unknown car density. We know the length of a car—say, about 4.0 m on the average. If we do the arithmetic, this means we could pack about 400. vehicles/mile bumper-to-bumper. The follow rule relates the distance between these vehicles to their speed. Each car effectively occupies its own footprint plus the separation distance between them. The car density is thus:

$$\text{Density} = \frac{1}{\left(\dfrac{\#\text{ car lengths} + 1}{400.}\right)} = \frac{400.}{\left(1 + \dfrac{\text{speed, [mph]}}{10.\text{ [mph]}}\right)}\text{[cars/mile]} \tag{13.6}$$

Combining Equations (13.4) and (13.6) we get:

$$\text{Capacity} = \frac{400. \times \text{speed, mph}}{\left(1 + \dfrac{\text{speed, [mph]}}{10.\ \text{[mph]}}\right)}\ [\text{cars/hr}] \tag{13.7}$$

Example 13.6

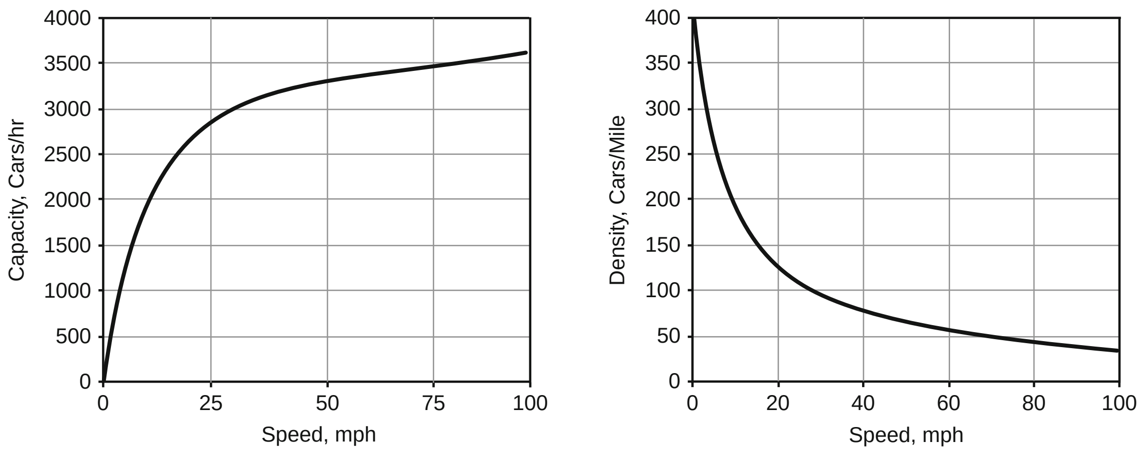
Using the follow rule, draw the full highway capacity diagram.

Need: Table and chart.

Know–How: Put the follow rule into a spreadsheet and graph.

Solve:

Speed, mph	Density, cars/mile	Capacity, cars/hr
0	400	0
10	200	2000
20	133	2667
30	100	3000
40	80	3200
50	67	3333
60	57	3429
70	50	3500
80	44	3556
90	40	3600
100	36	3636



Either of these two curves appropriately could be called a "highway capacity diagram." Notice what happens to the capacity as the car gains speed: It begins to level off and approach a constant value of 4,000 cars/hr. There is a gain of only 300 cars/hour as we increase the speed from 50 to 100 mph.

The Equation (13.7) for capacity developed earlier shows why this is occurring: The denominator of the equation is dominated by the spacing between vehicles and not by the footprint of the car. The former was expressed mathematically as speed in mph/10. mph » 1 so that at high speeds,

$$\text{Capacity} = \frac{400. \times 10.}{\text{speed, [mph]}} \times \text{speed, [mph]} = 40. \times 10^2 \text{ [cars/hr]}$$

Notice too that you can always deduce the third member of the triumvirate of speed, density, and capacity by manipulating the other two from Equation (13.4). For example, multiplying the density by the speed gives the capacity at every point (as per Figure 13.3), or you can get the density by dividing the capacity by the speed.

An up-to-date version of the follow rule is the **two-second rule**.[3] It is practiced as follows. First, mark the point on the side of the road that the car in front of you is passing right now. Next, count the number of seconds[4] that it takes you to reach that point. If you count less than two seconds, you are following too closely.

The two-second rule can be expressed mathematically by making the distance between cars a function of speed.

$$\textbf{Miles/car} = \boldsymbol{v} \times \boldsymbol{t}\ [\textbf{miles/hr}][\textbf{s}][\textbf{hr/s}] = \boldsymbol{v}\ [\textbf{mph}] \times \textbf{2/3600}\ [\textbf{hr}]$$

or, inverting this,

$$\textbf{The car density}, [\textbf{cars/mile}] = \textbf{1800}/\boldsymbol{v}$$

Equally,

$$\textbf{speed}\ [\textbf{mph}] = \textbf{1800/density}\ [\textbf{cars/mile}]$$

Hence, using Equation (13.1), capacity = speed × density, or dimensionally as [cars/hour] = [miles/hr] × [cars/mile] = 1,800—a fixed capacity independent of the car speed!

Example 13.7

Using the two-second rule, draw the full highway capacity diagram.

Need: Table and chart.

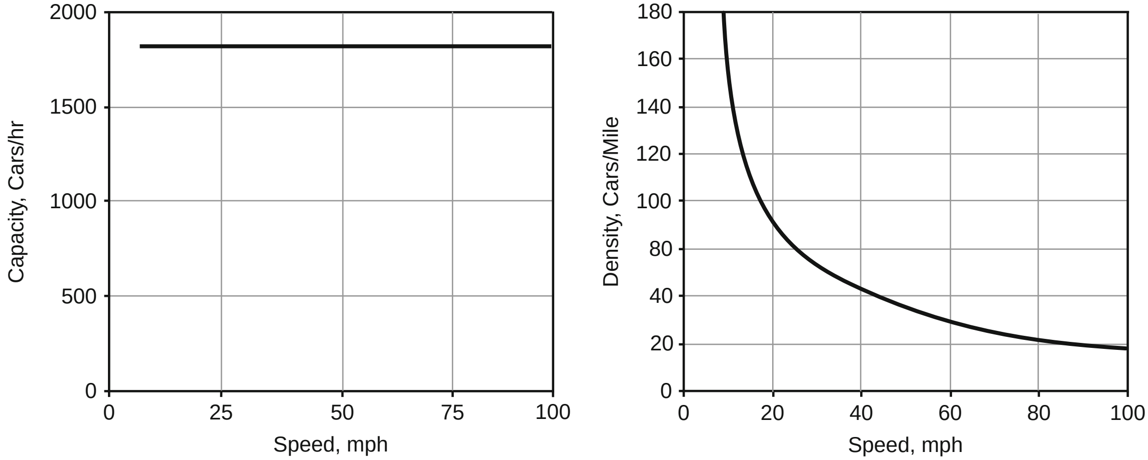
Know–How: Put the two-second rule into a spreadsheet and graph.

Solve:

Speed, mph	Density, cars/mile	Capacity, cars/hr
10	180	1800
20	90	1800
30	60	1800
40	45	1800
50	36	1800
60	30	1800
70	26	1800
80	23	1800
90	20	1800
100	18	1800

[3]This is less arbitrary than it sounds; the average human reaction time is about 0.7 s, leaving 1.3 seconds to decelerate from 65 mph (29 m/s) to zero. The average deceleration rate is thus (0. − 29.)/1.3 ≈ −2¼ g. This means you will feel as though you weigh 2¼ times your normal weight for 1.3 s—an unpleasant prospect!

[4]Count as "one, one thousand, two one thousand"—this will take about 2 seconds total to say in a normal speech pattern.



Notice that the two-second rule requires a greater distance between cars (as expressed by the calculated car density) than does the follow rule and, as just calculated, cuts the high speed highway capacity roughly in half! By the two-second rule, no matter how fast the cars are going, the number of cars that arrive at the destination each hour is the same! All that drivers accomplish by speeding up is to lengthen the distance between them and the surrounding cars (assuming they are all obeying either the follow rule or the two-second rule) and to lower the density of cars.

The implications are that for any superhighway, if the drivers observe the follow rule, the capacity of each lane for delivering cars is a little less than 4,000 cars per hour, or 1,800 cars per hour with the two-second rule, *no matter what the speed limit might be*. This insight enables engineers to solve practical problems such as the following.

Example 13.8

You are a highway engineer hired by the New York State Thruway Authority. About 5,000 citizens of Saratoga, New York, work in the nearby state capital, Albany, New York. They all have to arrive in Albany during the hour between 8 and 9. These workers all travel one to a car, and all drivers obey the two-second rule. How many highway lanes are needed between Saratoga and Albany?

Need: Number of highway lanes = _____ lanes.

Know–How: We have (apparently) just discovered the universal law of highways: A highway lane can deliver 1,800 safe drivers per hour to their destination using the two-second rule.

So, number of highway lanes = number of drivers/hr/(1,800 drivers/hr/lane).

Solve: Assume the number 5,000 /1,800 drivers/hr is uniform. Then the number of highway lanes = 3 lanes (to the proper number of significant figures!).

Not all drivers obey the two-second rule (surprise!). Highways (such as Interstate 87 that links Saratoga and Albany) do not merely link two cities but many cities. (I-87 runs from the Canadian border to New York

City.) Even within a small city, a highway typically has more than one on-ramp and off-ramp. Although some highways run between cities, others are beltways that ring a city.

To meet these more realistic conditions, more complicated versions of the traffic rules, as well as many additional considerations, are needed. Further, engineers have devised criteria for rules called level-of-service from A (major interstate highways) to F (rutted, damaged, under repair, etc., where one cannot exceed 10 mph).

13.7 THE ROTATIONAL KINEMATICS OF GEARS

In addition to the conventional concept of speed, engineers must also deal with the concept of **rotational speed**. Many familiar devices rely on rotary motion. You are surely familiar with bicycle wheels, steering wheels, and hard drives for computers, to name just a few. Although the subject of gears normally is reserved as a case in mechanical design (which it is), there is an aspect in which rightfully belongs in the study of kinematics. It rests on two expressions for speed, one relating to the linear equivalent of wheel rim speed and one to its equivalent in rotational speed.

Rotational speed is defined in two ways, the more familiar being N, the *revolutions per minute* (RPM) of a wheel. There is also a corresponding scientific unit of rotational speed in terms of circular measure, radians/*s*. Its symbol is the Greek lowercase letter omega (ω). There are 2π radians in a complete circle. Hence, $N = 60 \times \omega/2\pi$ [s/minute] [radians/s] [revolution/radian] = RPM; conversely:

$$\boldsymbol{\omega = 2\pi N/60} \textbf{ (in radians/s when } \boldsymbol{N} \textbf{ is in RPM)} \tag{13.8}$$

Angular speed ω is also related directly to **linear speed** v. Each revolution of a wheel of radius r covers $2\pi r$ in forward distance per revolution. Therefore, at N RPM the wheel's tangential speed is $v = 2\pi rN/60 = r\omega$ (in m/s if r is in meters). Hence,

$$\boldsymbol{v = r\omega} \tag{13.9}$$

In our previous chapter on mechanical engineering we developed models of how internal combustion engines worked. However, we did not pay any attention to the problem of linking up the engine speeds with a vehicle's wheels. We have seen Otto cycle engines that can sustain rotation speeds of approximately 1000 to 7000 RPM (as high as 12,000 RPM in very high-performance vehicles), but we have vehicles that are moving at linear speeds of, say, 115 km/h (32. m/s). Suppose the tire outer diameter is 0.80 m (radius of 0.40 m). What is the wheel's rotational speed?

In our current case, the wheels are rotating at a circular speed corresponding to the formula $r\omega = v = 32.$ m/s. Hence, $\omega = v/r = 32./0.40 = 80.$ radians/s or $80. \times 60/2\pi = 760$ RPM. Somehow the rotational speed of the engine must be transformed into the rotational needs of the wheels. How can these two different speeds of rotation be reconciled?

It is done by a mechanism called a transmission. A manual transmission is made of several gears.[5] A gear is simply a wheel with a toothed circumference (normally on the outside edge; see Figure 13.5).

A gear **set** or gear **cluster** is a collection of gears of different sizes, with each tooth on any gear having exactly the same profile as every other tooth (and each gap between the teeth being just sized to mesh). The teeth enable one gear to drive another—that is, to transmit rotation, from one gear to the other. **Note**: A simple gear pair as per Figure 13.5 *reverses* the rotational direction of the driven gear from that of the driving gear. You need at least three gears in a set of simple gears as per Figure 13.6 to transmit in the same direction as the original direction.

[5]In this book, all the gears are toothed and intermeshing; other types of gears such as ratchet wheels and "worm" gears are excluded from the present discussion.

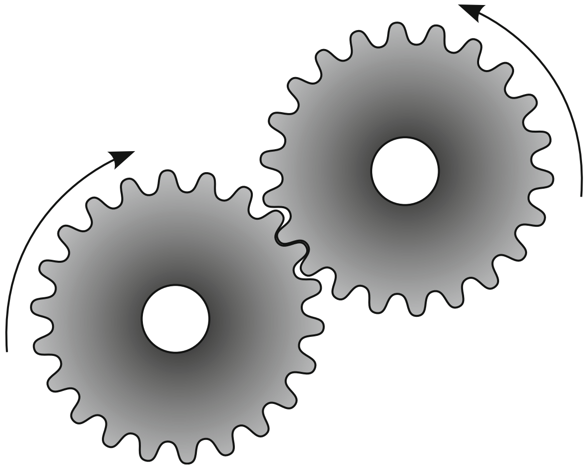


FIGURE 13.5 Intermeshing Gears

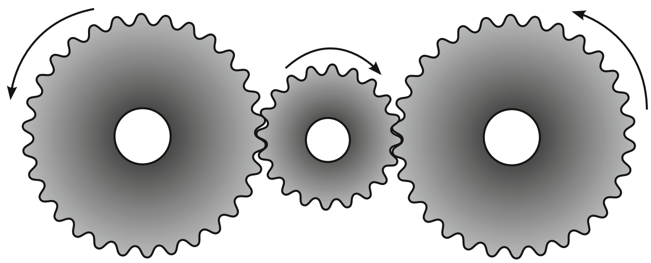


FIGURE 13.6 Nonreversing Gear Set

There exists a simple relation between the rates of revolution of two gears when one gear is driving the other. The essential kinematic feature is that they must have identical tangential or linear speed at their point of contact (or, at best, the gears will slip, and, at worst, the gear teeth will break). If gear 1 has radius r_1 and angular speed ω_1, and gear 2 has radius r_2 and angular speed ω_2, then the common tangential velocity is $\omega_1 r_1 = \omega_2 r_2$, so that $\omega_2/\omega_1 = r_1/r_2$. This relationship applies to all no-slip systems such as smooth friction wheels and toothed gears. The relationship can be described in terms of gear radii as noted earlier or equally in terms of the number of teeth, since the number of teeth must scale directly with the gear radius, given that the meshing teeth must be of equal size.

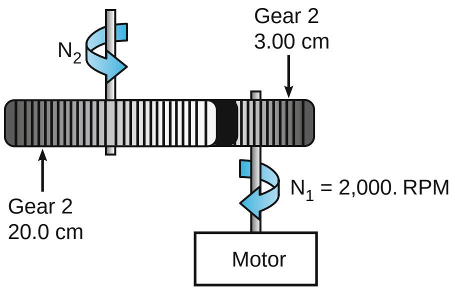
In the context of gears, it is more usual to think in terms of N, the RPM, rather than in terms of angular velocity ω, in radians per second, so that the **velocity ratio** for a simple gear train is:

$$\textbf{Velocity Ratio} = \boldsymbol{VR} = \frac{\textbf{Output rotation}}{\textbf{Input rotation}} = \frac{\boldsymbol{N_2}}{\boldsymbol{N_1}} = \frac{\boldsymbol{r_1}}{\boldsymbol{r_2}} = \frac{\boldsymbol{d_1}}{\boldsymbol{d_2}} = \frac{\boldsymbol{t_1}}{\boldsymbol{t_2}} \quad (13.10)$$

in which d stands for **diameter** and t stands for the **number of teeth** per gear. In other words, this relation, known as the velocity ratio[6] or ***VR***, tells us that to increase the RPM (make the driven gear turn faster than the driving gear) choose a driven gear that is smaller than the driving gear. To decrease the RPM (make the driven gear turn slower than the driving gear) choose a driven gear that is larger than the driving gear.

Example 13.9

An engine crankshaft is turning at 2000. RPM and is connected to a gear of radius 3.00 cm. That gear in turn is driving a gear of radius 20.0 cm. What is the RPM of the driven gear?



Need: RPM of driven gear, N_2.
Know: Speed of driving gear $N_1 = 2000.$ RPM.
Radius of driving gear (r_1) is 3.00 cm.
Radius of driven gear (r_2) is 20.0 cm.

A sketch as per the inset is recommended.

How: Use the VR equation $\frac{N_2}{N_1} = \frac{r_1}{r_2}$

Solve: $\boldsymbol{N_2} = 2000.[\text{RPM}] \times 3.00/20.0[\text{cm/cm}] = \mathbf{3.00 \times 10^2}$ **RPM**.

Note that this example has solved the problem posed earlier: how to reconcile a high-speed engine crankshaft with the lower speed demanded by the wheels. Figure 13.7 is a highly simplified semiconceptual diagram of how this is achieved in an automobile.

Compound gear sets—sets of multiple interacting gears on separate shafts, as shown in Figure 13.8, can be very easily treated using the gear ratio concept introduced earlier.

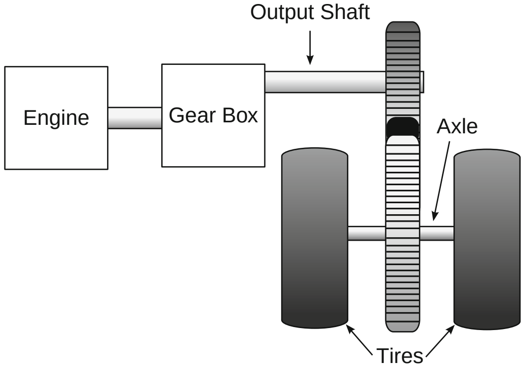


FIGURE 13.7 A Simple Gear Train

[6]Gear ratio is defined as the inverse of velocity ratio, $\boldsymbol{GR} = 1/VR =$ input rotation/output rotation.

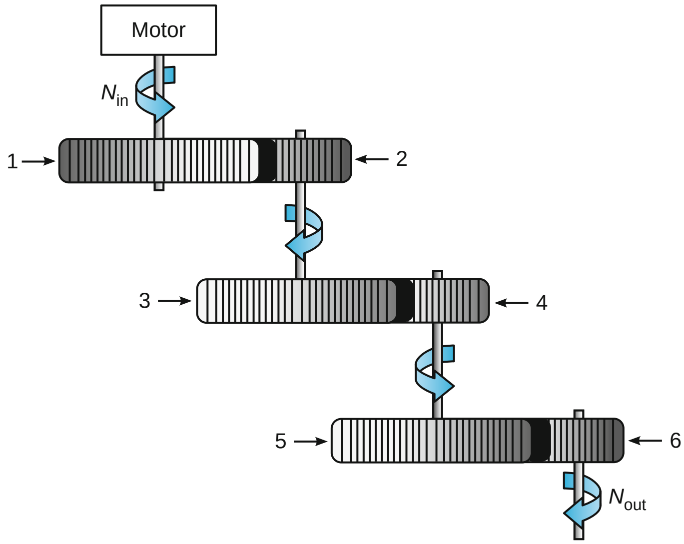


FIGURE 13.8 Compound Gear Train

What is the velocity ratio for the full set of compound gears? Note that some of the gears are *driven* and some are *drivers*; in addition, some of the gears are connected by internal shafts so that these turn at a common speed. The analysis is straightforward:

$$N_1/N_{in} = 1 \text{ (same shaft)}$$

$$N_2/N_1 = r_1/r_2$$

$$N_3/N_2 = 1 \text{ (same shaft)}$$

$$N_4/N_3 = r_3/r_4$$

$$N_5/N_4 = 1 \text{ (same shaft)}$$

$$N_{out}/N_5 = r_5/r_6$$

Therefore, the overall velocity ratio is:

$$VR \equiv N_{out}/N_{in} = N_1/N_{in} \cdot N_2/N_1 \cdot N_3/N_2 \cdot N_4/N_3 \cdot N_5/N_4 \cdot N_{out}/N_5$$

or

$$\boldsymbol{VR = 1 \cdot r_1/r_2 \cdot 1 \cdot r_3/r_4 \cdot 1 \cdot r_5/r_6 = \frac{r_1 \times r_3 \times r_5}{r_2 \times r_4 \times r_6}} \tag{13.11}$$

These relationships make it very easy to analyze compound gear trains.

$$VR = \frac{\text{Product of diameter of } \textit{driving} \text{ wheels}}{\text{Product of diameter of } \textit{driven} \text{ wheels}} \tag{13.12}$$

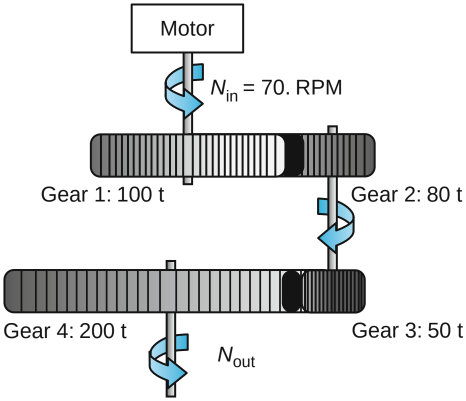
and

$$VR = \frac{\text{Product of number of gear teeth of } \textit{driving} \text{ wheels}}{\text{Product of number of gear teeth of } \textit{driven} \text{ wheels}} \tag{13.13}$$

Example 13.10

A 70. RPM motor is connected to a 100-tooth gear that couples in turn to an 80-tooth gear that directly drives a 50-tooth gear. The 50-tooth gear drives a 200-tooth gear. If the latter is connected by a shaft to a final drive, what is its RPM? The setup is essentially the same as the previous example; we'll go straight to **solve** for this reason.

Again, use a sketch to help visualize the problem.



$$\text{Overall } VR = N_{out}/N_{in} = \frac{\text{Product of Number of Teeth on "Driving" Gears}}{\text{Product of Number of Teeth on "Driven" Gears}}$$

Therefore, $N_{out}/N_{in} = (t_1 \times t_3)/(t_2 \times t_4) = (100 \times 50)/(80 \times 200) = 0.313$, and therefore

$$N_{out} = 0.313 \times 70. = 21.9 = \mathbf{22.\ RPM}$$

(Note that the number of gear teeth is a "counted" integer, and thus has infinite significant figures.)

In a comprehensive analysis of gears, more than kinematics is involved. We need to consider not only the strength of the gears themselves but how much power and force they are capable of transmitting. This subject is too complex in this introductory text and we will just briefly introduce the idea of **torque**, a central variable in gear analysis. Gears not only change the rotation speed, but also the twisting force, or torque, of the axle. We will merely note that you can't increase both the torque and the angular velocity (RPM) simultaneously using gears. That would violate conservation of energy—you would be getting something for nothing. In fact, the torque (T) transmitted is inverse to speed:

$$\frac{T_1}{T_2} = \frac{N_2}{N_1} = \frac{t_1}{t_2} = \frac{r_1}{r_2} \tag{13.14}$$

Thus, to apply high torque to a shaft you use large gears turned slowly by small intermeshing gears. However, you can use gears to achieve a desired combination of torque and RPM. For example, you can first use one set of gears to provide the wheels with high torque and low RPM for initial acceleration (first gear). Then you can shift to another set of gears providing the wheels with lower torque and higher RPM as the car speeds up (second gear). Then you can shift to a third gear combination that offers low torque and high RPM for cruising along a level highway at 65 miles per hour. In a modern automobile there may be four, five, or even six forward gears.

The set of gears that achieves these different gear ratios at the appropriate times for each is called the transmission. Today most American transmissions are automatic,[7] being based on a fluid coupling rather than mechanical gears. So drivers no longer experience gear ratios directly as they did in the days of the standard or "stick" shift,[8] when expertise with the left foot on the clutch (which temporarily disengages the gear train between engine and wheels) was as crucial a part of driving as was expertise with the right foot upon the accelerator.

SUMMARY

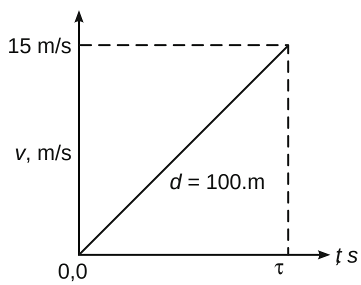
Engineers have designed everything from pyramids to proposed space colonies. One important thing they have designed is roads. Roads designed by engineers made the Roman Empire possible.[9] Railroads designed by engineers made it possible for the United States to remain one nation, from the Atlantic to the Pacific Oceans. Superhighways designed by engineers help make our modern automobile-based civilization possible. Perhaps, in the future, a radically different transportation system will make even safer and more convenient travel possible.

Kinematics is the study of motion without regard to the forces that produce the motion. The relationships among acceleration, speed, distance, and time are basically geometric (with some calculus in complicated cases). These relationships are used in a wide variety of applications, and the subject of kinematics is integral to a number of engineering fields such as mechanical, civil, mechatronic, and biomechanical. Kinematics is necessary for many engineers as part of their fundamental knowledge. This chapter has presented a versatile tool, the **speed versus time (v–t) diagram**. It has shown how to apply that tool to typical engineering problems, the design of highway on-ramps, as well as for the analysis for traffic flow on crowded highways. Less obviously an offshoot of kinematics, the motion of gear and gear trains that transmit power as needed through the use of **gears** is evaluated through **velocity ratios** and **gear ratios**.

EXERCISES

It is strongly suggested you use the v–t diagram for most of these exercises.

1. A car is alone at a red light. When the light turns green, it starts off at constant acceleration. After traveling 100. m, it is traveling at 15. m/s. What is its acceleration? **A: 1.1 m/s²**



[7]To learn how automatic transmissions work visit the web site http://www.edmunds.com/ownership/techcenter/articles/43836/article.html

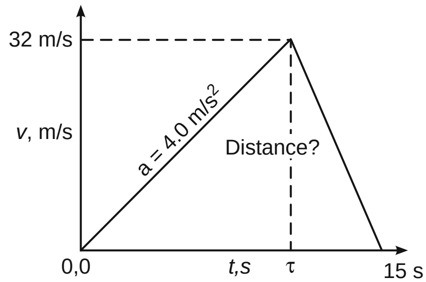
[8]Of course, many modern bicycles have from 15 to 24 gears from which the rider can choose, and appreciate their beneficial effects.

[9]Indeed, the remains of some can still be seen in Europe where some modern roads follow the same course originally laid by Roman engineers.

2. A car is alone at a red light. When the light turns green, it starts off at constant acceleration. After traveling exactly one-eighth of a mile, it is traveling at 30. mph. What is its acceleration in m/s^2?

3. An electric cart can accelerate from 0 to 60. mph in 15. sec. The Olympic champion sprinter Usain Bolt can run 100.0 m in 9.69 s. Who would win a 100.0 m race between this cart and the world champion sprinter? (**A: The sprinter wins by 15 m!**)

4. A car starts from a stop at a traffic light and accelerates at a rate of 4.0 m/s^2. Immediately on reaching a speed of 32. m/s, the driver sees that the next light ahead is red and instantly applies the brakes (reaction time = 0.00 s). The car decelerates at a constant rate and comes safely to a stop at the next light. The whole episode takes 15. seconds. How far does the car travel? (**Partial A:** See inset figure.)



5. For the previous problem, supply a table and draw a graph showing the distance, speed, and acceleration of the car versus time.

6. Based on your experience as a driver or a passenger in a car, estimate (with calculations) the maximum deceleration achieved by putting maximum pressure on the brakes when traveling at 30 mph.

7. A car leaves a parking space from a standing stop to travel to a fast-food restaurant 950 meters away. Along the journey it has to stop after 325 m at a stop sign. It has a maximum acceleration of 3.0 m/s^2 and a maximum deceleration of −10. m/s^2. It never exceeds the legal speed limit of 15. m/s. What is the least possible time it can take until the car comes to a full stop in front of the fast-food restaurant? (**A:70. s**)

8. You are an engineer designing a traffic light. Assume a person can see a traffic light change color from red to yellow, and it takes one second to respond to a change in color. Suppose the speed limit is 15. m/s. Your goal is to enable drivers always to stop after seeing and responding to the yellow light with a maximum deceleration of −5.0 m/s^2. How long should the yellow light last?

9. You are a driver responding to the traffic light in the previous exercise. If it was correctly designed according to that problem, at what distance from the light should you be prepared to make your "to stop or not to stop" decision? Assume you are a safe driver who neither speeds up to get through the yellow light nor stops more suddenly than the deceleration rate of −5.0 m/s^2.

10. Why are red lights on the top of traffic lights and green lights on the bottom?

11. The distance d in meters from which a person can see a stop sign is given by the formula $D = d/30$. (where D is the diameter of the stop sign in meters). Assume it takes a person one second to step on the brakes after seeing the stop sign. Assume that the brakes decelerate the car at a rate of −5.0 m/s^2. If the speed limit is 15. m/s, what should the diameter of the stop sign be? (**A: 1.3 m**)

12. Assume that the purpose of an on-ramp is to allow cars to accelerate from 15. mph to 60. mph. An off-ramp will allow cars to decelerate from 60. mph to 15. mph. Should an off-ramp be longer than, shorter than, or the same length as an on-ramp? Give a reason for your answer. (**Hint:** Do cars bunch more when accelerating or decelerating?)

13. Suppose the deceleration of a car on a level off-ramp is −3.0 m/s^2. How long would the off-ramp have to be to allow a car to decelerate from 60. mph to 15. mph? (**A: 110 m**)

14. An early proposal for space travel involved putting astronauts into a large artillery shell and shooting the shell from a large cannon.[10] Assume that the length of the cannon is 30. m and that the velocity needed by the shell to achieve orbit is 15,000 m/s. If the acceleration of the shell is constant and takes place only within the cannon, what is the acceleration of the shell in gees?

15. Suppose that a human body can withstand an acceleration of 5.0 gee, where 1 g is 9.8 m/s^2. How long would the cannon have to be in the previous exercise to keep the acceleration of the humans within safe limits? (**A: 2.3×10^6 m**)

16. You wish to cover a two-mile trip at an average of 30.0 mph. Unfortunately, because of traffic, you cover the first mile at just 15.0 mph. How fast must you cover the second mile to achieve your initial schedule?

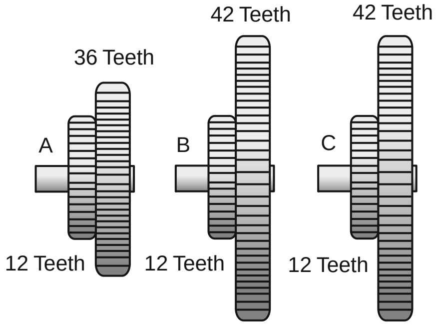
17. You are a TV reporter who rides in a helicopter to advise commuters about their travel time from their bedroom community to a city 20. miles away. You notice from your helicopter there are 160 cars per mile. Assume that all the drivers are observing the rule-of-thumb "follow rule" (see Figure 13.4). What should you tell your audience that the travel time between the cities will be? (**A: 1 hr, 11 min**)

18. This chapter has shown that for the two-second rule the number of cars per hour completing a trip is independent of speed. Consider now a slightly more realistic follow rule: mph = 1800/(cars/mile)$^{0.9}$ − 4. Find the speed for which the number of cars per hour is a maximum. (Ignore the speed limit up to 100 mph!) (**A: 40 mph**)

19. If you want to achieve a flow of 2,600 cars per hour in each direction on a highway, and the two-second rule applies, how many lanes must the highway have in each direction?

20. Imagine a smart car of the future with radar and computer control capabilities that make it safe to operate on today's highways with twice the speed *and* half the spacing of today's human-driven cars. Using the follow rule, compare the capacity in terms of cars/hr for smart cars with cars/hr for human-driven cars. (**A: Four times as much as for human driven cars**)

Exercises 21 through 26 concern a gear set consisting of up to the three gears as shown here. In each case the larger (36 or 42 tooth) and smaller (12 tooth) portions of each gear are fastened together so that they rotate at the same rate.



[10]A number of military satellites have been delivered to space using specially adapted naval cannons from the Vandenberg Air Force base in California.

21. Suppose an 1800.0 RPM motor is connected directly to gear A. The 12-tooth portion of gear A drives the 42-tooth portion of gear B. What is the rotational speed in RPM of the shaft of gear B? (**A: 514 RPM**)

22. If the drive gear in Exercise 21 is rotating *clockwise*, what is the direction of rotation of the driven gear?

23. Suppose gear A was driven by a motor at 1800. RPM. Suggest an arrangement of gears from the given gear set A, B, and C to achieve an output speed of 18,900 RPM.

24. Suppose the gear set is to be used with a motor rotating at a constant speed of 1000. RPM. What are the lowest and highest output speeds that can be developed with the given three gears A, B, and C? (**A: Lowest speed cluster = 81.6 RPM**; **highest speed cluster = 12,250 RPM**)

25. A motor drives a shaft at 1000. RPM with a torque of 35 newton-meters. A gear on that shaft has 42 teeth. That gear drives a second gear with 12 teeth. That second gear is connected to an output shaft. What is the torque produced by the output shaft? (**A: output torque = 10. Nm**)

26. Consider only two of the gears in the gear set, the 12–36 and 12–42 tooth gears. For any given input speed, how many different output speeds are possible? Illustrate by categorizing the number of configurations that (1) increase or decrease the final shaft speed, (2) increase or decrease the final torque, and (3) leave the final speed and torque unchanged compared to the drive motor.

27. You are an engineer with the responsibility for choosing the route for a new highway. You have narrowed the choice to two sites that meet all safety standards and economic criteria (call them Route A and Route B). Route B is arguably slightly superior in terms of both safety and economics. However, Route B would pass by the site of an expensive new house recently built by your favorite niece, and the proximity of the highway would severely depress the value of the house. Since your niece's last name is different from yours, and you have never mentioned her at the office, you would be unlikely to be "caught" if you chose Route A in order to save the value of her house. What do you do?

a. Choose Route A.
b. Choose Route B.
c. Ask to be relieved from the responsibility because of conflict of interest.

Use the Engineering Ethics Matrix.

28. You are an engineer on a team designing a bridge for a state government. Your team submits what you believe to be the best design by all criteria, at a cost that is within the limits originally set. However, some months later the state undergoes a budget crisis. Your supervisor, also a qualified engineer, makes design changes to achieve cost reductions that he believes will not compromise the safety of the bridge. You are not so sure, though you cannot conclusively demonstrate a safety hazard. You request that a new safety analysis be done. Your supervisor denies your request on the grounds of time and limited budget. What do you do?

a. Go along with the decision. You have expressed your concerns, and they have been considered.
b. Appeal the decision to a higher management level.
c. Quit your job.
d. Write your state representative.
e. Call a newspaper reporter and express your safety concerns.

Use the Engineering Ethics Matrix.

CHAPTER 14

Bioengineering

Source: © iStockphoto.com/Antonis Papantoniou

Bioengineering is the application of engineering, life sciences, and mathematics to define and solve problems in biology, medicine, healthcare, and other fields that deal with living systems. Bioengineering is a relatively new discipline that combines many aspects of traditional engineering fields such as chemical, electrical, materials, and mechanical engineering. Some examples of bioengineering include the design and development of:

- Devices that substitute for damaged body parts such as hearing aids, cardiac pacemakers, synthetic bone, and teeth
- Artificial kidneys, hearts, blood vessels, arms, legs, hips, knees, and other joints
- Medical imaging techniques (ultrasound, MRI, CT, and others)
- Engineered organisms for chemical and pharmaceutical manufacturing
- Blood oxygenators, dialysis machines, and diagnostic equipment

Bioengineering encompasses a number of specialties including biomedical engineering,[1] biotechnology, biological engineering, biomolecular engineering, biomechanics, biochemical, and clinical engineering. Each of these fields may differ slightly in their focus of interest, but they are all concerned with the improvement of human life.

14.1 WHAT DO BIOENGINEERS DO?

Bioengineers have a wide variety of career choices. Some may work alongside medical practitioners, developing new medical techniques, medical devices, and instrumentation for manufacturing companies. Hospitals and clinics employ clinical engineers to maintain and improve the technological support systems used for patient care. Engineers with an advanced bioengineering degree can perform biological and medical research in educational and governmental research laboratories.

Many bioengineers help people by solving complex problems in medicine and health care. Some bioengineering jobs combine several disciplines, requiring a diverse array of skills. Digital hearing aids, implantable defibrillators, artificial heart valves, and pacemakers are all bioengineering products that help people combat disease and disability. Bioengineers develop advanced therapeutic and surgical devices, such as a laser system for eye surgery and a device that regulates automated delivery of insulin.

[1]The term **biomedical engineering** sometimes is used synonymously with **bioengineering**.

In genetics, bioengineers try to detect, prevent, and treat genetic diseases. In sports medicine bioengineers develop rehabilitation and external support devices. In industry, bioengineers work to understand the interaction between living systems and technology. Government bioengineers often work in product testing and safety, where they establish safety standards for medical devices and other consumer products. A biomedical engineer employed in a hospital might advise on the selection and use of medical equipment or supervise performance testing and maintenance.

In biocommunication, bioengineers develop new communication systems that enable paralyzed people to communicate directly with computers using brain waves. Through bioinformation engineering, they are exploring the remarkable properties of the human brain in pattern recognition and as a learning computer. Through biomimetics bioengineers are trying to mimic living systems to create efficient designs. These areas range far beyond the material in this chapter, but their basic theme is similar.

By engineering analysis of the situations in which living matter might be exposed, and by characterizing, in engineering terms, the remarkable properties of living matter, knowledge is gained that improves safety and health. That understanding can be used to make life safer and healthier. Among the tasks undertaken by bioengineers are the design of safety devices, ranging from football helmets to seat belts to air bags; the development of prosthetic devices for use in the human body, such as artificial hip joints; the application of powerful methods for imaging the human body, such as computed axial tomography (or CAT scanning) and magnetic resonance imaging (or MRI); and the analysis and mitigation of possible harmful health effects on humans subject to extreme environments, such as the deep sea and outer space.

In this chapter you will be introduced to a few simple descriptions of human anatomy and of the effects of large forces on hard and soft human tissues. You will also learn (1) why automotive collisions can kill, (2) how to make a first approximation of the likelihood of damage during collisions to the human body using a **fracture criterion**, (3) how to predict the injury potential of a possible accident using a criterion that could be called **stress-speed-stopping-distance-area (SSSA)**, and (4) how to apply two other criteria for the effect of deceleration on the human body (the **30. *g* limit** and the **Gadd Severity Impact parameter**).

There are just two areas of human anatomy we will investigate: The first is to understand how serious blunt force trauma can affect the operation of the brain and other neurological tissue, and the second is to understand how bone protects internal soft tissues.

14.2 BIOLOGICAL IMPLICATIONS OF INJURIES TO THE HEAD

In many automobile accidents, the victims suffer severe head and neck injuries. Some of the accidents directly cause brain trauma, and others cause neck injuries. The kind of accident caused by severe overextension of the neck and associated tissues is called **whiplash**. Head and neck injuries are all too common in accidents where you suffer **high *g* decelerations** (meaning decelerations of many times that of gravity). What this means is that the victim absorbs very high decelerations on impact with another vehicle or with a stationary object. Still other victims are injured *within* the vehicle when they contact interior components of the passenger compartment such as the dashboard and the windshield.

We will first take a brief look at human anatomy from the neck up to understand what forces can do to neurological function. Figure 14.1 shows the essentials of the skull, the brain, the spinal column, and the spinal cord.

The skull protects the brain, which floats in a fluid-like layer. The base of the brain connects to the spinal cord through the spinal column, which provides protection for the spinal cord. The spinal column consists of individual bony vertebrae that surround the all-important spinal cord. The spinal cord is the essential “wiring” that takes instructions from the brain to the various bodily functions. The vertebrae are separated by

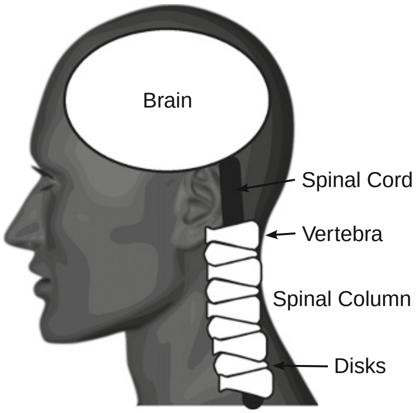


FIGURE 14.1 Part of the Human Nervous System

cartilaginous matter known as "disks" to provide flexibility and motion. Injury to the disks can result in severe pain and, in the case of the spinal cord, in paralysis. Injury to the spinal cord can lead to very severe bodily malfunctions and injury to the brain can cause a number of physical deteriorations or death. The system, as efficient as it is, can suffer a number of possible injuries during high g decelerations.

Consider the injuries that are illustrated in Figure 14.2. The brain will move relative to the skull in its fluid-like layer when experiencing high g's. In **hyperflexion**, the skull will move forward relative to the brain, causing damage to the occipital lobe (the back of the brain), and in **hyperextension** it will move in the opposite direction relative to the brain and damage its frontal lobe. Further, damage can also occur to the basal brain, a potentially devastating injury, since it may interrupt the nerve connections to the spinal cord. Also, there can be fractures to the vertebra and extrusion or rupture of the protective disks within the neck.

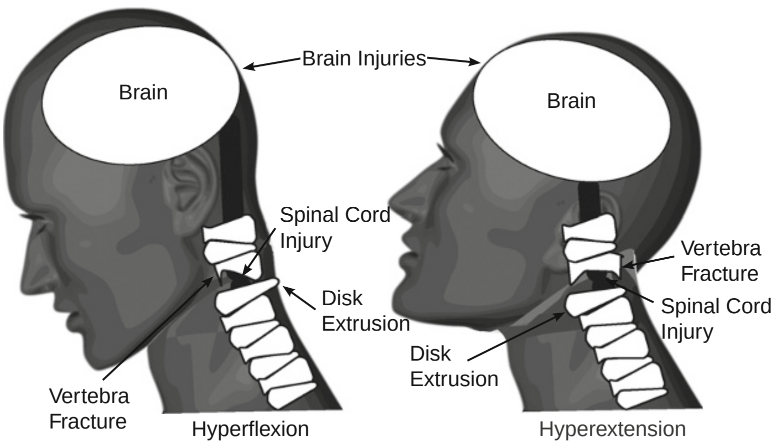


FIGURE 14.2 Brain Injuries Due to Extreme Whiplash

14.3 WHY COLLISIONS CAN KILL

What is it about collisions that can kill or severely injure people? The challenge of this chapter is to express the answer using engineering variables. By using the right numbers and units for the values of those variables, criteria can be identified that distinguish a potentially fatal crash from one from which you walk away.

One key variable in the bioengineering analysis of automobile safety is deceleration. Just as the "zero-to-sixty"[2] time of acceleration is a measure of a car's potential for performance, the "sixty-to-zero" time of deceleration is a first measure of an accident's potential for injury. But whereas the zero-to-sixty time typically is measured in seconds, the sixty-to-zero time may be measured in milliseconds—that is, in thousandths of a second.

Rapid deceleration causes injury because of its direct relation to force. As discussed in the section on Newton's Second Law (Chapter 2), force is directly proportional to positive or negative acceleration (i.e., deceleration). And, as discussed in Chapter 11, force divided by area is stress. As further discussed in Chapter 11, the stress–strain curve gives the yield strength of a material. Suppose that the biologic material is flesh, bone, or neurons (nerve cells found in the brain, the spinal column, and local nerves). If their yield strength is exceeded, serious injury or death can follow.

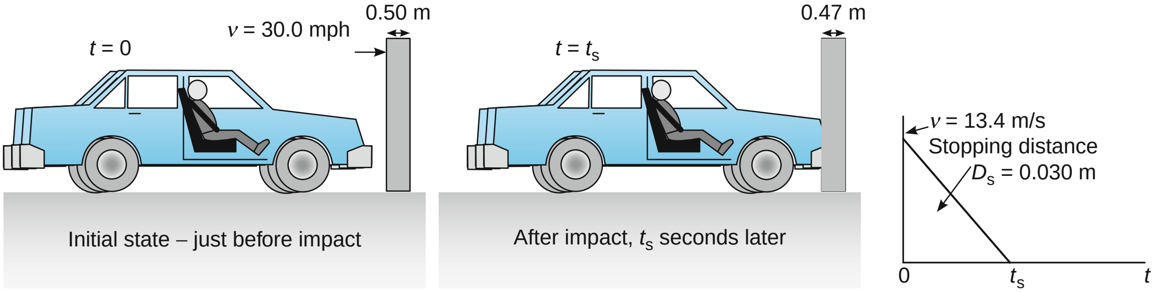
An engineering forensic analysis of collision injury combines the effects of acceleration with the properties of materials. So a first step in characterizing collisions is to determine the deceleration that occurs. We already have a tool for visualizing this: the $v-t$ graph.

Example 14.1

A car traveling at 30.0 miles per hour (mph) runs into a sturdy stone wall. Assume the car is a totally rigid body that neither compresses nor crumples during the collision. The wall "gives" a distance of $D_S = 0.030$ m in the direction of the collision as the car is brought to a halt. Assuming constant deceleration, calculate that deceleration.

Need: Deceleration = _______ m/s^2.

Know–How: First sketch the situation to clarify what is occurring at the impact. Then use the $v-t$ graph to describe the collision. A positive slope of the $v-t$ graph indicates acceleration, and a negative slope indicates deceleration.



Solve: We first need to calculate the stopping time, t_s, to decelerate from 30.0 mph (13.4 m/s) to zero.

If deceleration is constant, we know that stopping distance is equal to the area under the $v-t$ graph or $D_S = ½ \times v \times t_s$, where t_s is the stopping time. Or $D_S = ½ \times 13.4$ [m/s] $\times\ t_s$ [s] = 0.030 [m]. Therefore, $t_s = 0.060/13.4 = 0.0045$ s.

The slope of the $v-t$ graph is $\Delta v/\Delta t = (0 - 13.4)/(0.0045 - 0) = -2980 =$ **−3,000 m/s^2**. The negative sign indicates this is the **deceleration** rate.

[2]The term for the time to go from a complete standstill to 60 mph.

What does this deceleration mean? How high is it? One good way to understand its damaging potential is to compare it to the acceleration due to gravity. This nondimensional ratio is 3,000/9.8 = 310 g's (to two significant figures).

The use of acceleration in terms of g can be quite helpful. Referring to Chapter 2, what is the weight of a mass m on Earth? It is mg. What is the force on a mass accelerating by an amount a? It is ma. But let us simultaneously multiply and divide that force by g—that is,

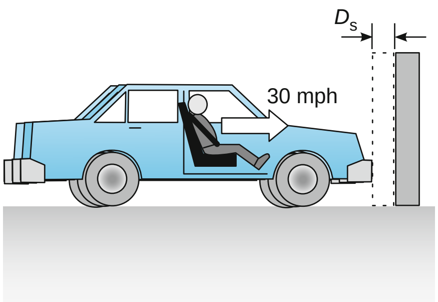
$$\boldsymbol{F = ma \times (g/g) = mg \times (a/g) = \text{the body's weight } (mg) \times \text{number of } g\text{'s } (a/g)} \tag{14.1}$$

In the impact described in Example 14.1, your head would weigh 310 times its usual weight (and the same for the rest of your body). An average human head has a mass of about 4.8 kg, so it will experience a force of about 4.8 [kg] × 9.8 [m/s^2] × 310 = 15,000 N or about 3,300 lbf! Your entire body will experience a similar force 310 times its normal weight.

You can also calculate the forces involved by kinetic considerations as opposed to the basically kinematic considerations just discussed. The force × distance to stop is equal to the kinetic energy that is to be dissipated. Let's assume the driver is belted into the car and therefore decelerates at the same rate as the car.

Example 14.2

The car and its belted driver of Example 14.1 suffer the same constant deceleration of 310 g or 3,000 m/s^2. What is the force the driver experiences during the collision if his or her mass is 75 kg? Assume the car is a totally rigid body that neither compresses nor crumples during the collision.



Need: Force stopping 75 kg driver on sudden impact = _______ N.
Know: The driver's weight is 75 [kg] × 9.8 [m/s^2] = 740 N and deceleration rate is 310 g.
How: $F = ma = mg \times a/g$.
Solve: $\boldsymbol{F}$ = 740 × 310 = $\mathbf{2.3 \times 10^5}$ **N** (or about 52,000 lbf!).

14.4 THE FRACTURE CRITERION

Under what conditions does a force cause living material to fail? That is, to crack, or break into pieces, or lose its ability to contain or protect fluid or soft structures. As we saw in Chapter 11, a material breaks when the stress on it exceeds its ultimate strength. Stress and strength are variables with units (N/m^2) that can be calculated from a description of the loading and the material's stress–strain diagram.

One aspect of the loading is the applied force, which was found in Example 14.2. Another aspect is the area over which that force is applied. This is needed to calculate the stress on the material in question.

Figure 14.3 shows an illustration of a bone. Bone is a natural composite material that consists of a porous framework made of the mineral calcium phosphate, interspersed with fibers of the polymeric material collagen. The calcium phosphate gives bone its stiffness, and the collagen provides flexibility. Figure 14.3 shows the internal structure of the body's all-important long bones that have a hard outer layer, spongy interior layers with axial channels built from bone cells known as osteoblasts, and a central core of marrow that is responsible for renewing blood supply and other important cells. Bone is **anisotropic**, which means that its properties are not the same in all directions. For example, it has different mechanical properties along its axis than it has perpendicular to the axis.[3]

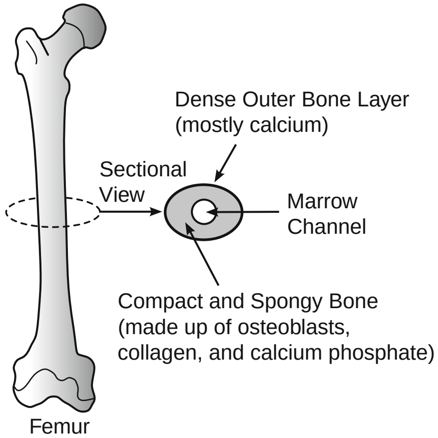


FIGURE 14.3 Structure of the Bodies Long Bones

The properties of bone can be approximated by a stress–strain diagram as in Figure 14.4.[4] Notice bone is more rigid and flexible along (i.e., in the longitudinal direction) the bone's length than crosswise (i.e., in its transverse direction). These properties are due to the dense outer bone layer being tough and with the compact and spongy interior bone providing surprisingly high flexibility. This model is a simplification of real bone behavior and is modeled as a perfectly elastic and perfectly plastic material. Similar data are shown in Table 14.1.

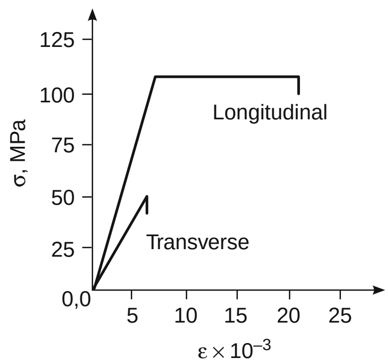


FIGURE 14.4 Assumed Bone Mechanical Properties

[3]Wood has similar anisotropy; you can split wood rather easily along the grain, but it is hard to break cross grain. The cellular patterns in wood are responsible for this useful behavior (for trees as well as for people).
[4]Adapted from A. H. Burstein and T. M. Wright, *Fundamentals of Orthopædic Biomechanics*. Williams and Wilkins, Baltimore, MD, 1994, p. 116.

Table 14.1 Approximate Mechanical Properties of Human Bone

Direction	Elastic Yield, MPa	Young's Modulus, GPa	Strain at Yield, %	Strain at Fracture %	Toughness, at Yield and Fracture, MJ/m^3
Longitudinal	110	16.	0.7	2.0	0.4 at yield and 1.8 at fracture
Transverse	60.	9.	0.7	0.7	0.2 at yield and fracture

Bone, like other materials, may fail in several different ways:

1. If bone is subjected to a local stress greater than its yield strength, it will take on a "permanent" deformation.
2. If that stress exceeds the bone's ultimate stress, the bone will break.
3. If the energy absorbed exceeds the toughness at yield × volume of affected bone, the bone will take a "permanent" set.
4. If the energy absorbed exceeds the toughness at fracture × volume of affected bone, the bone will break.

In general, bone failure modes 1 and 2 mostly apply to the situation in which the load is applied slowly. For a suddenly applied load, it is usually the toughness criterion that applies since bone toughness is measured by the strike of a heavy hammer in a sudden blow.

Most bones that fail in vehicular accidents are due to transverse stress rather than longitudinal stress[5] (for example, the ribs hitting the steering wheel if their owner is not seat-belted or air-bag protected). However, whiplash injuries may produce neck bone failure in the longitudinal direction. In general, you would prefer not to semi-permanently distort bone. Therefore, the elastic yield condition is the more conservative situation for a victim of an accident.

In another version of the accident as described in Examples 14.1 and 14.2, the driver is unbelted, so the skull hits and dents the dashboard. How does the stress on the skull compare with the elastic yield point of the bones in the skull? That depends on the area of the skull over which the force is applied. For a given force, the smaller the area, the higher the stress.

Example 14.3

This is an accident similar to that in Examples 14.1 and 14.2, except the driver is not wearing a seat belt. Assume the driver's skull dents the dashboard to a depth of 0.030 m. The area of contact between the driver's forehead and the dashboard is 3.0 cm × 3.0 cm = 9.0×10^{-4} m^2, and the properties of bone are those given in Table 14.1. Will the collision *fracture* the skull?

Need: Yield strength of skull exceeded = _______ (yes/no).

Know: Since the head suffers the same deceleration of 310 *g* given $D_S = 0.03$ m, we can use the results of Examples 14.1 and 14.2. Consequently, the force on impact is 2.3×10^5 N.

[5]Think of snapping a dead tree limb across your knee. The break occurs on the side away from your knee as fulcrum. That layer of branch is being *stretched* and it fails mostly in tension.

How: Stress is force per unit area and from Table 14.1, the transverse yield strength of the bone is 60. MPa.
Solve: Compare 60. $\times 10^6$ [N/m^2] with the applied stress on impact. That stress is

$$(2.3 \times 10^5)/(9.0 \times 10^{-4})[\text{N}]/[\text{m}^2] = 2.6 \times 10^8 \text{N/m}^2$$

which is considerably greater than 60. $\times 10^6$ N/m^2. So, yes, the bone will fracture, and the brain will be seriously damaged. Unfortunately, this driver is fatally injured.

14.5 THE STRESS–SPEED–STOPPING DISTANCE–AREA CRITERION

Let's generalize the results of Example 14.3. Collisions kill or injure because they involve a high rate of deceleration. That deceleration causes bodily contact with inflexible material, and the result is experienced as a force. The smaller the area over which that force is applied, the higher the stress on the bone or tissue. The higher the localized stress, the greater the likelihood that the material will fail in that region.

We can express this insight as a relationship among the variables. Consider first the generalized form of the $v-t$ diagram of a collision with a constant deceleration rate. Suppose a body of mass m is subjected to a constant deceleration a and stops in a distance D_S. Then, from Newton' Second Law, the force experienced by the body is $F = ma$, and $a = v/t$, so $F = mv/t$. For a constant deceleration, the stopping distance is $D_S = ½\ vt$, so $t = 2D_S/v$. Then the force experienced by the body is:

$$F = mv^2/2D_S \tag{14.2}$$

Now suppose that this force is experienced on an area A of the head or body with a hard surface. Then the stress can be calculated from Equation (14.2) as:

$$\textbf{Stress} = \boldsymbol{\sigma} = \textbf{F/A} = \boldsymbol{mv^2/2AD_S} \tag{14.3}$$

This relationship is called the **stress–speed–stopping distance–area (SSSA) criterion**. It states that, technically, it's not enough to say that "speed kills." What kills is the combination of high speeds, short stopping distances, *and* small contact areas!

Strategies for reducing the severity of a collision follow from these insights. By decreasing speed, increasing the stopping distance, and increasing the area of application of the force, the effects of the collision on living tissue can be reduced.

The presence of the v^2 term in the numerator of Equation (14.3) indicates that decreasing speed is a highly effective way of decreasing collision severity. Cutting speed in half will reduce the stress of a collision to one-fourth of its previous value (if everything else remains the same).

How might the other two terms in the SSSA be brought into play? Application of the area term is simple in concept, though more difficult in practice. The larger the area of contact between body and surroundings during a collision, the smaller the stress. This is one motivation behind the air bag. It is a big gas-filled cushion that can increase the area of contact by a large factor. Of course, that cushion must be deployed rapidly enough to be effective in a collision. It must also *not* be deployed unless a collision has just occurred. Only by conquering these two technical challenges (deployment in a few milliseconds when appropriate, nondeployment at all other times) did engineers turn the air bag into a valuable safety tool.

However, the area portion of the SSSA criterion is only part of the air bag story. A seat belt does indeed increase the area over which force is applied (compared to the forehead contacting the dashboard) and, as importantly, reduces the probability of an impact of the head with the dashboard. But that increase would not by itself account for the great lifesaving and injury-reducing potential of seat belts. A more important part of that potential involves that other term in the denominator, the stopping distance, D_S.

A more practically important contribution of seat belts is to *attach the driver to a rigid internal passenger shell while the rest of the car shortens by crumpling*. The high effectiveness of the seat belt as a safety device has been achieved only in combination with the design of a car that significantly crumples during a collision. Let's now assume that the car has such a crumple zone. This is illustrated in Figure 14.5.

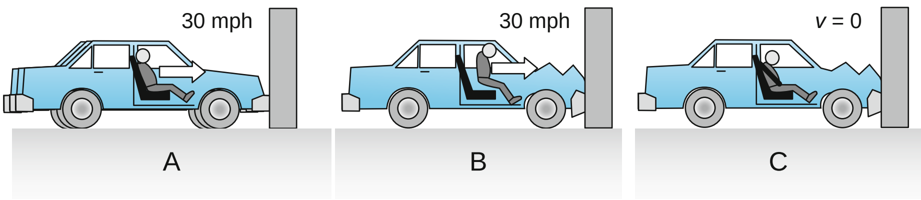


FIGURE 14.5 Seat Belts and Crumple Zone

Picture A shows the situation of driver, car, and wall during a 30 mph collision. The car rapidly crumples to a stop several milliseconds later, as shown in picture B. But the driver is still moving forward at the speed of 30 miles an hour. It is the collision of the driver's head and the windshield that brings the driver to a stop in a short distance on the order of a few hundredths of a meter. By contrast, when the driver is wearing a seat belt, the body stops with the car, as shown in picture C.

In the latter case, the driver gets the full benefit of the crumple of the car, which is on the order of a half to one meter. This hundredfold increase in stopping distance results, according to the SSSA criterion, in a reduction in stress to one-hundredth of its previous value. This can be the difference between life and death.

Example 14.4

A car traveling at 30.0 miles per hour (13.4 m/s) runs into a rigid stone wall. Assume the car's crumple zone results in a stopping distance of 0.60 m. Assume that the 75. kg driver is wearing a seat belt. Assume that the area of contact of seat belt and body is 4.0 cm $\times$ 30. cm $=$ 0.012 m^2. Determine the maximum stress and the number of g's the driver's body experiences.

Need: Maximum stress on the driver's body _______ N/m^2 and the number of g's.

Know–How: We could simply repeat the analysis of the previous examples, but the SSSA criterion provides a shortcut.

Solve: Stress, $\sigma = mv^2/2AD_S$. Substitute into the SSSA formula: $\boldsymbol{\sigma} = 75. \times 13.4^2/(2 \times 0.012 \times 0.60)$ [kg] $[\text{m/s}]^2[1/\text{m}^2][1/\text{m}] =$ **9.4×10^5 N/m²** $\approx$ **1 MPa** (which is less than the elastic yield of bone from Table 14.1). The number of g's is $9.4 \times 10^5 \times 0.012/(75. \times 9.8)$ $[\text{N/m}^2][\text{m}^2][1/\text{kg}][\text{s}^2/\text{m}] = 15$. (Previously it was 310, so survivability has been greatly enhanced.)

To sum up: Here is an answer to our original question, expressed in engineering variables and units. The principal way seat belts save lives is by attaching the driver securely to the inner shell of the car, enabling the driver to take advantage of the car's crumple zone of about 0.5–1 m. During a collision from 30. miles an hour, that strategy restricts deceleration to less than about 150 m/s^2, resulting in stresses on the body that are less than about 1 MPa and almost surely less than that required to break bones.

14.6 CRITERIA FOR PREDICTING EFFECTS OF POTENTIAL ACCIDENTS

The preceding sections have established that the effects of rapid deceleration on human bone and tissue may produce serious injury. We have just indicated one way to address that issue by translating deceleration into force, and force into stress. A comparison of maximum stress experienced by the body with the strength of bone provides a first criterion for the prediction of accident severity.

The effects of acceleration and deceleration on the human body go far beyond the potential to break bones. These effects range from the danger of blackouts of pilots experiencing very high acceleration or deceleration to the bone loss and heart arrhythmia experienced by astronauts exposed to the micro-gravitational forces of space flight for extended periods of time.

We will continue to focus on the effects of high accelerations/decelerations. Just how many g's can the human body stand? As a first approximation engineers drew on a wide range of experiences, such as those of pilots and accident victims, to arrive at the following *initial* criterion:

***The human body should not be subjected to* more than 30. *g*'s.**

At 30. g's and above, the damaging effects of acceleration or deceleration on the human body can range from loss of consciousness to ruptured blood vessels to concussion to the breaking of bone to trauma or death. This criterion still serves as an *initial* rule of thumb for the design of safety devices.

Example 14.5

Does a driver without a seat belt who experiences a 30.0 mile per hour (13.4 m/s) collision and is stopped in 0.10 m by a padded dashboard exceed the 30. g criterion? Does the role of a seat belt in taking advantage of the crumple zone to increase collision distance to 0.60 meter meet the 30. g criterion?

Need: Deceleration without seat belt = _______ (less than/equal/more than 30. g?).
Deceleration with seat belt = _______ (less than/equal/more than 30. g?).

Know–How: From Equation (14.2), $F = ma = mv^2/(2D_S)$ and therefore, in g's,

$$F/mg = ma/mg = a/g = mv^2/(2mgD_S) = v^2/(2gD_S)$$

Solve: Case 1, $D_S = 0.10$ m, $F/mg = (13.4^2)/(2 \times 9.8 \times 0.10) = 92.$ g, which is **considerably more than 30. *g*'s**.
Case 2: $D_S = 0.60$ m, $F/mg = (13.4^2)/(2 \times 9.8 \times 0.60) = 15.$ g, which is **less than 30. *g*'s**.

Consistent with our previous analysis, the combination of seat belt and vehicle crumple zone has reduced the driver's deceleration from a probably fatal 92. g to a probably survivable 15. g.

However, experiences under extreme conditions, such as high-speed flight, soon revealed that the 30. g criterion was seriously incomplete. In some cases, humans have survived accelerations far above 30. g. In other cases, accelerations significantly below 30. g caused serious injury.

The missing element in the simple 30. g criterion is **time**. Deceleration has many effects on human tissue, ranging from destruction at one extreme to barely noticeable restriction of blood flow at the other. Each of these effects has a characteristic time interval needed to take effect. Very high decelerations, measured in hundreds of g's, can be survived *if* the exposure is short enough. Deceleration at moderate g levels, on the other hand, can prove fatal if the exposure is long enough (see Figure 14.6).

In Figure 14.6, the axes are expressed in terms of the $\mathbf{log_{10}}$ of the variables. Not only can we get a scale covering several orders of magnitude by this neat trick, but also we can infer something important about the relationship. In fact, it is linear, not between the actual variables g and t_S but between log g and log t_S.

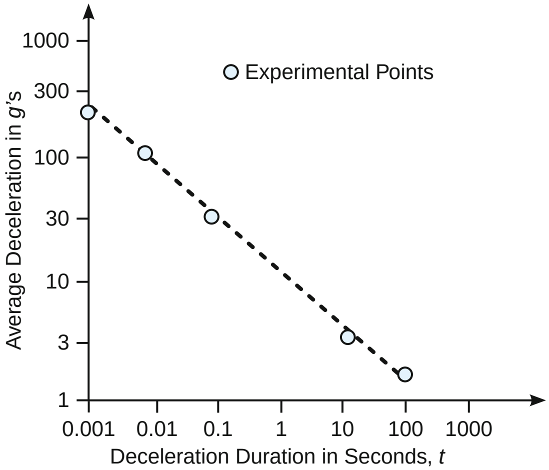


FIGURE 14.6 Injury Criterion Based on the Deceleration in g's and the Deceleration Duration Time, t (From: S. A. Berger et al. 1996. *Introduction to Bioengineering,* Oxford University Press.)

The dotted line represents a *rough boundary* between accelerations likely to cause serious injury or death (above and to the right of the line) and survivable accelerations (below and to the left of the line). Notice that the previous 30. g criterion is the point on that line corresponding to a duration of 0.1 seconds (the right order of magnitude for the duration of an automobile accident).

Figure 14.6 can also be expressed as a formula. A straight line on a log-log plot corresponds to the number on the x-axis raised to a certain power. In this case, the formula for the line dividing serious injury from survivability is:

$$a = (0.002 t_S)^{-0.4} \tag{14.4}$$

with a expressed in the nondimensional units of g's and with t_S expressed in seconds. The slope of the log/log graph is -0.4. This is *not* a basic scientific law. Rather it is an empirical relationship summing up the net effects of many physical and biological properties. It is, however, a useful guide. It is more convenient by rearranging it into the form:

$$a^{2.5} t_S = 500 \tag{14.5}$$

The quantity on the left of Equation (14.5) is the **Gadd Severity Index (GSI)**.[6]

$$\mathbf{GSI} = \boldsymbol{a^{2.5} t_S} \text{ (with } t_S \text{ in seconds and } a \text{ in } g) \tag{14.6}$$

The GSI in Equation (14.6) is a simplified numerical index that comes from assuming either constant deceleration or that a meaningful average deceleration can be measured. To use this equation, first calculate the GSI, and then compare the result to the number 500. If the GSI is greater than or equal to 500, there is a serious danger of injury or death.

[6]The Gadd Severity Index was introduced in C. W. Gadd, "Criteria for Injury Potential." National Research Council Publication #977, Washington National Academy of Sciences, 1961, pp. 141–145.

Example 14.6

Calculate the GSI for a driver without a seat belt who experiences a constant deceleration in a 30.0 miles per hour (13.4 m/s) collision and is stopped in 0.050 m by the dashboard.

Need: GSI = ________ (a number).

Know–How: From Example 14.1, we know that if deceleration is constant, the stopping distance is equal to the area under the $v-t$ graph or $D_S = 1/2 \times v \times t_s$, where t_s is the stopping time. So then, $t_s = 2D_s/v = 2 \times 0.050/13.4 = 0.0074$ s. Therefore, the deceleration is $a = \Delta v/\Delta t = (0 - 13.4)/(0.0074 - 0) = -1{,}800$ m/s^2, or $a = 180$ g's (neglecting the negative sign).

Solve: GSI $= a^{2.5}\, t_S = 180^{2.5} \times 0.0074 =$ **3,200** which is much greater than 500. This suggests that severe injury or death is likely.

Example 14.7

Calculate the GSI for a driver with a fastened seat belt who experiences a constant deceleration in a 30.0 miles per hour (13.4 m/s) collision and is stopped by a 1.0 m crumple zone.

Need: GSI = _______ (a number).

Know–How: From the development of Equation (14.2), the stopping time can be calculated from $t_S = 2D_S/v = 2(1.0)/13.4 = 0.15$ s, and the deceleration is $a = \Delta v/\Delta t = (0 - 13.4)/0.15 = -89.$ m/s^2, or $a = 9.1$ g's.

Solve: **GSI** $= a^{2.5}\, t_S = 9.1^{2.5} \times 0.15 =$ **37.** which is much less than 500. This suggests that severe injury or death is very unlikely.

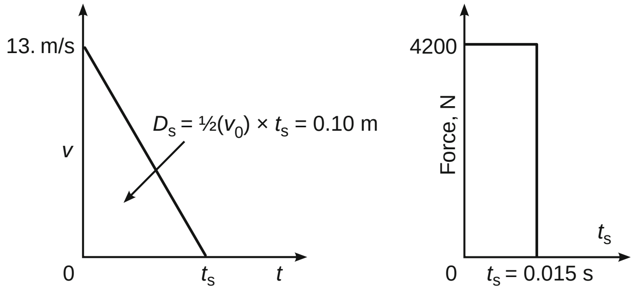
SUMMARY

Bioengineering applies the methods of engineering to living bodies, organs, and systems. This chapter illustrated just one aspect of bioengineering, the use of biomechanics and engineering analysis of biomaterials to improve safety. In this chapter you learned why collisions can kill, how to make a first approximation of the likelihood of collision damage to the human body using a fracture criterion, how to predict the injury potential of a possible accident using the **stress–speed–stopping distance–area (SSSA)** criterion, how to apply two criteria for the effect of deceleration on the human body (the **30. *g* limit** and the **Gadd Severity Impact** parameter), and how to analyze bioengineering problems.

EXERCISES

Exercises 1 and 2 concern the following situation: A car is traveling 30. mph and hits a wall. The car has a crumple zone of zero, and the passenger is not wearing a seat belt. The passenger's head hits the windshield and is stopped in the distance of 0.10 m. The skull mass is 5.0 kg. The area of contact of the head and the windshield is 0.010 m^2. Assume direct contact and ignore the time it takes the passenger to reach the windshield.

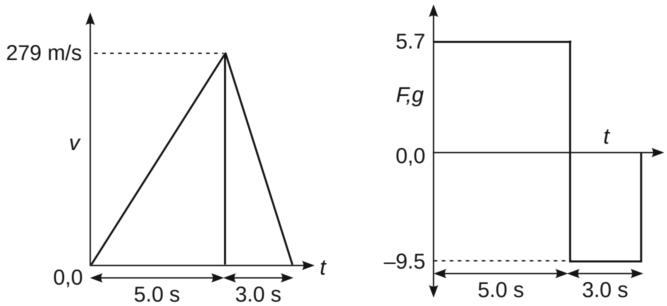
1. Provide a graph of $v-t$ of the collision of the skull and the windshield, and then graph the force experienced by the skull as a function of time. (**A: See below.**)



2. If the compressive strength of bone is 3.0×10^6 N/m^2, will the collision in the previous exercise break the skull?

Exercises 3 and 4 concern an experiment in the 1950s when Air Force Colonel John Paul Stapp volunteered to ride a rocket sled to test the resistance of the human body to *g* forces. The sled accelerated from 0 to 625 miles per hour in 5.0 seconds. Then the sled hit a water brake and decelerated in 3.0 seconds to a standstill. Assume that Stapp was rigidly strapped into the sled and that he had a mass of 75 kg.

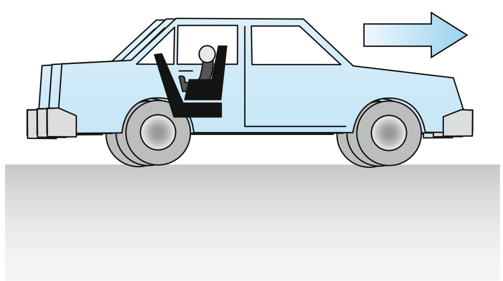
3. Prepare *v–t* and *F–t* graphs of Stapp's trip, and compute the *g* forces he experienced in the course of acceleration and deceleration. **(A: 5.7 *g*, −9.5 *g*; see below.)**



4. Using the force vs. time graph (Figure 14.6) for human resistance to *g* forces, predict whether Stapp suffered serious injury in the course of his record-breaking trip.

5. A tall person sits down on a sofa to watch TV. Assume that the center of gravity of the person falls 1.0 m with constant gravitational acceleration in the course of sitting down. The sofa compresses by 0.05 m. Assume constant deceleration. Determine the *g* forces experienced by the person in the course of this sitting down. (**A: 20. *g***—so be kinder to couch potatoes!)

Exercises 6 and 7 concern an infant's rear-facing safety seat as illustrated here.



6. A rear-facing child safety seat holds a child of mass 12. kg rigidly within the interior of a car. The area of contact between the seat and the child is 0.10 m^2. The car undergoes a 30. mph collision. The car's crumple zone causes the distance traveled by the rigid interior to be 1.0 m. Give the stress experienced by the child's body in terms of a fraction of the breaking strength of bone, assuming an infant's bones break at a stress of 10. MN/m^2. (**A:** As a fraction of breaking stress, $\mathbf{11 \times 10^{-3}}$, bones should hold and the infant should be safe.)

7. A rear-facing child safety seat holds a child of mass 25. kg rigidly within the rigid interior of a car. The area of contact between the seat and the child is 0.10 m^2. The car undergoes a 30. mph collision. The car has no crumple zone, but a harness attached to the car seat stops it uniformly within a distance of 0.30 m. According to the Gadd severity index, will the child sustain serious injury or death?

8. Consider a parachute as a safety device. When the parachute opens, the previously freely falling person typically has reached a speed of about 50. m/s. The parachute slows to a terminal speed of about 10. m/s in 1.3 s. Approximating this set of motions by a constant deceleration, what is the maximum g experienced by the parachutist? (**A:** $\mathbf{-3.1}$ $\boldsymbol{g}$)

9. In the previous problem, the force exerted by the parachute is spread by a harness in contact with 0.50 m^2 of the parachutist, and the parachutist has a mass of 75. kg. What is the force per unit area (stress) experienced by the person during the deceleration?

10. The parachutist in the previous two exercises hits the ground (still wearing the parachute!) and is stopped in a distance of 0.10 m. If this final deceleration is constant, calculate the Gadd Severity Impact of the landing. (**A: GSI = 370**)

11. A 75. kg person jumping from a 1.00×10^3 m cliff will reach a terminal speed of 50. m/s and uses a 1.00×10^2 m bungee cord to slow the descent. The bungee cord exerts a force F proportional to its extension, where F (in newtons) = (5.0 N/m) × (extension in m) and is designed to extend by 5.00×10^1 m in the course of bringing the user to a stop just above the ground. Is the maximum deceleration in g's experienced by the falling person more or less than the maximum deceleration experienced by a parachutist undertaking the same leap (excluding landing forces)?

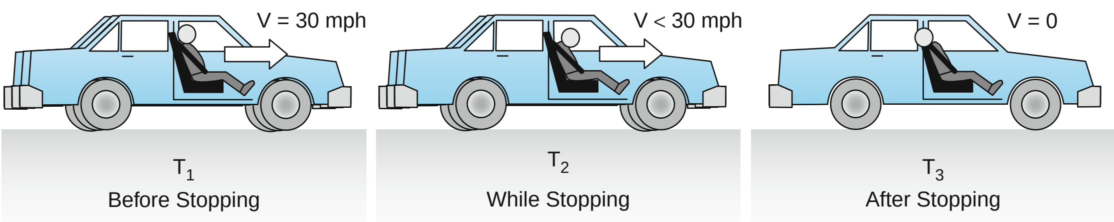
12. The air bag is designed to inflate very quickly and to subsequently yield if the driver hits it. A collision uniformly stops a car from 30. mph to 0.0 and then triggers the air bag. The driver is not seat belted and so hits the inflated air bag. This acts as a local crumple zone and consequently compresses by 0.20 meters as his head is brought to rest. According to the Gadd Severity Index, will the driver suffer serious injury? Assume constant deceleration of the driver's head after hitting the air bag. (**A: GSI = 380**, and no, the driver should not suffer serious injury or death.)

13. Two 100 kg football players wearing regulation helmets collide helmet to helmet while each is moving directly at each other at 10.0 m/s and come to a near instantaneous (<1 millisecond) stop. The area of contact is 0.010 m^2, and the helmets are each designed to provide a crumple zone of 0.025 m. What is the maximum stress exerted on each player? (**A:** $\mathbf{4.0 \times 10^7\ N/m^2}$)

14. A designer of football helmets has two options for increasing the safety of helmets but for economic reasons can implement only one. One option is to double the area of contact that will be experienced in a helmet-to-helmet collision. The other is to double the crumple distance experienced in a helmet-to-helmet collision. Which will be more effective in reducing the maximum stress? (**Hint:** Try the previous exercise first.)

15. A soccer player “heads” a wet 0.50 kg soccer ball moving directly toward him by striking it with his forehead. Assume the player initially moves his head forward to meet the ball at 5.0 m/s and the head stops after the ball compresses by 0.050 m during impact. Assume the deceleration of the head is constant during impact. Compute and comment on the calculated Gadd Severity Impact of heading a soccer ball under these conditions.

16. A car strikes a wall traveling 30. mph. The driver’s cervical spine (basically the neck) first stretches forward relative to the rest of the body by 0.010 m and then recoils backward by 0.020 m as shown below. Assume the spine can be modeled by a material with a modulus E = 10. GPa and a strength of 1.00×10^2 MPa. Will the maximum stress on the cervical spine during this whiplash portion of the accident exceed the strength of the spine? Assume a 0.15 m length of the cervical spine. (**A: Stress on cervical spine is greater than its tensile strength.**)



17. Which do you think has been more effective in reducing fatalities on American highways, seat belts or air bags? Give an engineering reason for your answer, containing variables, numbers, and units. (**Hint:** Recall the SSSA formula previously developed and consider what other safety element is designed into a modern automobile.) Then go to the Web and see if you were right.

18. As a bioengineer at the Crash Safety Test Facility of a major automobile company, you are asked to provide more data for the Gadd Severity Index (Figure 14.6). Your boss suggests using live animals, dogs and cats from the local pound, in hard impact tests and then inspecting them for injury. You know their injuries will be severe or fatal, and using dogs and cats seems cruel. What do you do?

a. Nothing. Live animals are used regularly in product testing, and besides, they will probably be killed in the pound anyway.

b. Suggest using dead animals from the pound, since their impact injuries probably don’t depend on whether or not they are alive.

c. Suggest using human cadavers since you really want data on humans anyway.

d. Suggest developing an instrumented human mannequin for these tests.

Use the Engineering Ethics Matrix format to summarize your conclusions.

19. You are now a supervisor in the bioengineering department of a major motorcycle helmet manufacturer. Your engineers are testing motorcycle helmets manufactured by a variety of your competitors. Motorcycle helmets contain an inner liner that crushes upon impact to decrease the deceleration of the head on impact. This liner material is very expensive and can be used only once (i.e., once the helmet sustains a single impact it must be replaced). Your company has developed an inexpensive liner that will withstand multiple impacts but is less effective on the initial impact than any of your competitors. The vice president for Sales is anxious to get this new helmet on the market and is threatening to fire you if you do not release it to the manufacturing division. What do you do?

a. Since your company has invested a lot of money in the development of this helmet, you should release it, and besides, if you don't, someone else will.
b. Recommend continued testing until your company's product is at least as good as the worst competitor's product.
c. Contact your company's legal department to warn them of a potential product liability problem and ask for their advice.
d. Go over the vice president's head and explain the problem to the company's president.

Use the Engineering Ethics Matrix format to summarize your conclusions.

20. During World War II, Nazi Germany conducted human medical experimentation on large numbers of people held in its concentration camps. Because many German aircraft were shot down over the North Sea, they wanted to determine the survival time of pilots downed in the cold waters before they died of hypothermia (exposure to cold temperatures). German U-boat personnel faced similar problems. In 1942, prisoners at the concentration camp in Dachau were exposed to hypothermia and hypoxia experiments designed to help Luftwaffe pilots. The research involved putting prisoners in a tank of ice water for hours (and others were forced to stand naked for hours at subfreezing temperatures) often causing death. Research in the pursuit of national interests using available human subjects is the ultimate example of questionable bioengineering. Since the Nazi scientific data were carefully recorded, this produces a dilemma that continues to confront researchers. As a bioengineer today, should you use these data in the design of any product (such as cold-weather clothing or hypothermia apparatus for open heart surgery)?
a. Since these experiments had government support and were of national interest at the time, they should be considered valid and available for scientific use now.
b. You should use these data, since similar scientific experiments have been conducted in other countries during periods in which national security was threatened, and these data are not questioned today. Even the United States conducted plutonium experiments on unsuspecting and supposedly terminally ill patients (some of whom survived to old age!).
c. This is just history and should have no bearing on the value or subsequent use of the data obtained.
d. Do not use the data.

Use the Engineering Ethics Matrix format to summarize your conclusions.

CHAPTER

Manufacturing Engineering

15

Source: © iStockphoto.com/Alexey Dudoladov

Virtually everything that we use at home, at work, and at play was manufactured. Manufactured goods are everywhere—aircraft, bicycles, electronics, coat hangers, automobiles, refrigerators, toys, clothing, cans, bottles, cell phones, and so on. Even the humble pencil and the paper clip[1] were triumphs of manufacturing processes.

15.1 WHAT IS MANUFACTURING?

You probably have a general idea about manufacturing processes, but let's look at the word itself. The word "manufacture" derives from two Latin words: *manu* (meaning "by hand") and *factum* (meaning "made"). We generally think of manufacturing taking place in a *factory*: an abbreviation of the eighteenth-century word "manufactory."

> *Manufacturing is the process of converting (either by manual labor or by machines) raw materials into finished products, especially in large quantities.*

Manufacturing covers a wide variety of processes and products. It involves the production of many different types of goods that range from food to microcircuits to airplanes to health care equipment. The number and complexity of the processes involved in the production of these items varies. Some products are basic such as flour, cheese, leather, and iron. Some are merely altered such as metal ingots, Portland cement, and chemicals like gasoline. Others are moderately changed like wire rods, metal pipes, glass bottles, soap, cloth, and paper, and things like vehicles, computers, medicines, and cell phones are elaborately transformed.

Manufacturing began more than 10,000 years ago when nomadic people settled in one geographic area and began to plant and grow food and domesticate animals. Skilled artisans did the manufacturing, and later the guild system (the predecessors of modern trade unions) evolved to protect their privileges and trade secrets.

Today's manufacturing engineers apply scientific principles to the production of goods. Manufacturing engineers design the processes and systems to make products with the required functionality, high quality, at the lowest price, and in ways that are environmentally friendly. Figure 15.1 illustrates a typical manufacturing process, from artisans to their market.

We will introduce machining processes including **drilling, lathe work milling, welding, extrusion, pultrusion, blow molding**, and **thermoforming**. Modern manufacturing methods include **just-in-time** or

[1]Petroski, H. *The Evolution of Useful Things: How Everyday Artifacts—From Forks and Pins to Paper Clips and Zippers—Came to Be as They Are*, 1992; and *The Pencil: A History of Design and Circumstance*, 1989; Alfred A. Knopf.

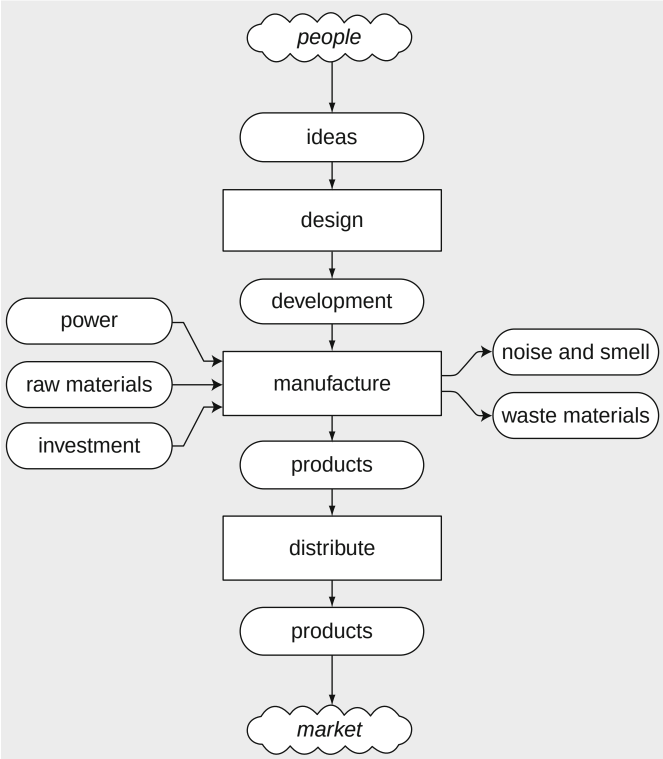


FIGURE 15.1 Typical Manufacturing Process

JIT inventory control, and **flexible, lean**, and **life cycle** manufacturing. A key concept in manufacturing is the recognition of **variability** and **tolerancing**, which we will describe by introducing **statistical methods** that include a relatively recent concept called **Six Sigma**. An important variable is called the **standard deviation** (**sigma**), which will be central to statistical methods.

15.2 EARLY MANUFACTURING

Before the Industrial Revolution, the production of goods for sale occurred in homes on farms and in villages to provide additional income. This "cottage" industry provided extra income to a farming family that would take in weaving, sewing, dyeing, and so forth, which then were sold to a retailer.

In 1776 the Scotsman Adam Smith[2] (1723–1790) published his famous book *An Inquiry Into the Nature and Causes of the Wealth of Nations*. This was one of the first studies on the production techniques in the colonial era.

[2]Adam Smith also is known for his explanation of how rational self-interest and competition can lead to economic well-being and prosperity without government intervention. The French expression *Laissez-faire* (roughly translatable as "to leave alone") became closely associated with his name. His work also helped to create the modern academic discipline of economics and provided one of the best-known rationales for free trade and capitalism.

In his book, he discussed how the division of labor could significantly increase production. One example he used was the making of pins. One worker could probably make only 20 pins per day. But if 10 people divided up the 18 steps required to make a pin, they could make 48,000 pins per day (240 times as many).

Eli Whitney (1765–1825) is most famous for his invention of the cotton gin.[3] However, the gin was a minor accomplishment compared to his perfection of the concept of *interchangeable parts*. At the time, the United States had limited human resources and scarce manufacturing labor. Whitney developed the concept of interchangeable parts at about 1799, when he took a contract from the US Army for the manufacture of 10,000 muskets at the then unbelievably low price of $13.40 each. Whitney is credited for inventing what is called the **American system of manufacturing**—the combination of power machinery, interchangeable parts, and the division of labor that would become one of the foundations of the industrial revolution.

15.3 INDUSTRIAL REVOLUTION

The Industrial Revolution[4] began in Great Britain during the last half of the eighteenth century and spread through Europe and the United States in the nineteenth century. In the twentieth century, industrialization extended to Asia and the Pacific Rim. Although modern manufacturing processes and economic growth continue to spread throughout the world, however, many people have yet to experience the benefits of improved economic and social conditions typical of an industrial revolution.

The key technology that brought about the Industrial Revolution was the invention and improvement of the steam engine. First invented by the English blacksmith Newcomen in the late seventeenth century, it was used initially to pump water out of flooded tin and coal mines. Increased coal production subsequently led to the smelting of iron, which in turn produced vast array of new metals and associated technologies.

The Industrial Revolution produced sweeping social changes,[5] including the movement of people to cities, the availability of a greater variety of material goods, and new ways of doing business. Goods that had been made in the home began to be manufactured at lower cost in a factory in a city. The resulting economic development combined with superior military technology to make the nations of Europe and the United States the most powerful in the eighteenth and nineteenth century world.

Between 1850 and 1900, machine tool technology advanced due, in part, to the development of engineering documentation through accurate detailed drawings. However, as products moved from one process to the next in factories, there were a number of unresolved questions: What should happen between processes, how should sequential processes be arranged, how should processes function as a system, and what was the best way for workers to carry out their tasks?

In the 1890s Frederick W. Taylor (1856–1915) began to study the process of how work was carried out. Taylor felt that the industrial management was poor, and that it could be formulated as an academic

[3]The term “gin” used here is a contraction of the word “engine,” since Whitney’s device was a creative machine (engine) for removing seeds from hand-picked cotton.

[4]The Industrial Revolution is called a revolution because it significantly and rapidly changed society. Over the course of human history, there has been only one other group of changes as significant as the Industrial Revolution. This is what anthropologists call the Neolithic Revolution, which took place in the later part of the Stone Age. In the Neolithic Revolution, people moved from social systems based on hunting and gathering to much more complex communities that depended on agriculture and the domestication of animals. This led to the rise of permanent settlements and, eventually, urban civilizations. The Industrial Revolution brought a shift from the agricultural societies created during the Neolithic Revolution to modern industrial societies.

[5]The excesses of the early industrial revolution were documented in Charles Dickens’ novels.

discipline. He thought that the best results would come from a partnership between management and workers. He felt that by working together there would be no need for trade unions. Taylor's "scientific" management process consisted of four principles:

1. Replace rule-of-thumb[6] methods with techniques based on a scientific study of the tasks.
2. Train workers rather than having them train themselves.
3. Provide detailed instruction and supervision for each worker.
4. Apply scientific management principles to planning the work.

Taylor's ideas became known as **Scientific Management**. The concept of applying science to management was useful; however Taylor unfortunately ignored the science of behavior. It seems that he had a peculiar attitude toward workers. He felt that they were not very bright, and generally incapable of understanding what they were doing (this was true even for rather simple tasks[7]).

In the early 1900s, Frank Gilbreth (1868–1924) developed a **time and motion study** and invented **process charting** to focus attention on the entire work process, including nonvalue-added steps that occur between production processes. Gilbreth became aware of these problems when he found ways to make bricklaying (his first trade) faster and easier. He later collaborated with his future wife, Lillian Moller Gilbreth[8] (1878–1972), to study the work habits of factory and clerical workers in several of industries to find ways to increase output and make their jobs easier. Lillian Gilbreth brought psychology into the mix by studying the motivations of workers and how their attitudes affect their work. She and her husband Frank were genuine pioneers in the field of industrial engineering.

Around 1910, Henry Ford developed the first all-inclusive manufacturing plan. As a farm boy in Michigan, he saw how animals were "disassembled" by a line of workers in slaughterhouses. By simply reversing this process, he created an "assembly" line process for manufacturing products. Ford took all the elements in a manufacturing system—people, machines, tooling, and supplies—and arranged them in a long continuous line to efficiently manufacture his early (Model T) automobiles. By the 1920s, you could buy a new Ford car for only $300 (about $3,400 in 2006 inflation-adjusted dollars). Ford was so successful that he quickly became one of the world's richest men. Ford is considered by many to be the first practitioner of just-in-time (JIT) and lean manufacturing.

After the 1920s the Ford system began to break down and an aging Henry Ford refused to change the system. Prosperity, the growth of labor unions, and competition from General Motors and Chrysler began to weaken Ford Motor Company. Annual model changes, multiple colors, and options were not embraced in Ford factories. General Motors developed new business and manufacturing strategies for managing and producing a variety of cars (Chevrolet, Pontiac, Buick, and Cadillac). By the mid 1930s, GM had passed Ford in dominating the automobile market. However, in the "Great Recession" that began in 2008, the finances of all of the US automobile manufacturers collapsed as demand for their products fell due to economic uncertainties and for concern about the future price of gasoline in the large models in which they specialized.

Edward Deming, an American quality control expert, was a consultant to the U.S. occupying forces in Japan after World War II. He introduced statistical quality control methods that allowed Japanese industries

[6]The term probably originated from woodworkers who used the length and width of their thumb as a "ruler" for measurement.

[7]"I can say, without the slightest hesitation," Taylor told a congressional committee, "that the science of handling pig-iron is so great that the man who is . . . physically able to handle pig-iron and is sufficiently phlegmatic and stupid to choose this for his occupation is rarely able to comprehend the science of handling pig-iron."

[8]She was the first woman elected into the National Academy of Engineering. She served as an advisor to Presidents Hoover, Roosevelt, Eisenhower, Kennedy, and Johnson on matters of civil defense, war production, and rehabilitation of the physically handicapped.

to manufacture very high-quality products. Deming's methods at first were not understood or appreciated by the rest of the world until Japan eventually became a world leader in exporting superior goods and technology. Motorola expanded Deming's statistical quality approach to reduce its defective products to an amazing one defect in 3.4 million. They called their process **Six Sigma**, which now has been adopted by much of European and American industry.

15.4 MANUFACTURING PROCESSES

Modern manufacturing can be divided into the following four main categories: subtractive, additive, continuous, and net shape.

15.4.1 Subtractive Processes

The **subtractive** process involves material removal via machining (e.g., turning, milling, boring, grinding, cutting, and etching). The three principal machining processes are classified as **turning, drilling**, and **milling**.

Turning is a process that uses a lathe (see Figure 15.2) to produce "solids of revolution." The lathe can be operated by a person or a computer. A lathe that is controlled by a computer is known as a CNC (**computer numerical control**) machine. CNC commonly is used with many other types of machine tools besides the lathe.

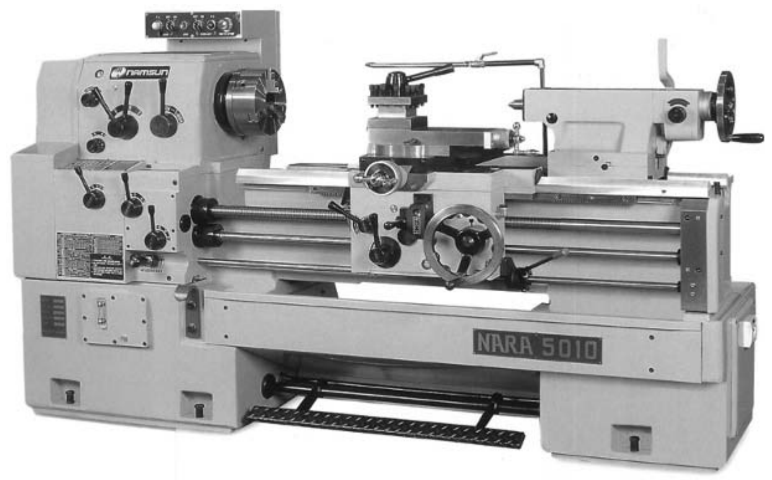


FIGURE 15.2 Typical Lathe

In a turning operation, the part is rotated (i.e., turned) while it is being machined. Turning produces straight, conical, curved, or grooved workpieces such as shafts, spindles, pins, and so forth. The majority of turning operations involve the use of a single point cutting tool such as that shown in Figure 15.3.

The material removal rate (MRR) is defined as the amount of material removed per unit time, and it typically has units of mm^3/min or in^3/min. In turning, each revolution of the workpiece removes a nearly ring-shaped layer of material. The material removed has a cross-sectional area equal to the axial distance the tool travels in one

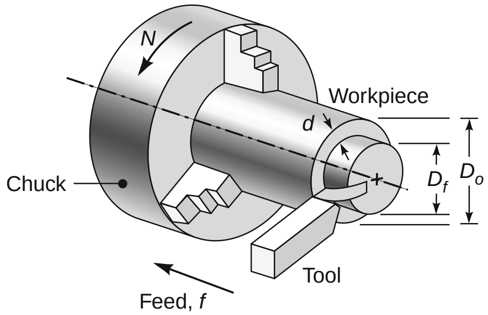


FIGURE 15.3 Basic Elements of a Turning Operation

revolution (called the **feed**, f) multiplied by the depth of the cut, d. Then the volume of material removed per revolution is approximately equal to the product of the circumference and the cross-sectional area of the workpiece, or:

$$\text{Material removed per revolution of the workpiece} = (\pi D_{avg}) \times (f \times d)$$

where D_{avg} is the **average diameter** of the workpiece during the cutting operation, or (see Figure 15.3):

$$D_{avg} = (D_o + D_f)/2 \tag{15.1}$$

If the workpiece is rotating at N RPM, then the material removal rate in turning is:

$$MRR_{turning} = \pi D_{avg} \times fd \times N \tag{15.2}$$

The **cutting speed**, $\boldsymbol{V_{cutting}}$, is the speed at which material is removed, and in turning it is:

$$V_{cutting} = \pi D_{avg} N \tag{15.3}$$

Then we can write the material removal rate as:

$$MRR_{turning} = f \times d \times V_{cutting} \tag{15.4}$$

In the design of manufacturing operations, the time required to machine a part directly impacts the manufacturing cost of the part. In turning, the cutting tool will have a **feed rate** (FR) in mm/minute or in/min determined by the tools' feed f times the rotational speed N of the workpiece, or:

$$FR = f \times N \tag{15.5}$$

If the axial cutting distance along the workpiece is L, then the time required to make the cut is:

$$t_{turning} = L/FR \tag{15.6}$$

Example 15.1

A shaft 6.0 inches long with an initial diameter of 0.50 inches is to be turned to a diameter of 0.48 inches on a lathe. The shaft rotates at 500. RPM on the lathe and the feed rate is 4.0 inches/minute. Determine:

1. the cutting speed
2. the material removal rate
3. the time required to machine the shaft

Need: $V_{cutting}$, $MRR_{turning}$, and $t_{machining}$.
Know: $L = 6.0$ inches, $D_o = 0.50$ inches, $D_f = 0.48$ inches, $D = 500.$ RPM, and the feed rate, $FR = 4.0$ in/min.
How: The cutting speed, material removal rate, and cutting time are given by Equations (15.3), (15.4), and (15.6).
Solve:

1. Equation (15.3) gives the cutting speed as $V_{cutting} = \pi D_{avg} N$, where
   $D_{avg} = (D_o + D_f)/2 = (0.50 + 0.48)/2 = 0.49$ inches.
   Then, $V_{cutting} = \pi(0.49)(500.) =$ **770 inches/minute.**
2. The depth of the cut is $d = (D_o - D_f)/2 = (0.50 - 0.48)/2 = 0.010$ inches, and the feed is $f = FR/N = 4.0/500.$ [in/min]/[rev/min] $= 0.0080$ in/rev. Then Equation (15.4) gives the MRR in turning as
   $\boldsymbol{MRR_{turning}} = \pi(0.49)(0.0080)(0.010)(500.)$ [in][in][in/rev][rev/min] $=$ **0.062 $\mathbf{in^3}$/min.**
3. The time required to machine the shaft is given by Equation (15.6) as:
   $\boldsymbol{t_{machining}} = L/FR = 6.0/4.0$ [in]/[in/min] $=$ **1.5 minutes.**

The power required by a machining operation, $P_{machining}$, is equal to the material's average machining energy per unit volume removed (see Table 15.1) multiplied by the material removal rate, or:

$$P_{machining} = (\text{Average Machining Energy per Unit Volume from Table 15.1}) \times MRR \tag{15.7}$$

and since the machining power is the product of the machining torque (in N·m or ft·lbf) and the rotational speed of the workpiece (in radians/minute[9]), then the machining torque is given by:

$$T_{machining} = P_{machining}/(2\pi N) \tag{15.8}$$

Table 15.1 Average Machining Energy Requirement per Unit Volume of Material Removed[10]

Material	Average Machining Energy per Unit Volume	
	$W \cdot s/mm^3$	$HP \cdot min/in^3$
Cast Iron	3.3	1.2
Carbon steels	5.5	2.1
Stainless steel	3.5	1.4
Aluminum alloys	0.70	0.28
Magnesium alloys	0.45	0.15
Nickel alloys	5.8	2.2
Titanium alloys	3.5	1.4

[9]Since there are 2π radians in one revolution (360 degrees), you can convert N in revolutions per minute (RPM) to radians per minute simply by multiplying it by 2π, as was done in Equation (15.8). Also, note that radians are dimensionless and do not appear in the final dimensions of the answer.
[10]Condensed from Table 21.2 in *Manufacturing Engineering and Technology*, 5th edition, by Kalpakjian and Schmid, Pearson Prentice Hall, 2006.

Example 15.2

Determine the machining power and torque required to machine the shaft in Example 15.1 if it was made from stainless steel.

Need: $T_{machining}$ and $P_{machining}$ for the shaft in Example 15.1.

Know: The material is stainless steel and from Example 15.1 we know that $MRR_{turning} = 0.062$ in^3/min and $N = 500.$ RPM.

How: Using Table 15.1 and Equations (15.7) and (15.8).

Solve: From Table 15.1, the average machining energy per unit volume of stainless steel removed is 1.4 Hp·min/in^3. Then, Equation (15.7) gives the machining power as:

$$\boldsymbol{P_{machining}} = 1.4(0.062)\ [\text{HP}\cdot\text{min/in}^3][\text{in}^3/\text{min}] = \mathbf{0.087HP}.$$

Then Equation (15.8) gives the cutting torque as:

$$T_{machining} = P_{machining}/(2\pi N) = 0.087/(2\pi \times 500)\ [\text{HP}]/[\text{rad/min}] = 2.8 \times 10^{-5}\ \text{HP}\cdot\text{min}.$$

Since 1 HP = 33,000 ft·lbf/min, then the machining torque is:

$$\boldsymbol{T_{machining}} = 2.8 \times 10^{-5}(33{,}000)\ [\text{HP}\cdot\text{min}][\text{ft}\cdot\text{lbf}/(\text{HP}\cdot\text{min})] = \mathbf{0.92\ ft\cdot lbf}.$$

Drilling is the process of using a cutting tool called a drill bit to produce round holes in materials such as wood or metal. The common twist drill shown in Figure 15.4 (the one sold in hardware stores) has a point angle of 118 degrees. This angle works well for a wide array of drilling operations. A steeper point angle (say, 90 degrees) works well for plastics and other soft materials. A shallower point angle (say 150 degrees) is suited for drilling steels and other tough materials. Drills with no point angle are used in situations where a blind, or flat-bottomed hole is required. Figure 15.5 shows a typical drilling machine, called a drill press.

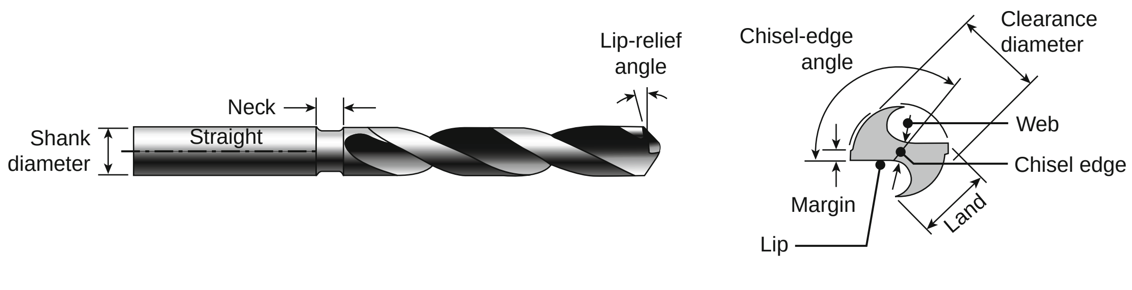


FIGURE 15.4 Characteristics of a Common Twist Drill

FIGURE 15.5 Typical Drill Press

Creating holes in a workpiece is a relatively common requirement. Drills are particularly good for making deep holes. The **feed**, f, of a drill is the distance the drill penetrates the workpiece in one revolution of the drill. The speed of the drill is the product of its feed and rotational speed, N, or:

$$V_{drill} = f \times N \tag{15.9}$$

The cross-sectional area of a drill with a diameter D is $\pi D^2/4$, and the material removal rate of a drill is the product of its cross-sectional area and its speed, or:

$$MRR_{drill} = (\pi D^2/4)fN \tag{15.10}$$

Equations (15.7) and (15.8) can also be used to calculate the drilling power and torque.

Example 15.3

You need to drill a 10. mm hole in a piece of titanium alloy. The drill feed is 0.20 mm/revolution and its rotational speed is 600. RPM. Determine:

1. the material removal rate
2. the power required
3. the torque on the drill

Need: MRR_{drill}, $P_{machining}$, and $T_{machining}$.
Know: $D = 10.$ mm, $f = 0.20$ mm/rev, and $N = 600.$ RPM.

How: The material removal rate is given by Equation (15.10). The power and torque are given by equations (15.7) and (15.8). The average machining energy per unit volume can be found in Table 15.1.

Solve:

1. Equation (15.10) gives: $\boldsymbol{MRR_{drill}} = (\pi D^2/4)fN = (\pi 10^2/4)(0.20)(600.)$ [mm^2][mm/rev][rev/min] $= 9{,}400$ mm^3/min $=$ **160 mm^3/s**.
2. From Table 15.1 for titanium alloys, the average machining energy per unit volume of titanium alloy removed is 3.5 W·s/mm^3.
   Then Equation (15.7) gives: $\boldsymbol{P_{machining}} = 3.5(160)$ [W·s/mm^3][mm^3/s] $=$ **560 W**.
3. From Equation (15.8) the torque on the drill is: $T_{machining} = P_{machining} / (2\pi N)$
   $= 560/(2\pi \times 600.)$ [W]/[radians/minute] $= 0.15$ W·min, $= 0.15(60)$ [W·min][s/min]
   $= 9.0$ W·s $= 9.0$ [J/s][s]$= 9.0$ J $=$ **9.0 N·m**.

Equations (15.5) and (15.6) can also be used to determine the time it takes to drill the hole to a specific depth. For example, if we were required to drill the hole in Example 15.3 to a depth of 50. mm, then the drill feed rate would be $FR = fN = 0.20(600.)$ [mm/rev][rev/min] $= 120$ mm/min, and the time required to drill the hole 50. mm deep would be $t_{machining} = L/FR = 50./120$ [mm]/[mm/min] $= 0.42$ minutes, or 25. seconds.

Milling is carried out on a machine tool called a "milling machine" that is used for the shaping of metal and other solid materials. It has a cutting tool that rotates about the spindle axis similar to a drill. However, a drill moves along only one axis, but on a milling machine the cutter and workpiece move relative to each other, generating a tool path along which material is removed along three axes. Often the movement is achieved by moving the workpiece while the cutter rotates in one place as shown in Figure 15.6. Milling machines may be manually operated, mechanically automated, or digitally automated via CNC (computer numerical control).

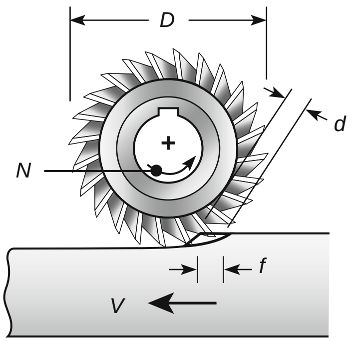


FIGURE 15.6 Characteristics of a Common Slab Milling Cutter

Milling machines (see Figure 15.7) can perform a vast number of operations, some of them with quite complex tool paths, such as slot cutting, planing, drilling, routing, and so forth.

FIGURE 15.7 Typical Milling Machine

Figure 15.6 illustrates the basic shape and operation of a milling cutter. The **cutting speed**, $V_{milling}$, of a cutter of **diameter** D rotating at N RPM is:

$$V_{milling} = \pi \times D \times N \tag{15.11}$$

The feed rate of the cutter is the rate at which its axis moves along the workpiece, and can be calculated from the **feed per tooth** f, the rotational speed of the cutter, N, and the **number of teeth on the cutter**, n, as:

$$FR = f \times N \times n \tag{15.12}$$

The material removal rate for a workpiece of **width** w, a **depth of cut** d, and a cutter **feed rate**, FR is:

$$MRR_{milling} = w \times d \times FR \tag{15.13}$$

The time required to mill a length L of material at a feed rate, FR, is:

$$t_{milling} = L/FR \tag{15.14}$$

Example 15.4

A flat piece of cast iron 4.0 inches wide and 15. inches long is to be milled with a feed rate of 20. inches/minute with a depth of cut of 0.10 inches. The milling cutter rotates at 100. RPM, and is wider than the workpiece. Determine:

1. the material removal rate
2. the machining power required
3. the machining torque required
4. the time required to machine the workpiece

Need: $MRR_{milling}$, $P_{machining}$, $T_{machining}$, and $t_{milling}$.

Know: The material is cast iron with a length of $L = 15.$ inches and a width of $w = 4.0$ inches. The feed rate (FR) $= 20.$ in/min, $N = 100.$ RPM, and the depth of the cut (d) $= 0.10$ inches.

How: Equation (15.13) provides the material removal rate. Equations (15.7) and (15.8) can be used to find the cutting power and torque, and Equation (15.14) gives the machining time.

Solve:

1. From Equation (15.13) we get: $\boldsymbol{MRR_{milling}} = wdFR$ $= (4.0)(0.10)(20.)$ [inches][inches][inches/min] $=$ **8.0 in^3/min**.
2. From Table 15.1 for cast iron, we find that the average machining energy per unit volume of cast iron removed is 1.2 HP·min/in^3. Equation (15.7) gives the machining power as: $\boldsymbol{P_{machining}} = 1.2(8.0)$ [HP·min/in^3][in^3/min] $=$ **9.6 HP**.
3. From Equation (15.8) the torque on the milling cutter is: $\boldsymbol{T_{machining}} = P_{machining}/(2\pi N) = 9.6/(2\pi \times 100.)$ [HP]/[radians/min] $= 0.015$ HP·min $= 0.015(33{,}000)$ [HP·min][ft·lbf/HP·min] $=$ **500 ft·lbf**.
4. Equation (15.6) gives the machining time as: $\boldsymbol{t_{machining}} = L/FR = 11/15.$ [in]/[20. in/minute] $= 0.75$ minutes $=$ **45 seconds**.

15.4.2 Additive Processes

The **additive** process is where material is added such as in joining (e.g., welding as in Figure 15.8, soldering, and gluing), rapid prototyping, stereolithography, 3D printing, and the application of composite layers of resin and fiber. Additive machining processes add material to a base object to create complex shapes. This is far less expensive than cutting an intricate product from a solid block of material by the subtractive process.

15.4.3 Continuous Processes

In the **continuous** process the product is produced continuously, such as in the extrusion of metals and plastics, and the pultrusion of composites. **Extrusion** is a continuous process of manufacture used to create objects of a fixed cross-sectional profile by pushing or drawing material through a die of the desired cross-section. In extrusion, a bar or metal or other material is forced from an enclosed cavity through a die orifice

FIGURE 15.8 Welding with an Oxy-Acetylene Torch

by a force applied by a ram (see Figure 15.9). The extruded product has the desired reduced cross-sectional area, and also has a good surface finish so that further machining is not needed. Extrusion products include rods and tubes with varying degrees of complexity in cross-section.

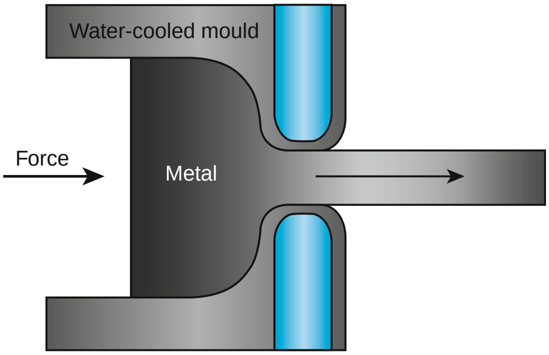


FIGURE 15.9 Extrusion Process

Pultrusion, shown in Figure 15.10, is a similar continuous process of manufacturing materials with constant cross-section except the material is **pulled** through a process or a die to its final shape. Pultrusion is the only continuous manufacturing process available for obtaining high quality composite profile, with good mechanical properties. Pultruded products normally are composed of high performance glass or carbon fibers embedded in a polymer matrix.

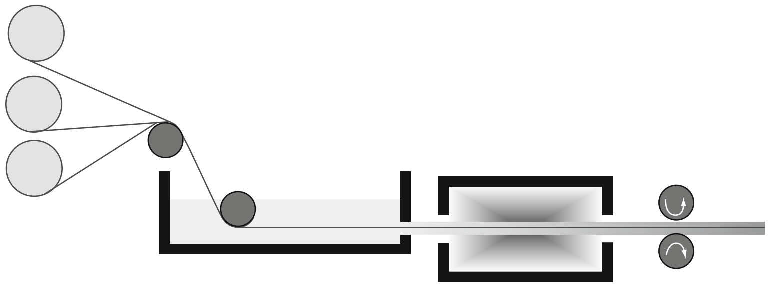


FIGURE 15.10 Typical Pultrusion Process

15.4.4 Net Shape Processes

A **net shape** process occurs when the output is at or near its final shape, such as stamping, forging, casting, injection molding, blow molding, and thermoforming.

Stamping is a metalworking process in which sheet metal is formed into a desired shape by pressing or punching it on a machine press. Stamping can be a single stage operation where only one stroke of the press produces the desired form, or could occur through a series of these stages. The most common stamping operations are piercing, bending, deep drawing, embossing, and extrusion.

Casting is a manufacturing process by which a liquid material is poured into a mold containing a cavity of the desired shape, and then allowed to solidify. The mold is then opened to complete the process. The casting process is subdivided into two distinct subgroups: expendable and nonexpendable mold casting. Casting most often is used for making complex shapes that would be otherwise difficult or uneconomical to make by other machining methods.

Injection molding is a manufacturing process used to make parts by injecting molten plastic at high pressure into a mold. Injection molding is used widely for manufacturing a variety of parts, from the smallest gears to entire automotive body panels. Injection molding is the most common method of production today, and is capable of tight tolerances.

The most commonly used thermoplastic (any remeltable plastic) materials are polystyrene, polypropylene, polyethylene, polycarbonate, and polyvinyl chloride (PVC is commonly used in extrusions such as pipes, window frames, and the insulation on wiring). Injection molding can also be used to manufacture parts from aluminum, zinc, or brass in a process called die casting. The melting points of these metals are, of course, much higher than those of plastics, but it is often the least expensive method for mass producing small metal parts.

Blow molding is a manufacturing process in which hollow plastic parts are formed. The blow molding process begins with melting the plastic and forming it into a tube-like piece with a hole in one end into which air can be injected. It is then clamped into a mold and air pressure pushes the plastic into the mold. Once the plastic has cooled and hardened the mold opens up and the part is ejected. Plastic soda bottles are made this way.

Thermoforming is a manufacturing process used with plastic sheets. Plastic sheets or film is converted into a finished part by heating it in an oven to its forming temperature, then stretching it on a mold and cooled. A thermoform machine typically utilizes vacuum to draw the plastic onto the mold in the forming process (see Figure 15.11).

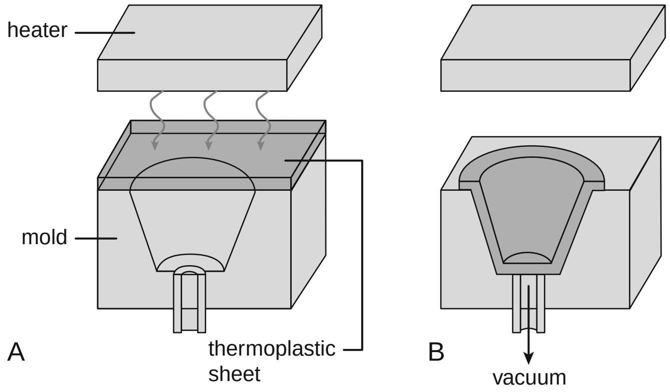


FIGURE 15.11A, B Thermoforming a Drinking Cup

Example 15.5

The energy needed to manufacture a reusable cup is significantly greater than that required to manufacture a disposable paper or foam cup as shown in the Table. For a reusable cup to be an improvement over a disposable cup on an energy basis, you have to use it multiple times. Determine the number of uses required for the energy per use of a reusable cup to become less than for a disposable cup (this is called the break-even point) if it takes about 180 kJ of energy to wash a single cup in a commercial dishwasher.[11]

Energy Required to Manufacture a Cup

Cup Type	kJ/Cup
Ceramic	14,000
Plastic	6,300
Glass	5,500
Paper	550
Foam	190

Need: The break-even point for a reusable cup.

Know: Manufacturing energy from the given table, and that it takes about 180 kJ of energy to wash a single cup in a commercial dishwasher.

How: If a cup is used 10 times, then each use costs one-tenth of the manufacturing energy. If it is used 100 times, then each use costs just one-hundredth of the manufacturing energy. However, to reuse a cup it has to be cleaned, and the energy required for each cleaning needs to be added to its manufacturing energy. The energy required per cup use can be computed from:

$$\textbf{Energy per use} = \textbf{Wash energy} + (\textbf{Manufacturing energy})/(\textbf{Number of uses})$$

Solve: Since it takes about 180 kJ of energy to wash a single cup in a commercial dishwasher, then the total amount of energy per cup use is given by the preceding equation. So how many times would you have to use a ceramic cup in order to break even with the manufacturing energy used in producing a foam cup?

[11]This work was done by Professor Martin B. Hocking at the University of Victoria, Canada.

$$\text{Foam cup energy per use} = 190 \text{ kJ/cup (from the Table on page 315)}$$

$$\text{Wash energy} = 180 \text{ kJ/cup}$$

$$\text{Manufacturing energy} = 14,000 \text{ kJ/cup (Ceramic)}$$

Then, 190 = 180 + 14,000/N, where N = number of uses (and washes) of the ceramic cup. Solving for N gives **N = 1,400**. Therefore, you would have to use and wash the ceramic cup 1,400 times in order to equal the manufacturing energy of a single foam cup.

The following table shows how the energy per use of the three reusable cups declines the more you use them.

Number of Times a Reusable Cup Would Need to Be Used to Break Even with a Disposable Cup		
Reusable Cup	**Paper**	**Foam**
Ceramic	38	700
Plastic	17	630
Glass	15	550

These results are extremely sensitive to the amount of energy the dishwasher uses for cleaning each cup. The energy calculation for the dishwasher, requiring 180 kJ/cup-wash, is barely less than the manufacturing energy of the foam cup, 190 kJ/cup. If an even slightly less energy efficient dishwasher was used, then the reusable cups would never have broken even with the foam cup.

In situations where cups are likely to be lost or broken and thus have a short average lifetime, disposable cups are the preferred option *if* energy usage is the main criterion.

15.5 MODERN MANUFACTURING

What makes a good manufacturing process? Today the overall performance of a company is often dictated by the design of its manufacturing process and facility. It turns out that there are four key elements that are important in designing a good manufacturing process.

The first key element is the **movement of materials**. A well-designed facility results in efficient material handling, small transportation times, and short queues. This, in turn, leads to low work-in-process levels, effective production management, decreased cycle times and manufacturing inventory costs, improved on-time delivery performance, and higher product quality. The second key item is **time**. The setup time plus process time from order to shipping is extremely important in meeting customer demands and keeping a constant flow of product.

The third key element is **cost**. Material, labor, tooling, and equipment costs must all be monitored and controlled. The fourth key element is **quality**. Customers today demand high quality products, so deviations from design specifications must be kept to a minimum.

Modern manufacturing philosophies such as JIT manufacturing, flexible manufacturing, lean manufacturing, and life cycle manufacturing contain all of these elements. They are discussed in detail in the following sections.

15.5.1 Just-in-Time Manufacturing

Just-in-time (JIT) manufacturing is easy to understand. Things are planned to occur *just-in-time*. For example consider your activities today. You left your room *just-in-time* to walk to your first class. That class ended *just-in-time* for you to go to your next class, which ended *just-in-time* for you to have lunch. Just-in-time processes may seem easy, but achieving them in a manufacturing environment is often very difficult.

In a manufacturing process, parts should arrive at the factory just-in-time to be distributed to the work stations, and they should arrive at a work station just-in-time to be installed by a worker. Finally, the factory should produce finished products just-in-time to be handed to a waiting customer. This will eliminate any inventory of parts and products, and in theory, JIT does not need *any* inventory of raw materials, parts, or finished products.

JIT originated in Japan with the Toyota motor company, and initially was known as the "Toyota Production System." After the Second World War, the president of Toyota wanted to compete aggressively with American automobile production. What he found was that one American car worker could produce about nine times as much as a Japanese car worker. By further studying the American automobile industry, they found that American manufacturers made large quantities of each item (parts, engines, car model, etc.) before changing to a new item. American car manufacturers also stocked all the parts needed to assemble their cars.

Toyota felt that this would not work in Japan because the small Japanese market wanted small quantities of many different car models. Consequently they developed a production system that provided parts to their workers only when they were needed (i.e., just-in-time). They analyzed the concept of "waste" in general terms that included wasted time and resources as well as wasted materials. They subsequently identified the following seven wasteful activities that could be minimized or eliminated:

1. **Transportation** (moving products that are not required to perform the processing)
2. **Inventory** (all components, work-in-progress, and finished product not being processed)
3. **Motion** (workers or equipment moving more than is required to perform the processing)
4. **Waiting** (workers waiting for the next production step)
5. **Overproduction** (production ahead of demand)
6. **Overprocessing** (due to poor product design)
7. **Defects** (the wasted effort involved in inspecting for and fixing defects)

There are a number of Japanese jargon terms associated with JIT that are in common use today. For example, the term **Andon** refers to trouble lights that immediately signal to the production line that there is a problem, and the production line is stopped until the problem is resolved. In the Toyota system, the Andon light indicating a stopped production line is hung from the factory ceiling so that it can be seen clearly by everyone. This raises the profile of the problem and encourages rapid attention to a solution. However, when General Motors instituted an Andon light, US workers were reluctant to take the responsibility for stopping a production line, and defective products were produced. General Motors later solved this problem by allowing workers to signal that they had a problem without stopping the production line.

The JIT philosophy involves the elimination of waste in its many forms, a belief that ordering and stocking costs can be reduced, and always striving to progress through continuous improvement. The elements of a JIT process typically include the following:

- Meet regularly with the workers (daily/weekly) to discuss work practices and solve problems.
- Emphasize consultation and cooperation with the workers rather than confrontation.
- Modify machinery to reduce setup time.
- Reduce buffer stock.
- Expose problems, rather than have them covered up.

It is not necessary to apply JIT to all stages of a process. For example, you could keep large stocks of raw material but operate the production process internally in a JIT fashion (hence eliminating work-in-progress stocks).

15.5.2 Flexible Manufacturing

A **flexible manufacturing system** (**FMS**) is a manufacturing system that contains enough flexibility to allow the system to rapidly react to production changes. This flexibility generally is considered to fall into two categories. The first is *machine flexibility*. This allows the system to be changed to produce new product types, and to change the order of operations executed on a part. The second is *routing flexibility*. This consists of the ability to use multiple machines to perform the same operation on a part, as well as the system's ability to absorb large-scale changes, such as in volume, capacity, or capability.

Most FMS systems consist of three main systems: a **material handling system** to optimize the flow of parts, a **central control computer** that controls material movement, and the **working machines** (often automated CNC machines or robots).

The use of robots in manufacturing industries has many benefits. Each robotic cell is connected with a material handling system, which makes it easy to move parts from one robotic cell to another. At the end of processing, the finished parts are routed to an automatic inspection cell and then removed from the Flexible Manufacturing System. This process is illustrated in Figure 15.12.

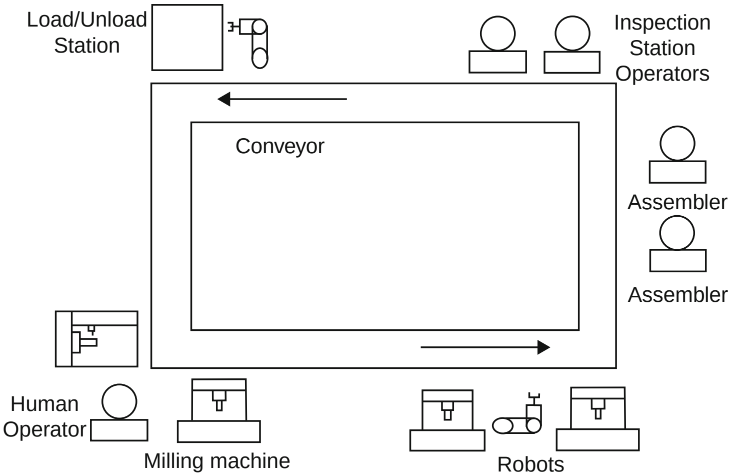


FIGURE 15.12 Flexible Manufacturing System

The main advantages of a FMS is its high flexibility in managing manufacturing resources like time and effort in order to manufacture a new product. The best application of an FMS is found in the mass production of small sets of products.

15.5.3 Lean Manufacturing

Lean Manufacturing is the optimal way of producing goods through the removal of waste and implementing flow, as opposed to batch processing. Lean manufacturing is a generic process management philosophy derived mostly from Toyota, and focuses mainly on reduction of the seven wastes originally identified by Toyota (see JIT, earlier).

Lean Manufacturing is focused on getting the right things, to the right place, at the right time, in the right quantity to achieve perfect work flow while minimizing waste, and being flexible and able to change. All of these concepts have to be understood, appreciated, and embraced by the workers who build the products and own the processes that deliver the value. The cultural and managerial aspects are just as important as the actual tools or methodologies of production itself. Lean Manufacturing tries to make the work simple enough to understand, to do, and to manage.

The main principles of Lean Manufacturing are zero waiting time, zero inventory, internal customer pull instead of push, reduced batch sizes, and reduced process times.

15.5.4 Life Cycle Manufacturing

Today the term "green" is synonymous with environmental sustainability. Sustainability is the development and application of processes, or the use of natural resources that can be maintained indefinitely. As public environmental awareness increased, manufacturing facilities began examining how their activities were affecting the environment.

To understand the environmental impact of a manufacturing process, you need to examine all the inputs and outputs associated with the entire existence of the product. This "cradle to grave" analysis is called life cycle analysis, or simply LCA. The concept of conducting a detailed examination of the life cycle of a product or a process in response to increased environmental awareness on the part of the public, industry, and government is relatively recent.

Just as living things are born, age, and die, all manufactured products have an analogous life cycle. Each stage of a product's development affects our environment from the way we use a product to the way we dispose of it when we are finished with it (see Figure 15.13). Looking at a product's life cycle helps engineers understand the connections between the Earth's natural resources, energy use, climate change, and waste disposal. Product life cycle analysis focuses on the processes involved in the entire production system, including:

- The design and functionality of the product
- The extraction and processing of raw materials
- The processes used in manufacturing
- Packaging and distribution
- How the product is used by the customers
- Recycling, reuse, and disposal of the product

FIGURE 15.13 Life Cycle Manufacturing

Engineers use LCA during product design and development and by manufacturers during the production stage. However, LCA has its greatest potential to reduce environmental impacts during the design stage when 70 percent of the total product cost is determined (see Figure 15.14).

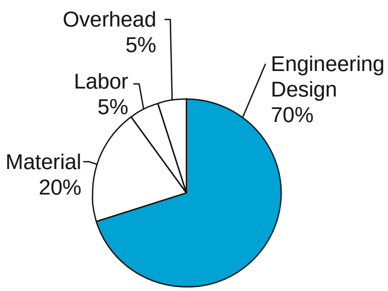


FIGURE 15.14 The Influence of Engineering Design on the Cost of Manufacturing

Life cycle analysis uses detailed environmental and energy estimates for the manufacture of a product, from the mining of the raw materials to production and distribution, end use and possible recycling, and disposal. LCA enables a manufacturer to measure how much energy and raw materials are used, and how much solid, liquid, and gas waste is generated at each stage of a product's life.[12]

15.6 VARIABILITY, DEMING, AND SIX SIGMA

It's instructive to have the students in the class measure the same thing with the same instrument. You will find that there will be slightly different results. For example, take a brand new unsharpened pencil and a metal ruler scribed in cm and mm. Each member of the class should measure the length of the pencil and record the number. Then, when all data have been collected, write all the measurements on the blackboard. Table 15.2 shows some typical numbers:

Table 15.2 Raw Measurements of Pencil Length

Length in mm	Length in mm	Length in mm
21.2	22.7	21.0
21.0	20.1	21.2
20.7	20.9	20.8
21.7	20.9	21.2
21.7	21.5	21.2

Note that not all the measured lengths are the same. Also, one person reported a number much greater than the other measurements (22.7 mm) and another measured 20.1 mm, which is quite a bit smaller than most of the other measurements. Outliers are quite common due to inattention or misreading of instruments.

[12] LCA normally ignores second-generation impacts, such as the energy required to fire the bricks used to build the kilns used to process the raw material.

Which reading is correct? Any, none, or some combination? Most of us would agree that the average or mean defined as the sum of all the measurements divided by the number of measurements has less bias than does any individual measurement.

Let's analyze what we have measured: the mean or average measured length is

$$\bar{x} = \left(\sum_{i=1}^{N} \frac{x_i}{N}\right) = 21.2 \text{ mm}$$

So, you say the pencil is 21.2 mm long. But is it? Suppose we had dropped the outlier of 22.7 mm. The mean of the remaining data is then 21.1 mm. Is that correct? Or should we drop the other outlier of 20.1 mm, giving a new mean of 21.3 mm. Again, is that correct?

Let's try something else. Subtract each measurement from the mean of 21.2 mm as in Table 15.3.

Table 15.3 Mean Value of Pencils Minus Actual Length

Mean–Measurement (mm)	Mean–Measurement (mm)	Mean–Measurement (mm)
0.0	–1.5	0.2
0.2	1.1	0.0
0.5	0.3	0.4
–0.5	0.3	0.0
–0.5	–0.3	0.0

If you add all these deviations they will total 0.0 (since, after all, that is what we mean by an average). So, let's try a different approach—plot these deviations as a frequency histogram, with the frequency of each measurement (say in 0.2 mm bins) versus the measured values as in Figure 15.15.

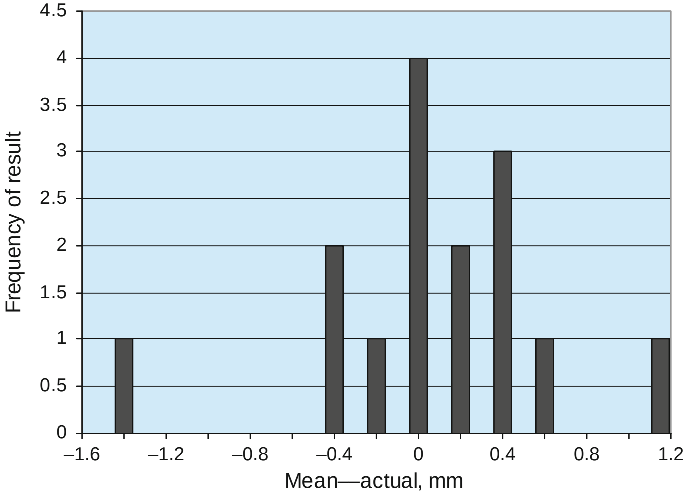


FIGURE 15.15 Histogram of Deviations from Mean

The histogram shows that the most frequent measurements are centered about zero so that many of the people in the class did think the actual length of the pencil is about 21.2 mm.

But is ignoring all those who did not measure exactly 21.2 mm the best we can do? Suppose we square these deviations—some of which are positive and some negative—so then each squared term is positive. An individual squared term is still a poor measure of all the data points. We need a measure that will eliminate the bias of a single measurement just as the mean reduced the bias of a single number. So add all the squares and take the square root of them and then divide by the number of points.[13]

$$\sigma = \frac{\sqrt{\sum_{i=1}^{N} (\overline{x} - x_i)^2}}{N} \tag{15.15}$$

The Greek letter sigma σ defined in Equation (15.15) is called the **standard deviation** of the data. In our case $\sigma = 0.6$ mm, but what does that mean?

Mathematical statistics was one of the methods introduced to the Japanese manufacturers by W. Edwards Deming after World War II to improve the quality of their products; it has been so successful that many large international companies have emulated them.

A typical analysis starts with a large sample generalization of the histogram in Figure 15.15 known as the Normal Error curve, the Bell curve, the Normal distribution curve, or the Gaussian[14] distribution shown in Figure 15.16 whose equation is $Frequency = \dfrac{Exp(-0.5Z^2)}{\sigma\sqrt{2\pi}}$.

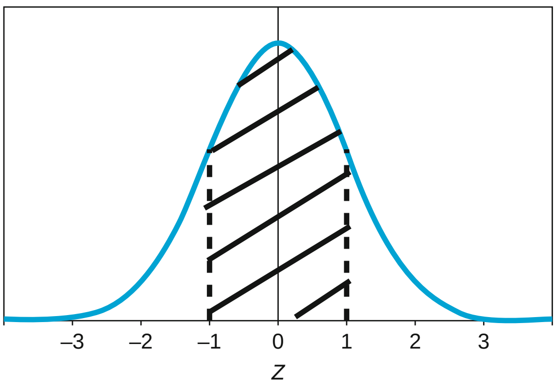


FIGURE 15.16 Normal Error Curve

The ordinate (vertical axis) is the probability of occurrence (the fraction of times the measurements will lie in some small range), also called the frequency. The abscissa is defined by the difference between a specific point x_i and the mean divided by the standard deviation, given by the symbol Z, i.e.,

$$Z = \frac{(x_i - \overline{x})}{\sigma} \tag{15.16}$$

[13]For mathematical reasons, if we use all N points of data to determine the mean of the data, we should divide by $N-1$ rather than N; however, for proper use of statistical methods, the number of samples should be large, and whether we divide by N or by $N-1$ should make no difference. If it does make a noticeable difference, you should reevaluate whether you should be using statistical methods at all! $N \sim \geq 30$ is usually sufficient.

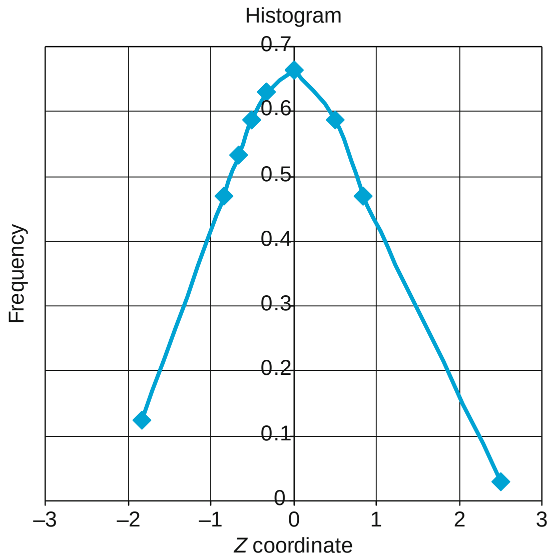
[14]Carl Friedrich Gauss (1777–1855) was one of the greatest mathematicians of all time.

Example 15.6

Convert the numerical data in Table 15.3 to Z coordinates. Then, using Equation (15.17), plot the frequency vs. Z.

$$Frequency = \frac{Exp(-0.5Z^2)}{\sigma\sqrt{2\pi}} \tag{15.17}$$

Need: The frequency of occurrence for a normal distribution of pencil length measurements.
Know: Equation (15.17) gives the required frequency of occurrence.
How: Using a spreadsheet.



Solve: Use a spreadsheet to calculate and plot the results.

This curve has the general characteristics of Figure 15.16, but why do we go to these lengths to plot it in this form? The reason is that with this coordinate system the area under this curve is 1. This means that there is 100% probability that the measurement will be somewhere under the curve if we were to go out to $Z = \pm\infty$. In this standardized form the abscissa can be thought of as the number of standard deviations corresponding to each point. The utility of this normalized coordinate is that:

- For $Z = \pm 1$ standard deviations ($\pm 1\sigma$), the area under the normal error curve, as shown in Figure 15.16, contains 68% of the data
- For $Z = \pm 2\sigma$ the area contains 95 percent of the data
- For $Z = \pm 3\sigma$ the area will contain 99.7 percent of the data

The areas under the tails of the normal error curve at an abscissa value greater than 1σ, 2σ, 3σ, and so forth contain the fraction of measurements that fail because they are too far from the mean. Every-day statistics (e.g., medical tests protocols, electorate polling, etc.) usually are accepted when 95 percent ($\pm 2\sigma$) of the data are included under the curve.

A Six Sigma[15] process means you are rejecting parts, processes, and such for which $Z = \pm 6\sigma$; this means that very few defective parts are out of tolerance since, at these large abscissa values, the area under the tails of the normal error curve is very small (about 2 per billion). This can have a profound effect on the cost of the acceptable units.

Example 15.7

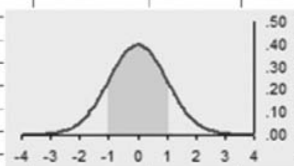
You manufacture 10,000 metal rods that you hope are 21.2 ± 0.2 mm long at a cost to you of $1.00 each. How much each do the *acceptable* rods cost you if their final machined measurements are normally distributed with a standard deviation of 0.6, 0.3, and 0.1 mm, respectively? In other words, what fraction of the rods' measurement is expected to be too far off the desired mean to accept? Note: Your spreadsheet will have several functions that will calculate areas under the normal error curve given *Z* specifications.

Need: Cost of the acceptable metal rods that are within the desired lengths between 21.0 and 21.4 mm and with a mean of 21.2 mm.

Know: Lengths are normally distributed with σ = 0.6, 0.3, and 0.1 mm, respectively.

How: Use fraction of acceptable rods between $Z = \pm 0.2/\sigma$ or 0.333, 0.667 and 2.0 corresponding to a mean length of 21.2 mm and standard deviations of 0.6, 0.3 and 0.1 mm, respectively.

	A	B	C	D	E	F	G	H	I
1	Mean, mm	Spread, mm	Std dev, mm	Z	Normsdist(Z)	Fraction passing	Number passing	Cost/rod	
2	21.2	0.2	0.6	0.333	0.631	0.261	2,611	$3.83	
3	21.2	0.2	0.3	0.667	0.748	0.495	4,950	$2.02	
4	21.2	0.2	0.1	2.00	0.977	0.954	9,545	$1.05	
5	21.2	-0.2	0.6	-0.33	0.369	0.261	2,611	$3.83	
6	21.2	-0.2	0.3	-0.67	0.252	0.495	4,950	$2.02	
7	21.2	-0.2	0.1	-2.00	0.023	0.954	9,545	$1.05	
8									
9		See:	http://www.exceluser.com/explore/statsnormal.htm						
10									
11	NORMSDIST(z)								
12									
13	NORMSDIST translates the number of standard deviations (z) into cumulative probabilities.								
14–20									
21	To illustrate:								
22									
23	=NORMSDIST(-1) = 15.87%								
24–26									
27	Therefore, the probability of a value being within one standard deviation of the mean is the difference between these values, or 68.27%.								
28	This range is represented by the shaded area of the chart.								



[15] Six Sigma is a registered service mark and trademark of Motorola, Inc. Motorola has reported over $17 billion in savings from Six Sigma as of 2006.

	A	B	C	D	E	F	G	H
1	Mean, mm	Spread, mm	Std dev, mm	Z	Normsdist(Z)	Fraction passing	Number passing	Cost/rod
2	21.2	0.2	0.6	=B2/C2	=NORMSDIST(D2)	=IF(D2<=0, 1-2*E2, 1-(1-E2)*2)	=10000*F2	=10000/G2
3	21.2	0.2	0.3	=B3/C3	=NORMSDIST(D3)	=IF(D3<=0, 1-2*E3, 1-(1-E3)*2)	=10000*F3	=10000/G3
4	21.2	0.2	0.1	=B4/C4	=NORMSDIST(D4)	=IF(D4<=0, 1-2*E4, 1-(1-E4)*2)	=10000*F4	=10000/G4
5	21.2	–0.2	0.6	=B5/C5	=NORMSDIST(D5)	=IF(D5<=0, 1-2*E5, 1-(1-E5)*2)	=10000*F5	=10000/G5
6	21.2	–0.2	0.3	=B6/C6	=NORMSDIST(D6)	=IF(D6<=0, 1-2*E6, 1-(1-E6)*2)	=10000*F6	=10000/G6
7	21.2	–0.2	0.1	=B7/C7	=NORMSDIST(D7)	=IF(D7<=0, 1-2*E7, 1-(1-E7)*2)	=10000*F7	=10000/G7

Solve: In Excel, the descriptions of the normal error curve are particularly obscure. For a clearer explanation of *Normsdist* and the *Normsdist* Excel functions see http://www.exceluser.com/explore/statsnormal.htm

Several points need to be made: Rows 2 through 4 actually answer the problem. Rows 5 through 7 are included to show that the normal error curve is distributed symmetrically about the origin. It's also slightly easier just to double the tail area of the normal error curve for a given negative Z to get the total (two-tailed) number of excluded metal rods.

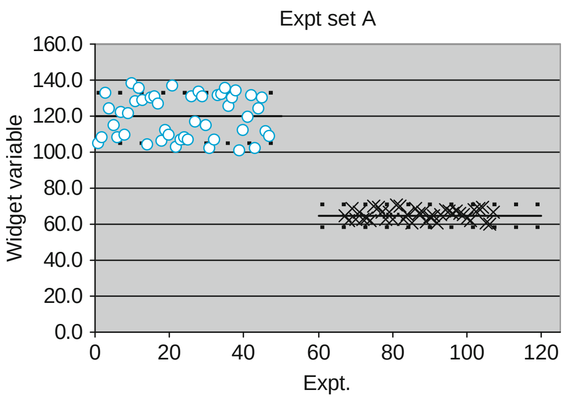
Of special interest in this problem is the cost of *acceptable* rods. The blanks cost \$1.00 each. If the machining is sloppy and you have to discard those with $\mathbf{Z \geq 0.333}$ **or** $\mathbf{Z \leq -0.333}$, the net cost of rods in specifications is **\$3.89 each**, since only 26.1 percent pass specifications. For $\mathbf{Z = \pm 0.667}$, only 49.5 percent pass and the cost is **\$2.02**. However, if you can machine them to $\mathbf{Z = \pm 2.0}$, with the fraction passing being 95.4 percent, the cost falls to **\$1.05** per rod.

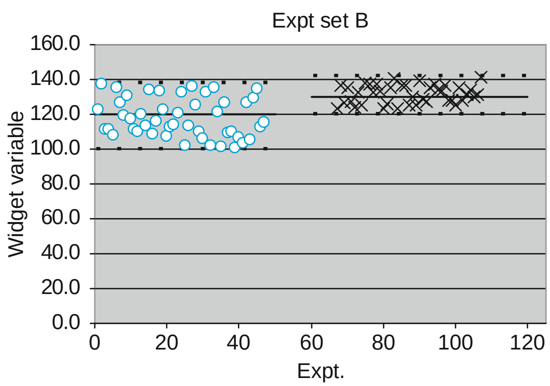
The message of Example 15.7 is very clear: quality is not only good for the consumer, but it is economical for the manufacture as well. In this respect it is twice blessed.

Example 15.8

Your boss has a self-proclaimed improvement in your company's most profitable line and, to prove it, you are given the opportunity to test the product before and after your boss's improvements. You test more than 50 of both the imputedly improved widgets and the existing widgets. The data are given graphically in the figure marked Expt set A. The open circles are the imputedly improved widgets and the crosses are the original ones. Has your boss really improved the widget? (The heavy line is the mean of the data and the dashed lines are $\pm 2\sigma$ bounds, meaning that 95% of the data are within those bounds.)

In a second trial marked set B, different widgets are subjected to comparative measurements. Again, the data with the open circles are your boss's imputed improvements and the crosses are the original widget. Again, do the data show a real improvement?





Need: Are the data different in experiment A ____ yes/no?
Are the data different in experiment B ____ yes/no?

Know: By eye, the means are different in both sets of experiments (137 vs. 73 in set A and 120 vs. 138 in set B).

How: 95% of the data are bounded by the two sigma lines.

Solve: In set A, the dashed 2σ bounds are clearly distinct, so we know that 95% of the data are different and the difference between the means is real (i.e., statistically significant). In set A, the boss is correct—the widgets are better and the difference in the means (137 vs. 73) is real.

In set B, the means (120 and 138) are apparently different (although by a rather small amount) but their respective scatter overlap; the difference in the means is due to scatter and not statistically different. (So don't irritate your boss by noting the mean for the original widget is higher than the "improved" version!)

These simple graphical illustrations are the basis of some mathematical tests that answer whether noisy data are statistically the same or different. It is valuable in assessing changes in processes as well as in manufactured widgets. In most cases, the graphical method shown here is either sufficient in itself to distinguish the real from the noise or is at least a way to visualize the situation before applying book formulae that accomplish the same thing.

SUMMARY

All products have some impact on the environment. Since some products use more resources, cause more pollution, or generate more waste than others, the aim is to identify those that are most harmful.

Even for those products whose environmental burdens are relatively low, the LCA should help to identify those stages in production processes and in use that cause or have the potential to cause pollution, and those that have a heavy material or energy demand.

Breaking down the manufacturing process into such fine detail can also be an aid to identifying the use of scarce resources, showing where a more sustainable product could be substituted.

Statistical methods are important for control of processes. Most processes are not exact but have a scatter about their **mean** given by Gauss's **normal distribution**. The most important variable that describes the scatter in the consistency of manufacturing processes is the **standard deviation**. It characterizes the shape of the graph showing the distribution of manufactured items, and thus the number of rejects in a normal distribution of manufactured objects. For a normalized variable with $\mathbf{Z \geq 1.0\sigma}$ **or** $\mathbf{Z \leq -1.0\sigma}$, 68% of the data are within specifications. For $Z = \pm 2\sigma$, 95% are and for $Z = \pm 3\sigma$, 99% are satisfactory.

EXERCISES

1. Discuss the building materials used by the three little pigs (straw, sticks, and bricks). Why were they chosen? Why did they fail? What was the environmental impact?

2. A **casting process** involves pouring molten metal into a mold, letting the metal cool and solidify, and removing the part from the mold. The solidification time is a function of the casting volume and its surface area, known as **Chvorinov's rule**.

$$\text{Solidification time} = K \times (\text{Volume}/\text{Surface area})^2$$

where K is a constant that depends on the metal. Three parts are to be cast that have the same total volume of 0.015 m^3, but different shapes. The first is a sphere of radius R_{sph}, the second is a cube with a side length L_{cube}, and the third is a circular cylinder with its height equal to its diameter ($H_{cyl} = D_{cyl} = 2R_{cyl}$). All the castings are to be made from the same metal, so K has the same value for all three parts. Which piece will solidify the fastest?

The following table gives the equations for the volume and surface area of these parts.

Object	Surface Area	Volume
Sphere	$4\pi R_{sph}^2$	$(4/3)\pi R_{sph}^3$
Cube	$6L_{cube}^2$	L_{cube}^3
Cylinder	$2\pi R_{cyl}^2 + 2\pi R_{cyl}H_{cyl} = 6\pi R_{cyl}^2$	$\pi R_{cyl}^2 H_{cyl} = 2\pi R_{cyl}^3$

(**A:** The cube will solidify the fastest and the sphere will solidify the slowest.)

3. For a particular casting process the constant K in Chvorinov's equation in Exercise 2 is 3.00 s/mm^2 for a cylindrical casting 150. mm high and 100. mm in diameter. Determine the solidification time for this casting.

4. **Tool wear** is a major consideration in machining operations. The following relation was developed by F. W. Taylor for the machining of steels:

$$VT^{n} = C$$

where V is the cutting speed; T is the tool life; the exponent n depends on the tool, the material being cut, and the cutting conditions; and C is a constant. If n = 0.5 and C = 400 mm/min$^{0.5}$ in the equation, determine the percentage increase in tool life when the cutting speed is reduced by 50%. (**A:** The tool life increases by 300%.)

5. For a particular machining operation n = 0.6 and C = 350 mm/min$^{0.4}$ in the Taylor equation given in Exercise 4. What is the percentage increase in tool life when the cutting speed V is reduced by (a) 25% and (b) 74%?

6. Using Taylor's equation in Exercise 4, show that tool wear and cutting speed are related by $T_2 = T_1(V_1/V_2)^{1/n}$.

7. A shaft 150. mm long with an initial diameter of 15. mm is to be turned to a diameter of 13.0 mm in a lathe. The shaft rotated at 750. RPM in the lathe and the cutting tool feed rate is 100. mm/minute. Determine:
a. The cutting speed
b. The material removal rate
c. The time required to machine the shaft

8. Repeat the calculations in Example 15.2 for a magnesium alloy being machined at 700. RPM.

9. Determine the machining power and cutting torque required to machine a magnesium shaft on a lathe if the material removal rate (MRR) is 0.1 in^3/min and N = 900. RPM.

10. Determine the machining time required to turn a 0.20 m long shaft rotating at 300. RPM at a tool feed of 0.20 mm/rev.

11. A lathe is powered by a 5.0 HP electric motor and is running at 500. RPM. It is turning a 1.0 inch cast iron shaft with a depth of cut of 0.035 in. What is the maximum feed rate that can be used before the lathe stalls?

12. A 2.0 inch diameter carbon steel shaft is to be turned on a lathe at 500. RPM with a 0.20 inch depth of cut and a feed of 0.030 in/rev. What is the minimum horsepower and torque that the lathe must have to complete this operation?

13. Suppose the material used in Example 15.3 was an aluminum alloy and the drill rotational speed was 500. RPM. Recalculate the material removal rate, the power required, and the torque on the drill.

14. A drill press with a 0.375 inch diameter drill bit is running at 300. RPM with a feed of 0.010 in/rev. What is the material removal rate?

15. You need to drill a 20. mm hole in a piece of stainless steel. The drill feed is 0.10 mm/revolution and its rotational speed is 400. RPM. Determine:
a. The material removal rate
b. The power required
c. The torque on the drill

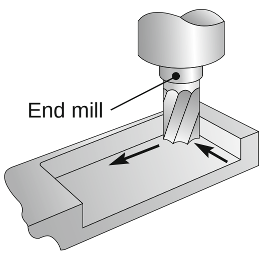
16. Repeat the calculations in Example 15.4 for stainless steel instead of cast iron. Use the same material dimensions and operating conditions, and compare the machining times for the two materials.

17. A flat piece of cast iron 100. mm wide and 150. mm long is to be milled with a feed rate of 200. mm/minute with a depth of cut of 1.0 mm. The milling cutter rotates at 100. RPM, and is wider than the workpiece. Determine:
a. The material removal rate
b. The machining power required
c. The machining torque required
d. The time required to machine the workpiece

18. A milling operation is carried out on a 10. inch long, 3.0 inch wide slab of aluminum alloy. The cutter feed is 0.01 in/tooth, and the depth of cut is 0.125 inches. The cutter is wider than the slab and the diameter is 2.0 inches. It has 25 teeth and rotates at 150 RPM. Calculate:
a. The material removal rate
b. The power required
c. The torque at the cutter

19. A part 275. mm long and 75. mm wide is to be milled with a 10-toothed cutter 75. mm in diameter using a feed of 0.10 mm/tooth at a cutting speed of $40. \times 10^3$ mm/min. The depth of cut is 5.0 mm. Determine the material removal rate and the time required to machine the part.

20. In the face-milling operation shown below the cutter is 1.5 inches in diameter and the cast iron workpiece is 7.0 inches long and 3.75 inches wide. The cutter has 8 teeth and rotates at 350 RPM. The feed is 0.005 in/tooth and the depth of cut is 0.125 inches. Assuming that only 70% of the cutter diameter is engaged in the cutting, determine the material removal rate and the machining power required. **Hint:** The width of the cut in face-milling is the percentage of the diameter of the cutting tool engaged in the cutting.



21. In Example 15.5, we did not take into account recycling or disposal costs or benefits. Do you think his conclusions would change if these were included?

22. The equation for a normal error curve is

$$\frac{\exp\left(\left(-\frac{1}{2}\right)\left(\frac{(x-\mu)}{\sigma}\right)^2\right)}{\sigma\sqrt{2\pi}}$$

or in spreadsheet script:

$$\exp(-0.5 * ((x - \mu)/\sigma) \wedge 2)/(\sigma * \text{sqrt}(2 * \text{pi}()))$$

Plot a normal distribution curve for the set of 100 numbers from 0 to 99.

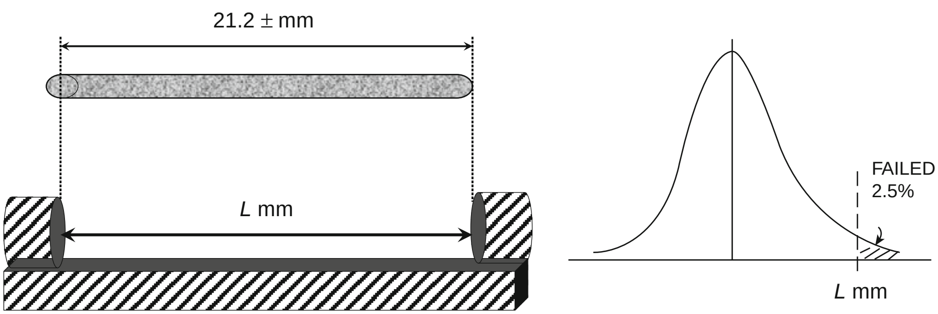
(**Hint:** Use the Chart–Column graphical representation. Observe its shape and calculate its mean and standard distribution. See Exercise 23.

23. Students often are confused by the difference between a normal distribution curve, as in Exercise 22, and a *random* distribution. Excel has a variable rand() that will generate random numbers between 0 and 1. (See the Excel Help menu to see how to use it.) Plot a random curve for the set of 100 numbers from 0 to 99.

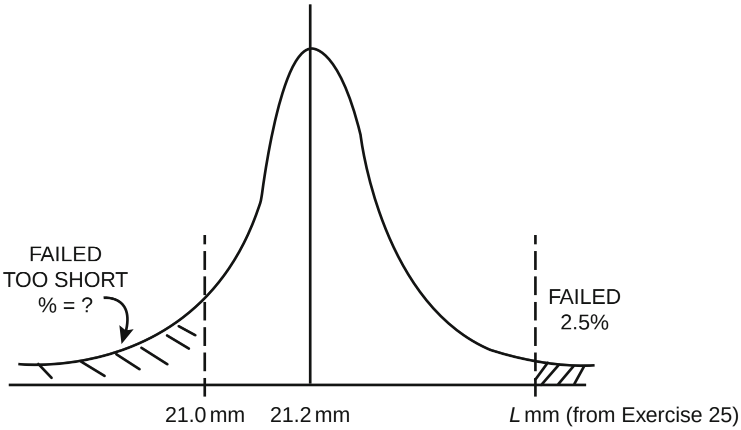
(**Hint**: Use the Chart–Column representation. Observe its shape and calculate its mean and standard distribution.)

24. Three hundred widgets are manufactured with the following normal error distribution statistics: mean = 123 units, standard deviation is (a) 23 units, (b) 32.5 units, and (c) 41 units. How many will measure less than 140. units?

25. We wish to fit steel rods (μ = 21.2 mm, σ = 0.20 mm) into the following jig in order to make sure we can meet specifications. The process requires that 97.5% of all rods are to be within specifications. (Presume that no rod is so short that it cannot pass specifications.) What is the appropriate size *L* mm of the jaws of the jig? (**Hint**: 95% of (two-sided) normally distributed data are found within $\pm$ 2σ from the mean and therefore 2.5% of the rods are too large.)



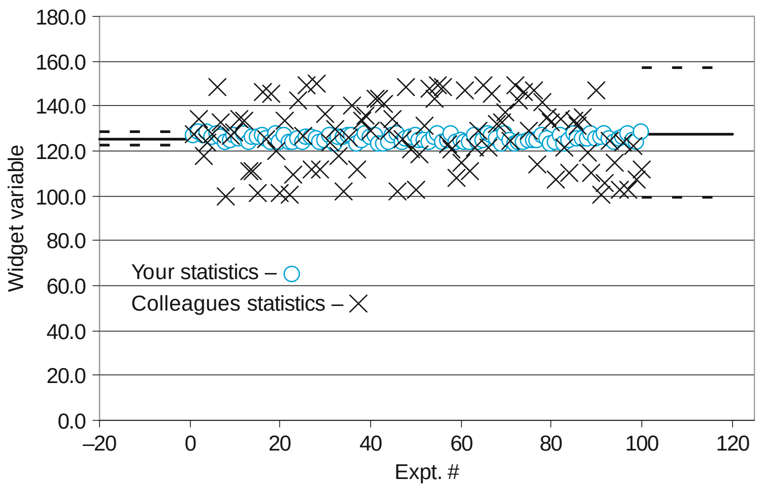
26. The jig in Exercise 25 is now OK, except we are getting too many undersized rods passing specifications. The manufacturing section of your company wants to eliminate all rods 0.2 mm or more that are undersized so that *L* = 21.0 mm is the lower cutoff. What fraction will now pass specifications?



27. You perform 100 tests on one of two prototypes of the X15-24 line of products and a colleague performs another 100 tests on the other prototype.

Test Set Statistic	Your Test	Colleague's Test
Mean, μ	125.7	123.5
Std Dev, σ	1.50	13.3

Graphing of the tests is similar with your data as circles and your colleagues as crosses:



Your colleague notes that because the means of the measurements are virtually identical, there is no difference between the two prototypes and that therefore hers should be used because it will be cheaper to manufacture (which you concede). What counterarguments can you muster?

28. Quality control is more personal when the products are hand grenades and it's your turn to learn how to throw one. You understand that once you pull the pin, if you hold on too long you might be severely injured or killed. However, if the fuse burns for too long after you have thrown it, the enemy might have time to pick it up and throw it back at you with results similar to holding it too long. Suppose the fuse burn time is designed to be 4.00 s with a standard deviation of 0.2 s, and the grenade has been made to Six Sigma standards. What is the variability you can expect on the time for it to detonate after pulling its safety pin?

29. You are a new quality control engineer at a company that manufactures bottled drinking water. All the bottles and the filling water are checked hourly to make sure that there are no contaminants. Monday morning you notice that for some reason the water that was used to wash the bottles before filling was not tested over the weekend. Now you have several carloads of product ready to be shipped. What do you do?

a. Have the shipment destroyed and start filling them over again.
b. Test random samples for the shipment and if they pass send the shipment.
c. Tell your supervisor and let him or her decide what to do.
d. Since the wash water has never been contaminated in the past, do nothing and release the shipment.

Use the Engineering Ethics Matrix.

30. As a young engineer, you have been told not to trust the machinists who work for you. They will make up excuses for bad parts that aren't true because they are lazy or incompetent. One day a machinist tells you his milling machine is "out of calibration," and that he can't make parts to specification unless the machine is repaired. What do you do?

a. Replace him with a more competent machinist.
b. Ask the machine shop supervisor to check the milling machine to see if there is a problem.
c. Ask the machinist to show you exactly what he means by demonstrating the problem.
d. Tell your supervisor you need a raise if you have to work with idiots like this.

Use the Engineering Ethics Matrix.

This page intentionally left blank

CHAPTER

16 Engineering Economics

Source: © iStockphoto.com/Vasiliy Yakobchuk

This chapter is applicable to *all* the fields of engineering since engineers design and make things that people buy. However, it is especially significant to Industrial Engineering, Systems Engineering, and Management Engineering, since these disciplines often are involved in the cost management of engineering systems.

16.1 WHY IS ECONOMICS IMPORTANT?

We all make economic decisions every day. Can I afford that steak dinner, a new laptop computer, an iPhone, or should I get a hamburger, continue to use my old desktop computer, and keep the old cell phone that I have used for years? Can I afford a new car? What about a mortgage for a new house? Engineers make similar decisions all the time to determine if what they are doing makes economic sense. Unlike personal economic decisions, the criteria used by engineers are multidimensional and have several components.

There are several financial terms with which an engineer should be familiar. These include models of the **cost of money** (i.e., **interest** charged), **principal**, **present value**, **first cost**, and **future worth** to name just a few. These terms are needed to assess the true cost of an investment, be that in buying a whole new factory or just buying a motor for an existing project such as a vacuum cleaner. Other criteria used to evaluate the economics of a new investment include **break even point (BEP)**, **process improvement (PI)**, and **return on investment (ROI)**. These terms are but a few of many standard methods used before an engineer expends large sums of money.

The answer to almost all economic decisions made for personal use rests on first cost. **First cost** is what you pay for the item when you buy it. Few of us think in terms of the purchase's life cycle costs such as repairs, depreciation, and the cost of borrowing money when making a purchase.[1] But if you are making a major business purchase, you better consider more than just the first cost. In the next section, we will concentrate on how the **cost of borrowing** may be as important as the cost of the item itself.

16.2 THE COST OF MONEY

Suppose you purchase a car for $10,000. Is that what it really costs you? Even the most elementary analysis suggests that this is not so. Because if you have $10,000 to spend, you already have this cash *somewhere*. Suppose it was not kept under your mattress, but in a bank in an interest bearing account earning 5 percent **simple** interest.

[1]Except possibly for a mortgage, because there the monthly amounts are usually large, relative to your income.

This means that each year you will have a deposit to your account of 5 percent of \$10,000, or \$500. Purchasing a car with your cash means that this income stops forever. If nothing else changes, your once \$10,000 now forgoes the addition of \$500 each and every year for a future worth of infinity[2] (should you live so long)!

So you see that money itself has a value, and a proper model of how to account for this value is important in an engineer's decision to buy a new car or anything else. It is very unusual to make cash purchases in a business (except for small petty cash purchases). If you want to make a significant purchase, a car for example, a business would probably borrow the money from a bank. Suppose the bank charges 5% simple interest on the loan. Then, once a year, you owe the bank \$500. With simple interest it's as if once a year you paid back the \$10,000 and gave the bank \$500 for the use of the \$10,000, and then you immediately borrowed another \$10,000 at the same interest for the next year. If you pay back the full amount of interest plus principal after five years, you will pay the creditor \$10,000 (the principal, P) plus \$500/year × 5 years = \$2,500 (the interest, I). Here you can see a finite cost of money—\$2,500 on a principal of \$10,000.

But, if you were a potential creditor of your business, and knew that you would have to wait five years before seeing a penny back, wouldn't you think that you have lent the business another \$2,500 spread over five years and on which you received nothing for that in return? The banker wants interest on this interest as well as on the original principal and so charges **compound** interest for the loan.

It works this way. Year 1 of the loan is \$10,000, year 2 is \$10,500, year 3 is \$11,025 (including 5% of \$10,500), and so on until in year 5 you repay the creditor/banker the sum of

$$P = \$10,000 + I$$

in which

$$I = \$500 + \$525 + \$551 + \$579 + \$608 = \$2,763$$

instead of just $I = \$2,500$. Note each yearly interest is $(1 + 0.05) \times$ the previous year's balance. Table 16.1 summarizes this situation for an interest rate of r.

Table 16.1 Generalizations of Interest Accrual Equations

Type of Interest	Period	Beginning of Period	End of Period	Future Worth, F
Simple[3]	1	P	$P + rP$	$P(1 + r)$
	2	P	$P + rP$	$P(1 + r) + rP = P(1 + 2r)$
	3	P	$P + rP$	$P(1 + 2r) + rP = P(1 + 3r)$
	...	...	...	...
	$\mathbf{N}$	$\mathbf{P}$	$\mathbf{P + rP}$	$\mathbf{P(1 + Nr)}$
Compound	1	P	$P(1 + r)$	$P(1 + r)$
	2	$P(1 + r)$	$P(1 + r) + rP(1 + r)$	$P(1 + r)^2$
	3	$P(1 + r)^2$	$P(1 + r)^2 + rP(1 + r)^2$	$P(1 + r)^3$
	...	...	...	...
	$\mathbf{N}$	$\mathbf{P(1 + r)^{N-1}}$	$\mathbf{P(1 + r)^{N-1} + rP(1 + r)^{N-1}}$	$\mathbf{P(1 + r)^N}$

[2]Scrooge may have believed in this because he did not spend what he earned.

[3]Although we also have included a simple interest table, it is for pedagogical reasons only because it is easy to understand and not because it is used in practical engineering economics.

Table 16.1 introduces new terms—the **future worth**, F, of the **present worth**, P. In other words, the future worth is what today's principal is worth based on some estimate of its future financial behavior. Virtually all business transactions uses the compound interest formula and other formulae derived from it.

Note: **In this chapter, in problems involving finances we will often ignore our rules concerning significant figures, and conform to the language of the financial industry.** We will normally express our calculations in whole dollars.[4]

Example 16.1

You wish to borrow $100,000 for 10 years at 5.0 percent annual interest. What is the difference in the cost of the loan if it is compounded yearly, monthly, or daily?

Need: Cost of borrowing $100,000 for 10 years at 5.0 percent under assumptions of 10 annual payment periods, 120 monthly periods, and 3,650 daily periods.

Know–How: The formulae for compound interest from Table 16.1 is $F = P(1 + r)^N$.

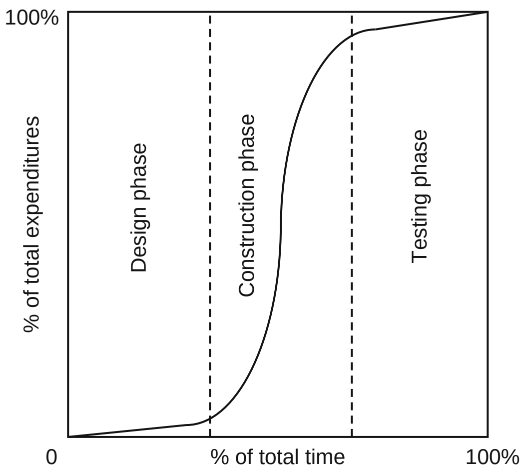
Solve:

1) $N = 10$, $r = 0.050$, therefore $F = P(1 + r)^N = \$100{,}000 \times (1 + 0.050)^{10} =$ **$162,889.**
2) $N = 120$, $r = 0.050/12 = 0.00417$, therefore
   $F = P(1 + r)^N = \$100{,}000 \times (1 + 0.00417)^{120} =$ **$164,701.**
3) $N = 3650$, $r = 0.050/365 = 1.37 \times 10^{-4}$, therefore
   $F = P(1 + r)^N = \$100{,}000 \times (1 + 1.37 \times 10^{-4})^{3650} =$ **$164,866.**

Notice that the price of the loan in Example 16.1 goes up when the interest is compounded more often. Note too that interest rates are nominally quoted in annual terms but they must be adjusted to reflect the daily or monthly periods.

Example 16.2

Nuclear power plants cost billions of dollars. This reflects the need for the highest possible confidence in the construction of the reactor. A spending plan for a large generic nuclear power plant is shown in the following figure.



A simplified S spending curve of expenditures vs. time for the construction of a nuclear power plant.

[4]However, some transactions, such as small loans, are compounded to the penny.

In a simplified case, each phase takes about one third of the total time to build the power plant, with the heaviest spending naturally occurring during the construction phase. A manufacturer of nuclear plants typically would get construction loans as needed to minimize the interest expense. But significant loan costs are incurred during the final test phase since the money already has been spent, and therefore borrowed, for prior activities that have yet to bring in any income.

Suppose the costs were $5 billion up to the beginning of the test phase for a 1,000 MW nuclear power plant. Suppose further the whole construction period is 12 years and thus the final test phase takes four years. If the interest rate is 12.0 percent with quarterly compounding, how much does the testing phase cost in finance charges alone?

Need: Testing phase finances charges = $______.

Know: Outstanding capital, P = $5 billion. Period of testing = 4 years. Costs are compounded over $N = 4$ quarters/year × 4 years = 16 quarters, and the interest rate $r = 12.0\%/4 = 3.0\% = 0.030$ per quarter.

How: Use the compound interest law from Table 16.1: $F = P(1 + r)^N$.

Solve: $F = P(1 + r)^N = \$5\times 10^9 \times (1 + 0.030)^{16} = \$8.02$ billion. The finance charges during the testing phase are then $= F - P = 8.02 - 5 =$ **$3.02 billion.**

Obviously, the financial expenditures during the testing phase swamps the actual costs of testing itself by a considerable amount. For a project with a long construction time upfront, the viability of the project may depend on financing charges more than any other single factor.

Rather than do the arithmetic on a calculator, any spreadsheet program you use almost certainly has many spreadsheets developed for financial analysis. Excel™ has the functions described in Table 16.2 (plus *many* other financial functions).

Table 16.2 Some Excel™ Financial Functions

Function	Use	Arguments
FV(*rate,nper,pmt,pv,type*) PV(*rate,nper,pmt,fv,type*) NPER(*rate, pmt, pv, fv, type*)[5]	Calculates future worth Calculates present worth Calculates how long it will take to increase *PV* to *FV*	■ *rate* = *r*. ■ *nper* = *N*. ■ *pmt* = periodic payments of *P* and *r*. ■ *pv* = *P*. ■ *fv* is optional; it is the future value that you'd like the investment to be after all payments have been made. If this parameter is omitted, the NPER function will assume that *fv* is 0. ■ *type* = 0 if the payment is at the start of the period. ■ *type* = 1 if the payment is at the end of the period.

Example 16.3

Repeat Example 16.1 using the Excel spreadsheet future worth function, FV.

Need: Cost of borrowing $100,000 for 10 years at $r = 5\%$ under assumptions of 1) $N = 10$ payment periods, 2) $N = 120$ payment periods, and 3) $N = 3{,}650$ payment periods.

[5]For example, suppose you deposit $10,000 in a savings account that earns an interest rate of 8%. To calculate how many years it will take to double your investment, use NPER as follows: = NPER(0.08, 0, –10000, 20000, 0). This will return an answer of 9.01, which indicates that you can double your money in about nine years.

Know–How: Use the *FV* function in Excel.
Solve: FV(*rate,nper,pmt,pv,type*)

	A	B	C	D
1	*P*	*r*, %		
2	$100,000	5.00%		
3				
4	**Case #**	***N***	***r* per period**	***FV***
5	1	10	5.00%	$162,889
6	2	120	0.42%	$164,701
7	3	3650	0.01%	$164,866

	A	B	C	D
1	*P*	*r*, %		
2	100000	0.05		
3				
4	**Case #**	***N***	***r* per period**	***FV***
5	1	10	=$B$2	=FV(C5,B5, ,-$A$2,1)
6	2	120	=B2/12	=FV(C6,B6, ,-$A$2,1)
7	3	3650	=B2/365	=FV(C7,B7, ,-$A$2,1)

Notice the convention that your principal (or, investment) is a negative number (–$A$2), which is logical but confusing. If you enter it as a negative number in $A$2, then you don't have to put a negative sign in front of $A$2 in the FV function. Also, notice that unless you are prepaying the owed principal or interest, you do not need to enter a number for *pmt* in the FV function. Just leave a blank space (or enter 0) for *pmt* so that, *pmt*, becomes , , as shown above.

Example 16.4

You want to reap $1,000,000 from an investment of $500,000. How many years will you have to wait if $r = 7.50\%$?

Solve: Use the Excel spreadsheet function NPER(*rate, pmt, pv, fv, type*).

	A	B	C	D
12	*P*	*r*,%	*FV*	$1,000,000
13	$500,000	7.50%		
14				
15		*N*, years	*r* per period	
16		**9.58**	7.50%	

	A	B	C	D
12	*P*	*r*, %	*FV*	1000000
13	500000	0.075		
14				
15		*N*, years	*r* per period	
16		**=NPER(B13, ,-A13, D12, 1)**	=$B$13	

Thus you need to wait just over 9.5 years to double your money.

To summarize what it really costs to make an engineering purchase, you need to include the present worth (the principal) P and the interest charges on that amount. Clearly you also have to add other costs such as labor costs, material costs, and overhead costs (such as the rent for your factory, the cost of power to keep the lights on, the cost of heat to keep the factory warm, and so on, to get the total costs).

16.3 WHEN IS AN INVESTMENT WORTH IT?

But how do you know if your project is worth the effort? In fact, most engineering businesses use one or more criteria to assess the value of a project. A major indicator occurs when you start to make a profit by selling enough of your widgets. This is called the **Break Even Point** (**BEP**), and it is a standard measure of a project's value. It has a simple definition:

BEP occurs when the project has earned back the cost it took to make it.

Example 16.5

Assume the cost of producing a new product is $1,000,000. Then the BEP occurs when net profit from the product reaches $1,000,000. Let's say the profit per widget is $1.00 and we're selling 1,000/day. What is the BEP in years?

Need: BEP = ______ years for a project costing $1,000,000 assuming you are selling 1,000/day with a profit of $1.00 per widget.

Know–How: Equate cost to total money stream.

Solve: 1,000 [widgets/day] × 1.00 [$/widget] × D [days] = $1,000,000. Solving for D gives:

$$D = 1{,}000 \text{ days} = \mathbf{2.74\ years}.$$

In Example 16.5 it will take 2.74 years to reach the BEP. Is this good enough? It depends on the industry. Many companies would prefer a BEP of 18 months or less.

Example 16.6

One "long horizon" industry is the electric power industry. In Example 16.2, our 1,000 MW nuclear power plant costs $8.02 billion, a very large sum of money. Yet its product sells for about 10 cents per kWh. What is its best possible[6] BEP after it starts to produce electricity?

[6]Ignore other significant costs such as labor and uranium fuel, refueling outage, repairs, etc.

Need: BEP = ______ years for a 1,000 MW nuclear power plant that costs $8.02 billion and selling electrical energy for 10.0 cents/kWh = 0.10 $/kWh.

Know: For ***t*** hours of operation a 1,000 MW (= $\$10^6$ kW) nuclear power plant can produce a maximum of 1,000,000 × ***t*** kWh of electricity.

How: Equate costs to total money stream produced by reactor.

Solve: 1,000,000 × ***t*** × 0.1 [kWh][$/kWh] = $\$8.02 \times 10^9$ [$], or
t = 8.02×10^4 [$][h/$] or **BEP** = **9.2 years** (assuming 8,760 hrs/yr)

Is the BEP in Example 16.6 good enough? Although nuclear power plants initially cost more than fossil fuel power plants, the advantages of uranium over fossil fuel will likely increase as world oil and gas supplies dwindle and concerns about greenhouse gases grow.

Example 16.7

An engineer proposes an improvement to an existing process. The cost required to make this process improvement (**PI**) is $100,000. Suppose the process makes 100,000 widgets/day. If the proposed process improvement saves one cent per unit ($0.01/unit), what is its BEP for the PI?

Need: BEP = ______ for a process improvement costing $100,000 assuming you are selling 100,000 widgets/day with a process improvement savings of $0.01/widget.

Know–How: You will save 100,000 [widgets/day] × 0.01 [$/widget] = $1,000/day with the new and improved widget manufacture.

Solve: The **PI BEP** is given by how long to recover your $100,000 investment.
This is 100,000/1,000/[$][days]/[$] = 100 days = **3.3 months**.

Is this PI BEP in Example 16.7 good enough? As a rule of thumb, a nominal BEP should be 18 months or less. And you would prefer a BEP to take less than 12 months. In Example 16.5, the new widget is a marginal investment. But in Example 16.7, a process improvement BEP is a definite go.

One other financial term that often is used to determine if an engineering investment is satisfactory is the annual return on investment or **ROI**. This is defined as

$$\boldsymbol{ROI\ (in\%) = \frac{Annual\ Return}{Cost\ of\ the\ Investment} \times 100}$$

In the simplest case, if an investment of $500,000 produces an income of $40,000 per year, its ROI = $40,000/$500,000 = 0.08 = 8%. Many companies would not invest for such poor returns on their money. Many successful large companies operate with ROIs of 15 percent or more.

As Table 16.3 shows, there is a big difference among the ROIs of companies. You should not be surprised to see that Exxon Mobil has the highest ROI in this table since its profits, and therefore the numerator of the ROI formula, is high while the processing cost for the equipment for converting crude oil to refined products is small relative to the cost of the processed material.

Table 16.3 ROI of Several Large Chemical Companies[7]

Company	ROI, Annual %
Dow Chemical	10.5
Exxon Mobil	22.4
DuPont	18.5
PPG Industries	20.2
Air Products	11.0
Eastman Chemical	10.9
W.R. Grace	9.8

Example 16.8

Calculate the annual ROI for the process improvement in Example 16.5. Assume factory is operating 300 days per year.

Need: ROI annual percentage for a project costing \$1,000,000 assuming you are selling 1,000 widgets per day at a profit of \$1.00 per widget for 300 days per year.

Know: Daily profit is 1,000 × 1.00 [widgets/day][\$/widget] = \$1,000/day.

How: Compare annual profits to the investment cost.

Solve: Profit (i.e., return) = 1,000 × 300 [\$/day][days/year] = \$300,000 \$/year.

Since the investment is \$1,000,000 then the **ROI** = \$300,000/\$1,000,000 = 0.30 = **30%**.

Is the ROI in Example 16.8 good enough? Almost surely yes! This is a wonderful place to invest \$1,000,000 (assuming all the assumptions and numbers are correct—such as the profit/widget, the number of days of operation/year, and the sales of widgets/day).

SUMMARY

This chapter was concerned with engineering economics, a discipline that is important to every branch of engineering, and particularly to industrial or management engineering. Terms broached in this chapter include **compound interest**, **present worth**, **principal**, and **future worth**. All are concerned with the value of money in making economic decisions. In some cases, this may be the dominant cost in a product rather than the investment in hardware, materials, and labor. Further guidelines are given in how to assess whether a proposed project will be commercially viable. These are the **break even point**, a measure of how long it takes to recover one's costs and **return on investment**, a ratio in percent of the annual profit divided by the investment that produces it. Successful engineering businesses typically want to see the **BEP** less than 18 months and the **ROI** greater than 15 percent.

[7]*Source:* Chemical Engineering Progress, July 2, 2007.

EXERCISES

1. Find the cost of borrowing \$100,000 in Example 16.1 if the interest is compounded hourly.
2. Rework Example 16.1 for a \$15,000 loan at a 10% annual interest.
3. You need to borrow \$12,000 to pay your tuition plus room and board. One bank offers to loan you the money for 10 years at 5.0 percent interest compounded annually. Another bank offers you the loan at 4.75 percent interest compounded monthly. Which bank has the better deal?
4. You are a big financial success and you want to purchase the Remlab Company for \$35 billion. You have \$5 billion in cash, but need to borrow the remaining \$30 billion from your friendly banker. You banker says fine, I'll lend you the money for 10 years, but at 7.5 percent annual interest compounded quarterly. How much interest will you pay to the bank over the life of this loan? (**A: \$33.1 B)**
5. What will be the testing phase finance charges in Example 16.2 if the nuclear power plant cost \$10 billion and everything else stays the same?
6. It turns out that the testing phase in Example 16.2 takes six years instead of four years, and the entire project takes 14 years. What will the finance charges be if everything else remains the same? (**A: \$5.2 B**)
7. Use the Excel spreadsheet FV function to determine the cost of borrowing \$100,000 for 10 years at 5 percent for 87,600 hourly payment periods.
8. Use the Excel spreadsheet FV function to determine the cost of borrowing \$100 billion for 10 years at 8 percent compounded quarterly.
9. Repeat Example 16.3 using the Excel spreadsheet FV function to determine the cost of borrowing \$3,000 on a credit card for 1 years at 28 percent compounded monthly.
10. Use the NPER Excel function to determine how long it will take to double your \$500,000 investment if you earn 7.5 percent interest? (**A: 9.58 years)**
11. Use the NPER Excel function to determine how long it will take to triple your \$500,000 investment if you earn 11.3 percent interest?
12. Use the NPER Excel function to determine how long it will take to increase an investment of \$3,500 to \$4,200 if you earn 3.5 percent interest?
13. How long will it take to double your money at an interest rate that is $<<$ 100% starting with the expression $F = P(1 + r)^N$? **Hint:**[8] $\ln(1 + r) \approx r$ if $r << 1$ and $\log(x)^N = N \log(x)$ (**A**: Years to double the investment is $70 \div r\%$, which is known as the **Rule of Seventy**[9])
14. Compare the answer to Exercise 10 with the Rule of Seventy described in Exercise 13.
15. Suppose the profit on each widget sold in Example 16.5 was \$10.00 instead of \$1.00. What would the BEP be if everything else remained the same?

[8] ln means the base of natural logarithms, $e = 2.718\ldots$.
[9] Sometimes called the Rule of Seven.

16. Suppose the new product in Example 16.5 has a profit of $5.00 per widget, but the sales projection is of off by 50 percent and the company is able to sell only 500 per day. What is the new break even point?

17. You are a manufacturing engineer and have been asked to project the costs required to produce a new line of products for your company. After contacting the appropriate equipment manufacturers and consulting with management about labor and overhead costs, you determine that the manufacturing line will cost $8.3 million to get into operation. The marketing department tells you that the company can expect to see a profit of $37.50 per item produced and that they can sell 1,500 items per day in big chain stores. You run into your boss at the coffee machine and she casually asks you what the break even point will be for your new line. What do you tell her?

18. The owners of the nuclear power plant in Example 16.6 want a break even point of five years. To do this, they decide to raise the price of the electricity they sell. For how much do they need to sell the electricity in order to meet the BEP? **(A: 18.3 cents/kWh)**

19. In Example 16.6 the owners of the nuclear power plant decide to use substandard materials to reduce the cost of building and testing the power plant from $8.02 billion to $7.35 billion. If everything else remains the same (and no one finds out what they've done), what will be their new BEP? Assume the selling price of electricity is 10. cents/kWh.

20. Another long horizon industry is the commercial aircraft industry. A Boeing 747-8 Freighter costs $300 million. A commercial air freight company purchases this plane and keeps it in the air 12 hours per day for 280 days per year. If the air freight company makes a profit of $4,500 per flight hour, how long before it breaks even with the cost of the plane?

21. In Example 16.7, it turns out that the new process does not save one cent per unit. It only saves 0.3 cents per unit. What is the new BEP, and was it worth the expense to make this improvement?

22. You are a manufacturing engineer and discover that you can purchase a machine that will replace a labor-intensive manual operation in your manufacturing line. The machine costs $1.5 million, but the labor savings costs are $7.80 per unit manufactured. Can you get a BEP less than 18 months if you make 8500 units per day?

23. An automobile manufacturer would like to add a new feature to their standard production model. The new feature will cost the manufacturer $8.67 per car to produce, and they expect to sell 100,000 cars per year. If the cars normally sell for $25,990 each, how much do they need to increase the price to reach a break even point in 18 months?

24. Calculate the annual return on investment (ROI) in for the process improvement in Example 16.5 if the factory operates 300 days per year, but sells only 500 widgets per day.

25. What is the ROI for the process improvement in Example 16.7 if the factory operates 300 days per year? Would you invest in this company?

Use the Engineering Ethics Matrix to analyze the following situations.

26. As a junior engineer in a small company you are involved in designing a production line for a new product. The company CEO believes that this new product will make a great deal of money and wants to get it into production as soon as possible. Your immediate supervisor wants to please the CEO and get things moving, so he starts ordering equipment and hires contractors to expand the building. While glancing at a copy of the new project your boss gave to the CEO, you discover an error. The break even point and the return on investment are both high by a factor of 10. When you ask your immediate supervisor about this, you are told that it is none of your concern. What do you do?

27. Now you own a large corporation and a recent economic downturn requires you to eliminate all dividends to your company stockholders. This is a highly confidential decision because it will negatively affect your stock prices. Several of your wealthy neighbors own a considerable amount of your company stock. At a recent picnic in your backyard one of them asks if she should sell some of your company's stock, considering how bad the stock market is doing in general. If you say "no," she could lose a considerable amount of money. What do you do?

This page intentionally left blank

PART

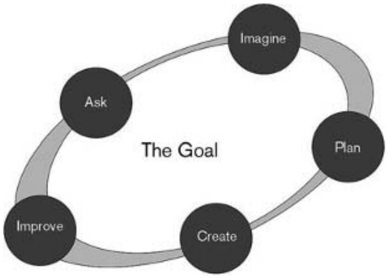
2

Hands-On

CHAPTER

Introduction to Engineering Design

17



Source: http://www.mos.org/eie/engineering_design.php

This chapter describes the **nature of engineering design**, explains a modern design philosophy called **Design for Six Sigma**, suggests some benefits of a **hands-on design project**, indicates the **qualities of a good designer**, and explains the need for a **systematic approach** that consists of an **eight-step design process**.

17.1 THE NATURE OF ENGINEERING DESIGN

In the course of creating new products, engineering design uses available technology to improve performance, lower cost, or reduce risk. For example, if the design of a bridge produces a new structure that is visually stunning with no consideration for its strength, this is design without the engineering. If, on the other hand, the designers of a new concept car use analysis or experiments to evaluate air drag, structural integrity, and manufacturability in addition to style when coming up with a new exterior design, this is engineering design.

Prior to the 1960s, the preceding description of engineering design might have been adequate, but this is no longer the case. The definition has been broadened to include the systematic thought processes and best practices that define the modern engineering design process. This systematic approach, which has become synonymous with engineering design, forms the heart of these design chapters.

Those of you experienced at tinkering with your own inventions in the basement or garage might say that you got along fine without engineering analysis and without training in the systematic approach. Such a view ignores the realities of engineering design. Usually, designers must search for the best possible design under severe conditions of limited time and limited resources (especially cost). For example, in most design situations there is barely enough time to produce a single prototype. The idea behind an engineering approach, as embodied in engineering analysis and the systematic approach, is to minimize the number of design iterations required to achieve a successful final design. Fewer design iterations means lower cost and shorter development times.

One of the greatest challenges of engineering design is the breadth of knowledge required of the designer. The diversity of topics covered in earlier chapters provides a hint of what a design might entail; some electromechanical designs could conceivably touch upon them all. In addition to those topics, there will be issues related to manufacturing, economics, aesthetics, ethics, teaming, government regulations, and documentation of the design, to name but a few. In the spirit of these design chapters, we propose that the best way to introduce the multifaceted nature of engineering design is to experience it for yourself through the following hands-on design exercise.

17.2 DESIGN PROBLEMS VERSUS HOMEWORK PROBLEMS

A design problem is unlike a traditional college homework problem. Homework problems like those in Part 1 of this textbook have specific, unique answers, and the student must find those answers using a logical approach such as the **need–know–how–solve** method. Design problems, on the other hand, do not have unique answers. There may be several answers (i.e., designs) that satisfy a design problem statement. There may also be a "best" answer based on critical requirements. For example, a design that minimizes manufacturing costs may not produce a very reliable product. If product reliability is a critical requirement of the customer, then minimizing manufacturing costs may not be the best design answer.

The challenge of Part 2 of this textbook is to transition you from solving well-formulated single-answer textbook problems to solving what we call open-ended engineering design problems. These problems are often complex and sometimes poorly formulated problems. You will find that excellent design work requires considerable creativity beyond that needed to solve textbook problems. In Part 2 of this text, we allow and encourage you to drop your academic inhibitions and explore the joy of engineering creativity.

The problem-solving process we have presented in Part 1 of this text (**need–know–how–solve**) works well for single answer textbook problems, but it needs to be expanded to encompass open-ended engineering design problems. In recent years the process of engineering design has been refined to produce more robust and economical designs. The process that most companies embrace today is a five-step process called Design for Six Sigma.

In Part 2 of this text we abandon the **need–know–how–solve** exercise solution method and adopt an eight-step design process based on the Six Sigma design methodology.

17.3 BENEFITS OF A HANDS-ON DESIGN PROJECT

A practicing engineer does not have to be an expert in machining or other basic manufacturing operations. Still, a basic understanding of the challenges involved in manufacturing a product is essential for producing a successful design. The best way to appreciate that fact at an early stage in your career is to manufacture a design yourself.

The lessons to be learned are universal. Don't expect your design to work on the first try. Leave a lot of time for testing. Complicated designs take a lot longer to build and have a lower probability of success. If you have a choice of manufacturing a part yourself or buying it, buy it. Many such lessons are foretold by the design principles and design for manufacture guidelines of later chapters. The consequences of violating those principles are best understood by experiencing the results of having done so.

There are other lessons to be learned from a hands-on design experience. For electromechanical systems with moving parts, that experience might be the only way to accurately evaluate the design. Also, students gain a sense of accountability by learning that it is not enough for a design to look good on paper. In order to actually be a good design, it has to lead to an end product that works.

In particular, we recommend that the hands-on design project should be done under a competition format, involving interactions between the machines. It is a natural motivator—the challenge of the task is heightened by having to deal with the unpredictability of your human opponent, and the other designs provide a relative scale against which to assess quality of performance.

17.4 QUALITIES OF A GOOD DESIGNER

These are the qualities of a good designer:

- **Curiosity about how things work.** Seeing other design solutions provides you with a toolbox of ideas from which you can draw when faced with a similar design challenge, so when you come across an unfamiliar device, try to figure out how it works. Take things apart; some companies actually do this and refer to it as "reverse engineering." Visit a toy store; the products there demonstrate creative ideas and new technology.

- **Unselfishness.** A key ingredient to effective teaming is suppressing ego and sacrificing personal comfort to serve the best interests of the team.
- **Fearlessness.** It takes a leap of faith to step into the unknown and create something new.
- **Persistence.** Setbacks are inevitable in the course of a design project. Remain resilient and determined in the face of adversity.
- **Adaptability.** Conditions during the design process are constantly evolving. For example, new facts may surface, or the rules of the design competition may change in some way. Be prepared to take action in response to those new conditions. In other words, if the ship is sinking, don't go down with the ship—redesign it.

17.5 HOW TO MANAGE A DESIGN PROJECT

Project management is a carefully planned and organized effort to accomplish a specific project. For example, design and construct a robot vehicle to be entered into a competition on a specific date.

Project management includes developing and implementing a plan that defines the project goals, and specifies how and by whom the goals will be achieved; identifying needed resources; and developing budgets and timelines. Project management is usually the responsibility of an individual team leader. When the project team members have been identified, the team (or the course instructor) will select one of its members as the project manager. This person has the responsibility for guiding the team design work in a professional, organized, and timely manner. The project manager is also responsible for meeting deadlines and that the team members are carrying fair work loads. Successful project management involves:

- **Understand the project's goals.** You should be able to state the goals of your project in a single sentence.
- **Engage all the team members.** Subdivide the work using functional decomposition to break the project down into individual work assignments, and make sure that everyone knows what he or she is responsible for to meet the project goals (see Ground Rule Number 2, later).
- **Keep the project moving.** Work methodically to meet your benchmarks; don't wait until the last minute and rush to meet a deadline.

17.6 TWO GROUND RULES FOR DESIGN

There are two important ground rules for design, the use of a **design notebook** and **effective teamwork**.

17.6.1 Ground Rule 1: Use a Design Notebook

When you are working on a design project and you want to write something down, the design notebook is the place to do it. There is no need for notepads, reams of paper, or sticky notes. The place to record your thoughts is in a permanently bound volume with numbered pages, a cardboard cover, and a label on the front cover identifying its contents. Every college bookstore has them, though they may be called laboratory notebooks.

As a starting engineer, now is the best time to start a career-long habit. Just how important the design notebook is can be explained in the case of Dr. Gordon Gould.[1]

> *On November 9, 1957, a Saturday night just given to Sunday, Gould was unable to sleep. He was 37 years old and a graduate student at Columbia University. For the rest of the . . . weekend, without sleep, Gould wrote down descriptions of his idea, sketched its components, projected its future uses.*
>
> *On Wednesday morning he hustled two blocks to the neighborhood candy store and had the proprietor, a notary, witness and date his notebook. The pages described a way of amplifying light and of using the resulting beam to cut and heat substances and measure distance. . . . Gould dubbed the process light amplification by stimulated emission of radiation, or laser.*

It took the next 30 years to win the patents for his ideas because other scientists had filed for a similar invention, although after Gould. Gould eventually won his patents and received many millions in royalties because he had made a witnessed, clear, and contemporaneous record of his invention.

The lesson is that patents and other matters frequently are settled in court for hundreds of millions of dollars by referring to a notebook that clearly details concepts and results of experiments. You must maintain that notebook in a fashion that will expedite your claim to future inventions and patents.

Another more immediate benefit of using a design notebook is that you will know that everything related to the project is in one place. Finding that key scrap of paper in a pile of books and papers on your desk after working for months on the project can be a rather time-consuming endeavor.

Here are some of the most important guidelines for keeping a design notebook:

- Date and number every page.
- *Never* tear out a page.
- Leave no blank pages between used pages. Draw a slash through any such blank pages.
- Include all your data, descriptions, sketches, calculations, notes, and so forth.
- Put an index on the first page.
- Write everything in real time—that is, do not copy over from scraps of paper in the interests of neatness.
- Write in ink.
- Do not use whiteout; cross out instead.
- Paste in computer output, charts, graphs, and photographs.
- Write as though you know someone else will read it.
- Document team meetings by recording the date, results of discussions, and assigned tasks.

17.6.2 Ground Rule 2: Team Effectively

Working in teams on a design project is both a joy and a challenge. Although there is a sense of security in knowing that others will be venturing into the unknown alongside you, the unpredictability of human interactions can be as perplexing as the design itself. To reduce the risk of ineffective teaming, rules of conduct will be presented in this section. These are well-accepted best practices based on observations of effective teams.

There are several advantages to attacking a design project in teams. First, design requires a wide range of skills and areas of knowledge. No one person is experienced enough to pursue every unfamiliar design challenge in isolation. Teaming provides an opportunity to expand the talents and life experiences that will be

[1] http://inventors.about.com/gi/dynamic/offsite.htm?site=http://www.inc.com/incmagazine/archives/03891051.html

brought to bear on the design problem. Second, if done right, teaming serves to keep personal biases in check. Third, more people should mean that more will get accomplished in a shorter period of time, although it is puzzling to often see team members standing by politely as one team member does all the work (especially during manufacturing). When best practices are followed, a team will be greater than the sum of its parts.

For design projects done during the freshman and sophomore years, three people are the ideal team size. Teams of two may not experience all the typical dynamics and so may not learn as much about teaming. With teams of four, it may be too easy for one team member to hide. Design teams at this level usually are not assigned a team leader by the instructor. Leadership typically emerges within the team. If a team leader is assigned, the role is *not* to be the boss, but rather to organize and facilitate participation by all team members.

Here are some teaming best practices:

- **Assign clear roles and work assignments.** A few things are best done as a team, such as brainstorming and evaluation of concepts. Most of the time it will pay off if everyone has his or her own assigned responsibilities and tasks to which he or she will be held accountable by the team. These tasks should be assigned or updated at the end of each and every team meeting.
- **Foster good communication between team members.** An atmosphere of trust and respect should be maintained in which team members feel free to express their ideas without retribution. That trust extends to allowing for civilized disagreement, delicately done so as not to suppress ideas or discourage participation. Everyone should participate in the discussions. Sometimes this means reaching out with sensitivity to the shy members of the team. If you succeed, you will have a team operating on all cylinders.
- **Share leadership responsibilities.** If there is a designated team leader, that person should empower the other team members with significant leadership responsibilities. This will give those students a strong sense of ownership in the project. At the same time, team members have to be willing to step forward to assume leadership roles.
- **Make team decisions by consensus.** Teams make decisions in one of three ways: (1) the team leader makes the decision, (2) discussions continue until everyone agrees (as in a trial by jury), or (3) after discussions are exhausted, the team takes a vote. Those who disagree with the outcome of the vote are then asked if they can put their opinions aside and move forward in the best interests of the team. In a college-level design project, the only ways to go are (2) and (3), which are both examples of decision making by consensus.

17.7 THE NEED FOR A SYSTEMATIC APPROACH

There are two main goals of the systematic approach to engineering design: (1) to eliminate personal bias from the process and (2) to maximize the amount of thinking and information gathering that is done up front, before committing to the final design. The result is fewer costly design changes late in the product development stages.

The engineering design process also provides a blueprint for design of complex systems. For example, you might be able to get along without a formalized design procedure when designing a new paper clip,[2] but when taking on the daunting task of designing a complex system like the space shuttle, brain gridlock can set in. The design process offers a step-by-step procedure for getting started, as well as strategies for breaking down complex problems into smaller manageable parts.

[2]But see just how difficult it originally was to perfect the paper clip: *The Evolution of Useful Things: How Everyday Artifacts—From Forks and Pins to Paper Clips and Zippers—Came to Be as They Are*, Henry Petroski, Alfred A. Knopf, Inc., 1992.

17.8 STEPS IN THE ENGINEERING DESIGN PROCESS

A systematic approach to engineering design that uses the elements of the Design for Six Sigma philosophy may be viewed as consisting of eight steps:

1. **Define the problem.**
2. **Generate alternative concepts.**
3. **Evaluate and select a concept.**
4. **Detail the design.**
5. **Design defense.**
6. **Manufacture and test.**
7. **Evaluate performance.**
8. **Prepare the final design report.**

These steps are shown in Figure 17.1.

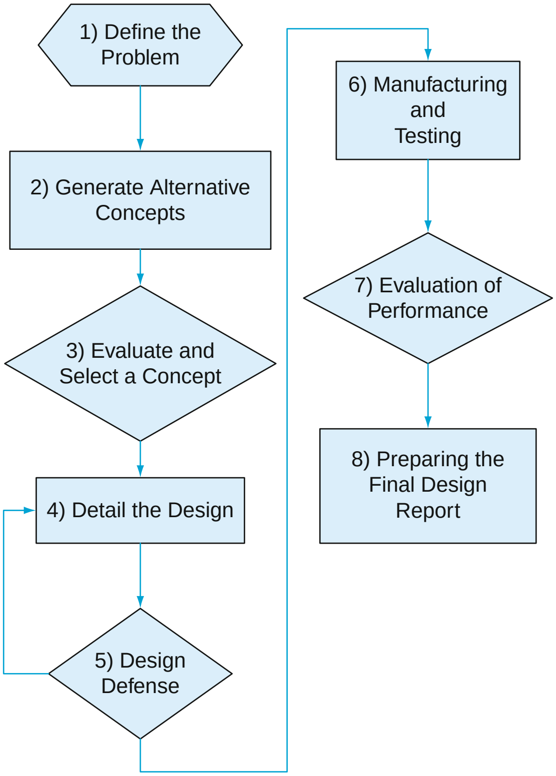


FIGURE 17.1 Design Process Flowchart

The next chapter presents two ground rules for engineering design. Subsequent chapters will treat each of the preceding eight steps in detail. At the end of each chapter there is a suggested milestone for successful completion of the step in the design process described in the chapter. Milestones are crucial for measuring progress toward the eventual goal of a successful design. After the eight steps have been described, this portion of the book concludes with a detailed example of an actual design competition.

17.9 HANDS-ON DESIGN EXERCISE: THE TOWER

Your first design objective is to build the **tallest tower** from the materials supplied. Since there are both individual and team winners here, this is both an individual and a team competition.

17.9.1 Setup

- Divide the class into about eight teams of three or more students per team.
- Each team should be provided with five sheets of 8.5×11 standard copier paper.
- The following material will be distributed **unevenly**:
  - Two teams each receive one roll of Scotch® tape.
  - Two teams each receive a roll of duct tape.
  - Two teams each receive one box of paper clips.
  - Two teams each receive one pair of scissors.

17.9.2 Rules

- Each team has just 10 minutes to build a tower.
- The final tower height measurement must be made by the instructor.
- Teams should indicate to the instructor when they are ready for a measurement.
- If the tower is composed of materials other than the supplied paper, Scotch® tape, duct tape, or paper clips, it will be disqualified.
- Tower must be stationary when the measurement is made.
- Tower must be built on a flat surface. Tower cannot lean against or be attached to any other surfaces (wall, table, etc.).
- Any team that intentionally knocks over another team's tower before it has been measured will be disqualified.
- After three minutes, **one** individual on each team will be offered 8 points to join another team.
- There are no other rules.

17.9.3 Scoring

- One point will be awarded for each inch of tower height. (Heights are to be rounded to the nearest inch.)
- The first team that finishes their tower and is ready to be measured will receive a bonus of 10 points.
- The team winner is the team with the highest point total.
- The individual winner is the student with the highest point total.

17.9.4 After the Exercise

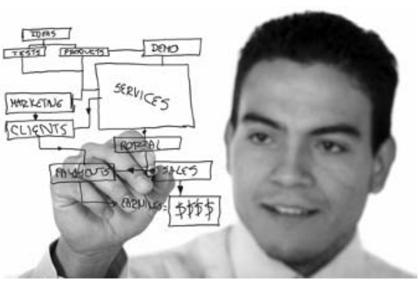
Discuss the importance of the following issues to the outcome of the competition:

- The quality of teaming among different teams. Were any of the extra materials (tape, paper clips, scissors) shared among teams?
- The quality of teaming within the team. Was everyone within the team given the opportunity to contribute ideas, or did one person dominate the decision making?
- Ethics. Was it ethical to jump to another team, not share materials, or copy the design of another team?
- Manufacturability. How important was it to have the right materials?

This page intentionally left blank

CHAPTER

18 Design Step 1: Defining the Problem



Source: © iStockphoto.com/Andresr

The design process begins when somebody, whom we shall refer to as the customer, expresses a need and so enlists the services of an engineer. The customer can be an individual, an organization, or the consuming public. Most customers are not engineers. It is up to the engineer to translate the customer's need into engineering terms. The result is cast in the form of a **problem definition** and a **list of specifications**.

18.1 PROBLEM DEFINITION

A problem definition states the design objective in one to three clear, concise sentences. For example, the problem definition addressed by Orville and Wilbur Wright at the turn of the twentieth century was *design a manned machine capable of achieving powered flight*.

This problem definition tells us that they wanted to design a flying machine subject to two constraints. First, it must carry a person, which rules out model aircraft. Second, an onboard power source must be used to take off, which eliminates the possibilities of leaping off a barn with hand-held wings and lighter-than-air craft such as a hot air balloon.

The problem definition is constructed in response to an expressed need. Failure to identify, understand, and validate the need prior to designing, is one of the most frequent causes of failure of the entire design process.

The customer's statement of need does not typically take the form of a problem definition. For example, consider the following statement of need from a fictitious client:

> **Need**: People who work at the Empire State Building are complaining about the long waits at the elevator. This situation must be remedied.

An engineer might translate this need into the following problem definition:

> **Problem Definition**: Design a new elevator for the Empire State Building.

Is this really a good problem definition? Is the main concern of the management at the Empire State Building to reduce average waiting times or to eliminate the complaints? When turning an expressed need into a problem definition, it is important to eliminate assumptions that unfairly bias the design toward a particular solution. A better, less-biased problem definition might be:

> **Improved Problem Definition**: Increase customer satisfaction with the elevators in the Empire State Building.

This would admit such solutions as a mirror on the elevator door or free coffee on the busiest floors.

As another example of an inadequate problem definition, consider the following: *Design a device to eliminate the blind spot in an automobile*. This proposed problem definition also contains an assumption that prematurely limits the designer. The word *device* rules out one solution that achieves the design goal (eliminating the blind spot) by simply repositioning the front and side mirrors.

A third example occurred in a design competition named Blimp Wars (see Figure 18.1). The goal was to *design a system to retrieve Nerf® balls from an artificial tree and return them to the blimp base*. Inclusion of the word *blimp* in the problem definition biased the students toward blimp designs. The alternative of an extendable arm that would span the distance between blimp base and the target balls was not considered.

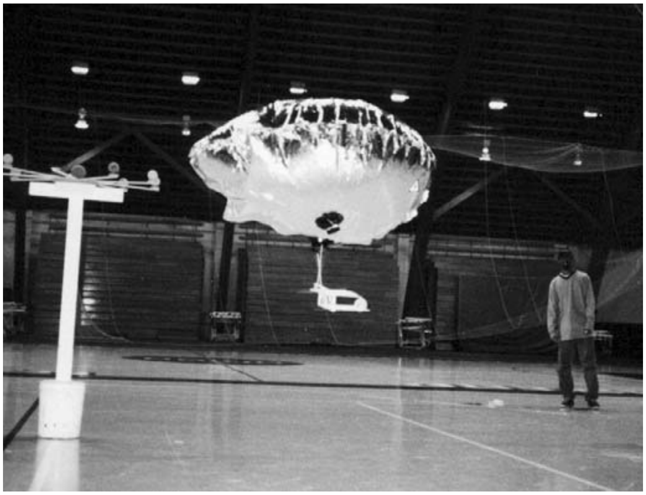


FIGURE 18.1 Blimp Returning to Base After Retrieving a Ball from the "Tree" of the Left

18.2 LIST OF SPECIFICATIONS

After translating the need into a problem definition, the next step is to prepare a list of specifications. The list of specifications includes both "demanded" design characteristics that must be present for the design to be considered acceptable and "wished for" design characteristics that are desirable but not crucial to the success of the final design. It is the usual practice to classify each specification as either a demand (D) or a wish (W). Don't confuse the two. If you treat a wish as if it was a demand, your design may become more complicated than is necessary.

Whenever possible, use numbers to express specifications. For example, instead of merely requiring that weight must be low, state, "Weight must be less than 10 pounds." Sometimes use of numbers is impossible. A quality such as "aesthetically pleasing" is difficult to quantify. However, use numbers wherever possible, even if at this early stage they seem like guesses. The numbers can be refined later as the design begins to take shape.

The specifications should be solution independent to avoid bias. For example, if you are designing a small mobile device, requiring that "the wheels must be made of rubber" will bias the design in two respects: in the use of wheels and in the choice of materials. Such decisions are reserved for later in the design process after careful consideration of alternatives.

Specifications come in the following categories:

- Performance
- Geometry
- Materials
- Energy
- Time
- Cost
- Manufacture
- Standards
- Safety
- Transport
- Ergonomics

These categories can also be used as headings by which to organize the list of specifications. Here is an example.

Example 18.1

The following problem definition was posed to three competing design teams.

Design and build a remote-controlled,[1] portable device that will play nine holes of golf at a local golf course with the fewest possible number of strokes. The instructor also supplied the following demands. It was left to the students to develop a complete list of specifications.

Demand (D) Specification

Must cost less than $600 (not including radio).
Must be remotely triggered.
Total number of radio-controlled servos[2] is eight.
Device cannot be touching the golf ball prior to remote triggering of the shot.
Entire device must form a single unit.
Must be portable.
Design must pass a safety review.
Ground supports must fit within a 3-foot circle.

Solution

The first step was to organize the demands under each heading. Then, using the headings as a guide, additional demand (D) or wish (W) specifications were formulated. The results follow.

Performance

D – Must be remotely triggered.
D – Device cannot be touching the golf ball prior to remote triggering of the shot.
D – Driving distance must be adjustable with a range between 15 and 250 yards.
D – Putting distance must be adjustable with a range between 0 and 15 yards.

---

[1]A common abbreviation in design for "remote control" or "remote controlled" is "RC."

[2]A servo is a control system to amplify a small signal into a large response. Typically it is an electric motor controlled by a small voltage.

D – Must operate on inclines of up to 45 degrees.
W – Must sink 95% of short putts (less than 3 feet).
W – Driving accuracy of ± 5 yards.

Geometry

D – Total number of radio-controlled servos is eight.
D – Entire device must form a single unit.
D – Ground supports must fit in 3-foot circle.

Materials

W – Materials must not degrade under expected range of weather conditions (including rain, snow, $30° \text{ F} < T < 90° \text{ F}$).

Time

D – Must be designed and manufactured in less than 14 weeks.

Cost

D – Must cost less than $600 (not including radio).

Manufacture

D – Must be manufactured using tools available in the machine shop.
D – Must be manufactured using machining skills available within the team.
W – Off-the-shelf parts and materials should be readily available.

Standards

D – Radio must adhere to FAA regulations.

Safety

D – Design must pass a safety review.

Transport

D – Must be portable.
W – Must fit in a car or small truck (for easy transport to golf course).

18.3 DESIGN MILESTONE: CLARIFICATION OF THE TASK

There are two versions of this milestone, depending on the format of the design project. If there is a design competition involved, main responsibility for producing the list of specifications shifts from the students to the instructor, as there is a need for everyone to operate under the same set of constraints. In either case, it is assumed that the instructor provides the problem definition.

For a General Design Project (Version A)

Assignment

1. Interview the customer. (In the case of a consumer product, conduct a product survey.)
2. Prepare a typed list of specifications.

For Design Competitions (Version B)

Assignment

1. Review the rules of the competition and ask the instructor for rule clarifications.
2. Prepare a typed list of design requirements to supplement those already appearing in the official rules of the competition. For example, set performance goals for your machine. You do not have to list requirements already appearing in the rules.

18.3.1 Design Competition Tips

- Probe the boundaries of the rules for wild, unconventional ideas.
- Avoid any temptation to bias the requirements toward a particular solution or strategy.
- Expect the list of supplemental requirements to be very short if the rules are well defined.

This page intentionally left blank

CHAPTER

Design Step 2: Generation of Alternative Concepts

19

Source: © iStockphoto.com/Linda Bucklin

Once the problem statement is in place, and the specifications have been listed, it is time to generate alternative concepts. By a *concept*, we mean an idea as opposed to a detailed design. The representation of the concept, usually in the form of a sketch, contains enough information to understand how the concept works but not enough information to build it. By *alternative*, we are requiring that the various proposed ideas must be fundamentally different in some way. The differences must go beyond appearance or dimensions. The usual rule of thumb in design courses is to generate at least three fundamentally different concepts.

In this chapter, four aspects of concept generation will be discussed: **brainstorming, concept sketching, research-based strategies**, and **functional decomposition**.

19.1 BRAINSTORMING

The most common approach for generating ideas is by brainstorming. As the term implies, you rely on your own creativity and memory of past experiences to produce ideas. Usually, team members will generate ideas on their own before meeting with the team for a brainstorming session.

Brainstorming is based on one crucial rule: *criticism of ideas is not allowed*. This enables each team member to put forth ideas without fear of immediate rejection. For example, a professor once recorded the brainstorming session of a small team of students. At one point, a student offered an idea, and another student referred to it as “stupid.” The voice of the first student was never heard from again during the session. Instead of a team of four, it had become a team of three.

It is important to devote some of the brainstorming time searching for bold, unconventional ideas. In the case of a design competition, this could mean searching for holes in the rules that could lead to ideas that the creators of the competition had not anticipated.

Only when brainstorming is complete should the team eliminate concepts that are not feasible, not legal, or not fundamentally different. After this weeding-out process, at least three concepts should remain. If not, more brainstorming is in order. The following example illustrates this step.

Example 19.1

Assuming the alternative concepts in Figure 19.1 were generated as part of an effort to design a new bat for Major League baseball, which concepts should be eliminated because they are not feasible, not legal, or not fundamentally different?

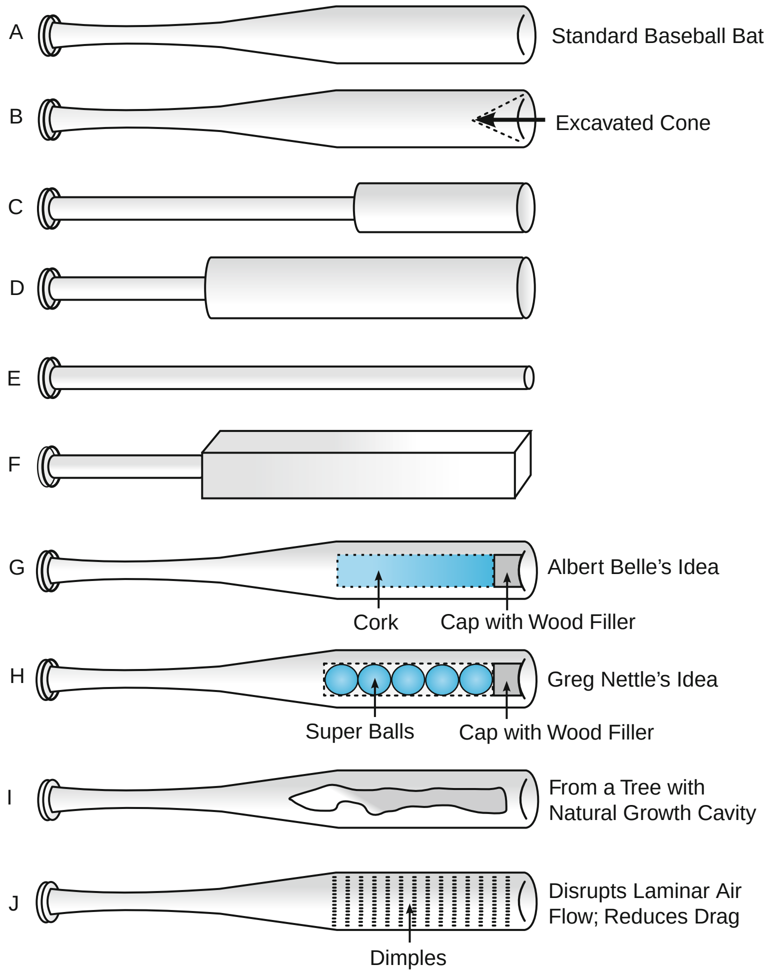


FIGURE 19.1 Alternative Concepts for a Major League Baseball Bat

Solution

Not feasible: E because it stands no chance of being competitive; I because it is too difficult to find in nature.

Not legal: F, G, H, and J.

Not fundamentally different from each other: C and D because basic shape is the same; only dimensions differ.

Therefore, the condensed list of viable alternatives consists of concepts A, B, and C.

19.2 CONCEPT SKETCHING

For an idea to be considered a feasible alternative concept, it must be represented in the form of a conceptual sketch. The goal in producing a concept drawing is to convey what the design is and how it works in the clearest possible terms. Any lack of clarity, such as failure to represent one of the subfunctions, will translate into doubts about the feasibility of the concept when it comes time to evaluate it.

At the same time, however, this is not a detailed design drawing. Dimensions and other details not relevant to understanding the basic nature of how the concept will work are left out.

It is best to proceed through two phases when generating a concept drawing. First, in the creative phase, hand-sketching is done freestyle and quickly, without regard for neatness or visual clarity. A few simple lines, incomprehensible to others, might be enough to remind you of your idea. Sketching is a means for both storing ideas and brainstorming others. The final outcome is a rough sketch of the concept. Second, in the documentation phase, the concept is neatly redrawn and labeled to facilitate communication with team members and project sponsors.

The final outcome is one or more sketches prepared with the following guidelines in mind:

- Can be hand-sketched or computer-generated.
- No dimensions. Remember, this is not a detailed drawing.
- Label parts and main features. If the drawing is hand-sketched, handwritten labeling is acceptable.
- Provide multiple views and/or close-up views if needed to describe how the design works.

The choice of views is up to you. Isometric views like those shown in Figures 19.2 and 19.3 convey a lot of information in a single picture. Most mechanisms can be described effectively using one or more two-dimensional views as in Figure 19.4. Despite their apparent informality, the quality of these drawings is crucial to fairly representing the designs during the evaluation process. In some cases they are the only source of evidence for judging if a design is likely to work.

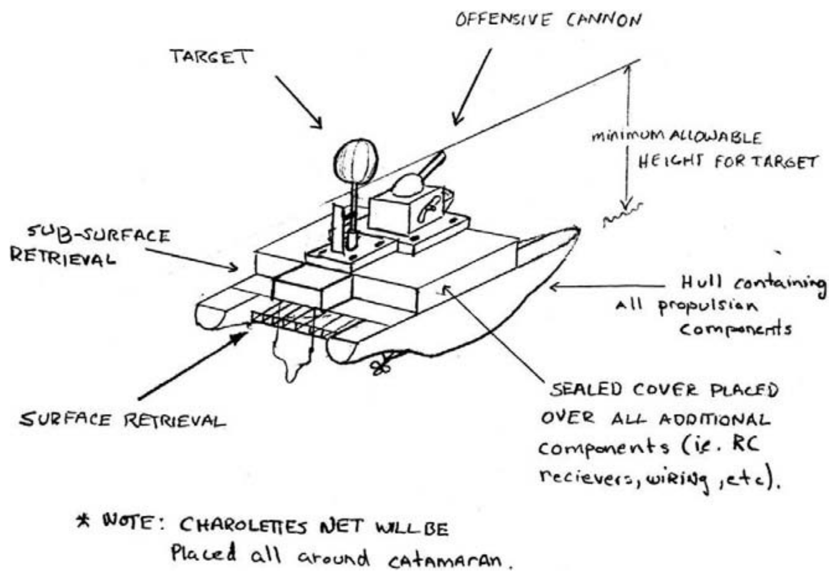


FIGURE 19.2 Concept Drawing of a Radio Controlled (RC) Boat for a Design Competition (hand-drawn isometric)

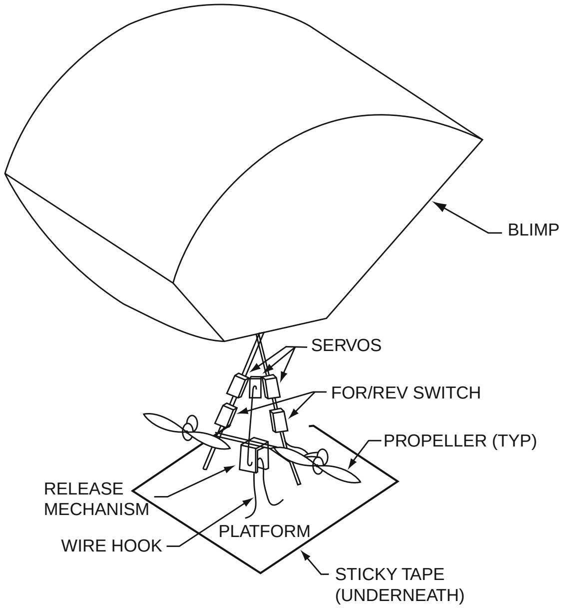


FIGURE 19.3 Concept Drawing of a Radio-Controlled (RC) Blimp (computer-generated isometric)

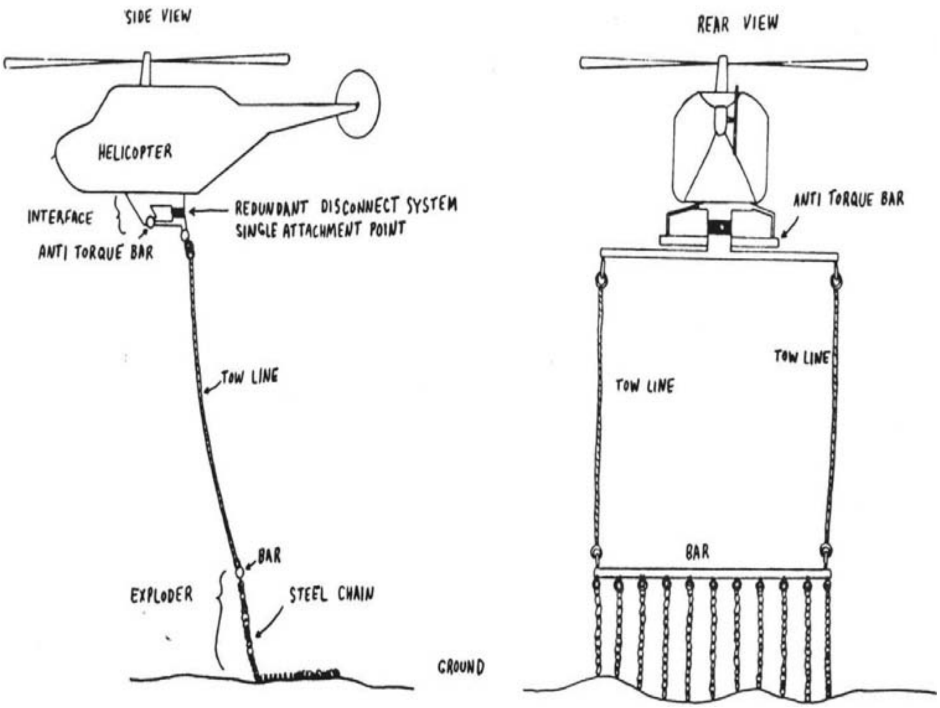


FIGURE 19.4 Concept Drawing of an Anti-Personnel Mine Clearing System (two hand-drawn views)

19.3 HANDS-ON DESIGN EXERCISE: THE TUBE

The design objective is to extract a golf ball from the bottom of a free-standing, open-ended mailing tube in the shortest possible time.

19.3.1 Setup

- Place a mailing tube vertically on the floor and drop a golf ball in the tube.
- Have a supply of the following materials: string, duct tape, Scotch® tape, 8.5×11 standard copier paper, and scissors.

19.3.2 Rules

- Limited to using the supplied materials.
- Scissors can be used for manufacturing.
- Everyone in the group can help in manufacturing, but only one person can extract the ball.
- Students are not allowed to handle the materials until it is time to test.
- Must manufacture the design shown on the concept drawing handed to the instructor.
- Time limit of 3 minutes to manufacture concept and extract ball.
- Cannot tip over the tube.
- Cannot touch the outside of the tube with anything.
- No forces can be applied to the inside of the tube in an effort to hold it vertical; accidental contact with inside of tube is okay as long as the tube does not tip over.
- Violation of any of the preceding rules will lead to immediate disqualification.

19.3.3 Procedure

1. First allow the students 3 minutes to individually brainstorm (encourage them to draw quick sketches of each of their concepts).
2. Then divide section into teams of four students per team.
3. Allow teams 10 minutes to collect ideas, brainstorm as a team, select their best concept, and give a sketch of their best concept to the instructor.
4. Instructor should walk around during brainstorming to remind teams to (a) generate multiple solutions before selecting one and (b) try to involve everyone in the process.
5. Allow teams 2 minutes to assign responsibilities for manufacture and test.
6. One at a time, give each team 3 minutes to manufacture their concept and attempt to extract the golf ball.
7. Team with the shortest retrieval time wins.

19.4 RESEARCH-BASED STRATEGIES FOR PROMOTING CREATIVITY

Some ideas are truly original, but most are drawn from past experience. The following strategies help you to look at old designs in order to generate new ones.

19.4.1 Analogies

One often used strategy is to look for analogous design situations in other unrelated fields. To do this, first you have to translate the design objective into an overall function that is general enough to widely apply.

For example, you may want to design a system to “climb a vertical wall” or “walk on two legs” or “move efficiently through the water.” Nature is filled with solutions to these problems (but because of their complexity, biological solutions usually have to be simplified and adapted before they can be of practical use). If you are designing a system to “throw an object,” a survey of ancient artillery could spark ideas.

19.4.2 Reverse Engineering

The basic strategy here is to acquire an existing product that is similar to a design you have in mind, take it apart, figure out how it works, and then either try to improve on it or adapt some of the ideas to your own design. Toy stores are a great place to search for small electromechanical devices that can be reverse engineered.

19.4.3 Literature Search

Web-based search engines are very effective at finding existing design solutions. For high-tech applications, you should also search books and the electronic databases for technical journals (e.g., the Science Citation Index).

19.5 FUNCTIONAL DECOMPOSITION FOR COMPLEX SYSTEMS

When confronted with a complex problem, it is frequently advantageous to break it down into smaller, simpler, more manageable parts. In the case of design, those smaller parts usually correspond to the individual functions (or tasks) that must be performed in order to achieve the overall design objective. This approach, known as **functional decomposition**, is the basis of the procedure described as follows for generating concept alternatives.

Step 1. Decompose the design objective into a series of functions.
Start out by decomposing the overall function into four or five subfunctions. Usually, verbs such as *move, lift*, and *control* are used in naming the functions. Figure 19.5 shows the functional decomposition of the remote-controlled golf machine of earlier examples. It is given in the form of a tree diagram, which is probably the most common form of representation. If more detail was needed, each of the subfunctions could be further broken down into its respective subfunctions.

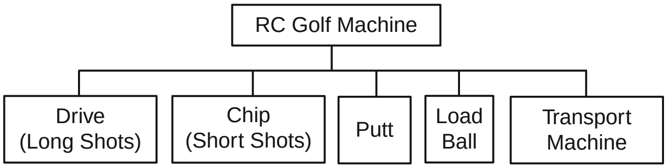


FIGURE 19.5 Functional Decomposition for a Design of a Remote-Controlled Golf Machine

When it is not readily apparent what the subfunctions are, it may help to think in terms of the sequence of tasks that must be performed by the design. The “sequential” functional decomposition for a design to assist disabled people into and out of a bathtub is shown in Figure 19.6.

It is very important that the functional decomposition be general enough to avoid biasing the design solution. For example, the separate drive and chip functions in Figure 19.5 may cause the design team to overlook the possibility of using the same device to fulfill both functions. If solution bias is unavoidable, introduce multiple functional decompositions.

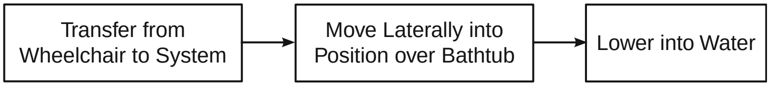


FIGURE 19.6 "Sequential" Functional Decomposition for a Design of a System to Aid the Disabled into and out of a Standard Bathtub

Step 2. Brainstorm on alternative concepts for each function and assemble the results in a classification scheme.

The classification scheme[1] is a two-dimensional matrix organized as shown in Table 19.1. The first column lists the functions resulting from the functional decomposition. The row of boxes next to each function name contains the corresponding design solutions that have been brainstormed. The design solutions are expressed using a combination of words and pictures, so be careful to draw the boxes large enough to accommodate small illustrations.

Table 19.1 Organization of the Classification Scheme

Concepts / Functions	Concept 1	Concept 2	Concept 3	Concept 4
Function A	A1	A2	A3	A4
Function B	B1	B2	B3	B4
Function C	C1	C2	C3	C4
Function D	D1	D2	D3	D4

Step 3. Combine function concepts to form alternative design concepts.

Table 19.2 demonstrates how one subfunction concept from each row of the classification scheme is selected to form a total concept. The same subfunction concept can be used with more than one total concept, though keep in mind that the idea is to generate fundamentally different design concepts. The only other rule when

Table 19.2 Combining of Compatible Sub-Function Concepts

Concepts / Functions	Concept 1	Concept 2	Concept 3	Concept 4
Function A	A1	A2	A3	A4
Function B	B1	B2	B3	B4
Function C	C1	C2	C3	C4
Function D	D1	D2	D3	D4

Total Concept I = A1 + B2 + C2 + D1
Total Concept II = A4 + B2 + C4 + D2

[1] Pahl, G. and W. Beitz. (1988). *Engineering Design—A Systematic Approach*. New York: Springer-Verlag.

deciding upon the best combinations is to be sure that the subfunction concepts being combined are compatible.

Step 4. Sketch each of the most promising combinations.
This is done in accordance with the rules previously presented for concept drawings. Remember that you must end up with drawings for at least three fundamentally different design concepts.

Example 19.2

Use functional decomposition to generate alternative concepts for a proposed remote-controlled blimp, capable of retrieving Nerf® balls from an artificial tree and returning them to blimp base (see Figure 18.1).

Solution

The first step was to produce the functional decomposition of Figure 19.7. Then concepts were brainstormed for each of the subfunctions, and the results were assembled in the classification scheme of Figure 19.8.

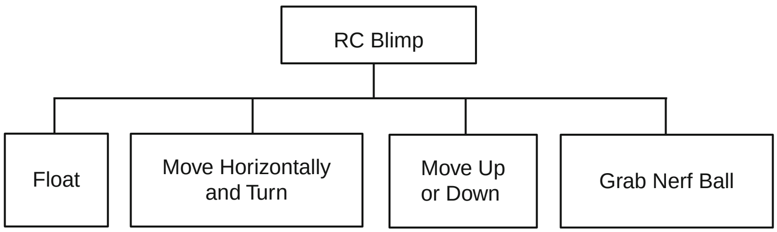


FIGURE 19.7 Functional Decomposition for the Design of a Remote-Controlled Blimp

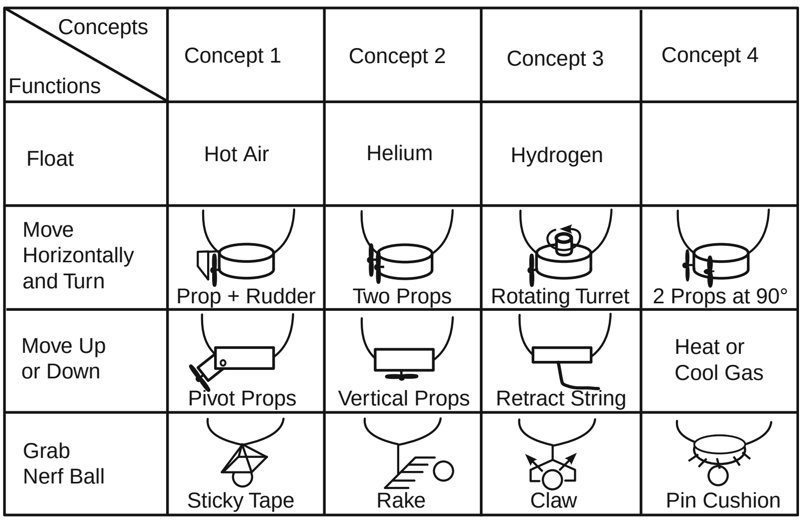


FIGURE 19.8 Classification Scheme for a Remote-Controlled Blimp

Total concepts were formed by combining compatible subfunction concepts. The three promising total concepts are:

Total Concept I = helium + 2 props + pivot props + sticky tape
Total Concept II = helium + rotating turret + vertical prop + rake
Total Concept III = helium + prop with rudder + string + claw

The final step is to represent each alternative design in the form of a concept drawing. The concept drawing for Total Concept I was shown in Figure 19.3.

19.6 DESIGN MILESTONE: GENERATION OF ALTERNATIVES

This milestone assumes the system to be designed is sufficiently complex (i.e., at least two subfunctions) to warrant the use of functional decomposition.

Assignment

1. For the functional decomposition given in class (or a modification of it that you are at liberty to propose), brainstorm to determine at least five feasible alternatives for each subfunction and assemble the results in a classification scheme.
2. Form three promising design concepts by combining compatible subfunction alternatives from your classification scheme.
3. Firm up your three design concepts by sketching them in the form of concept drawings. Functionality (i.e., how it works) should be clearly indicated in the drawings through the use of labeling and text.

Grading Criteria

Technical Communication

- Ideas are clearly presented
- Final concept drawings are neatly rendered

Technical Content

- All concepts are feasible, legal, and fundamentally different
- Concepts are presented in sufficient detail
- Requested number of concepts was generated

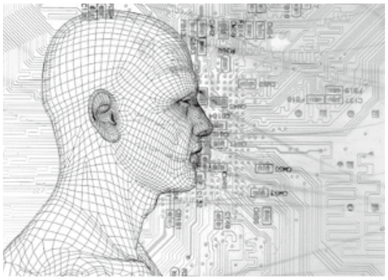
19.6.1 Design Competition Tips

- The goal is to generate three strong concepts.
- Search the boundaries of the rules for unusual ideas that could potentially dominate the competition. If you don't, someone else will.
- Include strategy as one of the items to be brainstormed in the classification scheme.
- Redraw your concept sketches to enhance clarity and neatness. The quality of the concept drawing, or lack of it, can do much to sway opinions when it comes time to judge the concepts.

This page intentionally left blank

CHAPTER

20 Design Step 3: Evaluation of Alternatives and Selection of a Concept



Source: © iStockphoto.com/Emrah Türüdü

Suppose you now have generated three concepts that will meet the problem definition and fulfill the specifications. Which one should you choose as the basis for your final design? There is no magic formula. However, Professor Nam P. Suh of MIT has provided two very helpful design principles for evaluating and improving concepts: **minimize information content** and **maintain the independence of functional requirements**.[1]

This chapter adds three additional considerations for evaluating alternatives: **ease of manufacture, robustness**, and **design for adjustability**. It then concludes with a method of pulling together all these ideas: the **decision matrix**.

20.1 MINIMIZE THE INFORMATION CONTENT OF THE DESIGN

When choosing among promising alternatives, the best design is often the one that can be uniquely specified using the least amount of information or, alternatively, can be manufactured with the shortest list of directions. This idea sometimes is stated as the *KISS* principle: *Keep It Simple, Stupid*.

There are a number of design guidelines that naturally follow. A few of the most notable ones are:

- Minimize the number of parts.
- Minimize the number of different kinds of parts.
- Buying parts is preferable to manufacturing them yourself.

20.2 MAINTAIN THE INDEPENDENCE OF FUNCTIONAL REQUIREMENTS

The functions considered in a functional decomposition provide the basis for Suh's second principle. This principle asserts that these functions should be independent of each other in a good design.

[1]Suh, N. P. (1990). *The Principles of Design*. New York: Oxford University Press.

A successful application of this principle is illustrated by the decoupled design in Figure 20.1. Independence of the functions "lift" and "move" was maintained by designing physically separate mechanisms for each action (scissors jack for lift, wheeled vehicle for move) and by performing the actions in sequence, rather that at the same time. First the scissors jack would lift the vehicle, the vehicle then would slide horizontally onto the next step, and finally the scissors jack would close upward and be pulled back underneath the vehicle. The coupled design employed four articulated arms, tanklike tracks on each arm, and a complicated motion to both lift and move at the same time. Though both machines performed admirably, the decoupled design had a much higher potential payload and was easier to build, since it required half as many motors.

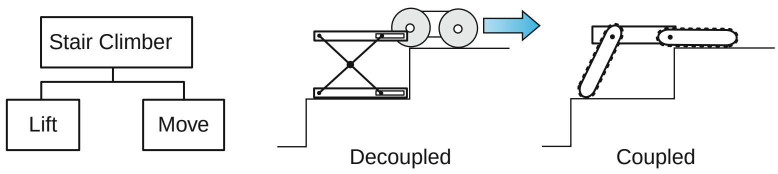


FIGURE 20.1 Two Concepts for a Stair-Climbing Machine—The First Concept (Center) Decouples the "Lift" and "Move" Functions but the Second Concept (Right) Does Not

The previous example suggests the following design guideline: **Seek a modular design**. A modular design is one in which the design solutions for each function have been physically isolated. The main advantage of a modular design is that the individual modules can be designed, manufactured, and tested in parallel, leading to much shorter product development times.

In looking for opportunities to improve a given design, the situation may arise in which Suh's two principles appear to be in conflict. For example, a design change aimed at increasing the independence of the functional requirements could result in greater complexity. Suh contends that any design change that either increases information or sacrifices the independence of the functional requirements should not be accepted. There always exists a less-coupled design with lower information content.

We now illustrate the application of these two design principles with an example.

Example 20.1

A head-to-head student design competition named Davy Jones's Treasure Trove was based on the following problem definition:

Design a system to retrieve surface (ping-pong balls) and subsurface (1 lb mass) objects from a swimming pool.

This was subject to the following major design requirements:

- It must fit in a 2 ft × 2 ft × 3 ft volume at the start of the competition.
- It must carry a target that disables the boat if struck by opponent.

Concept drawings of two of the student designs are shown in Figures 20.2 and 20.3.

Look at Figures 20.2 and 20.3 and evaluate the two student designs by identifying applications and violations of Suh's design principles. (**Hint**: Only relevant features have been labeled in the figures.)

Solution

Water Cannon Design

- Water cannon might be effective because aiming and steering are independent.
- Catapults will not be as effective because aiming is dependent on steering.

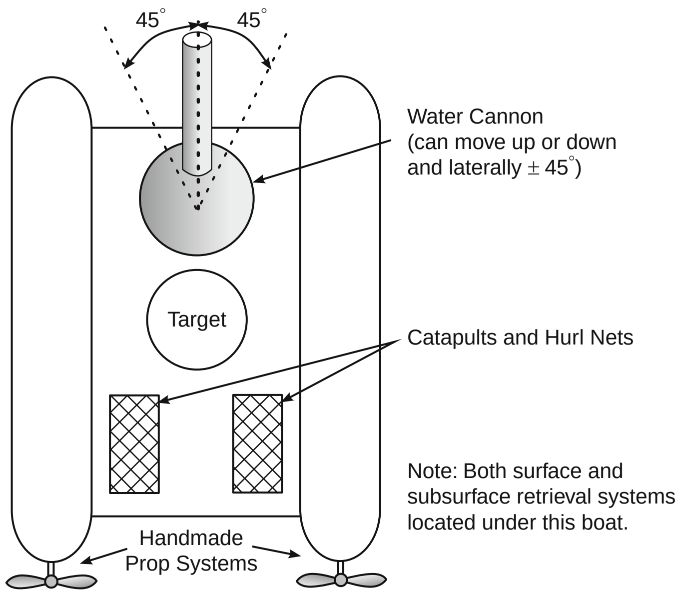


FIGURE 20.2 Water Cannon (top view)

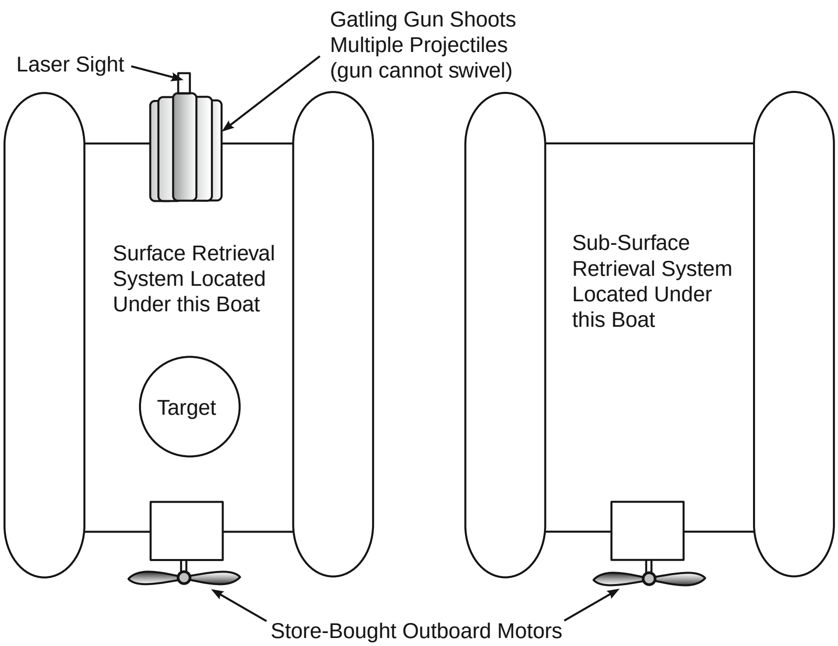


FIGURE 20.3 Twin Boat Design (top view)

- The boat should be very maneuverable because the two props serve to decouple the move and turn functions; that is, it should be able to turn on a dime.
- Manufacture of the prop systems could be needlessly time-consuming.

Twin Boat Design

- Use of two boats will be very effective because it decouples the two retrieval functions. One boat can collect the ping-pong balls while the other collects the 1 lb masses.
- Use of store-bought propellers will save time.
- Use of two nearly identical hull designs will simplify both design and manufacturing, thus saving more time.
- The boats will not be as maneuverable as the Water Cannon Design because the move and turn functions are not independent; that is, the single-prop design needs to be moving forward in order to turn.
- The Gatling gun will not be as effective as the water cannon because aiming is dependent on steering.

Final Note

The preceding characteristics provide accurate insight into how the boats actually performed. The Twin Boat Design won the competition largely on the strength of its dual retrieval system. Although the water cannon was far more effective than the Gatling gun, the Water Cannon Design was ultimately at the mercy of its handmade props, which took a lot of time to manufacture, left little time for testing, and proved to be unreliable.

20.3 DESIGN FOR EASE OF MANUFACTURE

There are clear advantages to going with a design that is easy to manufacture. If among competing design teams you are the first to complete manufacture of your design, the extra time can be used to test, debug, and optimize performance. For a commercial enterprise, first-to-market can mean a short-term monopoly in a fiercely competitive marketplace. Often, ease of manufacture goes hand-in-hand with lower costs. Thus, given the choice of two concepts, both of which satisfy the design requirements to the same degree, and where one is more difficult than the other to manufacture, it makes sense to choose the concept that is easier to manufacture.

At this stage in the design process, evaluation of ease of manufacture should be done at a level of abstraction consistent with the concept drawings. The counting of machining operations and assembly steps is reserved for a later time when the requisite level of detail in the design has been attained. Here the emphasis is on developing an impression of ease of manufacture as revealed through Suh's design principles.

A student or team with the formidable task of having to build a complex design themselves should be asking the following questions as each concept is evaluated:

Are there a large number of parts?

If there are a lot of parts that need to be made and assembled, it will take a long time to build.

Are there a large number of different kinds of parts?

For parts of comparable complexity, it takes less time to make two of the same part than it does two different parts because of reduced setup times.

Are there parts with complicated geometry?

These parts will take longer to make.

Can some parts be purchased?

This is not always an option in design competitions, but if it is, the time saved and the proven reliability of the prefabricated part usually justifies the purchase.

Do you have the skills to make all the parts?

Safest thing to do is to choose a concept that you know you can build.

Is it a modular design?

We noted earlier that modular components can be manufactured in parallel by subgroups within the design team, thus saving time.

Are there opportunities to simplify manufacture of the design by:

- Reducing the number of parts?
- Reducing the number of different kinds of parts?
- Simplifying the shape of some parts?
- Purchasing some parts?
- Redesigning the parts that are difficult to make?
- Modularizing the design?

20.4 DESIGN FOR ROBUSTNESS

Manufacturing errors, environmental changes, and internal wear can cause unexpected variations in performance. When the designed product is insensitive to these three sources of variability, the design is said to be **robust**. Engineers seek a robust design because performance of such a design can be predicted with greater certainty.

The designer must learn to expect the unexpected. All too often students conceive of a design while assuming ideal operating conditions. Yet, deviations from those ideal conditions can lead to less than ideal performance, as illustrated by the following example.

Example 20.2

For a design competition, students had to design machines that could accurately throw darts at a dartboard. The machines were powered by large falling masses. Yet, none of the machines was perfectly repeatable. For example, from 8 feet away, the best that the machine in Figure 20.4 could do was to keep the darts within a 1-inch circle. What factors contributed to this loss of dart-throwing accuracy?

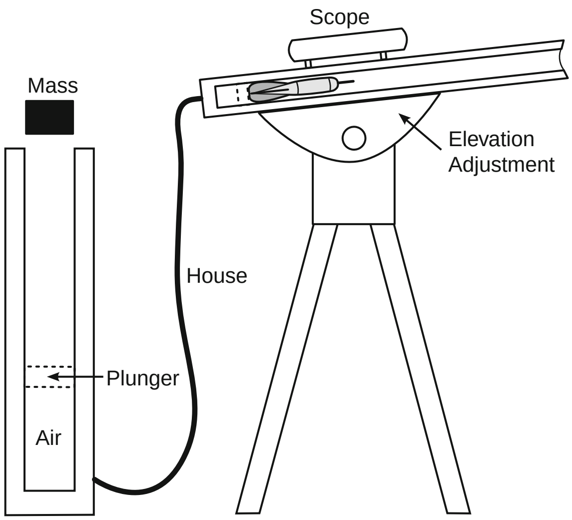


FIGURE 20.4 Concept Drawing of a Dart-Throwing Machine

Solution

Relevant manufacturing errors, environmental changes, and internal wear sites were brainstormed. The following list resulted:

Manufacturing Errors

- Small dimensional differences between darts in the set of three

Environmental Changes

- Small air currents
- Inexact repositioning of the plunger
- Inexact repositioning of the dart within the blow gun
- Inconsistent releases of the falling mass

Internal Wear

- Damage to the dart fins
- Blunting of the dart tip
- Damage to the dartboard

When evaluating concepts with respect to robustness, you should be asking yourself the following questions:

Will small manufacturing errors dramatically impair performance?

If parts have to be manufactured perfectly in order for the design to function properly, you should expect to run into problems. You want a design that will work even when part dimensions are a little off. This is one reason why gear sets are such a popular design choice. Small errors in center distance between mating gears do not change the gear ratio.

Will the design function properly over the full range of environmental conditions?

Environmental conditions subject to variation include applied forces, atmospheric conditions, and roughness of surface terrain. The expected range of relevant environmental conditions should be clearly defined in the list of specifications. If they are not there, now is a good time to include them.

In a head-to-head design competition, have the actions of the opposing teams been anticipated?

Those actions can contribute significantly to the variability of the environmental conditions. For example, an opposing machine can apply forces to your machine or alter the roughness of surface terrain by laying obstacles. Thus, you want to select a strategy/design combination that will perform well regardless of what the opposing teams may do.

20.5 DESIGN FOR ADJUSTABILITY

In engineering courses there is usually only enough time and resources to manufacture one design, and that design almost never performs as planned on the first try. Optimizing performance by building several designs is not an option. The only remaining course of action is to design adjustability into the initial implementation.

There are a number of ways to design for adjustability. One way is to design the system with modularity. This can serve to isolate required design changes to a single subsystem. In a mechanical system, dimensional adjustability can be attained by using nonpermanent fastening methods such as screw joints instead of a permanent method like epoxy.

Design for adjustability can be incorporated into the evaluation process by asking the following questions as each concept is reviewed:

What are the main performance variables?

Usually, only one or two of the most important variables need be considered. Typical performance variables are speed, force, and turning radius.

Can those performance variables be adjusted easily?

Common methods were described previously. Other methods are found by brainstorming and by examination of the governing equations.

Example 20.3

A motor-driven moving platform is a common feature of many small-scale vehicle designs. A top view of one such moving platform design is shown in Figure 20.5.

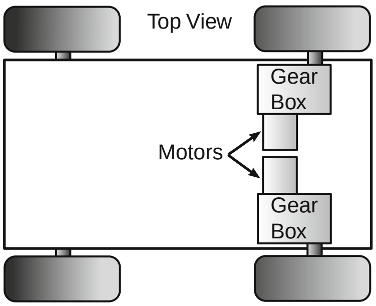


FIGURE 20.5 Moving Platform

Once manufacture of this design is complete, what adjustments can be made to:

a. Increase the speed of the moving platform?
b. Increase its pushing force?

Solution

1. Alternative methods for increasing the speed are:
- Decrease the gear ratio
- Increase the radius of the tires
- Increase the voltage from the power supply
- Switch in motors with higher RPM

2. Alternative methods for increasing the pushing force are:
- Increase the gear ratio
- Decrease the radius of the tires
- Increase the voltage from the power supply
- Switch to motors with higher peak torque

20.6 HANDS-ON DESIGN EXERCISE: WASTE BALL

20.6.1 Scenario

A company that uses radioactive substances for research sometimes has “spills” of spherical radioactive objects. When this happens, the radioactive objects must be transported to a waste container by the emergency team.

20.6.2 Design Objective

Design a method to transfer the radioactive substance (plastic ball) from the site of the spill to a waste container (small refrigerator) at another location.

20.6.3 Setup

- Divide the team, which consists of the entire class, into subfunctional groups of two or three students each. Each subfunctional group will be responsible for one leg of the transfer.
- Prior to the class, the instructor has to lay out the course. There must be as many different legs as there are subfunctional groups. The room in which the ball initially is placed and another room that contains the small refrigerator account for two of the legs. Other legs can consist of a corridor, a stairwell, an elevator, or an outdoor excursion. Try to make each challenge a little different to promote development of specialized designs by the subfunctional groups.
- Distribute the following materials to each subfunctional group:
  1 daily newspaper (or equivalent)
  1 roll of duct tape
  1 foam plate
  1 plastic cup
  1 pair of scissors (for construction only)
- The team will also receive:
  3 balls of string, which must be shared among the groups

20.6.4 Rules

1. Since the ball is radioactive, no one can be within 8 feet of the ball.
2. You must use only the materials provided.
3. For safety purposes, running is not allowed.
4. If necessary, doors must be safely held open by the teams and then closed immediately after waste passes through.
5. If during transport, the ball accidentally touches something besides the transport container (e.g., floor) a 30-second penalty will be imposed and the group carrying the ball must restart at the location where it received the handoff.
6. The team has 3 minutes per group to complete the design for the transport.
7. The class with the minimum transit time wins.

20.6.5 After the Exercise

- Assess team performance by comparing times to other sections.
- Discuss the quality of communication between subfunctional groups.
- What were the lessons learned?

20.7 THE DECISION MATRIX

The decision matrix promotes a systematic and exhaustive examination of concept strengths and weaknesses. The entire procedure, from selection of evaluation criteria to filling out the matrix, is designed to remove personal bias from the decision-making process. The results give a numerical measure for ranking alternatives and ultimately selecting the best concept.

20.7.1 Evaluation Criteria

The criteria by which the concepts should be judged are all contained in the list of specifications. To even qualify as a feasible concept, the expectation must be that all the design requirements designated as demands will be satisfied. Therefore, the ranking of the feasible concepts ultimately depends on the degree to which they fulfill the design requirements designated as wishes. However, at the conceptual level, qualities associated with both demands and wishes are included among the evaluation criteria owing to the uncertainty still associated with estimating their degree of fulfillment.

The design requirements selected to serve as evaluation criteria usually are reworded to indicate the desired quality. For example, instead of weight, cost, and manufacture, the corresponding evaluation criteria become *low weight, low cost*, and *easy to manufacture*.

Evaluation criteria should be independent of each other to ensure a fair weighting of requirements in the decision matrix discussed later. For example, low cost and ease of manufacture will be redundant and thus double counted if cost of labor is a significant fraction of total cost.

The number of evaluation criteria can vary depending on the situation. We suggest a level of detail consistent with the amount of detailed information available about the concept. For most hands-on student projects, five to seven of the most important evaluation criteria should suffice. *Easy to manufacture* and *low cost* are almost always included in this list.

20.7.2 Procedure for Filling Out a Decision Matrix

Step 1. Identify the evaluation criteria.

This step is described in the previous section.

Step 2. Weigh the evaluation criteria.

Weight values are assigned to each evaluation criterion in proportion to its relative importance to the overall success of the design; the larger the weight, the more important the evaluation criterion. Though not a mathematical necessity, it is usually a good idea to define the weights such that their sum is equal to 1, that is:

$$\sum_{n=1}^{N} W_n = 1 \tag{20.1}$$

in which N is the number of evaluation criteria. This constraint instills the view that weights are being distributed among the criteria and in so doing helps to avoid redundant criteria.

Step 3. Set up the decision matrix.

The organization of the decision matrix is illustrated in Table 20.1. The names of the concepts being evaluated are filled in at the top of each column. Likewise, the evaluation criteria and their assigned weights are written in the leftmost columns of the matrix. Scoring and intermediate calculations will be recorded within the subcolumns under each concept and then totaled at the bottom of the matrix.

Table 20.1 Organization of the Decision Matrix

		Concept A		**Concept B**		**Concept C**	
Evaluation Criteria	**Wt**	**Val_1**	**Wt × Val_1**	**Val_2**	**Wt × Val_2**	**Val_3**	**Wt × Val_3**
Criterion 1							
Criterion 2	↓		↓		↓		↓
Criterion 3							
Criterion 4							
Criterion 5							
Totals	1.0		O V_1		O V_2		O V_3

Step 4. Assign values to each concept.

Starting in the first row, each concept is assigned a value between 0 and 10 according to how well it satisfies the evaluation criterion under consideration. The values are assumed to have the following interpretation:

0 = *Totally useless* concept in regard to this criterion
5 = *Average* concept in regard to this criterion
10 = *Perfect* concept in regard to this criterion

and are recorded under the first subcolumn of each concept. This process is repeated for each criterion, going row by row to avoid bias. Usually assignment of values is based on a qualitative assessment, but if quantitative information is available, they can be assigned in proportion to known parameters.

Step 5. Calculate overall value for each concept.

For each concept-criterion combination, the product of the weight and the value is calculated and then recorded in the second subcolumn. After these calculations are completed, the overall value (OV) is computed for each concept using the following expression:

$$OV = \sum_{n=1}^{N} (W_n V_n) \tag{20.2}$$

which is equivalent to summing the second subcolumn under each concept heading. The OVs are recorded at the bottom of the matrix.

Step 6. Interpret the results.

The highest overall value provides an indication of which design is best. Overall values that are very close in magnitude should be regarded as indicating parity given the uncertainty that went into assignment of weights and values. The final result is nonbinding. Thus, there is no need to bias the ratings so as to obtain the hoped for final result. Rather, the chart should be regarded as a tool aimed at fostering an exhaustive discussion of strengths and weaknesses.

20.7.3 Additional Tips on Using Decision Matrices

1. Every member of the design team should individually fill out a decision matrix prior to engaging in team discussions. This will give everyone a chance to think about strengths and weaknesses ahead of time and thus make it more likely that they will be active participants at the team meeting.
2. Use the matrix to identify and correct weaknesses in a promising design. Give priority to the weaknesses that are most heavily weighted.
3. Feel free to create new alternatives by combining strengths from competing concepts.

Example 20.4

Recalling the remote-controlled golfing machines as discussed in the previous chapter, the three concepts appearing in Figures 20.6 through 20.8 have been proposed as best fulfilling the design requirements. Evaluate these three concepts by using the previously described procedures for filling out a decision matrix.

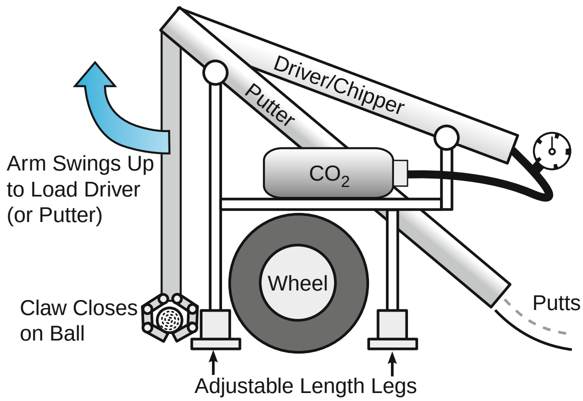


FIGURE 20.6 Concept Drawing of the "Cannon"

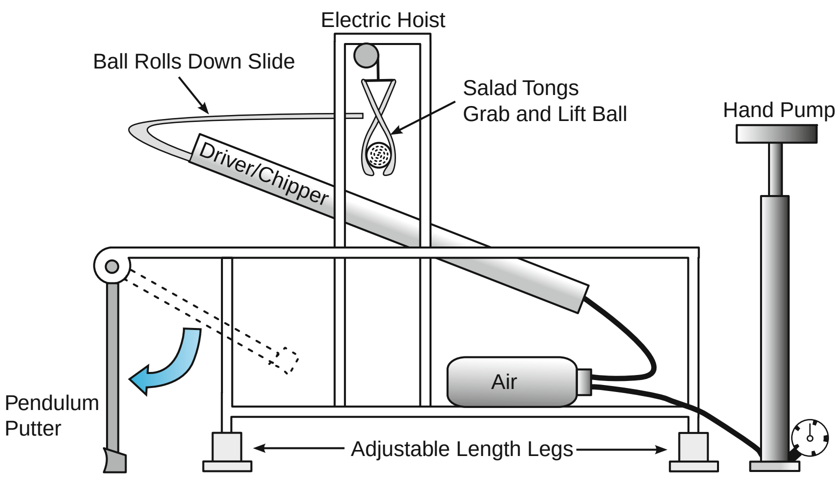


FIGURE 20.7 Concept Drawing of the "Original"

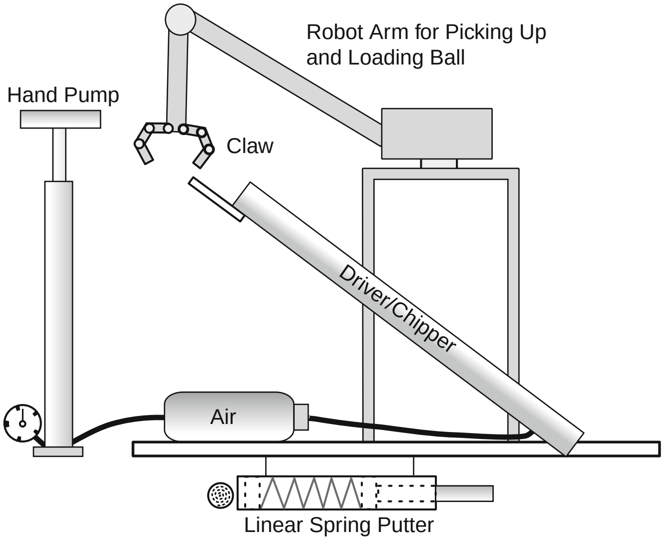


FIGURE 20.8 Concept Drawing of the "Robogolfer"

Solution

The following design requirements were selected to serve as evaluation criteria:

- Drives well
- Putts well
- Ball loader is robust (i.e., picks up ball off of all types of terrain)
- Easy to transport
- Easy to manufacture

Low cost, which usually appears, was not selected because all three concepts met the cost requirement and cost was not involved in the design competition scoring.

With these in hand, the decision matrix can be drawn up and weights assigned to each criterion. The drives and ball loader were considered equally important because one cannot work without the other. The drives/chips were weighted slightly higher than the putts because only 43 percent of all shots taken by golf professionals are putts. Transport is weighted low because it does not factor into scoring. Ease of manufacture is always important because of its impact on development times. The resulting weights are listed in the decision matrix of Table 20.2.

Then, proceeding one evaluation criterion at a time, the team analyzes the strengths and weaknesses of each concept in the context of the given criterion and assigns corresponding values to each concept in the decision matrix. The results of the analysis are presented in the following, and the values are recorded in Table 20.2.

Drives Well

All three drivers appear to be promising given the effectiveness of the notorious potato gun. However, since the CO_2 tank comes prepressurized and the hand-pumping is subject to a 60 s time limit on preshot preparation, the Cannon is likely to be firing the ball at higher pressures, and thus should have a greater range.

Table 20.2 Decision Matrix for the Remote-Controlled Golfing Machines

		Cannon		Original		Robogolfer	
Evaluation Criteria	**Wt**	**Val_1**	**Wt × Val_1**	**Val_2**	**Wt × Val_2**	**Val_3**	**Wt × Val_3**
Drives well	.25	9	2.25	8	2.00	8	2.00
Putts well	.20	4	0.80	4	0.80	8	1.60
Loader is robust	.25	6	1.50	6	1.50	9	2.25
Easy to transport	.05	9	0.45	5	0.25	5	0.25
Easy to manufacture	.25	5	1.25	7	1.75	3	0.75
Totals	1.0		6.25		6.30		6.85

Putts Well

The greens at the site of the competition will be severely sloped and slow. Therefore, the machines must be capable of executing long putts. Of the three machines, the Robogolfer is the most adjustable, since the springs can be easily replaced if the range proves inadequate. On the other hand, the potential energy of gravity powers the other two putters, and it will be difficult to increase starting heights once these machines are built. Therefore, there is greater risk associated with these putters.

Loader Is Robust

A wide range of lies is possible, from severe slopes to sand and divots. The Cannon and the Original address this issue by using legs that are adjustable in length. The robot arm of the Robogolfer is clearly the most flexible design and requires no setup time.

Easy To Transport

The rules require that only one student from the team may be used to transport the machine to the next shot location. The Cannon is the easiest to transport because only it has wheels.

Easy To Manufacture

The robot arm of the Robogolfer stands out as easily the most complicated system on any machine. As a three-degree-of-freedom mechanism, it requires three independently controlled motors. The Original's loader should be straightforward to manufacture. The salad tongs and the parts for the electric hoist can be easily purchased.

Discussion of Results

The Robogolfer is the clear winner on points. But the challenges involved in designing and manufacturing that robot arm should make you nervous (unless you have a robotics expert on your team). The decision matrix also revealed that the putters for the Cannon and the Original are weak concepts. If they are replaced by the Robogolfer's linear spring putter, the Original ends up with the most points.

The three concepts in Figures 20.6 through 20.8 correspond to actual student designs that were designed, manufactured, and tested. The Original team (so named because they were the first team to develop an air cannon) won the design competition. They compensated for their weak putter by chipping the long putts and adding a ramp to make the short putts. The Robogolfer team (who had a robotics expert) took longer than the Original and less time to test. The Cannon completed design and manufacture and thus had the longest drives and the shortest putts.

20.8 DESIGN MILESTONE: EVALUATION OF ALTERNATIVES

Successful completion of this milestone requires three strong design concepts, an open mind, and a lot of careful thought.

Assignment

1. Decide on five to seven evaluation criteria that will be used with a decision matrix to evaluate the three concepts from the previous milestone.
2. Assign weights to the evaluation criteria.
3. Fill out a decision matrix. One row at a time, discuss the strengths and weaknesses of all the concepts in the context of the given criterion, and then assign values by consensus before moving on to the next criterion.
4. Analyze the results of the decision matrix. Use the matrix to look for weaknesses and attempt to correct them by combining ideas from different concepts.
5. Select the best concept.
6. Document your evaluation process as per Example 20.4.

Grading Criteria

- Are weights and values accurate and fully justified?
- Were the results of the decision matrix interpreted thoughtfully when searching for and selecting the best concept?
- Were all three concepts strong designs?
- Is the documentation typed and clearly written?

20.8.1 Design Competition Tips

- There is no need to rig the results of the decision matrix to come out to the concept you want, as the results are nonbinding.
- Do not blindly obey the results of your decision matrix; the selection of evaluation criteria may have been flawed to begin with.
- Engage everyone in the decision-making process.
- Do not shy away from bold designs just because they are different from everyone else’s. Those differences could lead to victory at the final competition.

Design Step 4: Detailed Design

Source: © iStockphoto.com/Rui Jordao

The goal of this step in the design process is to specify the details of the design so that it can be manufactured. Those details are typically the dimensions and material composition of parts, as well as the methods used to join them. The decisions made during detailed design are guided by **analysis, experiments**, and **models** to reduce the risk that additional design changes will be needed later. The final results are documented in the form of **detailed drawings**.

21.1 ANALYSIS

Analysis refers to the application of mathematical models to predict performance. The role of analysis in freshman design projects is limited because the analytical capabilities of an engineering student are just starting to develop.

One calculation that may prove to be useful for small electromechanical devices is the determination of the optimal gear ratio. In developing this mathematical model we will draw from the equations on gearing and gear ratios in Chapter 13.

Assume that we want to determine the best overall gear ratio for the drive train of Figure 21.1. In Figure 13.7 of Chapter 13, this overall gear ratio corresponds to the combined gear ratio of the gear box

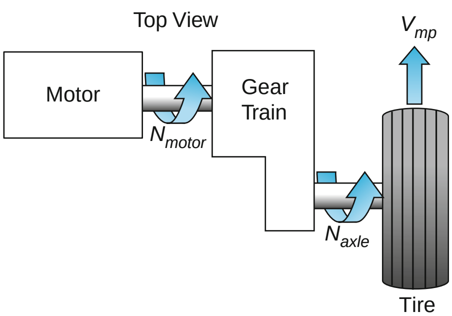


FIGURE 21.1 Schematic of the Drive Train for a Moving Platform

and the external pair of gears driving the axle. The overall gear ratio (GR) is equal to the reciprocal of the overall velocity ratio as expressed by the following equation:

$$GR = \frac{N_{motor}}{N_{axle}} \tag{21.1}$$

in which N_{motor} is the angular speed of the motor shaft (in RPM) and N_{axle} is the angular speed of the axle (in RPM). The linear speed of the moving platform (V_{mp}) is in turn related to the angular speed of the axle through the following relationship:

$$V_{mp} = \omega_{axle} R_{tire} \tag{21.2}$$

which is expressed here in terms of radians/s and R_{tire} is the radius of the driven tire. Changing units on angular speed to RPM in Equation (21.2), we obtain:

$$V_{mp} = \frac{2\pi N_{axle} R_{tire}}{60} \tag{21.3}$$

Rearranging Equation (21.3) to obtain an expression for N_{axle} and substituting the result into equation (21.1) leads to:

$$GR = \frac{2\pi N_{motor} R_{tire}}{60 V_{mp}} \tag{21.4}$$

where length units on R_{tire} and V_{mp} must be the same, and the time units of V_{mp} are seconds. Equation (21.4) can be used to calculate the overall gear ratio required to achieve a desired speed V_{mp}, but only if we know the value of N_{motor}.

With small DC motors, the determination of N_{motor} is not always a straightforward matter. To begin with, N_{motor} is linearly dependent on the rotational resistance, or torque, acting on the motor shaft as illustrated in Figure 21.2. Each point on the motor curve represents a different equilibrium state of the motor. For example, if the motor shaft is allowed to freely spin, it will have an angular speed equal to the no load angular speed denoted by N_{noload}. Conversely, if you grab the spinning motor shaft between your fingers and gradually increase the pressure, the shaft will stop spinning when the torque you are applying reaches a value equal to the stall torque, T_{stall}. Thus, to determine N_{motor}, we need to know N_{noload} and T_{stall} of the given motor, as well as the torque (T_{motor}) acting on the motor shaft. Herein lies the challenge, for the determination of T_{motor} is considered beyond the scope of this text and N_{noload} and T_{stall} are not always provided with the motor specifications.

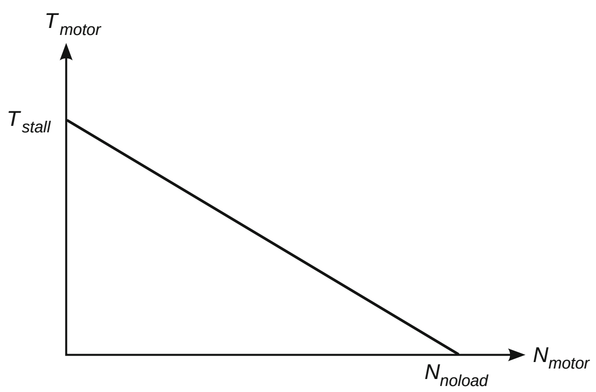


FIGURE 21.2 Typical Motor Curve for a Small DC Motor

Given these constraints, we recommend that you calculate gear ratio only if T_{motor} is close to zero for the application. This requires that the vehicle be very small and light (so that frictional losses are negligible) and that it only moves on a level plane without pushing against anything. Then, $N_{motor} = N_{noload}$, and Equation (21.4) reduces to:

$$GR = \frac{2\pi\, N_{noload}\, R_{tire}}{60\; V_{mp}} \tag{21.5}$$

Since frictional losses are hard to avoid, you can expect the moving platform to run at a speed that is smaller than the target value. If there are times during operation when T_{motor} is not negligible, such as when climbing a hill or pushing against an opponent, Equation (21.5) is no longer valid. All we can tell you is that the gear ratio will have to be larger than the Equation (21.5) prediction. How much larger will depend on the magnitude of the traction force on the driven tire.

When the value of N_{noload} is not provided, you may (at your own risk) assume an average value of 9000 RPM, given that N_{noload} for most small DC motors is in the range of 6,000 to 12,000 RPM. This will not work if your DC motor is a gearhead motor, which already has a built-in gear box. These will spin at much lower rates, and so your only recourse, if you want to use Equation (21.5), is to try to measure N_{noload}.

Example 21.1

We want to design a moving platform with a top speed of 0.500 ft/s on the flat. The motor, with specifications of T_{stall} = .210 oz-in and N_{noload} = 11,600 RPM, is already in hand, as are the 2.00 inch diameter tires. Determine the overall gear ratio required to achieve the desired speed.

Solution

Substituting into Equation (21.5) we get:

$$GR = \frac{2\pi N_{noload} R_{tire}}{60\; V_{mp}} = \frac{2\pi(11600)(1)}{60(6.00)}\,[\text{RPM}][\text{in}]/[\text{s/min}][\text{in/s}] = 202.$$

where V_{mp} and R_{tire} were expressed using the same length units (inches).

21.2 EXPERIMENTS

Physical experiments are a particularly effective way to reduce risk when working with small electromechanical systems. Because of the small scale, materials needed for the experiments can probably be scavenged, or at least obtained at low cost, and realistic forces can easily be applied. Also, physical experiments are often more accurate than idealized mathematical models at this scale.

Since the actual design has not been built yet, the subfunction being investigated may have to be idealized for the purposes of the experiment. For example, you might use cheaper materials or use your hands to create the motion. The errors introduced by these approximations will be tolerable if they are much smaller than the changes in performance being observed.

Knowing when to use experiments requires a keen awareness of the sources of risk in a design. This is no time for overconfidence; you can safely assume that if something can go wrong, it will. Thus, it is vital that you be able to distinguish between the aspects of the design about which you are sure and those about which you are not so sure. The latter are candidates for physical experiments.

The steps for formulating an experimental plan are as follows:

1. Identify aspects of the design and its performance about which you are uncertain.
2. Associate the aspects in step 1 with one or more physical variables that can be varied by means of simple experiments.
3. Carry out the experiments that will do the most to reduce risk within the available time frame.
4. If possible, document the results in the form of graphs or tables.

Example 21.2

A concept for a design competition named Dueling Duffers has been proposed and is shown in Figure 21.3. The object of this head-to-head competition is to be the first to deposit up to 10 golf balls into a hole in the center of the tabletop playing field of Figure 21.4.

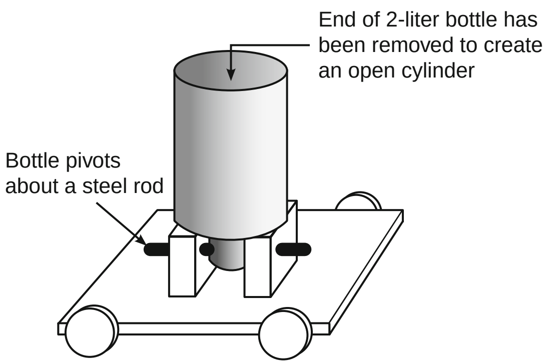


FIGURE 21.3 Proposed Design Concept for the "Dueling Duffers" Design Competition

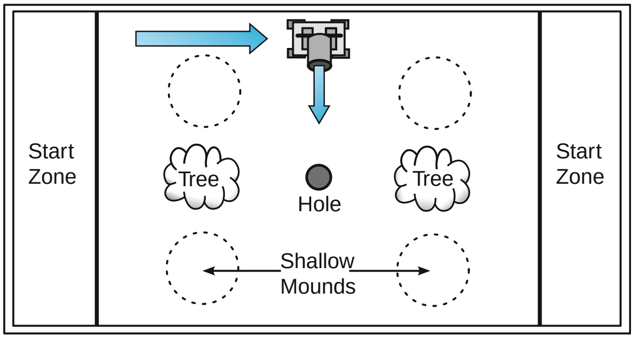


FIGURE 21.4 Playing Field for the "Dueling Duffers" Design Competition

These are the main features of the proposed design:

- It holds 10 golf balls in the top half of a 2-liter bottle.
- It uses the side rail to steer.
- When it is even with the hole, it dumps the golf balls in the general direction of the hole.

The path of the vehicle and the direction in which the balls are dumped are indicated in Figure 21.4. For this example, you are asked to (1) identify the main sources of risk and (2) propose experiments to address those sources of risk.

Solution

Sources of Risk/Uncertainty

Sources of risk are presented here in the form of questions, the answers to which are currently unknown. These are the same type of questions that will be asked by the jury at the oral design defense.

1. Will all 10 balls fit in the top half of the 2-liter bottle?
2. What is the optimal height from which to dump the balls?
3. Will all 10 balls drop into the hole when dumped in this manner?
4. Is it important to have the vehicle perfectly positioned before dumping the balls?
5. Will the machine travel slowly enough that its position can be easily controlled?

Proposed Experiments

The experimental setup in Figure 21.5 can be used to address each of the first four sources of risk just listed. The last item in the list is best handled by computing the overall gear ratio using Equation (21.5).

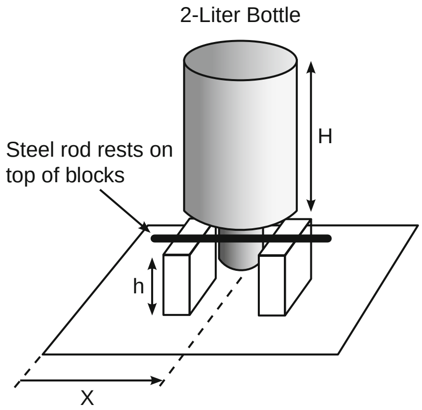


FIGURE 21.5 Experimental Setup for Establishing Key Dimensions and Reducing Risk

The bottle, the steel rod, and the two blocks of wood are all easily obtained. Assume the playing field is available for testing. The bottom of the plastic bottle will need to be cut off, and the steel rod will need to be inserted through the top of the bottle. There are no other parts that need to be joined. You can use your hands to keep the rod in place above the blocks, while taking care to let the bottle and balls fall under their own weight.

The numbers of the experiments described as follows correspond to the preceding numbered sources of risk.

1. The control variable is H (refer to Figure 21.5). Put 10 golf balls into the bottle and measure the minimum H required to hold 10 balls.
2. The control variable h determines how fast the balls will be rolling when they pass the hole. Increase h by inserting books under the blocks; decrease h by sawing the ends of the blocks. For each value of h, dump the balls three times and record the number of balls that drop into the hole. Document results by plotting h versus the average number of balls that dropped.
3. Here the concern is less with the speed and more with the distribution of the balls as they pass the hole. Control variables might be H or the manner in which the balls are packed within the bottle (e.g., a vertical divider could be used to keep the balls on the side of the bottle facing the hole). Again, use three trials for each value of the control variable and plot the control variable versus average number of balls that dropped.
4. The control variable is X. Perform three trials for each value of X, and plot X versus the average number of balls that dropped. The shape of this graph will provide the answer to question 4.
5. Calculate the gear ratio using Equation (21.5).

21.3 MODELS

Models are scaled replicas constructed out of inexpensive, readily available materials. In the case of small electromechanical devices, they often are constructed out of cardboard or foam board. Models are used to check geometric compatibility, establish key dimensions of moving parts, and to visualize the overall motion.

Typical examples are shown in Figures 21.6 and 21.7. The foam board model of the stair climbing device in Figure 21.6 was used to prove the feasibility of the design. Several other stair climbing concepts, mainly wheel and track designs, were regarded as feasible until models proved otherwise. In Figure 21.7,

FIGURE 21.6 Model of a Stair-Climbing Device

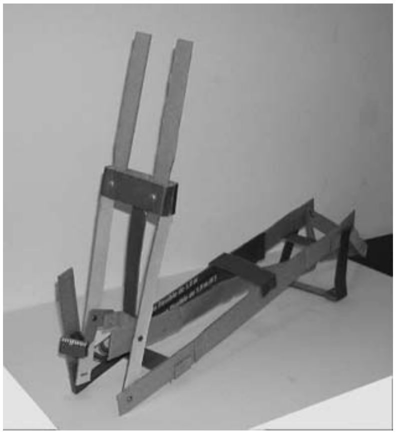


FIGURE 21.7 Model of a Dart-Throwing Device

a dart-throwing mechanical linkage was modeled in 3D using cardboard. All the links move, including the grip mechanism that releases the dart when the arm strikes a stopper.

21.4 DETAILED DRAWINGS

By definition, a detailed drawing will contain all the information required to manufacture the design. The drawings should be so complete that if you handed them off to someone unfamiliar with the design, that person would be able to build it.

The usual practice is to specify dimensions on multiple orthogonal views of the design. An isometric view sometimes also is provided to assist with visualization. In all, six orthogonal views are possible: front, back, left, right, top, and bottom. Three views, however, are most common. Figure 21.8 shows a detailed drawing with five orthogonal views.

Additional information such as material specification, part type, and assembly directions are conveyed through written notes on the drawings. Close-up views can be employed to clarify small features.

Although practicing engineers will generate drawings like Figure 21.8 using computer-aided design (CAD) software, first-year engineering students probably have not taken a CAD course yet. Therefore, we recommend that the usual standards for preparation of detailed drawings should be relaxed somewhat and replaced by the following set of guidelines:

- Drawings can be neatly hand drawn using ruler and compass.
- Drawings must be drawn to scale, though not necessarily full-scale.
- Drawings of at least two orthogonal views of the design should be prepared. An isometric view is not required, but close-up views should be used to clarify small features.
- Show hidden lines only when they will enhance clarity. These are dashed lines that are used to show edges that are not visible from the viewer's perspective.
- It is acceptable to show only essential dimensions—that is, key dimensions that either have a direct impact on performance or are needed to demonstrate that geometry constraints are satisfied.
- Use notes or labels to indicate material specification, part type, and assembly directions.

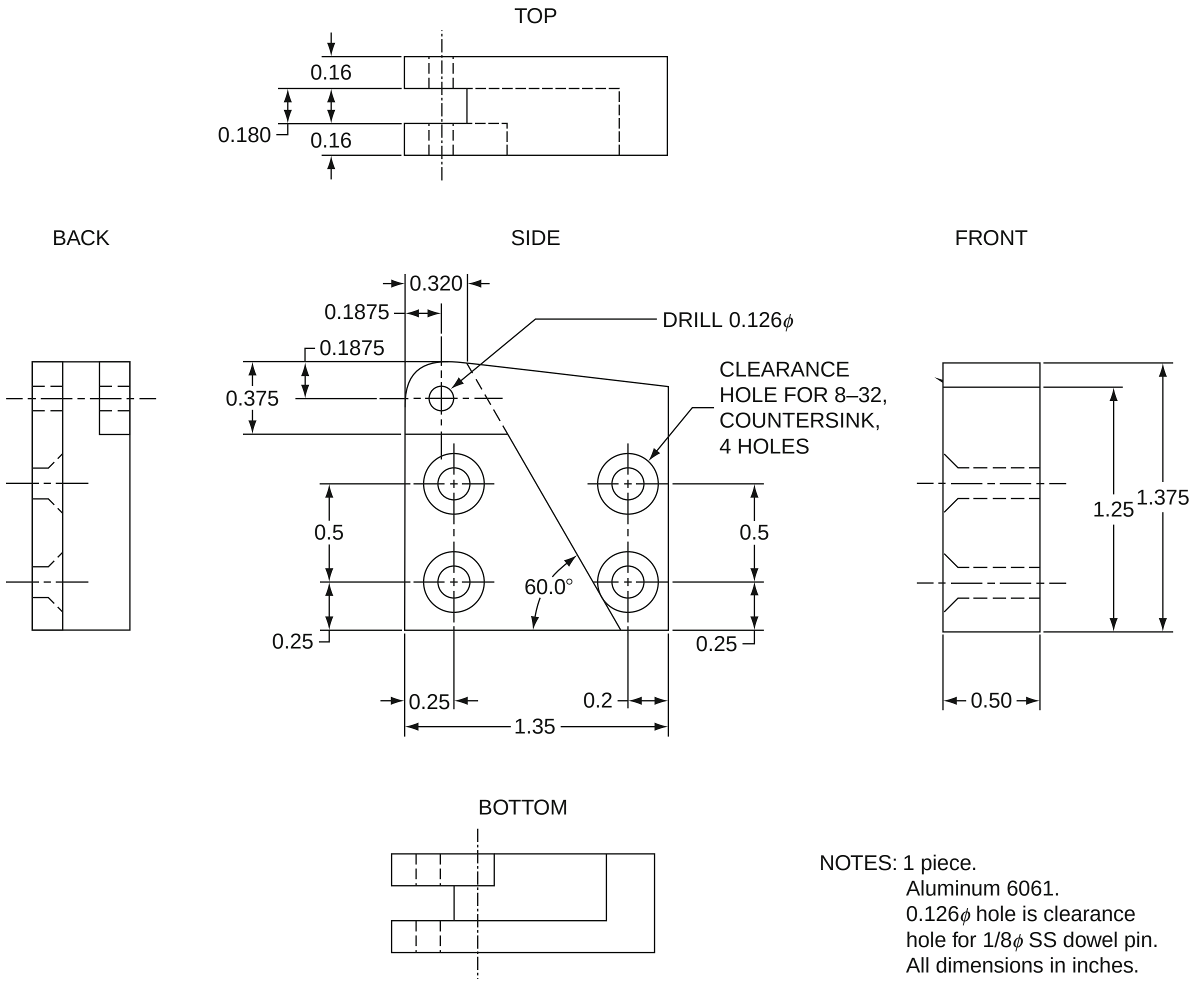


FIGURE 21.8 Five Views of the Right Toe of an Animatronic Eastern Gray Squirrel

- It is acceptable if the manufacturing details are incomplete. Students who lack the experience to fully specify them will have to discover those details by trial and error during building.
- If an electric circuit was designed, show it in the form a neatly hand-drawn but fully specified circuit diagram.

An example of a detailed drawing prepared in accordance with these guidelines is shown in Figure 21.9.

Although the freedom to leave out some dimensions and manufacturing details has been allowed, keep in mind that missing details amount to higher risk in the minds of those being asked to provide resources to the project. Thus, a design with fewer missing details may be viewed by the instructor as having lower risk.

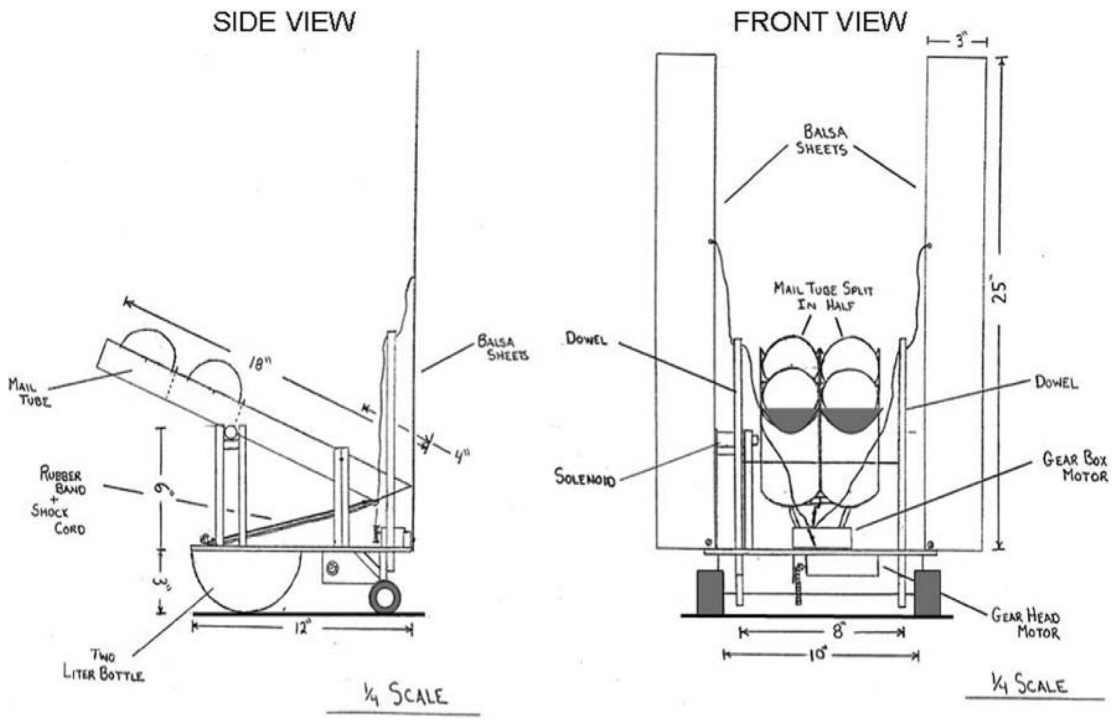


FIGURE 21.9 Hand-Drawn Detailed Drawing of a Competition Vehicle

21.5 DESIGN MILESTONE: DETAILED DESIGN

This milestone is all about reducing risk, not only in your own mind but in the minds of the jury at the upcoming oral design defense.

Assignment

1. Use analysis, experiments, and models to help establish dimensions and proof of concept.
2. Prepare detailed drawings of the design concept you selected.

Documentation

- Write the analysis details in the usual format, stating all assumptions.
- For each experiment: (1) state the purpose of the experiment, (2) describe the experimental procedure, (3) present results, and (4) state conclusions.
- Summarize useful information yielded by models, and turn in models.
- Attach hand-drawn detailed drawings.

Grading Criteria

- From examination of the detailed drawings, does the design have a chance of working?
- Have opportunities to reduce the level of risk (through analysis, experiments, and models) been fully exploited?
- Is there enough information in the detailed drawings to manufacture the design?
- What is the overall quality of the detailed drawings?

21.5.1 Design Competition Tips

- Time spent now on analysis, experiments, and models pays off later in fewer design iterations during manufacturing and testing.
- Of the four methods for reducing risk, good detailed drawings will reap the most rewards in a freshman design project.

CHAPTER

Design Step 5: Design Defense

22

Source: © iStockphoto.com/Jennifer Trenchard

Engineers must convince customers that a design is worth expenditures of money and the time of skilled people. In a student design project, this process of convincing customers is simulated by an **oral design defense**. The goal of this oral presentation is to win the confidence of the project sponsors, henceforth referred to as the **jury**.

In assessing a team's chances for future success, the jury will be searching for answers to the following questions:

- Did the team adhere to the systematic approach?
- How does the final concept work?
- What is the level of risk associated with this design?
- Do the students appear to be teaming effectively?

The jury's concerns suggest some strategies that should be effective. First, the organization of the presentation should parallel the steps in the design process, as shown in Table 22.1. This is your way of saying that you followed a systematic approach. Second, you should try to get the jury to understand how your final concept works as quickly as possible. This frees up more time during questions for alleviating concerns about the design. Third, you should anticipate that the jury will ask questions about potential sources of risk, and prepare evidence in advance that will quell those concerns. This evidence should be in the form of quality detailed drawings, results of calculations and experiments, and models. If you built models, bring them; if you conducted experiments, try to bring some evidence that indeed you did them. The strategies that serve to reduce risk, and their counterparts that don't, are summarized in Table 22.2

Meanwhile the jury will also be evaluating your teaming. They will base their impressions on the quality of your design and oral presentation (see Table 22.3 for some tips on delivery and visual aids). There are other telltale signs. For example, did everyone contribute equally to the presentation? Was everyone involved in answering questions? Did team members refer to themselves as "we" or "I" when citing accomplishments?

When answering questions, be forthright and honest. Failure to do so will lead to an unending chain of questions. If your response is an opinion and not a fact, state so, because one erroneous answer can damage your credibility and thus elevate the risk associated with your design.

Table 22.1 Suggested Organization of the Oral Design Defense

Organization	Slides
Title	1
Outline of Presentation	1
Problem Definition	1
Important Design Requirements	1
Alternative Concepts Not Selected	2
Final Concept ■ **Describe main features** ■ **Explain why you selected it**	1–2
Detailed Design ■ **Show main drawings** Explain how it works Explain how you will construct it	2
■ **Zero in on special features with close-up views**	1–2
■ **Present results of analyses, experiments, and models**	1–2
Summary ■ **Summarize strengths of the design** ■ **Quantify performance expectations (e.g., top speed)** ■ **Describe your strategy at the final competition**	1
Total Slides	**12–15**

Table 22.2 Strategies That Either Reduce or Amplify Risk

Risk Reducers ■ High-quality concept drawings and detailed drawings ■ Calculations, experiments, and models that establish proof of concept ■ Manufacturing details have been explained ■ Quality visual aids
Risk Amplifiers That Could Delay Manufacture ■ Poorly detailed drawings ■ One or more subfunctions obviously will not work ■ No thought given to manufacturing

Table 22.3 Oral Presentation Tips

Tips on Delivery
1. Do not read sentences directly off the slides.
2. Look at the audience.
3. Stand next to the screen when speaking.
4. Practice the presentation.
5. Be positive and dynamic.
Tips on Visual Aids
1. Avoid using too many words.
2. Use a font size that can be seen easily.
3. Include concept or detailed drawings.
4. Put a heading on each slide.
5. Show results of calculations or experiments.
6. View the projected images in advance to make sure all words and pictures will be visible to the audience.

22.1 DESIGN MILESTONE: ORAL DESIGN DEFENSE

To qualify to receive parts and materials for the manufacturing phase of a hands-on design project, a majority of the jurors must be convinced that your design will work. In the event such a consensus is not achieved, teams will be asked to revise their designs and resubmit at a later date.

Assignment

1. Prepare the visual aids for the oral design defense. You must use PowerPoint or an equivalent software package. Relevant drawings should be scanned.
2. Practice the presentation.
3. Deliver the oral presentation to a jury of evaluators.

Typical Format

- Eight minutes for the oral presentation; four minutes for questions.
- All team members participate in the presentation and in responding to questions.

Grading Criteria

- What is the level of risk associated with the design?
- What was the quality of the drawings and the other visual aids?

22.1.1 Design Competition Tips

- Grading tends to be proportional to the quality of the drawings.
- Evidence of the use of calculations, experiments, and models will do much to reduce the level of risk in the minds of the jurors. Visual representations of the results are more effective than just saying you did them.

This page intentionally left blank

CHAPTER

Design Step 6: Manufacturing and Testing

23

Source: Courtesty of Daimler-Chrysler

Manufacturing begins once the detailed design has been approved and ends when the machine is placed in the starting zone of the final competition. In between, the machine will undergo numerous modifications. Few new designs work on the first try. Manufacturing and testing will tend to take much longer than expected—probably from three to five times as long. This chapter begins with a summary of good **manufacturing and testing strategies**, and then moves on to describe **materials, joining methods**, and **hand tools**.

23.1 MANUFACTURING AND TESTING STRATEGIES

There are strategies that you can employ during manufacturing to minimize the time it takes to get the first prototype ready for initial tests. The extra time freed up for testing and design refinements can prove decisive in a design competition.

The following time-saving manufacturing strategies have been observed in successful teams:

- **Talk to a machinist.** Professional machinists are considered partners in the design process. What they lack in knowledge of the engineering design process and analysis, they make up for in manufacturing and practical experience. As a practicing engineer, you will be required to consult with a machinist before finalizing your detailed design.
- **Don't delay in getting started.** Take the leap of faith and begin manufacturing as soon as possible. Only then will the team gain a realistic sense of the manufacturing timeline.
- **Divide responsibilities so that team members can work in parallel on different subassemblies.** Otherwise, you may find the entire team standing around waiting for the same glue joint to dry.
- **Keep detailed drawings up to date.** If they are not up to date, only one person will know what the actual design looks like. That one person will end up doing most of the manufacturing while the other team members watch. With accurate drawings, team members can work in parallel.
- **Set and enforce intermediate deadlines.** Manufacturing can span several weeks. The instructor will set the big deadlines through the milestones; the teams should set the little ones in between.

Testing will be as important as manufacturing in preparing for a design competition. For example, three design teams were assigned the task of designing machines capable of playing 18 holes of miniature golf at a local course. The first team was stocked with experienced machinists, and so they built their machine out of thick steel parts. The second team had an expert welder, and so they welded together their machine out of steel beams and plates. The third team chose to make their machine out of wood. Three very different manufacturing skill sets, yet all three machines failed at the final competition for the same reason—not enough testing. The first team did not test their machine on synthetic grass before the competition. Their steel machine was heavy, which created large friction forces between its tank-line treads and the synthetic grass. When it attempted to turn, the treads broke, immobilizing their machine. The second team did not have time to test their steering mechanism due to last-minute modifications. As a result, they could not consistently maneuver into position for their putts within the time constraint. The third team completed manufacture and testing of their moving platform two weeks before the final competition. Over the next two weeks, while their putting mechanism was being made, they did not test the moving platform again until about 30 minutes before the start of the competition. Their machine never moved!

The lessons learned by these three teams apply to all design competitions. They are summarized in the following testing tips:

- Always leave a lot of time for testing.
- When conducting tests, try to simulate as closely as possible the conditions at the final competition. If these conditions are not known, test under a variety of conditions to insure robustness.

23.2 MATERIALS

When possible, the designs should be made of wood to facilitate manufacture and keep costs down. Manufacturing can be simplified still further by constraining the designs to be small (less than 1 ft^3) and lightly loaded. Under these conditions, balsa can be used as the main structural material.

Recommended materials for a small, lightly loaded electromechanical device are listed in Table 23.1 along with their relative attributes. Balsa is listed as easy to use with hand tools because it can be cut and shaped easily with a sharp knife. On the other hand, plastic tends to deform rather than shear cleanly under the action of cutting tools, so it is listed as hard to work with.

Table 23.1 List of Recommended Materials for a Small Electromechanical Design Project. Strength (Resistance to Breaking) and Stiffness (Resistance to Deformation) Are Normalized with Respect to Values for Balsa Wood.

Material	Relative Strength	Relative Stiffness	Ease of Manufacture	Useful Forms
Balsa	1	1	Very easy	Sheets, beams, blocks
Woods	5	4	Easy	Plywood sheet, wood dowels
Plastic	5	1	Hard	Prefabricated gears
Steel	50	80	Medium	Thin rods
Rubber	0.5	Very small	Very easy	Long strands

When selecting a material from Table 23.1, the strength and stiffness requirements of the given part also need to be considered. For example, if there are concerns about a part breaking under load, strength considerations will override ease of manufacture, leading us to use plywood instead of balsa. If a small-diameter axle requires high stiffness so that gears can remain engaged, then steel rods may be the best choice.

23.3 JOINING METHODS

For small-scale balsa and wood structures, the preferred methods for joining parts are adhesives, wood screws, and machine screws with nuts. Typical joint configurations employing these methods are shown in Figure 23.1. Use of tapes, especially duct tape, is frowned upon for their nonpermanence and poor aesthetics. Nails are not particularly compatible with balsa because of the large impact forces involved and the possibility of wood splitting.

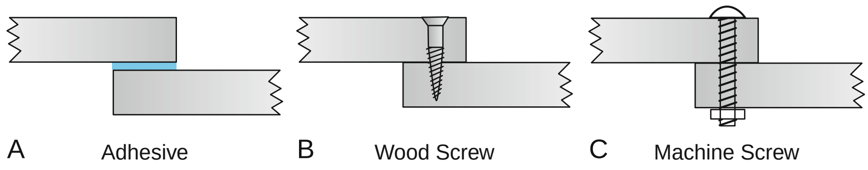


FIGURE 23.1A, B, C Different Joining Methods

Adhesives, in particular hot glue, are the method of choice when balsa is the predominant structural material. Hot glue cures quickly, and though essentially permanent for wood-to-wood bonds, metal-to-wood bonds can be adjusted or broken by heating. It is so general that it can be used to mount a motor to a plywood base. The main drawback of hot glue is its low strength, but this is usually not an issue for lightly loaded balsa structures. When it does become an issue—for example, when the surface area available for the glue joint is very small—one of the higher strength adhesives listed in Table 23.2 may be substituted.

Table 23.2 Common Adhesives

Adhesive	Typical Uses	Setting Time[a]	Curing Time[b]	How to Apply	Relative Strength
Wood glue	Wood, paper	8 hr	24 hr	Apply in liquid form direct from bottle.	13
Epoxy	Wood, metal	5 min to 12 hr	3 hr to 3 days	Comes in two tubes; mix equal amounts and apply with stick.	14
Hot glue	Almost anything	1 min	2 min	Place glue sticks in heated gun and apply with gun.	1

[a] *Setting time = time to harden*
[b] *Curing time = time to reach maximum strength*

On balance, wood screws are a less popular alternative for balsa-to-balsa joints. To reduce the chances of wood splitting, a pilot hole equal in diameter to the screw without the threads should be drilled prior to inserting the screw. This material removal plus the persistent possibility of wood splitting reduces the effective strength of the balsa members. In situations where adjustability is needed, the benefits of easy screw removal may override these costs.

Wood-to-wood joints are a different matter. Wood is much stronger than balsa, and so a higher strength joint is justified. In such cases, use of screws is often preferable to the long curing times of the higher-strength adhesives.

23.4 USEFUL HAND TOOLS

Small-scale electromechanical devices can be crafted using common hand tools, provided balsa or wood are the primary structural materials. In this section, we offer a short compendium of those tools for your easy reference. Our goal in presenting them is to make you aware of the alternative manufacturing operations at your disposal. We will not go into much detail on how to use them. Instead, we advise that you observe your classmates, talk to a machinist, or just give it a try.

For each tool, we will (1) show a picture of it (so that you can find it), (2) name it (so that you can ask for it), (3) describe its use, and (4) provide additional comments on usage.

23.4.1 Tools for Measuring

FIGURE 23.2 Tape Measure and Ruler

Name: Tape measure; steel ruler
Use: To measure length
Comments: Steel ruler can also be used as a straight edge when making straight cuts in balsa with the X-acto knife.

23.4.2 Tools for Cutting and Shaping

FIGURE 23.3 X-acto Knife

Name: X-acto knife (Hunt Manufacturing Co.)
Use: For cutting and shaping balsa
Comments: Blades come in different shapes and are replaceable.

FIGURE 23.4 Coping Saw

Name: Coping saw
Use: For cutting curves in wood
Comments: The blade can be mounted in the frame to cut on either the push or pull stroke. Unscrewing the handle relieves tension on the blade so that it can be removed, rotated up to 360°, or inserted through a small drilled hole in order to cut out a larger hole.

FIGURE 23.5 Hacksaw

Name: Hacksaw
Use: For making straight cuts in metal
Comments: It cuts on the push stroke, when blade is mounted correctly. Though designed for metal, it can also be used with wood. Blades are detachable and can be rotated to cut in four directions.

FIGURE 23.6 Dremel with Bits

Name: Dremel® (Robert Bosch Tool Corporation)
Use: For shaping and drilling holes in wood
Comments: This is an electric hand drill without the handle. Because of its shape, forces can be applied more easily when grinding/sanding.

23.4.3 Tools for Drilling Holes

FIGURE 23.7 Cordless Hand Drill

Name: Cordless hand drill
Use: For drilling holes in wood and plastic
Comments: Release trigger lock and pull trigger to start drilling. Power switch controls the direction of spin.

23.4.4 Tools for Joining Parts

FIGURE 23.8 Hot Glue Gun

Name: Hot glue gun
Use: For applying hot glue
Comments: Insert a glue stick in the back, wait 3 to 5 minutes for the gun to heat up, and then apply the glue by pulling the trigger. Handle with care as the glue can reach temperatures of 400°F.

FIGURE 23.9 Spring Clamps

Name: Spring clamps
Use: For holding parts together when waiting for an adhesive to set
Comments: If the grips do not open wide enough for your application, C-clamps are a good alternative.

FIGURE 23.10 Screwdriver

Name: Screwdriver
Use: For inserting and removing screws
Comments: The longer and thicker the handle, the easier it is to apply torque. The screwdriver in Figure 23.10 has a removable tip. You can pull out the tip, rotate it 180°, and reinsert it to switch from a regular screwdriver (⊖) to a Phillips head screwdriver (⊗).

FIGURE 23.11 Adjustable Wrench

Name: Adjustable wrench
Use: For holding nuts when tightening a screw
Comments: Thumb adjustment changes distance between jaws. Unlike pliers, there is no gripping force, so parts being held must have flat surfaces.

FIGURE 23.12 Claw Hammer

Name: Claw hammer
Use: For driving or removing nails
Comments: Nails should not be used to join balsa parts.

23.4.5 Tools for Wiring

FIGURE 23.13 Soldering Iron

Name: Soldering iron
Use: For attaching or connecting copper wires with solder
Procedure: (1) Create mechanical connection (e.g., by twisting wires together); (2) let iron heat up; (3) apply solder to tip of iron (called tinning); (4) heat wires with iron; (5) bring solder in contact with heated wires until solder melts and flows; and (6) let wires cool.

FIGURE 23.14 Long-Nosed Pliers

Name: Long-nosed pliers
Use: For bending wires, grasping parts, cutting wires, and reaching into tight places
Comments: Clamping forces are highest at the wire cutters and smallest at the tip of the nose.

FIGURE 23.15 Wire Cutters

Name: Wire cutters
Use: For cutting or stripping wires
Comments: You may not need this tool if your long-nosed pliers have a good wire cutter.

FIGURE 23.16 Wire Strippers

Name: Wire strippers
Use: For stripping insulation from wires, cutting wires, and shearing small diameter bolts
Comments: To strip insulation, lay wires across jaw in notch of same diameter as metal wire, and then pull wire while holding jaws compressed.

23.5 DESIGN MILESTONE: DESIGN FOR MANUFACTURE ASSESSMENT I

The **Design for Manufacture** (**DFM**) assessments are so-named because the teams that gave serious consideration to design for manufacture principles when first formulating their designs will have the best chance of doing well on these milestones.

This first DFM assessment should occur about halfway through the first manufacturing iteration. For a design project of the scale proposed in these chapters, that's about one week after receiving the box of parts.

Assignment

Make as much progress as possible toward completion of manufacturing and testing of your design.

Grading

Teams should bring their materials and a detailed drawing of their machine to the next design studio. Grading will be based on the amount of progress made.

100 points = manufacturing is more than 50% completed
80 points = manufacturing is between 25% and 50% completed
60 points = some progress made (but less than 25%)
0 points = no progress

23.5.1 Design Competition Tips

- Team members should work in parallel on different aspects of the design to accelerate progress.
- From here on the design project is a race to begin testing. Teams with the most testing time tend to win competitions.

23.6 DESIGN MILESTONE: DESIGN FOR MANUFACTURE ASSESSMENT II[1]

By the time this milestone is reached, teams should have completed their first manufacturing iteration and begun testing. For a freshman project, this could be as soon as one week after the previous DFM milestone.

The instructor will evaluate progress by conducting one (ideal) or more well-defined performance tests to determine whether or not the various subfunctions are working. Evaluation of how well they work is reserved for the next milestone. Subfunctions to be evaluated must be carefully selected so as not to be biased toward a particular design or strategy.

Assignment

Get all subfunctions working by the assigned due date.

Performance Test

To be defined by the instructor.

Grading

Team grades will be based on the level of functionality demonstrated in the best of three trials, and will be computed as follows:

If manufacturing is 95% completed, then:

$$Grade = B + \sum_{i=1}^{I}(W_i s_i)$$

where:

$Grade$ = assigned grade on a scale of 100 points
B = base grade (typically $B = 70$ points)
I = total number of subfunctions being evaluated
W_i = number of grade points associated with the ith subfunction and defined such that

$$100 = B + \sum\nolimits_{i=1}^{I} W_i$$

s_i = 1 (if the subfunction worked)
= 0 (if the subfunction did not work)

If manufacturing is less than 95% completed, then:

$$Grade = f\ B$$

where:

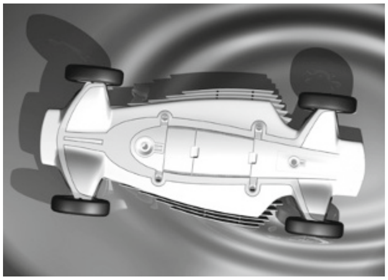
$Grade$ = assigned grade on a scale of 100 points
B = same base grade as earlier
f = fraction of manufacturing that was completed

[1]The instructor must tailor this milestone to the design project.

CHAPTER

Design Step 7: Performance Evaluation

24



Source: Courtesy of General Motors Corp.

Once the design has been manufactured, it is time to evaluate performance. If the machines are required to interact with each other, performance should be measured in two stages. First comes **individual performance testing**. The manufactured device is tested alone, under controlled conditions, to verify that it is capable of doing what the problem definition requires. Then comes the **final competition**. The device is tested against other machines in a series of head-to-head matches to determine the best overall design. The student grade resulting from performance evaluation will typically constitute up to 50 percent of the project grade.

24.1 INDIVIDUAL PERFORMANCE TESTING

Performance of a given machine in head-to-head matches may vary with the opponent. For example, an offensive or defensive strategy that works well against one opponent may not work against another. So the only way to test all the machines under the same set of conditions is to test them in isolation, without an opponent.

The basic approach is to measure one or more quantities, referred to as **metrics**, which are good predictors of success in the head-to-head matches. Typical metrics are time, speed, pushing force, or number of points scored against a stationary obstacle representing the opponent. Some basic rules when selecting metrics are (1) they should be easily measurable, (2) they should be continuously variable to maximize information content, and (3) they should not be biased toward particular design solutions.

The number of different physical tests required to measure all the metrics will depend on the choice of metrics. Ideally, you want to be able to design a single test that will measure all the metrics. Sometimes, each metric will require its own test.

The performance grade is the overall measure of performance expressed on a scale of 100. We recommend that it be computed as a weighted sum of the metric values as expressed in the design milestone at the end of this chapter.

The specifics of the performance tests often are not revealed to the students until a week before they take place. This is to prevent students from tailoring their machines to the performance tests instead of to winning the final competition, which is the real design objective.

24.2 THE FINAL COMPETITION

The final competition pits the machines against each other, usually in a series of head-to-head matches that may involve direct interactions between the machines. This is the ultimate test of the machines, as it is the only way to accurately evaluate the effectiveness of offensive and defensive strategies, robustness, durability, and the wisdom of past design choices.

However, it is probably best that the results of head-to-head competition not be linked to the performance grade, since each machine will be facing a different set of challenges. For example, one machine may not match up well against a particular opponent, or a prefabricated part may fail unexpectedly due to an accidental collision.

24.3 DESIGN MILESTONE: INDIVIDUAL PERFORMANCE TESTING[1]

Gradewise, this is the most important milestone. The testing regimen enforced by this milestone and the previous one will prepare the machines for the final competition.

Assignment

Optimize performance of your machine in preparation for individual performance testing.

Performance Test

To be defined by the instructor.

Grading

Team grades will be based on the quality of performance demonstrated in the best of three trials, and will be computed as follows:

$$Grade = B + \sum_{i=1}^{I} W_i \left(\frac{m_i}{M_i}\right)^n - \sum_{j=1}^{J} P_j$$

where:

$Grade$ = assigned grade on a scale of 100 points
B = base grade (typically $B = 70$ points)
I = total number of metrics
J = total number of penalties assessed for rules violations
W_i = number of grade points associated with the ith metric and defined such that

$$100 = B + \sum\nolimits_{i=1}^{I} W_i$$

m_i = measured value of the ith metric
M_i = best value of m_i recorded by any team in the class
n = 1 (if performance is directly proportional to m_i)
= −1 (if performance is inversely proportional to m_i)
P_j = number of grade points associated with the jth penalty

[1]The instructor must tailor this milestone to the design project.

If the *Grade* calculated is less than or equal to B, then

$$Grade = f\,B$$

where:

$Grade$ = assigned grade on a scale of 100 points
B = same base grade as above
f = fraction of manufacturing that was completed

CHAPTER 25

Design Step 8: Design Report

Source: © iStockphoto.com/Edyta Pawlowska

The design report documents the final design. It enables someone unfamiliar with a design to figure out how it works, evaluate it, and reproduce it. It is the final step in the design process. This chapter summarizes **organization** of the report and provides some **writing guidelines**.

25.1 ORGANIZATION OF THE REPORT

Like the oral design defense, the organization of a design report follows the steps of the design process (see Table 25.1). The report begins with a concise statement of the Problem Definition in the student's own words.

The Design Requirements section does not need to repeat the competition rules. Instead, it should begin with a short paragraph describing competition strategy. Then list all performance requirements—for example, "Must have a top speed of 1 ft/s"; "Must deposit at least 6 Ping-Pong balls"; "Must be able to steer."

Almost all the content of the Conceptual Design section should be available from previous milestones. In Section 3.1, present the sketches of your alternative design concepts and briefly describe how each one works. In Section 3.2, present your decision matrix and use the matrix as a vehicle to discuss the strengths and weaknesses of each concept. In Section 3.3, indicate the concept you selected and give your rationale.

The Detailed Design section describes the design that appeared at the final performance evaluation. New detailed drawings will have to be prepared. Place these new drawings in Section 4.1 along with text describing the operation and main features of the final design. Describe the overall design first and then zero in on the details of special features. If possible, include digital photographs of your machine. To create Section 4.2, retrieve the results presented at the oral design defense and add text. Section 4.3 is primarily a summary of the joining methods used.

The results of the performance tests are summarized in the Performance Evaluation section. Describe how your machine fared both during individual performance testing and at the final competition, being as quantitative about it as you can. Also compare performance predictions to actual results.

Table 25.1 Suggested Organization of the Design Report

Organization	Pages
Title and Authors	1
Table of Contents (with page numbers)	1
List of Individual Contributions to the Report	1
1. Problem Definition	0.5
2. Design Requirements	1
3. Conceptual Design	
3.1 Alternative Concepts	2
3.2 Evaluation of Alternatives	1
3.3 Selection of a Concept	0.5
4. Detailed Design	
4.1 Main Features and How It Works	3
4.2 Results of Analysis, Experiments, and Models	1
4.3 Manufacturing Details	1
5. Performance Evaluation	1
6. Lessons Learned	1–2
	Total = 15–16

The "Lessons Learned" section is an opportunity to reflect back on the design experience. Write it in the form of three paragraphs, with each paragraph dedicated to answering one of the following questions: (1) How would you redesign your machine to improve performance? (2) What general lessons did you learn about the design process? (3) What general lessons did you learn about teaming?

25.2 WRITING GUIDELINES

Use double spacing to leave room for instructor comments. Use the section and subsection headings of Table 25.1 and boldface them so that they stand out. Finally, figures should be embedded within the text (rather than placed at the end of the report) for easy reference. In addition:

1. Be concise.
2. Begin each paragraph with a topic sentence that expresses the theme or conclusion of the entire paragraph. The reader should be able to overview the entire report by reading just the topic sentences.
3. Generate high-quality concept drawings and detailed drawings to pass the "flip test." The first thing the instructor will do before reading the report will be to flip through the pages to examine the figures. The figures will form the instructor's first impression of the report.
4. Give each figure (i.e., drawing, graph, etc.) a figure number and a self-explanatory figure caption. For example: **Fig. 3: Decision matrix.** Figures should be numbered consecutively and should be referenced from the text using the figure numbers. For example: "The results of the comparison are summarized in the decision matrix of Figure 3."

5. Use a spelling checker. Spelling mistakes will cause the reader to question technical correctness.
6. Employ page numbers both in the text and in the table of contents.

25.3 DESIGN MILESTONE: DESIGN REPORT

This is a time both for documentation and for reflection.

Assignment

Prepare a design report in accordance with the guidelines of this chapter.

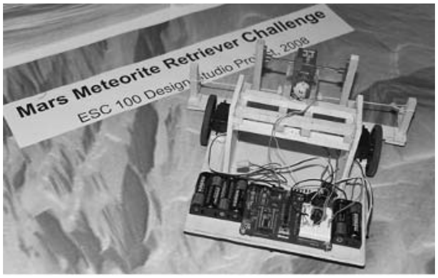
Grading Criteria

- Is the report complete?
- Does it pass the flip test?
- Does it use a solid writing style?
- Is it clear how the design works?
- Could someone unfamiliar with your machine manufacture it from the drawings and information in the report?

CHAPTER

Examples of Design Competitions

26



Source: Mars Challenge

In this chapter, we present a typical design competition along with one team's solutions to the first four milestones. Design competitions like the one described in this chapter have proven to be very successful with first- and second-year engineering students. The rules, tabletop playing field, and list of parts provided may be used as a template in defining similar competitions.

26.1 DESIGN COMPETITION EXAMPLE 1: A BRIDGE TOO FAR

This design project begins on the day that the instructor distributes the rules governing the design competition. This package of rules will typically consist of a statement of the design objective, a list of design requirements, a drawing of the playing field, and a list of parts and materials.

In the design competition named **A Bridge Too Far**, the objective is *to design and build a device to outscore your opponent in a series of head-to-head matches*. The playing field is shown in Figure 26.1. A team receives +1 point for every ball resting in its scoring pit at the end of play. In all there are 17 scoring balls. Six of the balls start out in the possession of the two teams; the remaining 11 balls have starting positions on the playing field as indicated in Figure 26.1. Other key rules are:

- The device must fit within a one-foot by one-foot by two-foot high volume.
- Parts and materials are limited to those listed in Table 26.1.
- Each device can have up to three independent tethered controls.
- There is a one-minute setup time.
- One game lasts 30 seconds.
- If there is a tie, both teams lose.

The parts and materials of Table 26.1 will not be supplied to the teams until they successfully defend their designs at the oral design defense. The complete set of rules is given at the end this chapter.

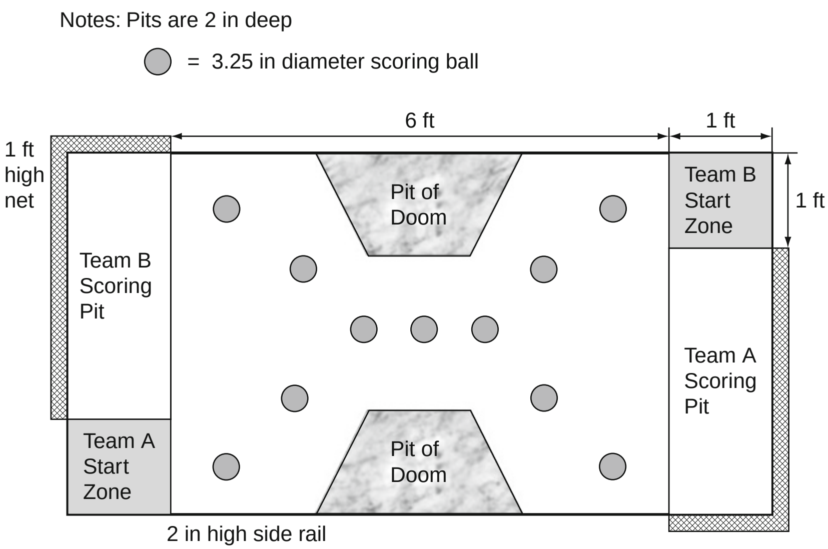


FIGURE 26.1 A Bridge Too Far Tabletop Playing Field

Table 26.1 The List of Parts and Materials for "A Bridge Too Far"

1	motor with adjustable gear box
1	gearhead motor
1	relay (for use with gearhead motor)
1	diode (for use with gearhead motor)
1	pull solenoid
1	set of 6 gears
1	wooden dowel, 3/8"×36"
2	balsa sheets, 1/16"×3"×36"
2	balsa beams, 3/8"×3/8"×36"
1	plywood sheet, 1/4"×12"×12"
10	super craft sticks
1	steel rod, 1/8"D×20"
1	piece of string, 10" long
2	metal hinges, 1" long
1	shock cord with hooks, 1/4"D×10" long
1	rubber band strip, 12" long
4	wheels with hubs, 2"D (for 1/8" shaft diameter)
1	mailing tube
1	2 liter (plastic) soft drink bottle (provided by team)

26.2 DESIGN MILESTONE SOLUTIONS FOR A BRIDGE TOO FAR

26.2.1 Design Milestone 1: Clarification of the Task

In the case of a design competition, clarification of the task is mostly the responsibility of the instructor. There were just two things left for the design team to do to complete this milestone.

First, the team directed the following questions to the rules committee:

1. Can a machine score from the Pit of Doom?
2. Can you drop obstacles to interfere with the other machine?
3. Can you score points by driving a machine loaded with balls into a scoring pit?
4. Can a machine attach itself to the other machine?
5. Can a machine expand beyond the 1-ft × 1-ft × 2-ft high dimensions once the game begins?

The answer to all of these questions is yes, since there is nothing against the proposed actions in the rules. Motivation behind the questions is to fully understand the design constraints and to probe for omissions that could lead to a design advantage.

Second, the team compiled a list of performance requirements that were specific to their design and their competition strategy:

D Must score at least four points
W Must score the first point within 10 seconds
D Must hinder the opponent's ability to score

The list is short to avoid solution bias. Later, after a design has been selected, other performance requirements can be added.

26.2.2 Design Milestone 2: Generation of Alternative Concepts

The functional decomposition upon which the team settled is shown in Figure 26.2. Consideration of offensive strategy is done through the subfunction "deliver balls."

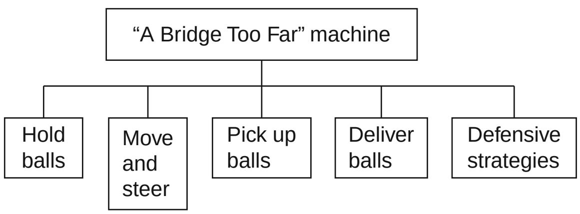


FIGURE 26.2 Functional Decomposition for "A Bridge Too Far"

The results of brainstorming each subfunction were collected in the classification scheme of Figure 26.3.

Functions \ Concepts	Concept 1	Concept 2	Concept 3	Concept 4	Concept 5
Hold balls	Place on machine	Keep under machine	Push balls in front	Container to put balls in	Drag behind machine
Move and Steer	2 motors	2 motors	1 motor steers	no steering	Don't move
Pick up balls	Don't pick up balls	wedge	scoop	plow	claw
Deliver balls	Push them into pit	Throw them into pit	Drive into pit with balls	Roll them down a ramp	Push opponent into pit
Defensive strategies	Block bridge	Ignore opponent	Throw extra balls into Pit of Doom	Pick up opponent	Throw obstacles

FIGURE 26.3 Classification Scheme for "A Bridge Too Far"

Compatible subfunction concepts were combined to form the four promising designs shown in the concept drawings of Figures 26.4 through 26.7.

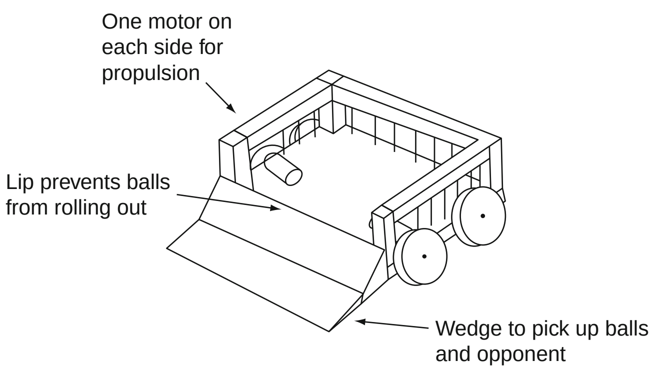


FIGURE 26.4 Concept Drawing of the "Wedge"

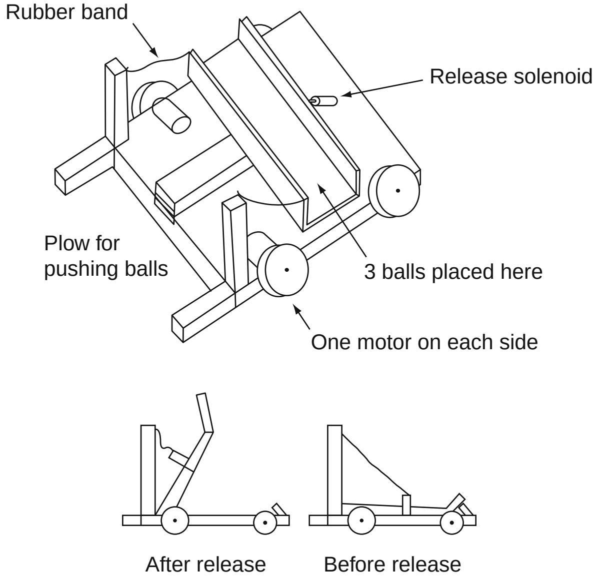


FIGURE 26.5 Concept Drawing of the "Catapult"

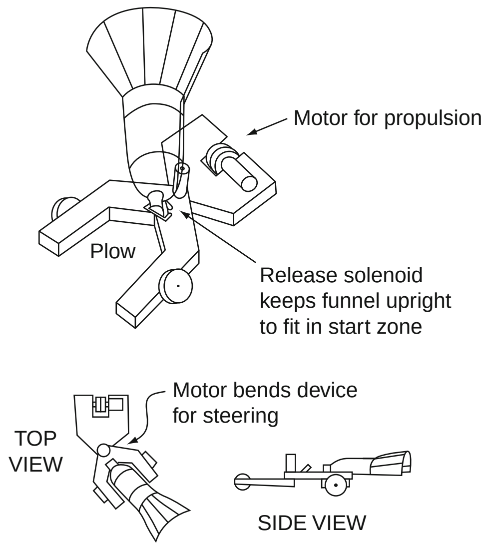


FIGURE 26.6 Concept Drawing of the "Snake"

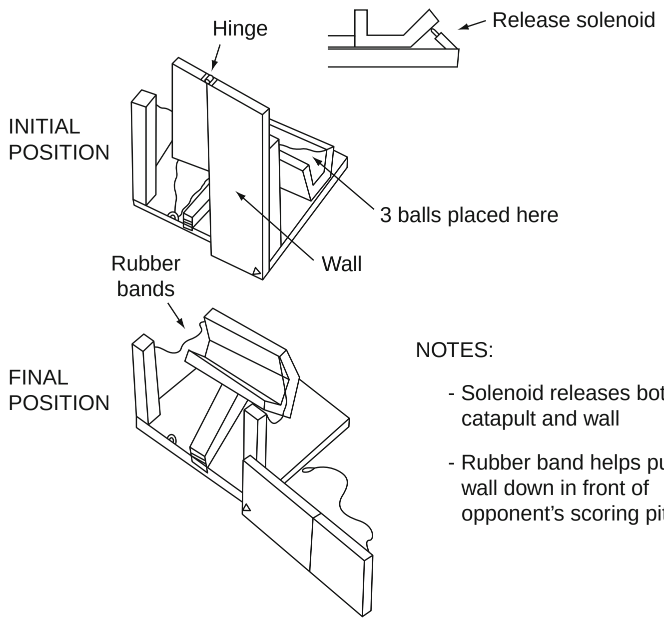


FIGURE 26.7 Concept Drawing of the "Wall"

26.2.3 Design Milestone 3: Evaluation of Alternative Concepts

The four alternative concepts developed in the previous milestone are illustrated in Figure 26.8 for easy reference.

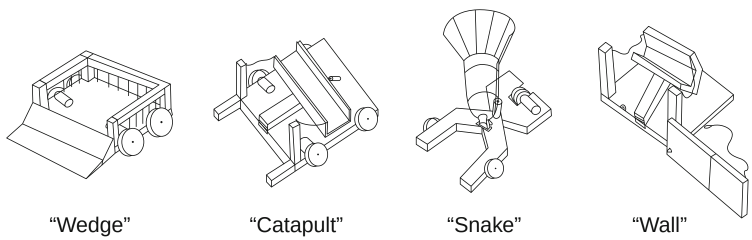


FIGURE 26.8 Alternative Design Concepts from "A Bridge Too Far"

The selection of evaluation criteria was tricky because a high-scoring machine that consistently gathers nine of the 17 balls should be ranked on the same level as a defensive machine that consistently wins by the score of 1 to 0. The evaluation criteria and their respective weights are shown in Table 26.2. The first two criteria both are needed to describe scoring potential because it is not enough that you can transport a lot of balls; you must be able to score with them with an opponent in your way. Their combined weighting of 0.4 is slightly higher than the weight of 0.3 assigned to defensive capabilities (third criterion), since you have to score to have a chance of winning. "Easy to manufacture" has its usual importance, but "low cost" was not included because it is not a factor in the competition. "Easy to control" appears because of the potential challenges involved in maneuvering across the bridge. The discussion of concept strengths and weaknesses is organized by evaluation criterion.

Table 26.2 Decision Matrix for "A Bridge Too Far"

Evaluation Criteria	Wt	Wedge		Catapult		Snake		Wall	
		Val_1	Wt × Val_1	Val_2	Wt × Val_2	Val_3	Wt × Val_3	Val_4	Wt × Val_4
Has a large payload	.2	6	1.2	5	1.0	8	1.6	2	0.4
Robust scoring capability	.2	5	1.0	9	1.8	3	0.6	7	1.4
Can disrupt opponent	.3	7	2.1	5	1.5	4	1.2	8	2.4
Easy to manufacture	.2	7	1.4	5	1.0	5	1.0	7	1.4
Easy to control	.1	7	0.7	6	0.6	3	0.3	9	0.9
Totals	1.0		6.4		5.9		4.7		6.5

Has a Large Payload

The Wall is rated low because it can score only with the three original balls. For the other three concepts, ratings are proportional to carrying capacity. Since the funnel design expands to be much larger than the start zone, it has a much larger carrying capacity than the other two.

Robust Scoring Capability

The two designs that launch the balls score highest here because it is much harder to block a ball that is thrown than one that is pushed. Also, these machines can score without having to cross the bridge.

Can Disrupt Opponent

The Wall is an obvious favorite for this category, since it is the only one that actively blocks the opponent's goal. The Wedge is designed to be able to pick up the opponent and drop it in the appropriate goal to gain more points, so it too can theoretically disrupt the opponent as well.

Easy to Manufacture

The Wedge is rated higher than the Catapult because it has fewer functions to build. The Wall also does well in this category because it does not involve any motors or corresponding drive-trains. In fact, it would be rated even higher were it not for the difficulties anticipated with sequencing the release of the wall and the catapult arm.

Easy to Control

The Wall simply requires flipping one switch, which is almost as easy as doing nothing. The Snake is slightly alien to most things people will have controlled and as a result would be hard to drive.

Discussion of Results

All the machines demonstrated strength in some areas. Observed weaknesses cannot be corrected without diminishing the attributes that made them strong. For example, the Snake's funnel, which gives it the largest payload, also makes it vulnerable to being pushed around by the opponent.

The Wall is the highest-rated design and also the boldest in that it dares to remain stationary. Students tend to shy away from designs like this one because it is different. This design was not selected because it was felt that its defensive capabilities have been overrated. The design may not be able to resist pushing by an opponent at the end of the wall.

The Wedge finally was selected as the best design. It is interesting that it is rated high even though it is not a clear winner in any of the categories. It compares very closely with the Catapult. The Wedge's simplicity, lifting potential, and large payload were the decisive factors.

26.2.4 Design Milestone 4: Detailed Design

As you may recall, the Wedge (shown on the left in Figure 26.9) was selected as the final concept. The original wedge design was dual purpose: (1) to lift and push the opponent's machine and (2) to allow balls to roll up into the holding bin above the moving platform.

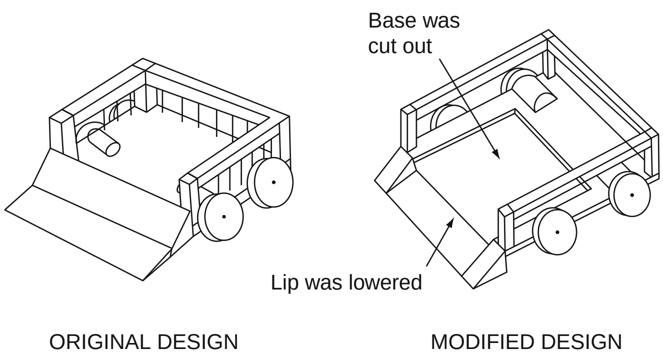


FIGURE 26.9 Design Modifications in Response to Experiments

Experiments

There were concerns that the forward speed of the vehicle might not be sufficient to allow the balls to roll up high enough into the bin. Experiments (see Figure 26.10) were conducted to check this out. Results showed that the concerns were justified. Acceptable dimensions for the wedge and holding bin were established based on the results of the experiments. The modified concept is shown on the right in Figure 26.9.

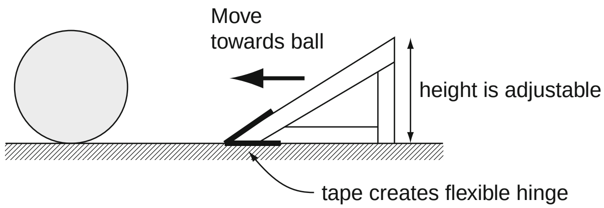


FIGURE 26.10 Experiment Used to Determine the Optimal Wedge Geometry

Analysis

Two identical gearhead motors were available for creating the drive trains for the left and right sides of the vehicle. The no-load angular speed (N_{noload}) at 24 V was listed in the catalogue as 145 RPM. To maintain a speed of 1.00 ft/s on the flat with tire diameters of 2.00 inches, the overall gear ratio needs to be (according to Equation (20.5) of Chapter 20):

$$GR = \frac{2\,\pi\,N_{noload}\,R_{tire}}{60\,V_{mp}} = \frac{2\,\pi\,(145)\,(1)}{60\,(12)} = 1.26$$

that we will round up to 2, based on the set of gears that was provided. With the pushing requirement of this design, it might seem that the actual gear ratio needs to be much higher. However, this particular motor is quite powerful; the motor shaft could not be visibly slowed by manually gripping the ends of the shaft. Given the power of the motor and the low weights of the vehicles, this gear ratio was deemed to be reliable even with the pushing requirement. The resulting drive train geometry is shown in Figure 26.11.

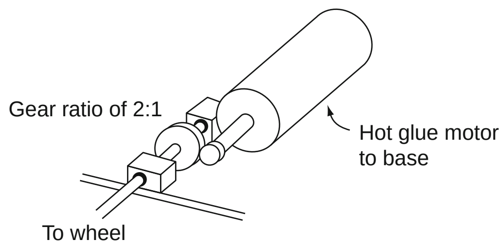


FIGURE 26.11 Close-up View of the Drive Train

Detailed Drawing

Two views of the final design are shown in the hand-drawn detailed drawing of Figure 26.12. Some manufacturing details are provided.

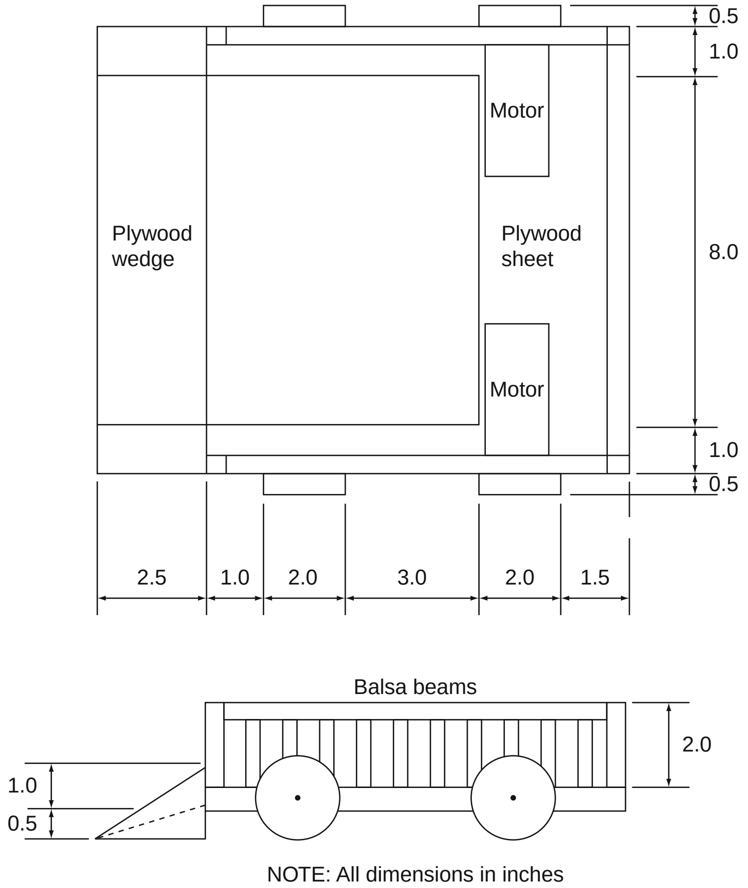


FIGURE 26.12 Detailed Drawing (two views)

26.3 OFFICIAL RULES FOR THE A BRIDGE TOO FAR DESIGN COMPETITION

26.3.1 Objective

Design and build a device to out-score your opponent in a series of head-to-head matches.

26.3.2 Constraints

1. Parts and materials that may be used in the construction of the device are limited to those defined in Table 26.1. Each team will be provided with a box containing all legal parts and materials.
2. The devices must be constructed entirely by the members of the team (i.e., if you lack the required expertise or tools to manufacture a certain part of the device, redesign it).
3. Each device can be loaded with up to three scoring balls. If a team chooses not to load all three balls, the discarded balls will be removed from the playing field.

4. When placed on the table at the start of the game, each machine must fit completely within its assigned one-foot by one-foot by two-foot high start zone.
5. An external power source will be made available to each device for use during the matches. It will consist of:
   - overhead wires that connect to the device (**Note**: These wires are considered to be part of the playing field.)
   - a control box with two forward-reverse-off switches and one on-off switch

26.3.3 The Game

1. Just prior to the game, there will be a one-minute setup time during which each team should:
   - place their device in the starting zone
   - attach and check the electrical connections
   - load their device with up to three scoring balls
2. Except for manipulation of the electrical control boxes located at one end of the playing field, no human interaction with the device (or playing field) is allowed once the game begins.
3. The game begins when indicated by the referee and ends 30 seconds later when the power is switched off.
4. The game, and all scoring, ends as soon as one of the following occurs:
   - all movement stops as a result of power being switched off
   - five seconds have elapsed since power was switched off

26.3.4 Scoring

1. At the end of the game, each team will receive 1 point for every ball in its scoring pit, regardless of which team caused the ball to fall into the pit. The team that scores the most points wins.
2. In all there are 17 scoring balls; six of the balls start out in the possession of the two teams, and the remaining 11 balls have starting positions on the playing field as indicated in Figure 26.1.
3. A ball is counted as lying in a scoring pit if an imaginary vertical line through the center of the ball lies within the boundary of the scoring pit. The referee can ascertain the status of a scoring ball at the end of the game by removing other balls from the scoring pit; if the ball in question then falls into the pit, it will count as lying in the scoring pit.
4. In the event of a tie:
   - If neither team has scored, both teams lose.
   - If both teams have scored, the winner will be the machine that has advanced farthest down the field as based on final positions. If this criterion proves indecisive, both teams will advance to the next round.

26.3.5 Other Rules

1. If the tethers (i.e., electrical connections) should entangle as a result of the machines passing each other, time will stop and both machines will be returned to their respective starting zones. Play will then continue at the signal of the referee. This situation can occur because each machine will be tethered to the nearest of two overhead rods running parallel to the length of the table. To avoid entanglement, keep to the right of the opposing machine when passing.

2. If devices permanently damage the playing field the balls will be disqualified.
3. Any attempt to intentionally inflict permanent damage upon an opponent's device will result in immediate disqualification. However, devices should be designed to hold up under expected levels of nonaggressive contact. For example, some pushing should be expected, but that pushing should not occur at significant ramming speeds.
4. Any device deemed unsafe will not be allowed to participate in the matches.
5. Implementation of any strategies that are not directly addressed in the rules but that are clearly against the spirit of the rules (e.g., intentionally interfering with the person at the controls) will lead to disqualification.
6. The rules committee has the final word on any interpretation of the rules.

26.4 DESIGN COMPETITION EXAMPLE 2: THE MARS METEORITE RETRIEVER CHALLENGE

Meteorites from Mars occasionally land on Earth. Those that land on Antarctica are particularly easy to spot because of their color contrast with ice. NASA is developing automated rovers to retrieve these meteorites. Your job is to implement such a rover on a simulated Antarctica: a wooden board 96" long and 48" wide. The meteorite will be simulated by a small object containing a light source. NASA has identified the most promising landing zone, simulated by a rectangle on the board 18" by 24". The meteorite can be anywhere within that landing zone.

Your *autonomous* vehicle will begin its trip at a Robot Depot, simulated by a 12" by 12" square on the board. Your challenge is to locate the meteorite, travel to it, pick it up, carry it without touching the ground at any point (to avoid contamination), and deposit the meteorite at the Meteorite Lab (another 12" by 12" square) (see Figures 26.13 and 26.14). Two completed student designs are shown in Figure 26.15

Vehicle starts at the Robot Depot and the meteorite is placed randomly within the landing zone.

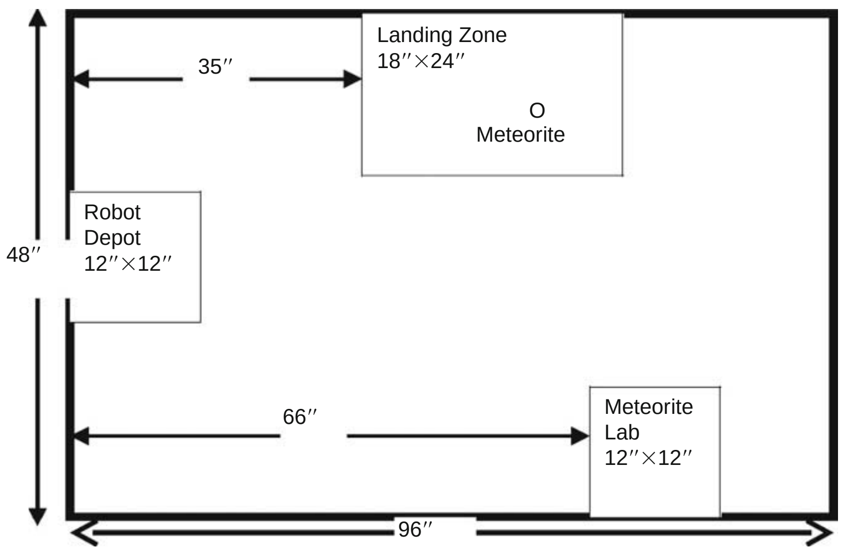


FIGURE 26.13 Layout of Simulated Antarctica.

FIGURE 26.14 The Actual Mars Meteor Retriever Challenge Game Board

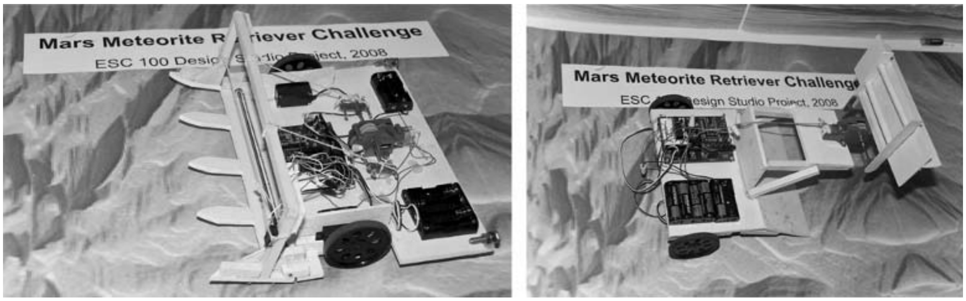


FIGURE 26.15 Two Student Vehicles

26.5 SOME DESIGN MILESTONES FOR THE MARS METEORITE RETRIEVER CHALLENGE

26.5.1 Design Milestone 3: Performance Testing

Grading

Team grades will be in proportion to level of functionality attained in the best of two tests, as defined by the following increasing levels of performance: The vehicle

- Is completely manufactured: 60%
- Moves when powered: 70%
- Moves with program control: 75%
- Navigates to the loading dock: 80%
- Navigates to the loading dock and touches meteorite under program control: 85%

- Lifts meteorite from the table surface and leaves the loading dock completely: 95%
- Lifts meteorite from the table surface and deposits light source in laboratory area: 100%

26.5.2 Design Milestone 4: Design Report

Purpose

To provide a concise, accurate and informative record of your design efforts. The report must be no more than 12 pages long, including cover page, table of contents, and drawings.

Organization

1. **Cover Page**
2. **Table of Contents**
3. **Introduction**: State the design objective and cite some of the design constraints.
4. **Strategy**: Describe your competition strategy.
5. **Conceptual Design**:
   - Discussion of alternatives with clearly labeled concept drawings. Discuss the strengths and weaknesses of each.
   - Include your functional decomposition, classification scheme, decision matrix, and the results.
6. **Detailed Design**:
   - How it works (i.e., details of the operation of each subfunction).
   - Include detailed drawings of the final design, and explain how it was put together.
   - Results of any calculations or experiments.
7. **Performance Evaluation**: Summarize the results of performance tests.
8. **Conclusions and Recommendations**: Discuss strengths and weaknesses of your design and your execution, and conclude with a discussion of lessons learned.

Grading

1. **Technical Communications: 50%**
   - The report is neat, concise, well organized, and includes all the items listed previously.
   - The report is free from spelling and grammatical errors.
   - The report is presented attractively.
2. **Technical content: 50%**
   - The report shows clear evidence that you understand and have followed the design methodology learned in the design studio.
   - The report describes how the design works clearly enough that a person who has never seen the device could produce a working machine from the descriptions and drawings in the report.

Oral Presentation

During the competition each team will give a two-minute presentation to the judges. This presentation must include a detailed sketch or picture of the final design, a brief description of design features, such as steering mechanism, pickup mechanism, and so forth that makes their design unique. You should also provide a brief description of a problem encountered and the way in which the problem was solved using teamwork. If you didn't encounter any difficulties in your design, discuss how teamwork made this possible.

Each team must practice their presentation and time it to ensure that it does not exceed two minutes. Your presentation will consist of two PowerPoint slides: a title page that includes the team members' names and team name (if you don't have a team name select one), and a slide with a detailed drawing of your final design.

26.6 OFFICIAL RULES FOR THE MARS METEORITE RETRIEVER CHALLENGE DESIGN COMPETITION

26.6.1 Objective

To construct an autonomous vehicle that can retrieve a (simulated) Mars meteorite from the (simulated) icy wastes of Antarctica.

26.6.2 Constraints

1. Parts and materials that may be used in the construction of the device are limited to those defined in Table 26.3. Each team will be provided with a box containing all legal parts and materials.
2. The devices must be constructed entirely by the members of the team (i.e., if you lack the required expertise or tools to manufacture a certain part of the device, redesign it).
3. When placed on the table at the start of the game, each machine must fit completely within its assigned one-cubic foot start zone.

Table 26.3 List of Parts and Materials for the Mars Meteorite Retriever Challenge

1	7/8"×11-7/8"×¼" Plywood
2	3/8"×3/8"×36" Bass wood beam
1	3/8"×36" Hardwood rod
6	Large Craft Sticks AKA tongue depressors
1	3"×36"×3/32" Balsa sheet
2	Parallax continuous rotation servo
2	Boe-Bot Wheels and Tires
1	Tamiza 4 speed Crank Axle Gearbox and Motor
1	Plastic ball caster
2	6-32 Threaded Rod 12" long with eight nuts and washers
2	4-AA battery holders
1	Parallax Board of Education microcontroller board with BASIC Stamp Processor Module
1	Box of paper clips
2	Light sensors
2	Contact switches

26.6.3 Scoring

The overall challenge will be to accumulate points for design quality and vehicle performance. The teams with the most points will be designated "superteams." Design quality points (up to 30) will be awarded by the judges for creativity, construction, and presentation. Vehicle performance points (up to 70) will be awarded for accomplishing the tasks of the challenge in the demonstration.

1. Teams will give a two-minute presentation about their design for the design quality judging.
2. Teams will demonstrate their vehicles. Vehicles will start from a designated Robot Depot.
3. At the judges' "go" signal, the teams will immediately operate any necessary switches and release their vehicles. No team member may subsequently make either direct or indirect contact with the vehicle or playing field until the "stop" signal is given.
4. Such contact, even if accidental, will result in immediate disqualification.
5. The vehicle will operate until it fully completes the task or until three minutes have elapsed. If three minutes have elapsed, the judges will give a "stop" signal and the team will terminate the vehicle's operation.
6. Points will be awarded for the demonstration of tasks as follows:
   - 5 points: Vehicle completely departs Robot Station
   - 5 points: Vehicle at least touches Landing Zone
   - 10 points: Vehicle touches meteorite
   - 10 points: Vehicle picks up meteorite
   - 10 points: Vehicle moves meteorite outside landing area (5 points for moving it but letting it touch the surface)
   - 10 points: Vehicle delivers meteorite to Meteorite Lab
   - 10 points: Vehicle deposits meteorite completely within Meteorite Lab
   - 10 points: Vehicle navigates completely outside Meteorite Lab

26.6.4 Rules

1. Vehicles must be a single unit.
2. Vehicles must at all times fit within a cube 12" on a side
3. No vehicle can damage or be attached to the playing field. All vehicles must move in a way that does no harm to the field.
4. All vehicles using batteries must include in their circuits a length of fuse wire between one pole of each battery pack and the next connected component. No vehicle may use more than eight AA batteries, or any other kind of battery.
5. The Board of Education microcontroller board must be mounted in a way that can be removed after the competition and returned undamaged.
6. Imaginative strategies in the spirit of the game are encouraged.
7. However, any strategies determined by the judges to be contrary to the spirit of the game will be excluded. Contestants have the responsibility of clearing with the judges before the competition any strategies that might possibly violate this rule.

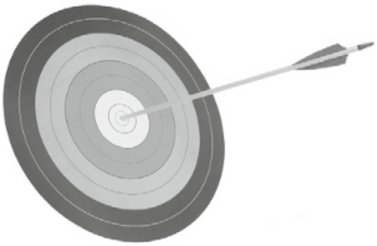
26.6.5 Additional Supplies

A circuit design and parts that will allow the Tamiya motor to be reversed will be provided to those who would like to implement it. Teams must purchase AA batteries for the microcontroller. Teams may purchase simple and inexpensive connectors, both mechanical (e.g., screws, nails) and electrical (e.g., simple contact switches). The simplicity and inexpensive nature of such additional supplies must be cleared with your instructor, and made known to and available to other contestants.

Teams who, in the opinion of the judges, are attempting to win the competition by exclusive use of expensive or otherwise difficult-to-access outside technology (i.e., components not made by the team itself from approved parts) will be disqualified.

CHAPTER 27

Closing Remarks on the Important Role of Design Projects



Source: © iStockphoto.com/Malcolm Romain

If you ask professors or students why they do design projects in engineering courses, you can expect to hear responses like this:

- They are motivational tools.
- They apply the analytical methods taught in courses.
- They help to develop written and oral communication skills.
- They teach teaming.

Indeed, these are all valuable outcomes of a design project, but each one can be achieved by some other means. The answer must lie elsewhere.

Part of the answer is found in the view that every engineering endeavor is ultimately about finding or designing a solution to an expressed need. The analytical methods, the teaming skills, and the rest are tools for achieving that goal—that is, they are the means to the end, not the end itself. Each design project offers a rare opportunity for students who spend most of their time deeply immersed in learning analytical methods to see the big picture.

The rest of the answer has to do with the real purpose behind these design chapters. Engineering design is at its core an unbiased and structured methodology for dissecting and solving complex problems. It is the way engineers should and must think. In contrast to analytical methods that are each limited to their own special class of problems, design methodology has universal applicability—to design, to research, to all fields of study. Design projects are the best way we know to exercise and develop this most fundamental of all engineering methods.

Hands-on design projects come closest to fully realizing these goals. They complete the design process, for as we have seen, it does not end with the detailed design; there will be design modifications to be made during manufacturing, testing, and the final performance evaluation. Students will learn the importance of design for manufacture principles by experiencing the results of having failed to heed them. They will also gain a sense of accountability by learning that it is not enough for a design to look good on paper—it has to *work*.

This page intentionally left blank

Postface

The old joke goes something like this: “One term ago, I couldn’t spell injuneer, but now I are one. . . .” Well, you are *not yet* an engineer, and if anything you should have developed an appreciation for the excitement and creativity of the profession. Here are some final thoughts:

1. Engineering is based on well-founded fundamental principles based in physics, chemistry, mathematics, and in logic, to name just a few skills.
2. The most general principles we have used are (1) definition of a Newtonian force unit, (2) conservation of energy, and (3) conservation of mass.
3. Engineering has multidisciplinary content, and the lines between each subdiscipline blur with other professions.
4. You will require sound-thinking skills in, say, Boolean algebra, as well as practical hands-on skills in the Design Studio.
5. Your Design Studio should have taught you that you will also need written and oral presentation skills.
6. Computer skills are essential to answer many kinds of practical engineering problems.
7. Engineering skills can be creative and intellectually rewarding as well as demanding.
8. You should now have some idea of what is meant by some of the subdisciplines of engineering and, for those who will continue to seek an engineering career, some idea of which of these subdisciplines appeals most to you.
9. We have offered you a practical way to ask if your behavior is ethical according to well-established canons. If you always act according to such canons, no matter the short-term temptations not to, you *will* come out ahead in the end.
10. For any Liberal Arts students who have enjoyed this course, you will take away an appreciation of what engineers do rather than the caricatures of them available in popular images.
11. Finally, we have enjoyed the privilege of guiding all of you! Thank you.

Index

Note: Page numberrs followed *f* indicates figures and *t* indicates tables.

J

K

L

M

N

T

U

V